FACULTY OF ENGINEERING
AND ARCHITECTURE

Continuous-Time Imprecise-Markov Chains: Theory and Algorithms

Thomas Krak

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Computer Science Engineering

Supervisors
Prof. Gert De Cooman, PhD* - Prof. Jasper De Bock, PhD* - Prof. Arno Siebes, PhD**

* Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University
** Universiteit Utrecht, the Netherlands

May 2021

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Continuous-Time Imprecise-Markov Chains: Theory and Algorithms

Thomas Krak

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Computer Science Engineering

Supervisors
Prof. Gert De Cooman, PhD* - Prof. Jasper De Bock, PhD* - Prof. Arno Siebes, PhD**

* Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University
** Universiteit Utrecht, the Netherlands

May 2021

GHENT
UNIVERSITY

ISBN 978-94-6355-486-2
NUR 993, 919
Wettelijk depot: D/2021/10.500/34

Members of the Examination Board

Chair

Prof. Hennie De Schepper, PhD, Ghent University

Other members entitled to vote

Prof. Koen De Turck, PhD, Ghent University
Prof. Enrique Miranda, PhD, Universidad de Oviedo, Spain
Max Nendel, PhD, Universität Bielefeld, Germany
Prof. Alain Sarlette, PhD, Ghent University

Supervisors

Prof. Gert De Cooman, PhD, Ghent University
Prof. Jasper De Bock, PhD, Ghent University
Prof. Arno Siebes, PhD, Universiteit Utrecht, the Netherlands

Address

Ghent University
Faculty of Engineering and Architecture
Department of Electronics and Information Systems
Foundations Lab for Imprecise Probabilities

Technologiepark Zwijnaarde 125
9052 Zwijnaarde
Belgium

Acknowledgements

This dissertation is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722734.

Contents

Preface

"If you can say 'I have learned' and 'I have loved',
you will also be able to say 'I have been happy'."

Arthur C. Clarke and Gentry Lee, "Rama II"

It was a beautiful summer's night in Prague when I was walking to my hotel, returning from a lavish banquet held in an old monastery. I was a first-year PhD student attending my first scientific conference. Accompanying me from the banquet was a friendly and enthusiastic French researcher. As we walked, he explained to me the idea behind something called "imprecise probabilities". I remember being confused, not entirely convinced, and yet thoroughly interested. At the time I found myself without a proper research project, although I had been reading up on Markov chains as a potential subject. The next day I suggested to my then-supervisor that perhaps I could combine these two subjects, working on Markov chains in the context of imprecise probabilities. She warned me that it would be technically challenging, but said that she was open to the idea. I vividly remember that on the return trip from the conference, my mind was already swirling with ideas. It felt like I had been handed a part of a map that read, in big bold letters, "HERE BE DRAGONS". How could I possibly resist?

Fast-forward seven years, and essentially none of those early ideas have worked out. However, I am now slightly less confused, much more convinced, and still thoroughly interested. Moreover, these previously unexplored lands turned out to be fertile ones, and I have found results that I could not even have conceived of when I set out on my journey. The research that I have performed these past years has culminated in the dissertation that You are now reading. In it, I develop and present my foundational work towards a theory of continuous-time Markov chains using imprecise probabilities.

I am greatly indebted to many people, without whom this dissertation would not exist. First and foremost, I want to thank my supervisors. The winding path that I took to reach this point has caused me to work at two universities, in three research groups, and under four supervisors. All of this has contributed to and enriched my experience, and so I would like to thank all those involved in chronological order.

To start, I want to thank Linda van der Gaag, who first hired me as a PhD student in the Decision Support Systems group at Utrecht University. I especially want to thank her for letting me stay on, when I later found myself without a project, and for allowing me the freedom to find my own research interests. Later, when she had become head of the department, I switched positions to the Algorithmic Data Analysis group—still at Utrecht University—where I continued my work under the supervision of Arno Siebes. I remain ever grateful to both Linda and Arno for enabling this change.

Next, I want to thank Arno, who abided my research that had relatively little to do with algorithms, and even less with data analysis. Despite my work being quite far removed from his own research, Arno has remained ever-curious, and always concerned with my professional development as a young researcher. During these years I had struck up a collaboration with researchers from Ghent University—who had independently started working on continuous-time imprecise-Markov chains—and I had learned that a position had opened up in Ghent on the subject we were collaborating on. So, when I found myself at the end of my contract with Utrecht University—but very much not at the end of my research—Arno didn't hesitate to write a letter of recommendation and to facilitate me in continuing my research at Ghent University. I remain ever thankful for this gracious act of selflessness.

At Ghent University, I continued my work at what is now called the Foundations Lab for Imprecise Probabilities, under the joint supervision of Gert de Cooman and Jasper De Bock. The position in question was that of an Early Stage Researcher on an EU Horizon 2020 project called UTOPIAE (*Uncertainty Treatment and Optimisation in Aerospace Engineering*), which would allow me to continue my theoretical research while exposing me to practically relevant applications and international collaborations. My thanks go out to Gert, for the warmth with which he welcomed me in his research group, and for his many words of wisdom. Many of my half-joking inquiries would be answered by him with insights of unanticipated depth. It has been an honour to work under one of the progenitors of the modern field of imprecise probabilities.

And last, but certainly not least, I want to thank Jasper. It was with him that I originally started the collaboration—when I was still based in Utrecht—that would determine this significant part of my research

career. And it was he who properly inaugurated me into the world of imprecise probabilities. Certainly many of the crucial ideas contained in this dissertation can be traced back to him. It is only fitting, then, that he would become my daily supervisor when I later moved to Ghent. In this capacity, he was always available to assist, critique, and provide critical insights. It was not rarely that he would solve, on the spot, problems that had stumped me for days. Ceaselessly striving for scientific quality, mathematical rigour, and didactic clarity, working with him has been enriching, sometimes infuriatingly frustrating, and always immensely stimulating. I owe him many thanks.

Next, I would like to express my sincere gratitude to the members of the Examination Board of this dissertation, for taking on the task of evaluating its scientific merit, and for their valuable feedback and suggestions. Any remaining errors in this work are entirely my own. I would also like to thank them for giving me the opportunity to present and defend my work in both our first meeting, and during the public defense of this dissertation.

My experience has also been greatly shaped by the many colleagues and co-workers that I've had the pleasure to meet over the years. Special thanks go out to Ad Feelders, who first took me under his wing when I was still an eager Master student. I remember barging into his office for the first time, informing him that I was interested in machine learning, and whether he knew of any good datasets to test some algorithms? He dryly suggested that perhaps I should just enrol in his courses. I ended up taking both of them, and Ad would wind up supervising two of my research projects, one of which being my Master thesis project. It was also Ad who introduced me to Linda. We would later become colleagues when I started as a PhD student.

I also feel that I should thank—by name—Sébastien Destercke, that French researcher who, all those years ago in Prague, enthused me with the concept of imprecise probabilities. It would not be the last conference at which we met.

Slightly less wordily—but certainly not with less affection—I want to thank all my other colleagues with whom I've shared my offices and lunches over the years. In an attempt at chronological ordering, my thanks go out to Merel, Steven, Arnoud, Janneke, Krzysztof, Hans, Lennart, Silja, Arthur, Stavros, Alexander, Meizhu, Natan, Simon, Michiel, Alain, Arne, and Floris. I particularly want to thank Stavros, Alexander, and Natan for our many interesting discussions and collaborations on the subject of imprecise-Markov chains. And I want to thank Krzysztof, for our friendship that developed first inside, and shortly after also outside, the office.

As part of my employment on the UTOPIAE project, I have spent three months on a research secondment to Strathclyde University in

Glasgow. My thanks go out to Massimiliano Vasile for hosting me during that period, and to Cristian Greco for our collaboration during that time. I also want to thank Frank Coolen for hosting me at Durham University for a short time, where I explored possible applications to reliability theory together with Daniel Krpelík. I want to thank Matthias Troffaes and Henna Bains, also of Durham University, for our collaborative work. I also feel compelled to mention my fellow Early Stage Researchers of the UTOPIAE project. It would not have been the same without them. My thanks go out to Gianluca, Cristian, Giulio, João, Elisa, Anabel, Zénó, Danda, Tatha, Bárbara, Margarita, Lorenzo, Dani, and Christian.

On a more personal note, I want to thank all the friends that I've made over the years. I have, over time, lost sight of some of them, which makes me hesitant to explicitly list their names. I am nevertheless immensely thankful for their impact on my life. I do, however, want to especially mention Frank, Luke, and Yvonne, who are some of my oldest and closest friends. My thanks also go out to Yvonne—again—and Ramison, for being godparents to my daughter.

Finally, I want to mention my family. My deepest gratitude goes out to my parents, Jeroen and Lea, who have always supported me in my endeavours. I know that it cannot have been easy. My thanks also go out to my younger brother Roeland, to my grandparents, extended family, and in-laws. They all helped shape my path. My special thanks go out to my parents—once more—and to my parents-in-law, Engelie and Pieter, for their practical support in recent times.

I want to thank my daughter, Emma, who has greatly enriched my life in the past two years. Her joyous presence has been a wonderful addition to our family, and fatherhood has been a marvellous experience. I cannot help but wonder if she will one day read these words, many years from now, curious what all the fuss was about.

Finally, I want to thank Maartje, my wife and very best friend. I most certainly could not have completed this dissertation without her. Mathematics is, I think, an inherently introspective, and at times solitary, occupation—yet she has made it slightly less so. She has helped me through the deepest valleys of motivation and witnessed my exuberance at breakthroughs and new discoveries. Throughout, she has always selflessly created the conditions that enabled me to finish this project. My dearest Maartje, I love you very much.

Tilburg, April 2021

Summary

Imprecise-Markov chains are mathematical models that generalise Markov chains using the theory of *imprecise probabilities*, which allows them to be used when (numerical) parameters are only partially specified and/or when assumptions like Markovianity may be unwarranted. Inferences computed with imprecise-Markov chains can be interpreted as being robust bounds on traditional inferences of interest, in that they are bounds with respect to all the possible variation that is implicit in the underspecification of these parameters and structural properties. In this dissertation, we develop the foundations for a theory of *continuous-time imprecise-Markov chains*: we discuss their definition, parameterisation, and interpretation; investigate many of their structural properties; develop basic inference algorithms for a large general class of problems; and prove a connection to existing work that allows us to obtain—essentially for free—many specialised and more efficient inference algorithms that were previously developed for *discrete-time* imprecise-Markov chains. In the remainder of this summary, we will further explain the concepts mentioned above, and sketch our main results and the strategy used to obtain them.

The basis of our theory is a formalisation of general *stochastic processes*; a stochastic process is a mathematical model that describes the behaviour of some dynamical system of interest as the state of this system evolves over time, in a manner that is uncertain. In particular, a stochastic process describes this uncertainty with a probability distribution over the realisations of the underlying system. Our formalisation is based on *full conditional probabilities* and *coherence*. This is slightly different from the more typical measure-theoretic framework, but has the advantage that it (i) endows the theory with a clear subjectivist and behavioural interpretation, and (ii) does not gratuitously impose technical assumptions that we do not need for the results in this work. We spend some effort on showing that at its core—despite the philosophical differences—our characterisation essentially agrees with what would be obtained using a measure-theoretic formalisation, at least for the problems that we are considering here.

A *Markov chain*, then, is a specific type of stochastic process: it assumes that, given the "current" state of the system, the uncertainty about the system's future behaviour does not depend on its historical behaviour, that is, what happened before the "current" point in time. This assumption is known as the *Markov property*, and it is this crucial property that leads to Markov chains being exceptionally tractable, while still being powerful enough to represent interesting dynamical behaviour; consequently, Markov chains have become very popular models throughout science and engineering. Due to the varying nature of the many applications of Markov chains, we may distinguish between several different types that have been developed over the years. For instance, in this dissertation we only consider systems with *finitely* many states. Another classification is based on the nature of the "time domain" along which the dynamical system evolves. *Discrete-time* Markov chains describe systems whose evolution occurs in discrete steps: at every point in time, there is an unambiguous "next" time point. Conversely, *continuous-time* Markov chains deal with systems for which the evolution occurs along a continuous time domain.

Imprecise-Markov chains are a generalisation of Markov chains that is based on the theory of imprecise probabilities. These models can be used when the (numerical) parameters of a Markov chain can only be partially specified—for example, when one only knows that these parameters should lie in a specific range that is "plausible"—or when assumptions like the Markov property do not apply. In this work we adopt the "sensitivity analysis" interpretation of imprecise probabilistic models. This means that we view an imprecise-Markov chain as a *set* of stochastic processes, all of whose members are in a specific sense *consistent* with what *is* known; for example, we might take it to be the set of all Markov chains whose numerical parameters lie in the "plausible" range of values. However, we might additionally include more complicated models in this set; for instance, general stochastic processes that do not satisfy the Markov property. Inferences computed from these models are formalised as *lower* and *upper* expectations; these are tight lower and upper bounds on some inference of interest, taken with respect to the entire set of stochastic processes that constitutes the imprecise-Markov chain.

Over the past decades, there has been a lot of development in the field of *discrete-time imprecise-Markov chains*, and this theory has matured to the point where many inference problems that can be analysed and tractably computed for (traditional) discrete-time Markov chains, can also be analysed and tractably computed for discrete-time *imprecise*-Markov chains. However, work on *continuous-time* imprecise-Markov chains has only begun in earnest much more recently.

It is our aim with this dissertation to develop the theory of

continuous-time imprecise-Markov chains to a point that is somewhat closer to the state-of-the-art of the discrete-time theory. In particular, it is our aim to develop the formal definitions, interpretations, basic qualitative properties, and elementary inference algorithms that are required to put future work in this field on an at least somewhat solid footing; and to provide for such future work the tools and techniques that can be used to attack more complicated problems. The thesis that we develop along the way is, essentially, that many of the fundamental results and discoveries from the discrete-time setting, carry over analogously to continuous-time models.

In order to establish this connection to the existing literature, we spend some effort on introducing and discussing crucial elements of the established theory of discrete-time (imprecise-)Markov chains, in terms of our current formalisation of stochastic processes. Of importance is the parameterisation of discrete-time Markov chains using (families of) transition matrices; a *transition matrix* is a row-stochastic matrix that, in this context, describes the probabilities that the system moves from any given state to any (other) state in a single time step. We subsequently discuss the parameterisation of discrete-time *imprecise*-Markov chains using (families of) *sets* of transition matrices. We discuss crucial properties of the lower and upper expectations corresponding to these models, and illustrate in particular that they satisfy (i) an *imprecise-Markov property* and (ii) a *law of iterated (lower/upper) expectation*. The imprecise-Markov property states, essentially, that although an imprecise-Markov chain can be a fairly complicated set of stochastic processes that need not be Markovian, the corresponding lower and upper expectation *are* history-independent in a manner that is analogous to the traditional Markov property. It is this property that motivates the terminology *imprecise-Markov chain*, and which also makes these imprecise probabilistic models fairly tractable. The law of iterated (lower/upper) expectation is, essentially, a decomposition property that lies at the heart of many efficient inference algorithms that have been developed in the literature. We discuss *lower transition operators* as being non-linear generalisations of transition matrices, and show that they are dual representations of sets of transition matrices. We illustrate how these lower transition operators provide an alternative characterisation of the lower expectations for discrete-time imprecise-Markov chains.

Having established some of the crucial properties of discrete-time imprecise-Markov chains, we then start the work on developing the machinery that we need to describe the more technically complicated *continuous-time* models. One of the main tools to do this are *transition rate matrices*; these are the continuous-time counterpart to the transition matrices that are used for discrete-time models, and they describe, intuitively, the speeds at which the underlying system moves between states. We also discuss multi-index families of transition matrices corresponding to continuous-time stochastic processes, and develop some machinery to manipulate such families. We introduce the notion of *outer partial derivatives* for such (families of) transition matrices. These are essentially set-valued generalisations of traditional derivatives, which may exist—in particular, be non-empty sets—even when the traditional derivatives do not. We characterise a type of continuous-time stochastic process, which we call *well-behaved*, for which such outer partial derivatives are always non-empty and compact sets of transition rate matrices. We illustrate how, in general, stochastic processes can behave rather pathologically, and that such behaviour is prevented by imposing this condition of well-behavedness; we spend the remainder of the dissertation focussing on well-behaved stochastic processes.

Next, we show how our formalisation of continuous-time Markov chains, which is in terms of full conditional probabilities and coherence, agrees in principle with more traditional characterisations. This is also notably the case for continuous-time Markov chains that are *(time-)homogeneous*, which essentially means that the description of the system's uncertain behaviour is the same at all points in time. Formally, we show that such processes are uniquely characterised by the specification of an *initial distribution*—a probability distribution specifying the uncertainty about the state in which the system starts—and a single transition rate matrix, and that the family of transition matrices corresponding to such a process is given by the *semigroup of transition matrices* that is *generated* by this transition rate matrix.

With these tools in hand, we then present our formalisation of continuous-time imprecise-Markov chains. Or rather, we introduce three distinct definitions: they are all parameterised using (i) a set of possible initial distributions and (ii) a set of transition rate matrices; and they are all sets of well-behaved continuous-time stochastic processes that are *consistent* with these parameters. This means that the initial distributions of their elements are contained in the parameterising set of possible initial distributions, and the outer partial derivatives of their families of transition matrices are contained in the parameterising set of transition rate matrices. However, the definitions differ in terms of the structural properties that we impose on their elements. Specifically, the three different versions correspond to (i) a set of time-homogeneous Markov chains, (ii) a set of—not necessarily time-homogeneous—Markov chains, and (iii) a set of general—not necessarily Markovian or time-homogeneous—stochastic processes.

We investigate structural and qualitative properties of these different sets of processes, and of their induced sets of corresponding transition matrices. We introduce the corresponding lower and upper expectations for these models, and investigate their properties. We derive sufficient conditions for the lower and upper expectations of all models to satisfy an *imprecise-Markov property*, which again motivates the terminology that they are *imprecise-Markov chains*. We argue how the most conceptually simple of our three definitions—the model containing only time-homogeneous Markov chains—is, unfortunately, the most difficult one to work with in practice. Intuitively, the issue is that the optimisation problem involved in the computation of its lower and upper expectations cannot really be simplified and made tractable, due to the strong constraint that the optimisation has to be taken over only time-homogeneous Markov chains; perhaps surprisingly, the additional degrees of freedom allowed by the other two models make them easier to work with. In fact, it is only for the conceptually most complicated model—the set of *all* well-behaved stochastic processes that are consistent with the parameterising sets—that we derive a *law of iterated (lower/upper) expectation*. As with discrete-time imprecise-Markov chains, this crucial property paves the way for efficient inference algorithms. We introduce lower transition operators—which we previously used in the discrete-time setting—also for continuous-time imprecise-Markov chains. We show how and under which conditions they provide an alternative characterisation of lower expectations also for continuous-time models.

We introduce *lower transition rate operators*, which are non-linear generalisations of transition rate matrices, and we show that they are dual to sets of transition rate matrices. We discuss how to evaluate such lower transition rate operators numerically, and in particular provide an efficient algorithm that can be used when this operator is the dual of a metric ball around a given transition rate matrix; this setting occurs naturally in sensitivity analysis contexts. We show how a given lower transition rate operator *generates* a *semigroup of lower transition operators*—in analogy with the semigroup of transition matrices generated by a transition rate matrix—and we show how and when this semigroup of lower transition operators coincides with the lower transition operators corresponding to an imprecise-Markov chain. We also present an algorithm to evaluate the elements of this semigroup numerically. This leads to a first practical algorithm for computing lower (and upper) expectations for imprecise-Markov chains: since we already established that lower transition operators form an alternative characterisation of the lower expectation of functions that depend on the state of the system at a single point in time, evaluating the elements of this semigroup numerically therefore amounts to computing the lower expectation of such functions. We combine this computational method with the law of iterated lower expectation, to derive a second, more general algorithm that can be used to compute arbitrary lower (and upper) expectations of functions that depend on the state of the system at arbitrarily, but finitely, many time points. We show that this algorithm works for the most imprecise of our definitions of continuous-time imprecise-Markov chains; and moreover, that it does *not* work, in general, when working with sets of (potentially non-homogeneous) Markov chains. However, this algorithm can still be used to compute conservative bounds on the lower (and upper) expectations of such sets.

We finish by establishing a strong connection between continuous-time and discrete-time imprecise-Markov chains. In particular, we show how the latter can be obtained as *restrictions* of the former, by taking into account only the time points in a given discrete time domain. We show how the lower (and upper) expectations for the original continuous-time imprecise-Markov chain coincide with the lower (and upper) expectations for this induced discrete-time imprecise-Markov chain. We illustrate the value of this result, by using it to translate to the continuous-time setting a very efficient inference algorithm for a particular class of functions, that has previously been derived in the literature for discrete-time imprecise-Markov chains, *without* having to explicitly re-derive it.

We conclude that the most promising of our three different definitions of a continuous-time imprecise-Markov chain is conceptually the most complicated one: the set of *all* continuous-time stochastic processes that are well-behaved and consistent with a given set of initial distributions and a given set of transition rate matrices. In general, this set will contain processes that are non-homogeneous and non-Markovian, and that are—individually—computationally and analytically very difficult to work with. Nevertheless, under some relatively mild assumptions on its parameters, the lower and upper expectations for this set satisfy an imprecise-Markov property; and it is for this set that we have been able to derive the most powerful results. Notably, many of the crucial results that make the analysis and computations for this model tractable, in general do not hold for the other two definitions that we considered: the set containing only time-homogeneous Markov chains, and the set containing only (not necessarily time-homogeneous) Markov chains. Specifically, it is only for our most imprecise model—which contains possibly non-homogeneous and non-Markovian processes—that we were able to derive a law of iterated lower expectation; a general inference algorithm for a large class of problems; and a connection with discrete-time imprecise-Markov chains that allows us to re-use, with minimal effort, many efficient and specialised algorithms that have previously been developed in the literature.

Samenvatting

Dutch Summary

Imprecieze Markovketens zijn wiskundige modellen die Markovketens veralgemenen door middel van de theorie van *imprecieze waarschijnlijkheden*. Hierdoor kunnen ze gebruikt worden wanneer (numerieke) parameters slechts gedeeltelijk gespecificeerd zijn en/of wanneer het aannemen van een Markoviaans karakter ongegrond zou zijn. Gevolgtrekkingen die verkregen worden met imprecieze Markovketens kunnen worden geïnterpreteerd als robuuste grenzen op traditionele gevolgtrekkingen waarin men geïnteresseerd zou zijn, in die zin dat ze rekening houden met alle variatie die impliciet mogelijk blijft door het slechts deels specificeren van parameters en structurele modeleigenschappen. In dit proefschrift ontwikkelen we een theorie van *imprecieze Markovketens in continue tijd*: we bespreken hun definitie, parameterisering en interpretatie; onderzoeken hun structurele eigenschappen; ontwikkelen eenvoudige algoritmen voor het uitrekenen van een grote generieke klasse van inferentieproblemen; en bewijzen een verband met bestaand werk waarmee we—zonder extra moeite—beschikking krijgen over veel gespecialiseerde en meer efficiënte algoritmen die in het verleden ontwikkeld zijn voor imprecieze Markovketens in *discrete tijd*. In de rest van deze samenvatting lichten we de bovengenoemde concepten verder toe, en schetsen we wat onze hoofdresultaten zijn en hoe we die verkregen hebben.

De basis van onze theorie is een formele benadering van generieke *stochastische processen*: een stochastisch proces is een wiskundig model dat het gedrag beschrijft van een dynamisch systeem waarvan de tijdsafhankelijke toestandsontwikkeling onzeker is. In het bijzonder beschrijft een stochastisch proces deze onzekerheid met een waarschijnlijkheidsverdeling over de mogelijke realisaties van dit systeem. Onze formele benadering is gestoeld op *complete conditionele waarschijnlijkheden* en *coherentie*. Deze aanpak verschilt enigzins van de meer gangbare maattheoretische benadering, maar heeft als voordeel dat het (i) onze theorie een duidelijke subjectivistische en gedragsgerichte interpretatie geeft, en (ii) dat het geen gratuite technische aannames introduceert die onnodig zijn voor het verkrijgen van onze resultaten. We besteden wat aandacht aan het laten zien dat—ondanks de filosofische verschillen—onze beschrijving in de kern overeenkomt met wat men zou verkrijgen met een maattheoretische aanpak, althans voor de problemen die we hier beschouwen.

Een *Markovketen* is een specifiek soort stochastic proces: het maakt de aanname dat, gegeven de "huidige" toestand van het systeem, de onzekerheid over het toekomstige gedrag van het systeem niet afhangt van het historische gedrag, dat wil zeggen, van wat er gebeurd is vóór het "huidige" tijdstip. Deze aanname staat bekend als de *Markoveigenschap*, en het is deze cruciale eigenschap die maakt dat Markovketens bijzonder handelbaar zijn, en toch krachtig genoeg om interessante dynamica te beschrijven; hierdoor zijn Markovketens zeer populaire modellen geworden in de (ingenieurs)wetenschappen. Vanwege de verschillen tussen de vele mogelijke toepassingen van Markovketens, kunnen we een onderscheid maken tussen verschillende soorten, die door de jaren heen zijn ontwikkeld. Bijvoorbeeld, in dit proefschrift beperken we ons tot systemen met een *eindig* aantal mogelijke toestanden. Een ander onderscheid is gebaseerd op het soort *tijdsdomein* waarover het systeem zich evolueert. Markovketens in *discrete tijd* beschrijven systemen waarvan de evolutie plaatsvindt met discrete stappen: op elk moment is er een welbepaald "volgend" tijdstip. Daarentegen beschrijven Markovketens in *continue tijd* systemen waarvan de evolutie plaatsvindt over een continu tijdsdomein.

Imprecieze Markovketens zijn een veralgemening van Markovketens die zich baseert op de theorie van imprecieze waarschijnlijkheden. Deze modellen kunnen gebruikt worden wanneer de (numerieke) parameters van een Markovketen slechts gedeeltelijk bepaald kunnen worden—bijvoorbeeld als men slechts weet dat deze parameters in een gegeven bereik liggen, dat "aannemelijk" wordt geacht—of wanneer aannames zoals de Markoveigenschap niet van toepassing zijn. In dit werk hanteren we de "sensitiviteitsanalyse"-interpretatie van imprecieze-waarschijnlijkheidsmodellen. Dit wil zeggen dat we een imprecieze Markovketen zien als een *verzameling* van stochastische processen, waarvan de elementen allemaal op een welbepaalde manier *consistent* zijn met wat *wel* geweten is; we zouden bijvoorbeeld kunnen spreken over de verzameling van alle Markovketens waarvan de numerieke parameters in een gegeven bereik van "aannemelijke" waarden liggen. Echter, we zouden ook meer gecompliceerde modellen in zo een verzameling kunnen opnemen; bijvoorbeeld, algemenere stochastische processen die niet over de Markoveigenschap beschikken. Gevolgtrekkingen die men met zo een model kan verkrijgen, ook wel *inferenties* genoemd, worden geformaliseerd als *onder-* en *bovenverwachtingswaarden*; dit zijn nauwe onder- en bovengrenzen op traditionele inferenties—probabilistische verwachtingswaarden—met betrekking tot de volledige verzameling van stochastische processen waaruit de imprecieze Markovketen bestaat.

Over de afgelopen decennia heeft er een hoop ontwikkeling plaatsgevonden op het gebied van *imprecieze Markovketens in discrete tijd*, en deze theorie heeft het punt bereikt waarop veel inferentieproblemen die geanalyseerd en efficient uitgerekend kunnen worden met (traditionele) Markovketens in discrete tijd, ook geanalyseerd en efficient uitgerekend kunnen worden met *imprecieze* Markovketens in discrete tijd. Echter, het serieus bestuderen van imprecieze Markovketens in *continue tijd* is pas veel recenter begonnen.

Het is ons doel met dit proefschrift om de theorie van imprecieze Markovketens in continue tijd te ontwikkelen tot een punt dat iets dichter ligt bij de huidige stand van zaken van de theorie in discrete tijd. In het bijzonder is het ons doel om de formele definities, interpretaties, basiseigenschappen, en elementaire inferentiealgoritmen te ontwikkelen die nodig zijn om toekomstig werk in dit veld een solide houvast te bieden; en om voor zulk toekomstig werk de gereedschappen en technieken te voorzien die nodig zijn om ingewikkeldere problemen op te lossen. De thesis die we gaandeweg ontwikkelen is, in de kern, dat veel van de fundamentele resultaten en ontdekkingen uit de context in discrete tijd, analoog over te dragen zijn naar modellen in continue tijd.

Om de koppeling met de bestaande literatuur te maken, besteden we wat aandacht aan het introduceren en bespreken van cruciale onderdelen van de gevestigde theorie van (imprecieze) Markovketens in discrete tijd, in termen van onze huidige formalisering van stochastische processen. Van speciaal belang is de parameterisering van Markovketens in discrete tijd door (families van) transitiematrices; een *transitiematrix* is een rij-stochastische matrix die, in deze context, de waarschijnlijkheden beschrijft voor het systeem om vanuit een gegeven toestand naar eender welke (andere) toestand over te gaan, in een enkele tijdstap. Vervolgens bespreken we de parameterisering van *imprecieze* Markovketens in discrete tijd door (families van) *verzamelingen* transitiematrices. We bespreken cruciale eigenschappen van de onder- en bovenverwachtingswaarden die bij deze modellen horen, en illustreren in het bijzonder dat ze voldoen aan (i) een *imprecieze Markoveigenschap* en (ii) een *wet van herhaalde onder- en bovenverwachtingswaarden*. De imprecieze Markoveigenschap stelt, in essentie, dat hoewel een imprecieze Markovketen een vrij complexe verzameling kan zijn die bestaat uit niet noodzakelijk Markoviaanse stochastische processen, de bijbehorende onder- en bovenverwachtingswaarden *wel* geschiedenisonafhankelijk zijn op een manier die volledig analoog is aan de traditionele Markoveigenschap. Het is deze eigenschap die de terminologie *impre-*

cieze Markovketen motiveert, en het is deze eigenschap die zulke modellen relatief hanteerbaar maakt in hun gebruik. De wet van herhaalde onder- en bovenverwachtingswaarden is, in de kern, een decompositie-eigenschap die ten grondslag ligt aan vele efficiënte inferentiealgoritmen die zijn ontwikkeld in de literatuur. We beschrijven *ondertransitieoperatoren* als niet-lineaire veralgemeningen van transitiematrices, en we laten zien dat dit duale representaties zijn van verzamelingen transitiematrices. We illustreren hoe deze ondertransitieoperatoren een alternatieve beschrijving geven voor de onderverwachtingswaarden van imprecieze Markovketens in discrete tijd.

Nadat we een aantal cruciale eigenschappen van imprecieze Markovketens in discrete tijd hebben vastgesteld, beginnen we met de ontwikkeling van de wiskundige instrumenten die we nodig hebben om de technisch ingewikkeldere modellen in *continue tijd* te kunnen beschrijven. Een van de belangrijkste gereedschappen om dit te doen zijn *transitietempomatrices*; dit zijn de tegenhangers in continue tijd van de transitiematrices gebruikt voor modellen in discrete tijd, en in intuïtieve zin beschrijven ze het tempo waarmee het onderliggende systeem zich tussen toestanden beweegt. Ook bespreken we meervoudig geïndexeerde families van transitiematrices die bij stochastische processen in continue tijd horen, en ontwikkelen we een aantal technieken om dit soort families te manipuleren. We introduceren *partiële-afgeleideverzamelingen* voor dit soort (families van) transitiematrices. Dit zijn in essentie veralgemeningen van traditionele afgeleiden, die kunnen bestaan—in het bijzonder, die niet-lege verzamelingen kunnen zijn—ook als de traditionele afgeleiden niet bestaan. We karakteriseren een type stochastisch proces in continue tijd, waarvan we zeggen dat het zich *goed gedraagt*, waarvoor zulke partiële-afgeleideverzamelingen altijd niet-lege en compacte verzamelingen transitietempomatrices zijn. We illustreren hoe algemene stochastische processen behoorlijk pathologisch gedrag kunnen vertonen, en dat dit soort gedrag wordt vermeden wanneer we opleggen dat een proces zich goed moet gedragen; we beperken ons in de rest van het proefschrift tot processen die zich goed gedragen.

Hierna laten we zien dat onze formele benadering van Markovketens in continue tijd, die in termen is van complete conditionele waarschijnlijkheden en coherentie, in de kern overeenkomt met meer gangbare karakteriseringen. Dit is ook in het bijzonder het geval voor Markovketens in continue tijd die *homogeen* zijn, wat in essentie wil zeggen dat de beschrijving van de onzekerheid over het gedrag van het onderliggende systeem op ieder moment hetzelfde is. We tonen formeel aan dat dit soort processen uniek bepaald worden door het specificeren van een *beginverdeling*—een waarschijnlijkheidsverdeling die de onzekerheid representeert over de toestand waarin het systeem begint—en een enkele transitietempomatrix, en dat de familie van transitiematrices horende bij zo een proces overeenkomt met de *halfgroep van transitiematrices* die wordt *voortgebracht* door deze transitietempomatrix.

Met deze gereedschappen kunnen we dan eindelijk onze formele benadering van imprecieze Markovketens in continue tijd introduceren. Of beter gezegd, we introduceren drie verschillende definities: ze worden allemaal geparametriseerd door (i) een verzameling van mogelijke beginverdelingen en (ii) een verzameling transitietempomatrices; en het zijn allemaal verzamelingen van stochastische processen in continue tijd die zich goed gedragen en die *consistent* zijn met deze parameterverzamelingen. Dit wil zeggen dat de beginverdelingen van hun elementen bevat zijn in de parametriserende verzameling van mogelijke beginverdelingen, en dat de partiële-afgeleideverzamelingen van hun families van transitiematrices bevat zijn in de parametriserende verzameling transitietempomatrices. De definities verschillen echter in de structurele eigenschappen die we opleggen aan hun elementen. De drie definities zijn (i) een verzameling die bestaat uit homogene Markovketens, (ii) een verzameling van—niet noodzakelijk homogene—Markovketens en (iii) een verzameling van generieke—niet noodzakelijk homogene noch Markoviaanse—stochastische processen.

We bestuderen structurele en kwalitatieve eigenschappen van deze verschillende verzamelingen processen, en van hun geïnduceerde verzamelingen transitiematrices. We introduceren de bijbehorende onder- en bovenverwachtingswaarden voor deze modellen, en onderzoeken hun eigenschappen. We vinden voldoende voorwaarden voor de onder- en bovenverwachtingswaarden van deze modellen om te voldoen aan een *imprecieze Markoveigenschap*, wat opnieuw de terminologie motiveert dat dit *imprecieze Markovketens* zijn. We beargumenteren waarom de conceptueel eenvoudigste van onze drie definities—het model dat uitsluitend bestaat uit Markovketens die homogeen zijn—helaas het lastigste model is om mee te werken. De intuïtieve reden is dat het optimalisatieprobleem dat komt kijken bij het bepalen van de onder- en bovenverwachtingswaarden niet eenvoudiger (en handelbaarder) gemaakt kan worden, vanwege de sterke beperking dat deze optimalisatie moet plaatsvinden over een verzameling die uitsluitend bestaat uit homogene Markovketens; het is misschien verrassend, maar de extra vrijheidsgraden die de andere twee modellen toestaan, maken ze uiteindelijk makkelijker in hun gebruik. Sterker nog, het is alleen voor ons conceptueel meest ingewikkelde model—de verzameling van *alle* stochastische processen die zich goed gedragen en consistent zijn met de parametriserende verzamelingen—dat we in staat zijn een *wet van herhaalde onder- en bovenverwachtingswaarden* te bepalen. Net als voor imprecieze Markovketens in discrete tijd, ligt deze wet ook hier ten grondslag aan de ontwikkeling van efficiënte inferentiealgoritmen. We introduceren ondertransitieoperatoren—die we eerder gebruikten voor modellen in discrete tijd—nu ook voor imprecieze Markovketens in continue tijd. We laten zien hoe, en onder welke voorwaarden, ze een alternatieve representatie vormen van de onderverwachtingswaarden van modellen in continue tijd.

We introduceren *ondertransitietempo-operatoren*, wat niet-lineaire veralgemeningen zijn van transitietempomatrices, en we laten zien dat ze duale representaties zijn van verzamelingen transitietempomatrices. We bespreken hoe ondertransitietempo-operatoren numeriek geëvalueerd kunnen worden, en voorzien in het bijzonder een efficiënt algoritme dat gebruikt kan worden wanneer deze operator de duale is van een metrische bal rond een gegeven transitietempomatrix; dit geval dient zich natuurlijkerwijs aan in de context van sensitiveitsanalyse. We tonen hoe een ondertransitietempo-operator een *halfgroep van ondertransitieoperatoren voortbrengt*—naar analogie met de halfgroep van transitiematrices die wordt voortgebracht door een transitietempomatrix—en we tonen hoe en onder welke voorwaarden deze halfgroep van ondertransitieoperatoren samenvalt met de ondertransitieoperatoren horende bij een imprecieze Markovketen in continue tijd. We voorzien een algoritme waarmee de elementen van deze halfgroep numeriek geëvalueerd kunnen worden. Dit leidt tot een eerste praktisch bruikbaar algoritme voor het berekenen van (boven- en) onderverwachtingswaarden voor imprecieze Markovketens in continue tijd: aangezien we al hadden vastgesteld dat de ondertransitieoperatoren een alternatieve representatie zijn van de onderverwachtingswaarden van functies die van de toestand op één enkel tijdstip afhangen, is het numeriek evalueren van elementen van deze halfgroep eigenlijk hetzelfde als het uitrekenen van de onderverwachtingswaarden van zulke functies. We combineren deze rekenmethode met de wet van herhaalde onderverwachtingswaarden, zodat we een tweede, algemener algoritme krijgen voor het uitrekenen van onder- en bovenverwachtingswaarden van functies die van de toestand op een willekeurig, maar eindig, aantal tijdstippen afhangen. We tonen dat dit algoritme werkt voor onze meest imprecieze definitie van imprecieze Markovketens in continue tijd; en in het bijzonder, dat het *niet* werkt, in het algemeen, voor verzamelingen die slechts bestaan uit (mogelijk niet-homogene) Markovketens. Echter, dit algoritme kan wel gebruikt worden om conservatieve grenzen op de onder- en bovenverwachtingswaarden voor deze andere verzamelingen uit te rekenen.

We eindigen met het beschrijven van een sterk verband tussen imprecieze Markovketens in continue en discrete tijd. In het bijzonder tonen we hoe modellen in discrete tijd verkregen kunnen worden als *beperkingen* van modellen in continue tijd, door slechts rekening te houden met tijdpunten in een gegeven discreet tijdsdomein. We laten zien hoe de onder- en bovenverwachtingswaarden voor de oorspronkelijke imprecieze Markovketen in continue tijd samenvallen met de onder- en bovenverwachtingswaarden van zo een geïnduceerde imprecieze Markovketen in discrete tijd. We illustreren het nut van dit resultaat, door het te gebruiken om een vertaling te maken naar continue tijd, van een zeer efficiënt inferentiealgoritme voor een specifieke klasse functies, dat eerder ontwikkeld werd voor imprecieze Markovketens in discrete tijd, *zonder* dat we dit algoritme opnieuw expliciet hoeven af te leiden.

We concluderen dat de meest veelbelovende van onze drie verschillende definities van imprecieze Markovketens in continue tijd het conceptueel meest ingewikkelde model is: de verzameling van *alle* stochastische processen in continue tijd die zich goed gedragen en die consistent zijn met een gegeven verzameling beginverdelingen en een gegeven verzameling transitietempomatrices. Deze verzameling zal in het algemeen processen bevatten die niet homogeen noch Markoviaans zijn, en die—individueel—zeer moeilijk te analyseren en berekenen zijn. Echter, onder een aantal relatief milde aannames over de parameters, voldoen de onder- en bovenverwachtingswaarden van deze verzameling aan een imprecieze Markoveigenschap; en het is ook voor deze verzameling dat we onze meest krachtige resultaten hebben kunnen verkrijgen. In het bijzonder zijn veel van de cruciale resultaten die de analyse en het uitrekenen van inferenties voor dit model handelbaar maken, in het algemeen niet van toepassing voor onze andere twee definities: de verzameling die uitsluitend bestaat uit homogene Markovketens, en de verzameling die bestaat uit (niet noodzakelijk homogene) Markovketens. Het is slechts voor ons meest imprecieze model—de verzameling die ook niet-homogene en/of niet-Markoviaanse processen bevat—dat we een wet van herhaalde onder- en bovenverwachtingswaarden konden bepalen; een algemeen inferentiealgoritme konden afleiden voor een grote algemene klasse van inferentieproblemen; en een gelijkenis konden vaststellen met imprecieze Markovketens in discrete tijd die ons in staat stelt om, zonder verdere moeite, vele gespecialiseerde en efficiënte inferentiealgoritmen uit de bestaande literatuur over te nemen voor imprecieze Markovketens in continue tijd.

1

Introduction

"Hello, my friend. Stay awhile, and listen..."

Deckard Cain, Diablo

Imprecise-Markov chains [22,45,48,57,103] are mathematical models that generalise Markov chains [49, 55, 82] using the theory of imprecise probabilities [3, 114], which allows them to deal with partially specified parameters and weakened structural assumptions. These models can be used whenever obtaining a complete, precise specification of a traditional model is infeasible, and/or when Markovian assumptions are unwarranted or in doubt. Inferences from these models can be interpreted as being "robust", or "cautious", with respect to the variation implicit in the underspecification of the parameters.

Over the past several decades, the theory of *discrete-time* imprecise-Markov chains has matured significantly; both in terms of theoretical and foundational results, e.g. [19, 20, 22, 45, 47, 57, 69, 101], and in the development of efficient inference algorithms, e.g. [18, 21, 107]. However, substantial effort towards *continuous-time* imprecise-Markov chains has only started to be made much more recently [17, 61, 103].

In this dissertation we present our foundational work to develop a theory of continuous-time imprecise-Markov chains: we introduce and discuss their representation, interpretation, and basic qualitative properties; provide fundamental algorithmic solutions to elementary inference problems; and prove a correspondence to discrete-time imprecise-Markov chains with which we can leverage existing (algorithmic) solutions to more advanced inference problems.

1.1 Motivation and Overview of Results

A Markov chain [49,55,82] is a particular type of *stochastic process* [28]: a probabilistic model that describes the uncertain behaviour of some dynamical system of interest as this system evolves over time. In particular, a Markov chain is a stochastic process that satisfies the eponymous *Markov property*. This means that at any point in time, and given the system's "current" state, the model's description of the system's "future" behaviour does not depend on its "past" behaviour. These models are named after Andrey Markov [73], who realised that such models are the simplest generalisation of an independent-trials process that is powerful enough to represent interesting structural properties of the underlying dynamical system, while still being tractable enough to describe and use in practice [49].

Because there are many distinct types of underlying systems that might be modelled, Markov chains occur in various flavours: we may distinguish between *discrete-time* and *continuous-time* Markov chains; and may make a distinction based on the cardinality of the set of possible states. In this dissertation we restrict ourselves entirely to systems with *finitely* many states.

A discrete-time Markov chain, then, is a stochastic process that satisfies the Markov property, and for which time evolves in discrete steps: at every point in time there is a well-defined "next" time point. As time evolves in this manner, so does the state of the underlying system; these state changes are "uncertain", and a Markov chain describes this uncertainty using a probability distribution over the different possible realisations of the system. Since their inception, discrete-time Markov chains have become ubiquitous as a probabilistic model throughout science and engineering, with applications in fields like queueing theory [2,7], natural language analysis [74], DNA sequence decoding [11], Webpage ranking [84], information theory [99], and mathematical finance [91].

A continuous-time Markov chain, unsurprisingly, models systems with a continuous time domain; at every point in time t, we can consider a time point $t+\Delta$ that lies an arbitrary amount $\Delta \geq 0$ of time in the future. Conceptually, there is not much difference with discrete-time models: a continuous-time Markov chain specifies a probability distribution over the possible realisations of the system. However, due to the fundamental differences between these time domains, there are additional technical difficulties in representing and reasoning with continuous-time models. As with their discrete-time counterparts, these models have found applications in fields as disparate as queueing theory [2,7], mathematical finance [31,91,95], epidemiology [30,52,67], and system reliability analysis [6,36,116].

The Markov property can essentially be seen as a simplifying assumption that is imposed to obtain a tractable model. Another such assumption that is often encountered in this context is that of *time-homogeneity* [38, 49, 55, 82].[1] Roughly speaking, this means that the model's description of the system's uncertain behaviour is the same at all time points. Under these conditions, it turns out that Markov chains are exceptionally easy to parameterise. Regardless of the time domain, one always needs an *initial distribution*: a probability distribution that describes the uncertainty about the state in which the system starts. For discrete-time Markov chains, it suffices to additionally supply the *transition probabilities* between all pairs of states (see e.g. [82, Chapter 1]); essentially, for each state, one needs to provide the probabilities that the system moves from that state to any (other) state in a single time step. For continuous-time Markov chains, one instead needs to supply the *transition rates* between all pairs of states (see e.g. [82, Chapter 2]); these can essentially be interpreted as the speeds with which the system moves between these states.

One other—often implicit—assumption that is required to use these models in practice, is that one can provide the (numerical) values of these parameters exactly. Under the above conditions, a Markov chain is an analytically and computationally tractable model that can be used to answer probabilistic or statistical queries of interest about the underlying system. Throughout this dissertation, we will use the general terminology that the model is used to make (or compute) *inferences* about some quantity of interest. Such an inference is typically the probability of some event occurring—e.g. the probability $P(X_t = x)$ that the (uncertain) state X_t of the system at time t will be equal to the state x—or the conditional expectation $\mathbb{E}[f\,|\,C]$ of a function f whose value $f(\omega)$ depends on the (uncertain) realisation ω of the system, given the occurrence of some event C.

The theory that we develop in this dissertation is predicated on, and motivated by, situations where the above assumptions break down. For instance, a practitioner may find it difficult to supply numerically exact parameters, because they were estimated from few or un-

[1] There appears to be no consensus in the literature on how to call this property. Doob [28] refers to these models as having "stationary transition probabilities". The authors of [49, 55] make this assumption implicitly in their definition of a Markov chain; Howard [49] refers to models that do *not* satisfy it as "time-varying", and Kemeny and Snell [55] as "Markov processes". Norris [82], in contrast, uses the terms "Markov chain" and "Markov process" to distinguish between Markov models with, respectively, at most countably and uncountably many states; but makes this assumption implicitly in both definitions. Throughout this dissertation, we follow e.g. Grimmett and Stirzaker [38] in using the term "homogeneous", or "time-homogeneous" when we want to be explicit.

reliable data, or obtained from conflicting expert elicitation. The underlying system of interest may exhibit temporal correlations that are too complex to warrant assumptions of Markovianity. Parameter estimates that were obtained at different times—for example from multiple experiments—may differ too substantially to make time-homogeneity plausible. Under any or all of these more difficult conditions, can we still construct and employ models that remain analytically and computationally tractable? Crucially, if we need to introduce approximations to capture this more complicated structure, can we guarantee in some way that the inferences obtained from such models are in some sense "robust" against these deviations from the traditional assumptions?

These questions, it turns out, can be answered in the affirmative—with some caveats. Of course, it would be unreasonable to expect to obtain numerically precise and reliable answers, from a model whose parameters are (perhaps partially) unknown, and whose structural assumptions are expected or assumed to be wrong. However, we may aim for a more reasonable goal: to obtain the most informative answers that are consistent with what *is* known, while being conservative, or "cautious", with respect to those things that are *not* known. To achieve this goal, we turn to the theory of *imprecise probabilities* [3, 114].

In his seminal work [114], Walley introduced his theory of imprecise probabilities; a mathematical framework for representing and reasoning under uncertainty which generalises and subsumes traditional probability theory. The basis of this theory is a subjectivist and behavioural formalisation of the notion of uncertainty, which traces its roots to the work of Ramsey [86], De Finetti [24], and Williams [117]. We refer to [3] for an excellent introductory treatment and overview of work in the general field of imprecise probabilities. There are several mathematically equivalent ways to interpret this theory; throughout this dissertation, we will adopt the "sensitivity analysis" interpretation [114, Section 2.10], which is reminiscent of e.g. Huber's work [50] on robust statistics.

Under this interpretation, an imprecise-probabilistic model can be seen as a *set* $\mathscr{P}$ whose elements are all (traditional, "precise") probabilistic models. This set $\mathscr{P}$ is taken to include all precise models that are deemed, in some sense, to be "plausible". In our current setting, for example, we might collect in $\mathscr{P}$ all homogeneous Markov chains whose parameters are obtained from (possibly conflicting) expert elicitation. That is, if the assessments of several experts would lead to several distinct models, we could include all of these models in $\mathscr{P}$. We need not stop there, however, and could include in $\mathscr{P}$ also *non*-homogeneous Markov chains, whose parameters can vary freely over time; provided that at each point in time they are consistent with at least one expert's assessment, say. In a similar vein, we might relax the Markov assumption of the elements of this set, and include models whose behaviour depends intricately on their history. Constructions like the above lead to the notion of an *imprecise-Markov chain* [22, 45, 48, 57, 61, 103]; the imprecise-probabilistic generalisation of traditional Markov chains.[2]

With such a model, the inferences of interest are no longer simply expectations, as above, but rather (conditional) *lower* and *upper* expectations. These are defined, respectively, as

$$\underline{\mathbb{E}}[f\,|\,C] := \inf_{P\in\mathscr{P}} \mathbb{E}_P[f\,|\,C] \quad \text{and} \quad \overline{\mathbb{E}}[f\,|\,C] := \sup_{P\in\mathscr{P}} \mathbb{E}_P[f\,|\,C]\,,$$

where, for all $P \in \mathscr{P}$, $\mathbb{E}_P[f\,|\,C]$ is the usual conditional expectation of the function f, conditional on the event C, with respect to the probabilistic model P. In words, the lower expectation is the infimum of—the tightest possible lower bound on—the expectations of all precise models that are deemed to be plausible. Conversely, the upper expectation is the supremum of—the tightest possible upper bound on—these expectations. Consequently, provided that $\mathscr{P}$ was constructed to contain all plausible variations of the model that one is trying to obtain—in terms of numerical parameters and/or structural assumptions—these lower and upper expectations bound the range of values that the inference of interest could plausibly take. As an aside, it is useful to note that the lower and upper expectation are mathematically conjugate: it holds that $\overline{\mathbb{E}}[f\,|\,C] = -\underline{\mathbb{E}}[-f\,|\,C]$. As a matter of convenience, therefore, we can restrict any discussion to either of these two objects; the analogous properties carry over through this relation. In this dissertation, we phrase our results mostly in terms of lower expectations.

Many interesting questions arise from this rough description of imprecise-Markov chains and this general inference problem. How might we best parameterise and represent such models, mathematically or in practical applications? To what extent can or should we include non-homogeneous and/or non-Markovian processes in these sets? What properties are satisfied by these lower expectations, and under which conditions? When can these lower expectations be tractably

[2] As a brief digression on terminology, we want to point out that these objects are usually called "imprecise Markov chains" in the available literature; see e.g. [22, 61, 103]. That is, most authors do not hyphenate the adjectives "imprecise" and "Markov". However, the terminology derives from that fact that these models satisfy an *imprecise version of the Markov property*; see e.g. Propositions 3.25_{110} and 5.28_{203}. As such, we feel that "imprecise-Markov", as a compound adjective, should be hyphenated. In particular for this current work, where we also deal with further adjectives—i.e. we discuss both discrete-time and continuous-time imprecise-Markov chains—this helps to disambiguate the target of the adjective "imprecise". In previous work [61] we ourselves used the terminology "imprecise continuous-time Markov chain", something that we now regret as a somewhat unfortunate precedent.

computed; to what extent does this depend on the type of inference? What about properties in the limit where time goes to infinity? And so on and so forth...

We would go so far as to say that the substantial existing literature on discrete-time imprecise-Markov chains has effectively settled most of these issues for systems with a discrete time domain, at least for inferences that depend on the state of the system at finitely many time points; the setting where inferences depend on the state at infinitely many time points—e.g. the computation of time averages or hitting times—remains an area of active research. In particular, the seminal work of De Cooman and Hermans [20] established that, for certain types of discrete-time imprecise-Markov chains, the computational complexity of solving a wide class of inference problems is effectively similar to that observed for traditional Markov chains. Colloquially, we may say that while there is some computational overhead to the use of imprecise-probabilistic models, this overhead does not depend on the particular inference that is being solved; and consequently, many inferences that can be computed efficiently for traditional Markov chains, can also be computed efficiently for discrete-time imprecise-Markov chains. The thesis that we develop in this dissertation is, at its core, that such results carry over, essentially identically, to systems with a continuous time domain.

Concretely, in this dissertation we develop a number of mathematical tools that allow us to describe (collections of) continuous-time stochastic processes that may or may not be Markovian. We introduce multiple distinct definitions of continuous-time imprecise-Markov chains; all sets of such stochastic processes, but different in the structural assumptions of their elements. We investigate properties of these sets and, dually, of their corresponding lower (and upper) expectations. We argue and show that, under some conditions, these properties mirror those of discrete-time imprecise-Markov chains.

Arguably the two most important properties that we derive for the lower (and hence upper) expectations are sufficient conditions for these models to satisfy (i) an *imprecise-Markov property* and (ii) a *law of iterated (lower) expectation*. The imprecise-Markov property states, essentially, that while the imprecise-probabilistic model may be a set of general, non-Markovian stochastic processes, the corresponding conditional lower expectation *is* history-independent in a manner analogous to the Markov property. This leads to many simplifying properties. The law of iterated (lower) expectation is a decomposition property that, as we show, is instrumental to the development of efficient computational methods. Both of these properties are also known to hold for, and are central to, the established theory of discrete-time imprecise-Markov chains.

In addition to the above, we derive an alternative characterisation of the lower expectations in terms of *lower transition operators*. Such lower transition operators are, again, also central to discrete-time imprecise-Markov chains. We use these lower transition operators to derive a general computational (i.e. algorithmic) method to compute arbitrary lower (and hence upper) expectations for continuous-time imprecise-Markov chains, at least for functions that depend on the state of the system at finitely many time points. Additionally, we derive a specialised augmentation of this algorithm that can be used for efficient computations in models where the parameter set is obtained by taking a region (i.e. a metric ball) around a precise numerical parameter estimate; this setting may arise naturally in, e.g., a sensitivity analysis context.

Finally, we prove how certain inferences in continuous-time imprecise-Markov chains can be reduced to inferences in discrete-time imprecise-Markov chains. We do this by, essentially, taking the restriction of the continuous-time model to the time points on which this inference depends. We show how this reduction can be used to leverage any number of more advanced algorithms from the discrete-time imprecise-Markov chain literature, for use with continuous-time imprecise-Markov chains. In particular, we illustrate its use in obtaining a very efficient inference algorithm for a particular class of problems, without having to explicitly re-derive this result for the continuous-time setting.

1.2 Related Work

As should hopefully already be clear from our discussion in Section 1.1$_{30}$, we base ourselves in large part on previous work done on discrete-time imprecise-Markov chains. A relatively recent overview of such results can be found in Reference [48]; in what follows, we present only a brief historical account. Discrete-time imprecise-Markov chains were introduced by Hartfiel, who called them *Markov set chains* [44–46]. He took as his models sets of (potentially non-homogeneous) Markov chains. Similar models were later studied by Škulj [101]. A different type of discrete-time imprecise-Markov chain was considered by Kozine and Utkin [57], and Campos *et al.* [10]; they took as their models sets of homogeneous Markov chains.

It was the work by De Cooman and Hermans [20] that first introduced a third type of discrete-time imprecise-Markov chain; under our interpretation, these models can be seen as sets of general—not necessarily homogeneous or Markovian—stochastic processes. They developed these models through a connection with Shafer and Vovk's formalisation of *game theoretic probabilities* [97, 98]. It was this work that ultimately paved the way for efficient inference algorithms for wide classes of problems; we provide a (non-exhaustive) list of subsequent work through References [18, 19, 21, 22, 47, 58, 64, 69, 71, 72, 106, 107].

As in the discrete-time case, work on continuous-time imprecise-Markov chains also traces its early roots back to Hartfiel [43], who investigated set-valued solutions to (sets of) differential equations that are closely connected to (precise) continuous-time Markov chains. As far as we are aware, there has not been much related work in the intervening time until Škulj [102, 103] revisited the problem almost three decades later. This work led to a first application by Troffaes *et al.* [110], and Lopatatzidis *et al.* [70] investigated some computational aspects in a special case relevant to queueing theory. It was around this time that we started the work that forms the basis of this dissertation, which would be published two years later [61]. In parallel, De Bock [17] analysed the long-term (ergodic) behaviour of the machinery that we were developing.

Due to the time that has passed since then, a situation has developed where a fair amount of work has been done that is based on, or at least strongly related to, the theory that we present in this dissertation. We nevertheless feel that this is "related work", so we would like to provide some pointers to this literature for the convenience of interested readers. We refer to Section 1.5_{42} further on for an overview of subsequent work in which we were ourselves involved. Of particular note is the work by Erreygers *et al.* [33, 35] on the construction of continuous-time imprecise-Markov chains with a method known as *lumping*, where parameter-underspecification (and hence model imprecision) is a consequence of a reduction from an intractably large (but finite) state space to one with a more manageable size. This has led to a number of more applied publications about continuous-time imprecise-Markov chains [35, 92, 108]. We should also mention the work by Erreygers and De Bock on computational methods [32] that improve the ones that we presented in [61], and their work towards a generalisation of the theory to systems with countably infinitely many states [34]. Krpelík *et al.* [66] have investigated an application in the context of reliability theory. Recent work by Škulj [104] investigates possible improvements for fundamental computational aspects of elementary inferences.

A somewhat recent discovery is that there appears to be a closely related—but independently developed—line of research in the field of mathematical finance, based on Peng's [85] work on nonlinear expectations and, in particular, "nonlinear Markov chains". His theory of nonlinear expectations provides a generalised formalism for representing uncertainty that is very similar to, but subtly different from, Walley's theory of imprecise probabilities. Of particular note in this context is Nendel's work on nonlinear Markov chains [78, 79], as is the work by Denk *et al.* [25, 26].

Moreover, work that is mathematically related to, but conceptually very different from, what we present here, can be found in the fields of (continuous-time) *Markov decision processes* [40] and *controlled Markov chains* [41]. There, too, one deals with parameters that may vary in some pre-specified set. However, such sets there represent possible behaviours of the underlying system that can be actively chosen or controlled, e.g. in such a manner as to optimise some objective function of interest. Nevertheless, in this context we could alternatively interpret lower and upper expectations as corresponding to inferences which an external party—e.g. "nature"—tries to steer towards the best, or worst, possible outcome, within the set of possible behaviours of the underlying system. Because this interpretation is substantially different from the "model uncertainty" view that we adopt here, we will not really consider this connection in the remainder of this dissertation.

Finally, a practically important question is how one obtains the parameters that describe a continuous-time imprecise-Markov chain. To this end, we briefly consider some methods based on perturbations of the parameters of precise continuous-time Markov chains, which seems natural from the sensitivity-analysis point of view. However, more complicated methods are largely outside the scope of this work, and from a theoretical standpoint we generally simply assume that the parameters are given. For some examples of related work in the literature that addresses this question to some extent, we refer to the lumping methods [33, 35] already mentioned above, as well as to our own preliminary research into estimating these parameters from data [62].

1.3 Navigating This Dissertation

This dissertation is divided into eight chapters and two global appendices, not counting sections like the summaries, the list of symbols and terminology, and the bibliography; the latter two sections can be found at the back. Most chapters are divided into multiple sections, which may themselves be divided into subsections. Many of the chapters in this work have their own appendices, in which we have collected proofs, technical results, and explanations which we feel would detract from the discussion in the main text. There are also many, many references used in this dissertation. References to external work are numbered and delimited by square brackets, and refer to work in the literature that is listed in the bibliography; for example, Reference [3] is an excellent introductory treatment and overview of the general theory of imprecise probabilities.

For ease of navigation, internal references are provided with a subscript that displays the page on which the referent can be found; for instance, Proposition 5.28_{203} can be found on Page 203. To make them easily distinguishable, references to equations are delimited by parentheses; for example, Equation $(5.18)_{208}$ states the law of iterated (lower) expectation for continuous-time imprecise-Markov chains. Rather than explicitly list the page number, we use the subscript symbols $\curvearrowleft$ or $\curvearrowright$ when the referent can be found on, respectively, the previous or subsequent page. In the special case where the referent is on the same double-page spread as the reference, we omit the subscript; after all, you can already see the start of Section 1.4 without turning the page.

1.4 Overview of the Chapters

Let us now present a brief overview of the content and main results of the chapters in this work, as well as indicate which parts are based on the literature and which parts constitute original work. After this section, we will in the current introductory chapter— Chapter 1_{29}— present an overview of our publications and, for ease of reference, again discuss which parts of this dissertation represent substantial new work. We then conclude this chapter with some elementary mathematical preliminaries.

In Chapter 2_{45} we introduce the probabilistic framework that will form the foundation of this work. In Section 2.1_{46} we discuss full conditional probabilities [29] and the notion of coherence [4, 24, 87, 109, 117, 118]. We formalise a notion of conditional expectations for these objects, based on the concept of coherent conditional previsions [24, 90, 109, 114, 117], and inspired by the notion of linear extension discussed in Reference [109]. We explore some connections to measure-theoretic expectations [105] by extending a result from Reference [109]. In Section 2.2_{58} we formalise some notions of time, state, and function spaces, and in Section 2.3_{64} we combine these concepts to formally define stochastic processes with both discrete and continuous time domains; the characterisation of discrete-time processes is inspired by Reference [69], and that of the continuous-time ones is based on our original work published in Reference [61]. We conclude this chapter with some elementary discussion and properties of conditional expectations specifically for stochastic processes.

Chapter 3_{83} then takes a short digression into the theory of (imprecise-)Markov chains in discrete time. As we have repeatedly claimed in Section 1.1_{30}, this well-established theory has strong connections to the theory of continuous-time imprecise-Markov chains that we will develop in this work, and many of the results that we obtain are inspired by, and can be seen as extensions of, results from this theory. Due to the comparatively straightforward intuition of the discrete time domain that is used here, this chapter also lets us ease the reader into some of the more involved concepts that we will encounter later. We start in Section 3.1_{85} by stating some general properties of discrete-time processes, that were initially proved in [69]. In Section 3.2_{89}, we discuss discrete-time (precise) Markov chains [39, 54, 82, 96]—of particular importance here is their parameterisation using (families of) *transition matrices*. We consider various different characterisations of discrete-time imprecise-Markov chains in Section 3.3_{101}, based on previous work in References [10,20–22,27,44–46,48,57,69,101]. Of central importance here, is their characterisation using (families of) *sets* of transition matrices. We investigate properties of their corresponding lower and upper expectation operators, by discussing some results that conceptually originate from [20–22, 48, 69]. In Section 3.4_{116} we discuss *lower transition operators* [22,23,48], which are essentially nonlinear generalisations of the linear maps represented by transition matrices, and we demonstrate how they can be seen as dual representations of *sets* of transition matrices. In Section 3.5_{121}, we discuss some results based on References [22, 48], that show how these lower transition operators form an alternative characterisation of the lower expectations of discrete-time imprecise-Markov chains.

Many of the subtleties of working with continuous-time stochastic processes are discussed in Chapter 4_{143}. This chapter is mostly based on original work that was previously published in Reference [61]. In Section 4.1_{144} we introduce the notion of *well-behaved stochastic processes*; intuitively, processes that cannot move between states instantaneously. In Section 4.2_{148} we introduce an operator-theoretic framework for describing the behaviour of such processes. The basis of this framework are multi-index families of transition matrices induced by continuous-time stochastic processes. We then spend some effort on developing machinery that lets us construct, manipulate, and combine such families. As discussed in Section 4.3_{150}, a crucial building block of these families are *semigroups of transition matrices* [111] generated by *transition rate matrices*; these are essentially matrices containing the *transition rates* that constitute the parameters of homogeneous continuous-time Markov chains [82]. In Sections 4.4_{156} and 4.5_{158}, we develop the machinery to manipulate such families of transition matrices. We conclude this chapter with Section 4.6_{166}, where we discuss generalised derivatives of the time-evolving transition probabilities of general continuous-time stochastic processes, which are central to our definition of continuous-time imprecise-Markov chains.

In Chapter 5_{181}, we finally introduce the continuous-time imprecise-Markov chains that are the subject of this dissertation; it is mostly based on original work that was previously published in Reference [61]. We start in Section 5.1_{182} by introducing continuous-time (precise) Markov chains, based on our formalism of full and coherent conditional probabilities. In Section 5.2_{188}, we introduce three distinct definitions of what could reasonably be called continuous-time imprecise-Markov chains. These three definitions are sets of processes with increasingly weaker structural assumptions on their elements; they correspond to sets of homogeneous Markov chains, sets of—possibly non-homogeneous—Markov chains, and sets of general—possibly non-homogeneous and non-Markovian—stochastic processes. We spend some effort in analysing structural properties of these sets of processes. The content of Section 5.3_{194} is largely novel work that has so far been unpublished, and we there investigate structural properties of the induced sets of transition matrices. We conclude this chapter with Section 5.4_{198}, in which we provide a discussion about the corresponding lower (and upper) expectations of these imprecise-Markov chains. Here we derive some first results about sufficient conditions for these models to satisfy an *imprecise-Markov property*, as well as a *law of iterated lower expectation*.

In Chapter 6_{259} we introduce lower transition operators—which we previously discussed for discrete-time models—also in the continuous-time context. This chapter is mostly original work, containing in part some new and unpublished results, but being otherwise largely based on Reference [61]. In Section 6.1_{260}, we use the structural properties of the induced sets of transition matrices, as well as the law of iterated lower expectation, to derive an alternative characterisation of the lower expectations of our most imprecise type of continuous-time imprecise-Markov chains, in terms of these corresponding lower transition operators. In Section 6.2_{265} we introduce *lower transition rate operators*, which are to transition rate matrices as lower transition operators are to transition matrices; and we establish a duality between such lower transition rate operators and *sets* of transition rate matrices. In Section 6.3_{269} we use these lower transition rate operators to construct a *semigroup of lower transition operators*, and we discuss how this construction has a connection to previous work in References [17, 78, 81, 103]. We present an algorithm to evaluate the elements of this semigroup, and present some new, unpublished, results for when the parameters are obtained from a perturbation model, in which case this evaluation can be performed fairly efficiently. In Section 6.4_{279}, we then derive sufficient conditions for this semigroup of lower transition operators to coincide with lower transition operators corresponding to our continuous-time imprecise-Markov chains. This yields an improved sufficient condition for our most imprecise type of continuous-time imprecise-Markov chain to satisfy an imprecise-Markov property. In Section 6.5$_{284}$, we use the correspondence between the lower expectations of continuous-time imprecise-Markov chains, their lower transition operators, and the generated semigroup of lower transition operators, to develop a general computational (i.e. algorithmic) method to compute arbitrary inferences that depend on the state of the system at finitely many time points. We conclude this chapter with Section 6.6$_{290}$, in which we both illustrate the use of this computational method on a numerical example, and also establish that this computational approach only works for sets of general stochastic processes. In particular, we provide a novel, previously unpublished, example that demonstrates that the lower expectations of this model can differ from the lower expectations for sets of Markov chains.

The final technical chapter of this dissertation is Chapter 7$_{335}$, and it mostly contains novel, previously unpublished, results. There, we discuss a connection between discrete-time and continuous-time imprecise-Markov chains. In particular, in Section 7.1$_{336}$ we show how the parameters of a continuous-time imprecise-Markov chain can also be used to describe a discrete-time imprecise-Markov chain. We demonstrate in Section 7.2$_{338}$ that this characterisation is such, that the lower expectations of these two models coincide. We demonstrate the use of this result in obtaining an especially efficient inference algorithm for a particular class of functions, that was previously derived for discrete-time imprecise-Markov chains—without explicitly re-deriving it for the continuous-time setting. We conclude this chapter with Section 7.3$_{345}$, in which we provide this discrete-time imprecise-Markov chain with an alternative characterisation, as being the set of *restricted* continuous-time stochastic processes corresponding to the elements of the continuous-time imprecise-Markov chain with the same parameters.

We conclude with Chapter 8$_{363}$, in which we summarise the main results, provide some outlook on future work, and give our closing thoughts on the thesis developed in this dissertation.

In addition to the above chapters in the main text, we have included two appendices to this dissertation. In Appendix A$_{369}$ we present and summarise some well-known technical definitions and results that are crucial for analysis in (finite-dimensional) normed vector spaces, on which we rely throughout this dissertation. This appendix is mostly based on results from References [9,51,100]. Appendix B$_{391}$ contains a number of technical inequalities on which we rely in multiple chapters, but which presented some difficulty to prove in the chronological order of their use.

1.5 Publications and New Results

This dissertation represents the culmination of the research performed during my time as a PhD student. This research has resulted in the publication of nine papers and one book chapter. One of these papers was published as a journal article, and it is only this paper that forms the basis of my dissertation:

- Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017 [61].

This dissertation presents the core results from this publication, but with more discussion and contextualisation. Moreover, a fairly substantial part of this dissertation consists of new results that have not been published previously. Of particular note are the following, which I feel are substantial contributions in their own right and which I hope in the future to publish as (part of) research articles:

- The results in Section 5.3_{194} about sets of transition matrices induced by continuous-time imprecise-Markov chains;
- The efficient computational method presented by Proposition 6.21_{278}, for evaluating the lower transition rate operator obtained from a metric ball around a single transition rate matrix;
- The results in Section $6.6.2_{297}$ demonstrating that lower expectations for sets of (non-homogeneous) Markov chains can differ from those for sets of general stochastic processes;
- The correspondence between discrete-time and continuous-time imprecise-Markov chains presented in Chapter 7_{335}.

In addition to the above, I have (co-)authored six full-paper contributions related to imprecise-Markov chains, which were presented at—and included in the proceedings of—international conferences. Although the results from these publications are not included in this dissertation, I list them here in reverse chronological order:

- Thomas Krak. Computing expected hitting times for imprecise Markov chains. ***Accepted for publication** in Proceedings of UQOP 2020, forthcoming* [58].
- Natan T'Joens, Thomas Krak, Jasper De Bock, and Gert de Cooman. A recursive algorithm for computing inferences in imprecise Markov chains. *Lecture Notes in Artificial Intelligence, Vol. 11726 (Proceedings of ECSQARU 2019)*, pages 455–465, 2019 [107].

- Thomas Krak, Natan T'Joens, and Jasper De Bock. Hitting times and probabilities for imprecise Markov chains. *Proceedings of Machine Learning Research, Vol. 103 (Proceedings of ISIPTA 2019)*, pages 265–275, 2019 [64].

- Matthias Troffaes, Thomas Krak, and Henna Bains. Two-state imprecise Markov chains for statistical modelling of two-state non-Markovian processes. *Proceedings of Machine Learning Research, Vol. 103 (Proceedings of ISIPTA 2019)*, pages 394–403, 2019 [108].

- Thomas Krak, Alexander Erreygers, and Jasper De Bock. An imprecise probabilistic estimator for the transition rate matrix of a continuous-time Markov chain. *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*, pages 124–132, 2018 [62].

- Thomas Krak, Jasper De Bock, and Arno Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. *Proceedings of Machine Learning Research, Vol. 62 (Proceedings of ISIPTA 2017)*, pages 193–204, 2017 [60].

Additionally, I have written a chapter contribution to a book:

- Thomas Krak. An introduction to imprecise Markov chains. In Massimiliano Vasile, editor, *Optimization Under Uncertainty with Applications to Aerospace Engineering*, chapter 5. Springer Nature, 2021 [59].

In addition to my work on imprecise-Markov chains, I have (co-)authored two papers on unrelated topics; I have included them here for completeness:

- Thomas Krak and Ad Feelders. Exceptional model mining with tree-constrained gradient ascent. *Proceedings of SIAM ICDM 2015*, pages 487–495, 2015 [63].

- Thomas Krak and Linda C. van der Gaag. Knowledge-based bias correction – a case study in veterinary decision support. *Frontiers in Artificial Intelligence and Applications (Proceedings of ECAI 2014)*, pages 489–494, 2014 [65].

1.6 Mathematical Preliminaries

Let us conclude this introductory chapter with some preliminary discussion of mathematical notation and terminology on which we will rely throughout this dissertation.

We denote the reals as $\mathbb{R}$, the non-negative reals as $\mathbb{R}_{\geq 0}$, the positive reals as $\mathbb{R}_{>0}$, and the negative reals as $\mathbb{R}_{<0}$. For any $c \in \mathbb{R}$, $\mathbb{R}_{\geq c}$, $\mathbb{R}_{>c}$, and $\mathbb{R}_{<c}$ have a similar meaning. The positive and non-negative integers are denoted by $\mathbb{Z}_{>0}$ and $\mathbb{Z}_{\geq 0}$, respectively. The rationals are denoted by $\mathbb{Q}$.

Infinite sequences of quantities will be denoted $\{a_i\}_{i \in \mathbb{Z}_{>0}}$, possibly with limit statements of the form $\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c$, which should be interpreted as $\lim_{i \to +\infty} a_i = c$. If the elements of such a sequence belong to a space that is endowed with an ordering relation, we may write $\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c^+$ or $\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c^-$ if the limit is approached from above or below, respectively.

When working with suprema and infima, we will sometimes use the shorthand notation $\sup\{\cdot\} < +\infty$ to mean that there is some $c \in \mathbb{R}$ such that $\sup\{\cdot\} < c$, and similarly for $\inf\{\cdot\} > -\infty$.

For any set A and any superset C of A, we use $\mathbb{I}_A$ to denote the indicator of A, defined for all $a \in C$ by $\mathbb{I}_A(a) := 1$ if $a \in A$ and $\mathbb{I}_A(a) := 0$, otherwise. If A is a singleton $A = \{a\}$, we instead write $\mathbb{I}_a := \mathbb{I}_{\{a\}}$.

Finally, for any finite or countably infinite set A, a *probability mass function on A* is a map $p : A \to \mathbb{R}$ such that $\sum_{a \in A} p(a) = 1$ and $p(a) \geq 0$ for all $a \in A$.

2

Foundations of Stochastic Processes

"Who are you?"

This was not an encouraging opening for a conversation.

Alice replied, rather shyly, "I–I hardly know, sir, just at present–at least I know who I was *when I got up this morning, but I think I must have been changed several times since then."*

Lewis Carroll, "Alice's Adventures in Wonderland"

Consider some system or object whose state X_t evolves in some uncertain manner as time t progresses. This describes in rough terms the concept of a *stochastic process*, which is ultimately the kind of mathematical object that we aim to study in this dissertation. To formalise these ideas, we will need some machinery to describe what, exactly, we mean by notions like "state", "uncertain", and "time". To this end, we develop in this chapter the foundational elements that we will need throughout the remainder of this work. We start in Section 2.1↷ by introducing the basic formalism to represent uncertainty, on which we will rely throughout: this is the framework of *full conditional probabilities* and *coherence*. This provides a language for talking about "things"—in the broadest sense—whose behaviour we are uncertain about in one way or another. Notably, this framework is distinct from the measure-theoretic one that is most often employed in the literature; we discuss some of the important differences as they come up, and provide the interested reader with a behavioural interpretation of this formalism in Appendix 2.A$_{80}$ to this chapter.

In Section 2.2$_{58}$ we introduce some notation and conventions for reasoning about points in time, about states that a system can be in, and about functions of these states. We tie these concepts together in Section 2.3$_{64}$, where we use this framework for reasoning about uncertainty to formally define the notion of a stochastic process. As we will see, this provides a mathematical model for reasoning about the behaviour of some abstract underlying system of interest, which evolves in time in a manner about which we are uncertain.

We end this chapter with Section 2.4$_{71}$, where we formally discuss the *inferences* that we want to make about the system of interest; these are statements about quantities of interest that depend on the state or behaviour of the system. We will discuss how these inferences are represented as (conditional) expectations of functions, taken with respect to the coherent (conditional) probabilities that we introduced earlier.

2.1 Full Conditional Probabilities and Coherence

Let us first turn our attention to how to represent and reason about *uncertainty*; broadly speaking, this is a particular epistemological state that a subject can be in. That is, we take uncertainty to be a subjective notion that derives from a subject's knowledge (or rather, lack thereof) about some topic to which this uncertainty pertains. To illustrate this, consider a variable X that takes values ω in some non-empty—possibly infinite—*outcome space* Ω. A subject may be uncertain about the actual value of X, in the sense that she does not know it. An example of this, which we borrow from Walley [114], is the throwing of a thumb tack,[1] and the question of whether it lands pin-up or pin-down. Then $\Omega = \{\text{pin-up}, \text{pin-down}\}$ are orientations that the thumb tack can be in, and the variable X represents the resulting orientation after the throw. For various reasons, it seems reasonable for a subject to say that she is uncertain about X before learning the outcome of the throw.[2]

[1] Another classical example is the flipping of a coin, but we feel that this brings with it an implied and unneeded—indeed, undesired—sense of symmetry.

[2] It is not really our intention to give an exhaustive philosophical overview of what it might *mean* to be uncertain, or where one's uncertainty might derive from, or in which way quantified degrees of uncertainty could derive from particular epistemological states. Instead, we simply proceed axiomatically with Definition 2.1. Nevertheless, the closely related notion of *coherence*, which is introduced formally in Definition 2.2$_{48}$, can be provided with a behavioural interpretation that we sketch in Appendix 2.A$_{80}$. If one insists on a label, this would provide the formalism that we use with a behavioural subjectivist semantics. We refer to Appendix 2.A$_{80}$ and the references contained therein for further information.

To formalise this mathematically, we call any subset E of Ω an event, we use $\mathscr{E}(\Omega)$ to denote the set of all such events, and we let $\mathscr{E}(\Omega)_{\supset\emptyset} := \mathscr{E}(\Omega) \setminus \{\emptyset\}$ be the set of all non-empty events. A subject's uncertainty about the value of X may then be described by means of a full conditional probability [29].

Definition 2.1 (Full Conditional Probability). *A full conditional probability P is a map from $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ to $\mathbb{R}$ that satisfies the following axioms. For all $A, B \in \mathscr{E}(\Omega)$ and all $C, D \in \mathscr{E}(\Omega)_{\supset\emptyset}$:*

F1: $P(A|C) \geq 0$;

F2: $P(A|C) = 1$ *if* $C \subseteq A$;

F3: $P(A \cup B|C) = P(A|C) + P(B|C)$ *if* $A \cap B = \emptyset$;

F4: $P(A \cap D|C) = P(A|D \cap C)P(D|C)$ *if* $D \cap C \neq \emptyset$.

For any $A \in \mathscr{E}(\Omega)$ and $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$, we call $P(A|C)$ the probability of A conditional on C. Also, for any $A \in \mathscr{E}(\Omega)$, we use the shorthand notation $P(A) := P(A|\Omega)$ and then call $P(A)$ the probability of A.

A number of additional properties follow readily from F1–F3.

Proposition 2.1. *Let P be a full conditional probability. Then for all $A \in \mathscr{E}(\Omega)$ and all $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$, it holds that*

F5: $0 \leq P(A|C) \leq 1$;

F6: $P(A|C) = P(A \cap C|C)$;

F7: $P(\emptyset|C) = 0$;

F8: $P(\Omega|C) = 1$.

Proof. Consider any $A \in \mathscr{E}(\Omega)$ and $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$. It then follows from F2 and F3 that

$$P(A|C) = P(A \cup C|C) - P(C \setminus A|C) = 1 - P(C \setminus A|C) \tag{2.1}$$

and

$$P(A \cap C|C) = P(C|C) - P(C \setminus A|C) = 1 - P(C \setminus A|C). \tag{2.2}$$

F5 follows from Equation (2.1) and F1. F6 follows from Equations (2.1) and (2.2). F7 follows from F3, by letting $B := \emptyset$. F8 follows from F2. □

Basically, $F1_{\curvearrowleft}$–$F4_{\curvearrowleft}$ are just the standard rules of probability. However, there are four rather subtle differences with the more traditional approach. The first one is that a full conditional probability takes conditional probabilities as its basic entities: $P(A|C)$ is well-defined even when $P(C) = 0$. The second difference, which is related to the first, is that Bayes's rule—$F4_{\curvearrowleft}$—is stated in a multiplicative form; it is not regarded as a definition of conditional probabilities, but rather as a property that connects joint- and conditional probabilities. The third difference is that we consider all events, and do not restrict ourselves to some specific subset of events—such as a σ-algebra. The fourth difference, which is related to the third, is that we only require finite additivity—$F3_{\curvearrowleft}$—and do not impose σ-additivity.

The "full" in full conditional probability refers to the fact that the domain of P is the complete set $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. At first sight, this might seem unimportant, and one might be inclined to introduce a similar definition for functions P whose domain is some subset $\mathscr{C}$ of $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. However, unfortunately, as our next example illustrates, such a definition would not guarantee the possibility of extending the function to a larger domain $\mathscr{C}^*$, with $\mathscr{C} \subseteq \mathscr{C}^* \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$.

Example 2.1. Let $\Omega = \{1,2,3,4,5,6\}$ be the set of possible values for the throw of a—possibly unfair—die and let $\mathscr{C} := \{(E_o,\Omega),(E_e,\Omega)\} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, where the events $E_o = \{1,3,5\}$ and $E_e = \{2,4,6\}$ correspond to an odd or even outcome of the die throw, respectively. The map $P\colon \mathscr{C} \to \mathbb{R}$ that is defined by

$$P(E_o) := P(E_o|\Omega) = 2/3 \text{ and } P(E_e) := P(E_e|\Omega) = 2/3$$

then satisfies $F1_{\curvearrowleft}$–$F4_{\curvearrowleft}$ on its domain. However, if we extend the domain by adding the trivial couple (Ω,Ω), it becomes impossible to satisfy $F1_{\curvearrowleft}$–$F4_{\curvearrowleft}$, because $F2_{\curvearrowleft}$ and $F3_{\curvearrowleft}$ would then require that

$$1 = P(\Omega|\Omega) = P(E_o|\Omega) + P(E_e|\Omega) = 2/3 + 2/3 = 4/3,$$

which is clearly a contradiction. ◊

In order to avoid the situation in this example, that is, in order to guarantee the possibility of extending the domain of a conditional probability in a sensible way, we use the concept of coherence [4,24,87,109,117,118].

Definition 2.2 (Coherent conditional probability)**.** *Let P be a real-valued map from $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ to $\mathbb{R}$. Then P is said to be a* coherent conditional probability *on $\mathscr{C}$ if, for all $n \in \mathbb{N}$ and every choice of $(A_i,C_i) \in \mathscr{C}$*

and $\lambda_i \in \mathbb{R}$, $i \in \{1,\dots,n\}$,[3]

$$\max\left\{\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \,\middle|\, \omega \in C_0\right\} \geq 0,$$

with $C_0 := \cup_{i=1}^{n} C_i$.

The interested reader is invited to take a look at Appendix 2.A$_{80}$, where we provide this abstract concept with an intuitive gambling interpretation. However, for our present purposes, this interpretation is not required. Instead, our motivation for introducing coherence stems from the following two results. First, if $\mathscr{C} = \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, then coherence is equivalent to the axioms of probability, that is, properties F1$_{47}$–F4$_{47}$.

Theorem 2.2 ([87, Theorem 3])**.** *Let P be a real-valued map from* $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ *to* $\mathbb{R}$. *Then P is a coherent conditional probability if and only if it is a full conditional probability.*

Secondly, for coherent conditional probabilities on arbitrary domains, it is always possible to extend their domain while preserving coherence.

Theorem 2.3 ([87, Theorem 4])**.** *Let P be a coherent conditional probability on* $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. *Then for any* $\mathscr{C} \subseteq \mathscr{C}^* \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, *P can be extended to a coherent conditional probability on* $\mathscr{C}^*$.

In particular, it is therefore always possible to extend a coherent conditional probability P on $\mathscr{C}$, to a coherent conditional probability on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. Due to Theorem 2.2, this extension is a full conditional probability. The following makes this explicit.

Corollary 2.4. *Let P be a real-valued map from* $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ *to* $\mathbb{R}$. *Then P is a coherent conditional probability if and only if it can be extended to a full conditional probability.*

[3] Many authors replace the maximum in this expression by a supremum, and also impose an additional inequality, where the maximum—supremum—is replaced by a minimum—infimum—and where the inequality is reversed [4, 5, 87]. This is completely equivalent to our definition. First of all, if the maximum is replaced by a supremum, then since n is finite and because, for every $i \in \{1,\dots,n\}$, $\mathbb{I}_{A_i}$ and $\mathbb{I}_{C_i}$ can only take two values—0 or 1—it follows that this supremum is taken over a finite set of real numbers, which implies that it is actually a maximum. Secondly, replacing the maximum by a minimum and reversing the inequality is equivalent to replacing the λ_i in our expression by their additive inverses, which is clearly allowed because the coefficients λ_i can take any arbitrary real value.

Proof. First assume that P can be extended to a full conditional probability P^*. Theorem 2.2$_{\curvearrowleft}$ then implies that P^* is a coherent conditional probability, and therefore, since P is the restriction of P^* to $\mathscr{C}$, it follows from Definition 2.2$_{48}$ that P is a coherent conditional probability.

Conversely, if P is a coherent conditional probability on $\mathscr{C}$, it follows from Theorem 2.3$_{\curvearrowleft}$ that P can be extended to a coherent conditional probability P^* on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, which, because of Theorem 2.2$_{\curvearrowleft}$, is a full conditional probability. $\square$

Note, therefore, that if P is a coherent conditional probability on $\mathscr{C}$, we can equivalently say that it is the restriction of a full conditional probability. Hence, any coherent conditional probability on $\mathscr{C}$ is guaranteed to satisfy properties F1$_{47}$-F4$_{47}$. However, as was essentially already illustrated in Example 2.1$_{48}$, and as our next example makes explicit, the converse is not true.

Example 2.2. Let Ω, $\mathscr{C}$, E_o, E_e and $P\colon \mathscr{C} \to \mathbb{R}$ be defined as in Example 2.1$_{48}$. Then as we have seen in that example, P satisfies F1$_{47}$-F4$_{47}$ on its domain $\mathscr{C}$. However, P is not a coherent conditional probability on $\mathscr{C}$, because if it were, then according to Corollary 2.4$_{\curvearrowleft}$, P could be extended to a full conditional probability. Since $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ includes (Ω,Ω), the argument at the end of Example 2.1$_{48}$ implies that this is impossible. The same conclusion can be reached by verifying Definition 2.2$_{48}$ directly; we will demonstrate this in what follows.

Let $A_1 := E_\text{e}$ and $A_2 := E_\text{o}$, and let $\lambda_1 := \lambda_2 := -1$. Then $(A_1,\Omega),(A_2,\Omega) \in \mathscr{C}$ and $\lambda_1,\lambda_2 \in \mathbb{R}$. Now consider the function $G\colon \Omega \to \mathbb{R}$, defined for all $\omega \in \Omega$ as

$$G(\omega) := \sum_{i=1}^{2} \lambda_i \mathbb{I}_\Omega(\omega)\big(P(A_i \,|\, \Omega) - \mathbb{I}_{A_i}(\omega)\big)\,.$$

According to Definition 2.2$_{48}$, to show that P is not coherent it suffices to show that $\max_{\omega\in\Omega} G(\omega) < 0$. To this end, fix any $\omega \in \Omega$. It clearly holds that $\mathbb{I}_\Omega(\omega) = 1$. Moreover, note that $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \Omega$; hence we have that either $\omega \in A_1$, or $\omega \in A_2$, and therefore either $\mathbb{I}_{A_1}(\omega) = 1$ and $\mathbb{I}_{A_2}(\omega) = 0$, or $\mathbb{I}_{A_1}(\omega) = 0$ and $\mathbb{I}_{A_2}(\omega) = 1$. Hence, in either case, we find that

$$\sum_{i=1}^{2} \lambda_i \mathbb{I}_\Omega(\omega)\big(P(A_i \,|\, \Omega) - \mathbb{I}_{A_i}(\omega)\big) = -\big(P(A_1 \,|\, \Omega) + P(A_2 \,|\, \Omega) - 1\big) = -1/3\,,$$

whence $G(\omega) < 0$ for every $\omega \in \Omega$; it follows that P is not coherent. $\Diamond$

2.1.1 *Coherent Previsions and Conditional Expectations*

In this section we provide a general definition for conditional expectations taken with respect to the coherent conditional probabilities introduced above. Conceptually, these conditional expectations are analogous to their counterparts in a more traditional probabilistic framework; for a given coherent conditional probability P, a conditional expectation with respect to P is a functional $\mathbb{E}_P$, where for all $f : \Omega \to \mathbb{R}$, we are aiming for a definition of the form

$$\mathbb{E}_P[f \,|\, C] := \int_\Omega f(\omega) P(\mathrm{d}\omega \,|\, C)\,. \tag{2.3}$$

Which is to say, the expected value of some real-valued function f of the possible outcomes Ω, is a convex combination of this function's values; where the value $f(\omega)$ in every possible realisation ω is weighted by the (potentially infinitesimal) probability of that realisation occurring.

The obvious difficulty is how to rigorously define the integral on the right-hand side of this expression. The approach that we choose will determine the interpretation of our notion of (conditional) expectation, and will determine for which functions it is well-defined. For instance, as we know from traditional (measure-theoretic) probability theory, if we were to interpret these integrals in the Lebesgue sense,[4] then we cannot really have a definition that works for *any* function; this is why attention is then restricted to measurable functions.

In our present setting, where we are working with coherent conditional probabilities that can be given a natural behavioural interpretation—again, see Appendix 2.A$_{80}$—it seems appropriate to use the notion of coherent (conditional) *previsions* [24, 90, 109, 114, 117] to set up the general definition of the expectation operators that we are after. This concept of prevision is mathematically analogous to the (measure-theoretic) notion of expectation, but is provided with a behavioural interpretation that is strongly connected to the notion of coherence for probabilities. As was the case with coherent conditional probabilities, the notion of coherence for previsions can be understood to stipulate constraints that a rational agent must follow when making decisions under uncertainty.

Here and in what follows, we will only consider conditional previsions (and expectations) on the set $\mathbb{B}$ of real-valued functions on Ω that

[4] Supposing that we could even do that; since we are explicitly not imposing σ-additivity, P need not necessarily be a measure; see e.g. [54, 105]. Since integration in the Lebesgue sense is typically defined with respect to a measure (*ibid.*), this approach may encounter some difficulty. If one insists on this interpretation, there are however ways around this; see e.g. [109, Section 8.6] for how to define this integral when $P(\cdot \,|\, C)$ is defined on an algebra.

are bounded, meaning that $\sup f < +\infty$ and $-\infty < \inf f$ for all $f \in \mathbb{B}$. The key concept moving forward will, as mentioned, be that of a coherent conditional prevision:

Definition 2.3 ([87, Definition 1]). *Let E be a map from $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ to $\mathbb{R}$. Then E is called a* coherent conditional prevision *on $\mathscr{D}$ if, for all $n \in \mathbb{Z}_{>0}$ and every choice of $(f_i, C_i) \in \mathscr{D}$ and $\lambda_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$, it holds that*[5]

$$\sup \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(E[f_i \,|\, C_i] - f_i(\omega)\big) \,\middle|\, \omega \in C_0 \right\} \geq 0\,,$$

with $C_0 := \cup_{i=1}^{n} C_i$.

A coherent conditional prevision satisfies the following properties.

Proposition 2.5 ([87, Section 2]). *Let E be a coherent conditional prevision on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. Then, for all $(f, C), (g, C) \in \mathscr{D}$, all $\lambda \in \mathbb{R}$, and all $D \in \mathscr{E}(\Omega)_{\supset\emptyset}$,*

E1: $\inf_{\omega \in C} f(\omega) \leq E[f \,|\, C] \leq \sup_{\omega \in C} f(\omega)$

E2: $E[f + g \,|\, C] = E[f \,|\, C] + E[g \,|\, C]$ *whenever* $(f + g, C) \in \mathscr{D}$

E3: $E[\lambda f \,|\, C] = \lambda E[f \,|\, C]$ *whenever* $(\lambda f, C) \in \mathscr{D}$

E4: $E[\mathbb{I}_D f \,|\, C] = E[f \,|\, C \cap D] \cdot E[\mathbb{I}_D \,|\, C]$

whenever all of $(\mathbb{I}_D f, C)$, $(f, C \cap D)$, $(\mathbb{I}_D, C) \in \mathscr{D}$.

In the previous section, we discussed the crucial Theorem 2.3$_{49}$ stating that any coherent conditional probability on $\mathscr{C}$ can be coherently extended to an arbitrary larger domain $\mathscr{C} \subseteq \mathscr{C}^* \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. An analogous property is true for coherent conditional previsions.[6]

Theorem 2.6 ([87, Theorem 4]). *Let E be a coherent conditional prevision on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. Then for any $\mathscr{D} \subseteq \mathscr{D}^* \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$, E can be extended to a coherent conditional prevision on $\mathscr{D}^*$.*

It will be useful to introduce the following. For any $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, we let

$$\mathscr{D}_{\mathscr{C}} := \left\{ (\mathbb{I}_A, C) \,\middle|\, (A, C) \in \mathscr{C} \right\}.$$

[5] In contrast to our remark in the footnote in Definition 2.2$_{48}$, we cannot here replace the supremum with a maximum; to see this, observe that the image of each f_i under C_0 need not be closed.

[6] It follows relatively straightforwardly from Proposition 2.7 further on that Theorem 2.6 implies Theorem 2.3$_{49}$. In fact, we used this implicitly to provide the reference for Theorem 2.3$_{49}$.

Note that $\mathscr{D}_{\mathscr{C}} \subset \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. This allows us to state the following correspondence between coherent conditional probabilities and coherent conditional previsions.

Proposition 2.7. *Let P be a real-valued map on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, and let E be a real-valued map on $\mathscr{D}_{\mathscr{C}}$ such that $E[\mathbb{I}_A|C] := P(A|C)$ for all $(A,C) \in \mathscr{C}$. Then E is a coherent conditional prevision if and only if P is a coherent conditional probability.*

Proof. On the domain $\mathscr{D}_{\mathscr{C}}$ of E, Definitions 2.2$_{48}$ and 2.3 coincide. ☐

This correspondence between coherent conditional probabilities and coherent conditional previsions can be generalised; we will use the following definition.

Definition 2.4. *Let P be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, and let $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. Then a coherent conditional prevision E on $\mathscr{D}$ is said to* correspond *to P, if $E[\mathbb{I}_A|C] = P(A|C)$ for all $(A,C) \in \mathscr{C}$.*

The next result tells us that for any coherent conditional probability P on $\mathscr{C}$, there always is a corresponding coherent conditional prevision on arbitrary domains that include $\mathscr{D}_{\mathscr{C}}$.

Proposition 2.8. *Let P be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, and choose any $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. Then there is a coherent conditional prevision E on $\mathscr{D}$, that corresponds to P.*

Proof. Let $\tilde{E}$ be the real-valued map on $\mathscr{D}_{\mathscr{C}}$ such that $\tilde{E}[\mathbb{I}_A|C] := P(A|C)$ for all $(\mathbb{I}_A, C) \in \mathscr{D}_{\mathscr{C}}$. It follows from Proposition 2.7 and Definition 2.4 that $\tilde{E}$ is a coherent conditional prevision that corresponds to P. Let E be any coherent extension of $\tilde{E}$ to $\mathscr{D}$; such an E exists due to Theorem 2.6. Because E extends $\tilde{E}$ and $\tilde{E}$ corresponds to P, it holds that $E[\mathbb{I}_A|C] = \tilde{E}[\mathbb{I}_A|C] = P(A|C)$ for all $(A,C) \in \mathscr{C}$, whence E corresponds to P by Definition 2.4. ☐

We want to emphasise that the previous result establishes that a corresponding coherent conditional prevision always exists, but not that it is unique. In fact, there are typically many corresponding coherent conditional previsions, provided that the domain $\mathscr{C}$ of P is small enough, and the domain $\mathscr{D}$ of E large enough. It is worth mentioning, then, that if the domain of P is sufficiently large, it will typically have a unique corresponding coherent conditional prevision—this is in particular the case when P is a full conditional probability; see also [109, 117] for further discussion. However, even a coherent conditional probability that is only defined on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, will typically uniquely determine the value of the corresponding coherent conditional prevision on *some* functions in $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$. In fact, this is by definition the case on $\mathscr{D}_{\mathscr{C}}$. In what follows, we will define the *conditional expectation* corresponding to a given coherent conditional probability P, exactly for those functions for which the corresponding coherent conditional prevision is uniquely determined. We will collect these functions in the set $\mathscr{D}_P \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$, which constitutes the domain of definition for the conditional expectations $\mathbb{E}_P$ corresponding to P.

Definition 2.5. *Let P be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$. We collect in $\mathscr{D}_P \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ all pairs $(f,C) \in \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ for which there is some $\mathbb{E}_P[f \,|\, C] \in \mathbb{R}$ such that*

$$\mathbb{E}_P[f \,|\, C] = E[f \,|\, C],$$

for all coherent conditional previsions E that correspond to P and that are defined on any $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $(f,C) \in \mathscr{D}$ and $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}$.

We call the map $\mathbb{E}_P : \mathscr{D}_P \to \mathbb{R} : (f,C) \mapsto \mathbb{E}_P[f \,|\, C]$ the conditional expectation corresponding to P. *We use the shorthand $\mathbb{E}_P[f] := \mathbb{E}_P[f \,|\, \Omega]$ for all $(f,\Omega) \in \mathscr{D}_P$ to denote the (unconditional) expectation.*

In light of the above discussion, the following result should not be surprising.

Lemma 2.9. *For any coherent conditional probability P on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ it holds that $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}_P$.*

Proof. Fix any $(f,C) \in \mathscr{D}_{\mathscr{C}}$, and let E be any coherent conditional prevision on $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ that corresponds to P; then trivially $(f,C) \in \mathscr{D}$. Since $(f,C) \in \mathscr{D}_{\mathscr{C}}$, there is some $(A,C) \in \mathscr{C}$ such that $(f,C) = (\mathbb{I}_A, C)$. Because E corresponds to P it holds that

$$E[f \,|\, C] = E[\mathbb{I}_A \,|\, C] = P(A \,|\, C).$$

Because this is true for all coherent conditional previsions E that correspond to P, it follows from Definition 2.5 that $(f,C) \in \mathscr{D}_P$ and, in particular, that $\mathbb{E}_P[f \,|\, C] = P(A \,|\, C)$. Because this is true for all $(f,C) \in \mathscr{D}_{\mathscr{C}}$ it follows that $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}_P$. $\square$

Moreover, as the next result shows, the conditional expectation $\mathbb{E}_P$ corresponding to any coherent conditional probability P is, itself, a coherent conditional prevision; in particular, therefore, it satisfies Properties E1$_{52}$–E4$_{52}$.

Proposition 2.10. *For any coherent conditional probability P on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, its corresponding conditional expectation $\mathbb{E}_P$ is the unique coherent conditional prevision on $\mathscr{D}_P$ that corresponds to P.*

Proof. By Lemma 2.9 it holds that $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}_P$. By Proposition 2.8$_{53}$ it follows that there is a coherent conditional prevision E on $\mathscr{D}_P$ that corresponds to P. Due to Definition 2.5, it follows that $\mathbb{E}_P[f\,|\,C] = E[f\,|\,C]$ for all $(f,C) \in \mathscr{D}_P$, which implies that $\mathbb{E}_P = E$. Because E is a coherent conditional prevision on $\mathscr{D}_P$, it follows that $\mathbb{E}_P$ is also a coherent conditional prevision on $\mathscr{D}_P$.

To show the uniqueness, let E' be any coherent conditional prevision on $\mathscr{D}_P$ that corresponds to P. Then it follows from Definition 2.5 that $E'[f\,|\,C] = \mathbb{E}_P[f\,|\,C]$ for all $(f,C) \in \mathscr{D}_P$, and hence $E' = \mathbb{E}_P$. □

As is probably not surprising, the domain $\mathscr{D}_P$ is the largest set on which P has a unique corresponding coherent conditional prevision.

Corollary 2.11. *For any coherent conditional probability P on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\emptyset}$, the set $\mathscr{D}_P$ is the largest subset of $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ on which there is a unique coherent conditional prevision that corresponds to P.*

Proof. Consider any $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ for which there is a unique coherent conditional prevision E_P on $\mathscr{D}$ that corresponds to P. We will show that $\mathscr{D} \subseteq \mathscr{D}_P$.

To this end, fix any $(f,C) \in \mathscr{D}$. Let E be any coherent conditional prevision that corresponds to P and that is defined on $\mathscr{D}' \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}'$ and $(f,C) \in \mathscr{D}'$. Let E^* be any coherent conditional prevision on $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ that extends E, which exists by Theorem 2.6$_{52}$. Let E_P^* be the restriction of E^* to $\mathscr{D}$. Then it follows from Definition 2.3$_{52}$ that E_P^* is a coherent conditional prevision on $\mathscr{D}$, which implies that $E_P^* = E_P$ because E_P is the unique coherent conditional prevision on $\mathscr{D}$. Therefore, and because E_P^* is the restriction of E^* and E^* is the extension of E, it follows that

$$E[f\,|\,C] = E^*[f\,|\,C] = E_P^*[f\,|\,C] = E_P[f\,|\,C]\,.$$

Because this is true for all coherent conditional previsions E that correspond to P and that are defined on any $\mathscr{D}' \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}'$ and $(f,C) \in \mathscr{D}'$, it follows from Definition 2.5 that $(f,C) \in \mathscr{D}_P$. Because $(f,C) \in \mathscr{D}$ is arbitrary we conclude that $\mathscr{D} \subseteq \mathscr{D}_P$. □

This approach to define conditional expectations only on this domain, generalises to the conditional case the concept of *linear extension* discussed in [109, Chapters 8 and 9]. We emphasise that this only provides a definition of conditional expectation for functions whose coherent conditional prevision is uniquely determined by the specification of P. Indeed, one might be interested in providing a more general definition that covers a larger part of the domain $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$—or even consider *unbounded* functions—but the rationality criteria behind coherence do not provide us with enough guidance to uniquely determine a “sensible” value of those other functions’ conditional expectation; this is essentially what Corollary 2.11$_{\curvearrowleft}$ says. Possible approaches to extend this definition would be to impose further criteria that one might deem desirable, e.g. certain continuity properties of the resulting conditional expectation, but the current definition suffices for the results of this work. We briefly revisit these issues in Chapter 8$_{363}$.

Next, let us point out that $\mathscr{D}_P$ always includes a particular class of functions, whose conditional expectation is explicitly given by the simple formula in Equation (2.4) below. As we will discuss in Section 2.4$_{71}$, these functions play an important role in this work.

Proposition 2.12. *Let P be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ and let $\mathbb{E}_P$ be its corresponding conditional expectation. Consider any $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$ and $n \in \mathbb{Z}_{\geq 0}$ and, for all $i \in \{1, \ldots, n\}$, consider any $\lambda_i \in \mathbb{R}$ and $A_i \in \mathscr{E}(\Omega)$ such that $(A_i, C) \in \mathscr{C}$. Then for $f := \sum_{i=1}^n \lambda_i \mathbb{I}_{A_i}$ we have that $(f, C) \in \mathscr{D}_P$, and*

$$\mathbb{E}_P[f \,|\, C] = \sum_{i=1}^{n} \lambda_i P(A_i \,|\, C)\,. \tag{2.4}$$

Proof. Let E be any coherent conditional prevision corresponding to P that is defined on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}$ and $(f, C) \in \mathscr{D}$; such an E exists by Proposition 2.8$_{53}$. Then it holds that

$$E[f \,|\, C] = E\left[\sum_{i=1}^{n} \lambda_i \mathbb{I}_{A_i} \,\middle|\, C\right] = \sum_{i=1}^{n} \lambda_i E[\mathbb{I}_{A_i} \,|\, C] = \sum_{i=1}^{n} \lambda_i P(A_i \,|\, C)\,,$$

where we used Properties E2$_{52}$ and E3$_{52}$ for the second step, and that E corresponds to P for the last step. Since this holds for any coherent conditional prevision E corresponding to P, on any $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}$ and $(f, C) \in \mathscr{D}$, it follows from Definition 2.5$_{54}$ that $\mathbb{E}_P[f \,|\, C] = \sum_{i=1}^n \lambda_i P(A_i \,|\, C)$ and $(f, C) \in \mathscr{D}_P$. □

2.1.2 *Connection to Measure-Theoretic Expectations*

It is worth pointing out that, intuitively, the functions for which Proposition 2.12 provides a simple expression for their conditional expectation are, essentially, the functions which are known as the *simple functions* that are measurable with respect to $\mathscr{C}$. Unfortunately, the concept of measurability—typically a measure-theoretic notion, see e.g. [105, Definition 1.4.32]—is a bit difficult to translate to our current setting where $\mathscr{C}$ is a structureless domain of conditional events. Reference [109, Section 1.8] discusses a notion of measurability for finitely-additive probabilities, but still assumes that the underlying domain is an algebra, which is a structural assumption that we are not currently using.

In an attempt to provide intuition, rather than rigour, let us mostly ignore the part where we are dealing with conditional expectations. So, simply fix $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$ and let $(\Omega, \Sigma_C, \mu_C)$ be a probability space, with Σ_C a σ-algebra over Ω and μ_C a probability measure on the measurable space (Ω, Σ_C). Following [105, Definition 1.4.32], f is called Σ_C*-measurable*, if $f^{-1}(U) \in \Sigma_C$ for all open $U \subseteq \mathbb{R}$. According to [105, Definition 1.4.35], f is called *simple* if it is Σ_C-measurable and if it only takes on finitely many values, say $\lambda_1, \ldots, \lambda_n$ with $n \in \mathbb{Z}_{>0}$. Setting $A_i := f^{-1}(\{\lambda_i\})$ for all $i = 1, \ldots, n$, it is easily seen that f then admits a representation of the form $f = \sum_{i=1}^{n} \lambda_i \mathbb{I}_{A_i}$, and, following [105, Definition 1.4.34 and Definition 1.4.37], it holds that

$$\int_\Omega f \, \mathrm{d}\mu_C = \sum_{i=1}^{n} \lambda_i \mu_C(A_i)\,,$$

where the integral on the left-hand side is understood in the usual (Lebesgue) sense to be taken with respect to the measure μ_C. Provided then that also $(A_i, C) \in \mathscr{C}$ for all $i = 1, \ldots, n$, we obtain the correspondence with the conditional expectation formula from Equation (2.4), where the terms $P(A_i \,|\, C)$ replace the terms $\mu_C(A_i)$ in the sum above.

If we want to obtain a workable notion of measurability in our present setting, we have to be careful about which preimages of f we require to be in $\mathscr{C}$; because $\mathscr{C}$ is structureless—as opposed to e.g. Σ_C above, which is a σ-algebra—we cannot simply work with preimages of open sets. With the aim of obtaining the result in Proposition 2.13 below, let us simply say that a function $f : \Omega \to \mathbb{R}$ is $\mathscr{C}$*-measurable conditional on* $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$, if the level sets $\{f \geq t\} := \{\omega \in \Omega : f(\omega) \geq t\}$ of f satisfy $(\{f \geq t\}, C) \in \mathscr{C}$ for all $t \in [\inf f, \sup f]$.

Let us now conclude this section with a straightforward integral formula for the conditional expectation of arbitrary bounded functions that are measurable in this sense, and that should look recognizable to readers that are mostly familiar with measure-theoretic probability. This result is a straightforward generalisation of [109, Theorem 8.18] to the conditional case; the heavy lifting for this result is performed there.

Proposition 2.13. *Let P be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$ and let $\mathbb{E}_P$ be its corresponding conditional expectation. Suppose that $f \in \mathbb{B}$ is $\mathscr{C}$-measurable conditional on $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$. Then $(f, C) \in \mathscr{D}_P$, and*

$$\mathbb{E}_P[f \,|\, C] = \inf f + \int_{\inf f}^{\sup f} P\big(\{f \geq t\} \,\big|\, C\big)\, \mathrm{d}t\,, \tag{2.5}$$

where the integral is understood in the Riemann sense.

Proof. Fix $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$ such that $f \in \mathbb{B}$ is $\mathscr{C}$-measurable conditional on C. Let $\tilde{E}$ be any coherent conditional prevision that corresponds to P and that is defined on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}$ and $(f, C) \in \mathscr{D}$. Let E be a coherent extension of $\tilde{E}$ to $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$, which exists by Theorem 2.6$_{52}$. Because $\tilde{E}$ corresponds to P and because E extends $\tilde{E}$, clearly also E corresponds to P. Consider the map $E_C : \mathbb{B} \to \mathbb{R}$ that is defined, for all $g \in \mathbb{B}$, as $E_C(g) := E[g \,|\, C]$. Then, because E is a coherent conditional prevision, and using [109, Definition 4.11 and Theorem 4.12], E_C is a coherent (*unconditional*) prevision on $\mathbb{B}$. Let $\mu : \mathscr{E}(\Omega) \to \mathbb{R}$ be defined as $\mu(A) := E_C(\mathbb{I}_A)$ for all $A \in \mathscr{E}(\Omega)$. Then, according to [109, Definition 8.21 and Corollary 8.23], μ is a probability charge on $\mathscr{E}(\Omega)$, and E_C is the unique coherent prevision on $\mathbb{B}$ such that $E_C(\mathbb{I}_A) = \mu(A)$ for all $A \in \mathscr{E}(\Omega)$. Therefore, and because $\mathscr{E}(\Omega)$ is an algebra of sets, it follows from [109, Theorem 4.42 and Theorem 8.18] that

$$E_C(f) = \inf f + \int_{\inf f}^{\sup f} E_C\left(\mathbb{I}_{\{f \geq t\}}\right) \mathrm{d}t\,.$$

Now, because f is $\mathscr{C}$-measurable conditional on C, we obtain for any $A := \{f \geq t\}$ with $t \in [\inf f, \sup f]$ that $P(A \,|\, C) = E[\mathbb{I}_A \,|\, C] = E_C(\mathbb{I}_A)$, where we used that E corresponds to P in the first equality, so that we obtain

$$\tilde{E}[f \,|\, C] = E[f \,|\, C] = E_C(f) = \inf f + \int_{\inf f}^{\sup f} P\big(\{f \geq t\} \,|\, C\big)\, \mathrm{d}t\,,$$

where for the first equality we used that $(f, C) \in \mathscr{D}$ and that E extends $\tilde{E}$. Because the coherent conditional prevision $\tilde{E}$ corresponding to P and its domain $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}} \subseteq \mathscr{D}$ and $(f, C) \in \mathscr{D}$ are arbitrary, we conclude from Definition 2.5$_{54}$ that $(f, C) \in \mathscr{D}_P$ and that $\mathbb{E}_P[f \,|\, C] = \tilde{E}[f \,|\, C] = \inf f + \int_{\inf f}^{\sup f} P(\{f \geq t\} \,|\, C)\, \mathrm{d}t$. □

2.2 Time, State, and Function Spaces

This section introduces some notation that will be fundamental in the remainder of this work. In Section 2.2.1, we introduce notation for, and operations on, (finite) sequences of time points. Notation for state spaces is introduced in Section 2.2.2$_{61}$, and elementary concepts for functions on these state spaces are discussed in Section 2.2.3$_{62}$.

2.2.1 *Time Domains and Sequences of Time Points*

Fundamental to any discussion about dynamical systems is a notion of *time*, i.e., a dimension along which the states of such systems evolve. A *time domain* is simply an index set for the trajectory of a dynamical system. We will reserve the symbol $\mathbb{H}$ to denote a generic time domain, for which we use the following definition.

Definition 2.6 (Time Domain)**.** *A* time domain *is a set* $\mathbb{H} \subseteq \mathbb{R}_{\geq 0}$ *such that* $\min \mathbb{H}$ *exists and* $\sup \mathbb{H} = +\infty$.

In this definition, the requirements that $\min \mathbb{H}$ exists and that $\sup \mathbb{H} = +\infty$ are made to guarantee that the system's initial state is well-defined, and that "time never stops".

Although the above provides a general definition, we will restrict attention to two particular types of time domains in this work. The first of these is when $\mathbb{H} = \mathbb{R}_{\geq 0}$, which we call the *continuous time domain*. A dynamical system (resp. stochastic process, Markov chain, *etcetera*) with time domain $\mathbb{H} = \mathbb{R}_{\geq 0}$ is called a *continuous-time dynamical system* (resp. stochastic process, Markov chain, *etcetera*). The other type that we consider are *discrete* time domains. Although there is no unique discrete time domain, the prototypical one is arguably $\mathbb{H} = \mathbb{Z}_{\geq 0}$. We will use the following general definition in the remainder of this work.

Definition 2.7 (Discrete Time Domain)**.** *Let* $\mathbb{D}$ *be a time domain such that there is a strictly monotone bijection from* $\mathbb{Z}_{\geq 0}$ *to* $\mathbb{D}$*. Then* $\mathbb{D}$ *is called a* discrete time domain.

Here, the existence of a bijection from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$ guarantees that $\mathbb{D}$ is countably infinite, which is to say, that it is indeed a discrete set. The requirement that such a bijection (can be chosen to) be strictly monotone is largely made for convenience; given a time point $t \in \mathbb{D}$, it ensures the existence of a unique "next" time point in $\mathbb{D}$:

Lemma 2.14. *Let* $\mathbb{D}$ *be a discrete time domain. Then for any* $t \in \mathbb{D}$*, there is a unique* $s \in \mathbb{D}$ *for which* $t < s$ *and such that there is no* $r \in \mathbb{D}$ *for which* $t < r < s$ *(using the natural ordering of* $\mathbb{D}$ *as a subset of* $\mathbb{R}_{\geq 0}$*).*

Proof. Let τ be a monotone bijection from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$, which exists by Definition 2.7. Fix any $t \in \mathbb{D}$. Then because τ is a bijection from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$, there is some unique $n \in \mathbb{Z}_{\geq 0}$ such that $\tau(n) = t$. Let $s := \tau(n+1)$. Then $s \in \mathbb{D}$ is well-defined because τ is a bijection, and $t < s$ because $n < n+1$ and τ is strictly monotone.

Now fix any $r \in \mathbb{D}$; we will show that it does not hold that $t < r < s$. Clearly, we may proceed under the assumption that $t \neq r$ and $s \neq r$. Then there is some unique $m \in \mathbb{Z}_{\geq 0}$ such that $\tau(m) = r$. Because $r \neq t$ it follows from the fact that τ is a bijection that $m \neq n$. Similarly, $m \neq n+1$ because $r \neq s$. We now consider two remaining cases. First, suppose that $m < n$. Then also $r = \tau(m) < \tau(n) = t$ because τ is strictly monotone, and hence it does not hold that $t < r < s$. For the other case, suppose that

$n+1 < m$. Then $s = \tau(n+1) < \tau(m) = r$ because τ is strictly monotone, and then also not $t < r < s$.

To show the uniqueness, let s' be any element of $\mathbb{D}$ such that $t < s'$, and such that there is no $r \in \mathbb{D}$ for which $t < r < s'$. Because there is no $r \in \mathbb{D}$ for which $t < r < s$, it follows that in particular either $s' \leq t$, or $s \leq s'$. Because $t < s'$ by assumption, we must have $s \leq s'$. Since by assumption it does not hold that $t < s < s'$, it follows that $s = s'$. □

The following proposition establishes that the bijection in Definition 2.7↶ is unique.

Proposition 2.15. *Let $\mathbb{D}$ be a discrete time domain, and let τ and τ' be two strictly monotone bijections from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$. Then $\tau = \tau'$.*

Proof. Because $\mathbb{D}$ is a time domain, there is some $c \in \mathbb{H}$ such that $\min \mathbb{D} = c$. This implies that $\tau(0) = c = \tau'(0)$ because τ and τ' are strictly monotone bijections from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$. Now suppose *ex absurdo* that $\tau \neq \tau'$. Then there is some $n \in \mathbb{Z}_{\geq 0}$ such that $\tau(n) \neq \tau'(n)$. Let us suppose without loss of generality that $\tau(n) < \tau'(n)$.

Because τ' is a bijection, there is some $m \in \mathbb{Z}_{\geq 0}$ such that $\tau'(m) = \tau(n)$. Moreover, because τ' is strictly monotone, and because $\tau'(m) = \tau(n) < \tau'(n)$, we know that $m < n$. Because also τ is strictly monotone, this implies that $\tau(m) < \tau(n)$. Hence we conclude that also $\tau(m) < \tau'(m)$ with $m < n$. We can now keep repeating this argument, finding $k < m$ such that $\tau(k) < \tau'(k)$, and so on, until we reach the conclusion that $\tau(0) < \tau'(0)$. This is a contradiction with the argument at the beginning of this proof. □

Later on, we will need to make explicit use of this unique bijection. We use the following terminology in the remainder of this work.

Definition 2.8. *Let $\mathbb{D}$ be a discrete time domain, and let $\tau : \mathbb{Z}_{\geq 0} \to \mathbb{D}$ represent the unique strictly monotone bijection from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$, whose existence is guaranteed by Definition 2.7↶ and Proposition 2.15. Then we call τ the* canonical time index for $\mathbb{D}$, *and, for all $n \in \mathbb{Z}_{\geq 0}$, we denote its value in n as τ_n.*

Analogous to the continuous-time setting, when $\mathbb{D}$ is a discrete time domain, we call a dynamical system (resp. stochastic process, Markov chain, *etcetera*) with time domain $\mathbb{H} = \mathbb{D}$ a *discrete-time dynamical system* (resp. stochastic process, Markov chain, *etcetera*).

Because most of this work is concerned with continuous-time systems, we will take the case $\mathbb{H} = \mathbb{R}_{\geq 0}$ to be the implicit default. Hence, where appropriate, we will in the remainder simplify notation if no confusion should arise.

Moving on, for the remainder of this section, fix any time domain $\mathbb{H}$. We next discuss some ways to manipulate finite sequences of time points; a *time point* is simply an element of $\mathbb{H}$. A finite sequence of time points is either empty, or of the form $u{:=}t_0,t_1,\dots,t_n$, with $n\in\mathbb{Z}_{\geq 0}$ and, for all $i\in\{0,\dots,n\}$, $t_i\in\mathbb{H}$. These sequences are taken to be ordered, meaning that for all $i,j\in\{0,\dots,n\}$ with $i<j$, it holds that $t_i\leq t_j$. Let $\mathscr{U}^{\mathbb{H}}$ denote the set of all such finite (or empty) sequences in $\mathbb{H}$ that are *non-degenerate*, meaning that for all $u\in\mathscr{U}^{\mathbb{H}}$ with $u=t_0,\dots,t_n$, it holds that $t_i\neq t_j$ for all $i,j\in\{0,\dots,n\}$ such that $i\neq j$. We also define $\mathscr{U}^{\mathbb{H}}_{\supset\emptyset}:=\mathscr{U}^{\mathbb{H}}\setminus\{\emptyset\}$.

For any non-empty finite sequence u of time points, let $\max u:=\max\{t_i : i\in\{0,\dots,n\}\}$. For any time point $t\in\mathbb{H}$, we then write $t>u$ if $t>\max u$, and similarly for other inequalities. If $u=\emptyset$, then $t>u$ is taken to be trivially true, regardless of the value of t. We use $\mathscr{U}^{\mathbb{H}}_{<t}$ to denote the subset of $\mathscr{U}^{\mathbb{H}}$ that consists of those sequences $u\in\mathscr{U}^{\mathbb{H}}$ for which $u<t$, and similarly for other inequalities. Moreover, for any $u_1,u_2\in\mathscr{U}^{\mathbb{H}}$ such that $u_1\neq\emptyset$ and $u_2\neq\emptyset$, we write $u_1<u_2$ if $\max u_1<\min u_2$. If either u_1 or u_2 is empty then $u_1<u_2$ is taken to be trivially true.

Since a sequence $u\in\mathscr{U}^{\mathbb{H}}$ is a subset of $\mathbb{H}$, we can use set-theoretic notation to operate on such sequences. The result of such operations is again taken to be ordered. For example, for any $u,v\in\mathscr{U}^{\mathbb{H}}$, we use $u\cup v$ to denote the ordered union of u and v. In particular, for any $s\in\mathbb{H}$ and any $u\in\mathscr{U}^{\mathbb{H}}_{<s}$ with $u=t_0,\dots,t_n$, we use $u\cup\{s\}$ to denote the sequence $t_0,\dots,t_n,s$. In the trivial case that u is empty, we of course simply have that the sequence $u\cup\{s\}$ equals s.

Finally, for the special case that $\mathbb{H}=\mathbb{R}_{\geq 0}$, we consider finite sequences of time points that partition a given time interval $[t,s]\subset\mathbb{R}_{\geq 0}$, with $t,s\in\mathbb{R}_{\geq 0}$ such that $t\leq s$. Such a sequence is taken to include the end-points of this interval. Thus, the sequence is of the form $t=t_0<t_1<\cdots<t_n=s$. We denote the set of all such sequences by $\mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$, and note that this set never contains the empty sequence. Since we take these sequences to be non-degenerate, it follows that $\mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,t]}$ consists of the single sequence $u=\{t\}$.

For any $u\in\mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$ with $t<s$ and $u=t_0,\dots,t_n$, we define the sequential differences $\Delta^u_i:=t_i-t_{i-1}$, for all $i\in\{1,\dots,n\}$. We then use $\sigma(u):=\max\{\Delta^u_i : i\in\{1,\dots,n\}\}$ to denote the maximum such difference, which is also called the *mesh* of u. For notational convenience, for any sequence $u\in\mathscr{U}_{[t,t]}$, which as discussed above consists of a single time point $u=\{t\}$, we let $\sigma(u):=0$.

2.2.2 *States and Joint State Spaces*

Throughout this work, we will consider some fixed finite *state space* $\mathscr{X}$. A generic element of this set is called a *state* and will be denoted by x.

For a given time domain $\mathbb{H}$, we will often find it convenient to explicitly indicate the time point $t \in \mathbb{H}$ that is being considered, in which case we write $\mathscr{X}_t := \mathscr{X}$ to denote the state space at time t, and x_t to denote a state at time t. This notational trick also allows us to introduce some notation for the joint state at (multiple) explicit time points. For any finite sequence of time points $u \in \mathscr{U}^{\mathbb{H}}$ such that $u = t_0, \ldots, t_n$, we use

$$\mathscr{X}_u := \bigtimes_{t \in u} \mathscr{X}_t$$

to denote the joint state space at the time points in u. A joint state $x_u \in \mathscr{X}_u$ is a tuple $(x_{t_0}, \ldots, x_{t_n}) \in \mathscr{X}_{t_0} \times \cdots \times \mathscr{X}_{t_n}$ that specifies a state x_{t_k} for every time point t_k in u. Note that if u only contains a single time point t, then we simply have that $\mathscr{X}_u = \mathscr{X}_{\{t\}} = \mathscr{X}_t = \mathscr{X}$. If $u = \emptyset$, then $x_\emptyset \in \mathscr{X}_\emptyset$ is a "dummy" placeholder, which typically leads to statements that are vacuously true.

2.2.3 *Functions, Norms, and Operators*

We collect all real-valued functions $f : \mathscr{X} \to \mathbb{R}$ on $\mathscr{X}$ in the set $\mathscr{L}(\mathscr{X})$. Note that, if we were to fix an ordering on $\mathscr{X}$, then $\mathscr{L}(\mathscr{X})$ could be identified with the space $\mathbb{R}^{|\mathscr{X}|}$. However, we do not really need this identification and therefore proceed without explicitly fixing such an ordering. It should also be noted that $\mathscr{L}(\mathscr{X})$ is a vector space under the usual operations of addition and scalar multiplication. For that reason, we will in the sequel use the terms "function" and "vector" interchangeably when referring to elements of $\mathscr{L}(\mathscr{X})$.

For any time domain $\mathbb{H}$ and any $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$, we use $\mathscr{L}(\mathscr{X}_u)$ to denote the set of all real-valued functions on $\mathscr{X}_u$. We equip $\mathscr{L}(\mathscr{X}_u)$ with the supremum norm, so, for any $f \in \mathscr{L}(\mathscr{X}_u)$ the norm $\|f\|$ is defined as

$$\|f\| := \|f\|_\infty := \max_{x_u \in \mathscr{X}_u} |f(x_u)| .$$

Note that this is agnostic about the particular time domain that is being used; what matters to determine $\mathscr{X}_u$—and therefore f and $\|f\|$—is only the cardinality of u, which is finite for any $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$; the semantics of the time domain do not matter. Moreover, observe that we have $\|f\| = \max\{|f(x)| : x \in \mathscr{X}\}$ for any $f \in \mathscr{L}(\mathscr{X})$ as a special case.

Linear maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ will play an important role in this work, and we collect them in the set $\mathbb{M}$. For any $T \in \mathbb{M}$, we use the notation $Tf \in \mathscr{L}(\mathscr{X})$ to denote the image of $f \in \mathscr{L}(\mathscr{X})$ under T. As is well-known—and as we explain in detail in Appendix A.3383—these linear maps can be equivalently represented using *matrices*, whose x,y-entry $T(x,y)$ is identified by $T(x,y) := T\mathbb{I}_y(x)$, where $\mathbb{I}_y$ is the indicator of y, and with $x,y \in \mathscr{X}$. For any $x \in \mathscr{X}$, we use $T(x,\cdot)$ to denote the x-row of this matrix representation of T.

Because the state space $\mathscr{X}$ is taken to be fixed, we will always consider square, real-valued matrices with dimension $|\mathscr{X}| \times |\mathscr{X}|$. In terms of this matrix representation, it holds that $Tf(x) = \sum_{y\in\mathscr{X}} T(x,y)f(y)$ for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$, which shows that Tf is simply the matrix-vector product of T with f. Analogously, for any $T,S \in \mathbb{M}$ and any $x,y \in \mathscr{X}$ it holds that $TS(x,y) = \sum_{z\in\mathscr{X}} T(x,z)S(z,y)$, which represents the composition TS of T and S in terms of their matrix product. We will henceforth identically and interchangeably refer to the elements of $\mathbb{M}$ as "matrices" without cause for confusion. The symbol I will be reserved throughout to refer to the identity matrix; this is the identity map on $\mathscr{L}(\mathscr{X})$ and satisfies $I(x,x) = 1$ and $I(x,y) = 0$ for all $x,y \in \mathscr{X}$ with $x \neq y$.

Because we will also be interested in non-linear maps, we consider as a more general case operators that are *non-negatively homogeneous*. An operator T from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ is non-negatively homogeneous if $T(\lambda f) = \lambda Tf$ for all $f \in \mathscr{L}(\mathscr{X})$ and all $\lambda \in \mathbb{R}_{\geq 0}$. We emphasise that this includes matrices as a special case.

For any non-negatively homogeneous operator T from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, we consider the induced operator norm

$$\|T\| := \sup\big\{\|Tf\| : f \in \mathscr{L}(\mathscr{X}), \|f\| = 1\big\}. \tag{2.6}$$

For any matrix $T \in \mathbb{M}$, it is well-known—and proved in Proposition A.32389 for completeness—that then

$$\|T\| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)|. \tag{2.7}$$

We note that the space $\mathbb{M}$ equipped with this norm is a finite-dimensional normed vector space under the usual operations of addition and scalar multiplication, and therefore in particular, due to Proposition A.7374, it is a *Banach space*; a *complete* metric space under the metric induced by this norm. This implies that any sequence in $\mathbb{M}$ that is convergent with respect to this norm, has a limit that is also in $\mathbb{M}$. For these and other concepts that are needed for the analysis in the normed spaces $\mathscr{L}(\mathscr{X})$ and $\mathbb{M}$, and on which we rely intensively throughout this work, we refer to Appendix A369. Moreover, we there also present some results about (real-valued) linear functionals on $\mathscr{L}(\mathscr{X})$, which are relevant in some of our proofs.

To conclude this section, we note that the norms introduced above satisfy the following properties; Reference [17] provides a proof for the non-trivial ones.

Proposition 2.16. *For all $f,g \in \mathscr{L}(\mathscr{X})$, all T,S from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ that are non-negatively homogeneous, all $\lambda \in \mathbb{R}$ and all $x \in \mathscr{X}$, we have that*

N1: $\|f\| \geq 0$

N2: $\|f\| = 0 \Leftrightarrow f = 0$

N3: $\|f+g\| \leq \|f\| + \|g\|$

N4: $\|\lambda f\| = |\lambda|\,\|f\|$

N5: $|f(x)| \leq \|f\|$

N6: $\|T\| \geq 0$

N7: $\|T\| = 0 \Leftrightarrow T = 0$

N8: $\|T+S\| \leq \|T\| + \|S\|$

N9: $\|\lambda T\| = |\lambda|\,\|T\|$

N10: $\|TS\| \leq \|T\|\,\|S\|$

N11: $\|Tf\| \leq \|T\|\,\|f\|$

2.3 Stochastic Processes as Coherent Conditional Probabilities

We will now use the machinery introduced in the previous sections to define stochastic processes formally. In particular, a stochastic process with time domain $\mathbb{H}$ is simply a coherent conditional probability on a specific domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$, or equivalently, the restriction of a full conditional probability to this domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$—see Definition 2.12$_{68}$ further on. However, before we get to this definition, let us first provide some intuition.

Basically, a stochastic process describes the behaviour of a system as it moves through the—finite—state space $\mathscr{X}$ over the time domain $\mathbb{H}$. A single realisation of this movement is called a path or a trajectory. We are typically uncertain about the specific path that will be followed, and a stochastic process quantifies this uncertainty by means of a probabilistic model, which, in our case, will be a coherent conditional probability. These ideas are formalised as follows.

A *path* ω is a function from $\mathbb{H}$ to $\mathscr{X}$, and we denote with $\omega(t)$ the value of ω at time $t \in \mathbb{H}$. For any sequence of time points $u \in \mathscr{U}^{\mathbb{H}}$ and any path ω, we will write $\omega|_u$ to denote the restriction of ω to $u \subset \mathbb{H}$. Using this notation, we write for any $x_u \in \mathscr{X}_u$ that $\omega|_u = x_u$ if, for all $t \in u$, it holds that $\omega(t) = x_t$.

An outcome space $\Omega_{\mathbb{H}}$ of a stochastic process with time domain $\mathbb{H}$ is a set of paths. Some authors impose properties on these paths that may depend on $\mathbb{H}$; some commonly considered choices are to let $\Omega_{\mathbb{Z}_{\geq 0}}$ be the set of all paths, and to let $\Omega_{\mathbb{R}_{\geq 0}}$ be either the set of all paths, the set of all right-continuous paths [82], or the set of all càdlàg paths (right-continuous paths with left-sided limits) [90]. However, our results do not require such a specific choice. For the purposes of this work, we will use the following definition:

Definition 2.9. *For any time domain* $\mathbb{H}$*, an* outcome space $\Omega_{\mathbb{H}}$ *of a stochastic process with time domain* $\mathbb{H}$ *is a set of paths* $\omega : \mathbb{H} \to \mathscr{X}$ *that satisfies:*

$$(\forall u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset})(\forall x_u \in \mathscr{X}_u)(\exists \omega \in \Omega_{\mathbb{H}})\ \omega|_u = x_u\,. \tag{2.8}$$

Thus, an outcome space $\Omega_{\mathbb{H}}$ must be chosen in such a way that, for any non-empty finite sequence of time points $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$ and any state assignment $x_u \in \mathscr{X}_u$ on those time points, there is at least one path $\omega \in \Omega_{\mathbb{H}}$ that agrees with x_u on u. This assumption, although fairly minimal, is nevertheless crucial for many of our results and we will often refer to it in our proofs. That said, in the remainder of this work we will typically assume implicitly that the outcome space $\Omega_{\mathbb{H}}$ for a stochastic process with time domain $\mathbb{H}$ is fixed and chosen arbitrarily, provided that Equation (2.8) is satisfied.

For any set of events $\mathscr{E} \subseteq \mathscr{E}(\Omega_{\mathbb{H}})$, we use $\langle\mathscr{E}\rangle$ to denote the algebra that is generated by them. That is, $\langle\mathscr{E}\rangle$ is the smallest subset of $\mathscr{E}(\Omega_{\mathbb{H}})$ that contains all elements of $\mathscr{E}$, and that is furthermore closed under complements in $\Omega_{\mathbb{H}}$ and finite unions, and therefore also under finite intersections. Moreover, for any $t \in \mathbb{H}$ and $x \in \mathscr{X}$, we define the elementary event

$$(X_t = x)_{\mathbb{H}} := \{\omega \in \Omega_{\mathbb{H}} \colon \omega(t) = x\}\,,$$

and, for any $u \in \mathscr{U}^{\mathbb{H}}$, we let

$$\mathscr{E}^{\mathbb{H}}_u := \{(X_t = x)_{\mathbb{H}} \colon x \in \mathscr{X}, t \in u \cup \mathbb{H}_{>u}\}$$

be the set of elementary events whose time point either follows or belongs to u, and we let $\mathscr{A}^{\mathbb{H}}_u := \langle\mathscr{E}^{\mathbb{H}}_u\rangle$ be the algebra that is generated by this set of elementary events. It will be useful to characterise a bit more explicitly the structure of such algebras of events. To this end, we give the following two results.

Lemma 2.17 ([80, Proposition 1.2.2])**.** *Let* Ω *be a set and let* $\mathscr{C}$ *be a collection of subsets of* Ω*, i.e. let* $E \subseteq \Omega$ *for all* $E \in \mathscr{C}$*. For any* $E \subseteq \Omega$*, let* $E^c := \Omega \setminus E$ *denote the complement of* E *in* Ω*. Now let*

$$\mathscr{C}_0 := \mathscr{C} \cup \left\{E^c : E \in \mathscr{C}\right\} \cup \{\emptyset, \Omega\}$$

be the set containing all elements of $\mathscr{C}$*, their complements, and* $\emptyset$ *and* Ω*, let*

$$\mathscr{C}_1 := \left\{\cap_{i=1}^n E_i : E_i \in \mathscr{C}_0 \text{ for all } i = 1, \dots, n,\ n \in \mathbb{Z}_{>0}\right\},$$

be the set of all finite intersections of elements of $\mathscr{C}_0$*, and let*

$$\mathscr{A} := \left\{\cup_{i=1}^n E_i : E_i \cap E_j = \emptyset, E_i, E_j \in \mathscr{C}_1 \text{ for all } i,j \in \{1,\dots,n\},\ n \in \mathbb{Z}_{>0}\right\},$$

be the set of all finite unions of pairwise disjoint elements of $\mathscr{C}_1$.

Then $\mathscr{A} = \langle \mathscr{C} \rangle$ is the algebra generated by $\mathscr{C}$.

Essentially, the above result tells us that we can write events in a generated algebra in a kind of "normal form", *viz.* as unions of intersections of elementary events and their complements (and the entire and empty set). For the specific kind of algebras that we consider here, we can use this result to obtain even simpler expressions for their elements, as shown by the following result.

Proposition 2.18. *Let $\mathbb{H}$ be a time domain, and fix any $u \in \mathscr{U}^{\mathbb{H}}$ and any $A \in \mathscr{A}_u^{\mathbb{H}}$. Then there are $v \in \mathscr{U}^{\mathbb{H}}$ and $S \subseteq \mathscr{X}_v$ such that $v \subset u \cup \mathbb{H}_{>u}$ and $A = \cup_{x_v \in S}(X_v = x_v)_{\mathbb{H}}$.*

Proof. Because $\mathscr{A}_u^{\mathbb{H}} = \langle \mathscr{E}_u^{\mathbb{H}} \rangle$, it follows from Lemma 2.17↶ that there is some $n \in \mathbb{Z}_{>0}$ and, for all $i = 1, \ldots, n$, some $n_i \in \mathbb{Z}_{>0}$ and, for all $j = 1, \ldots, n_i$, some $E_{i,j}$ such that either $E_{i,j} \in \{\emptyset, \Omega_{\mathbb{H}}\}$, $E_{i,j} \in \mathscr{E}_u^{\mathbb{H}}$, or $E_{i,j}^c \in \mathscr{E}_u^{\mathbb{H}}$, for which $A = \cup_{i=1}^n \cap_{j=1}^{n_i} E_{i,j}$. Moreover, for all $E_{i,j}$ there are $t_{i,j} \in u \cup \mathbb{H}_{>u}$ and $S_{i,j} \subseteq \mathscr{X}$ such that $E_{i,j} = \cup_{x \in S_{i,j}}(X_{t_{i,j}} = x)_{\mathbb{H}}$. In particular, if $E_{i,j} \in \mathscr{E}_u^{\mathbb{H}}$ then $E_{i,j} = (X_{t_{i,j}} = x)_{\mathbb{H}}$ for some $x \in \mathscr{X}$, whence in that case $S_{i,j} = \{x\}$. On the other hand, if $E_{i,j}^c \in \mathscr{E}_u^{\mathbb{H}}$ then $E_{i,j} = (X_{t_{i,j}} = x)_{\mathbb{H}}^c$ for some $x \in \mathscr{X}$, and then $S_{i,j} = \mathscr{X} \setminus \{x\}$. For the final two cases, if $E_{i,j} = \emptyset$ then $S_{i,j} = \emptyset$, and if $E_{i,j} = \Omega_{\mathbb{H}}$ then $S_{i,j} = \mathscr{X}$; and in both of these latter cases $t_{i,j} \in u \cup \mathbb{H}_{>u}$ may be taken arbitrarily.

Now let $v \in \mathscr{U}^{\mathbb{H}}$ be the ordered union of all $\{t_{i,j}\}$, with $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$; then clearly $v \subset u \cup \mathbb{H}_{>u}$. Fix any $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$, and let $S_{i,j}^* := \left\{x_v \in \mathscr{X}_v : x_{t_{i,j}} \in S_{i,j}\right\}$. Then it holds that

$$E_{i,j} = \bigcup_{x \in S_{i,j}} (X_{t_{i,j}} = x)_{\mathbb{H}} = \bigcup_{x_v \in S_{i,j}^*} (X_v = x_v)_{\mathbb{H}}\,.$$

Hence it follows that $A = \cup_{i=1}^n \cap_{j=1}^{n_i} \cup_{x_v \in S_{i,j}^*}(X_v = x_v)_{\mathbb{H}}$. Now let $S_i := \bigcap_{j=1}^{n_i} S_{i,j}^*$. Let us now show that

$$\bigcap_{j=1}^{n_i} \bigcup_{x_v \in S_{i,j}^*} (X_v = x_v)_{\mathbb{H}} = \bigcup_{x_v \in S_i} (X_v = x_v)_{\mathbb{H}}\,.$$

To see this, first fix any $\omega \in \cup_{x_v \in S_i}(X_v = x_v)_{\mathbb{H}}$. Then there is some $x_v' \in S_i$ such that $\omega|_v = x_v'$. Because $x_v' \in S_i$ it follows that $x_v' \in S_{i,j}^*$ for all $j = 1, \ldots, n_i$, which implies that $\omega \in \cup_{x_v \in S_{i,j}^*}(X_v = x_v)_{\mathbb{H}}$ for all $j = 1, \ldots, n_i$. This implies that ω belongs to $\cap_{j=1}^{n_i} \cup_{x_v \in S_{i,j}^*} (X_v = x_v)_{\mathbb{H}}$. Conversely, fix any $\omega \in \cap_{j=1}^{n_i} \cup_{x_v \in S_{i,j}^*} (X_v = x_v)_{\mathbb{H}}$. Then for all $j = 1, \ldots, n_i$ it holds that $\omega \in \cup_{x_v \in S_{i,j}^*}(X_v = x_v)_{\mathbb{H}}$, which implies that there is some $x_v^{(j)} \in S_{i,j}^*$ such that $\omega|_v = x_v^{(j)}$. Let $x'_v := \omega_v$. Then $x_v^{(j)} = \omega|_v = x'_v$ for all $j = 1, \dots, n_i$. Hence it follows that $x'_v \in S^*_{i,j}$ for all $j = 1, \dots, n_i$, and therefore that $x'_v \in S_i$. Because $\omega|_v = x'_v$ it follows that ω belongs to $\cup_{x_v \in S_i} (X_v = x_v)_{\mathbb{H}}$. Hence we have found that

$$A = \bigcup_{i=1}^{n} \bigcap_{j=1}^{n_i} \bigcup_{x_v \in S^*_{i,j}} (X_v = x_v)_{\mathbb{H}} = \bigcup_{i=1}^{n} \bigcup_{x_v \in S_i} (X_v = x_v)_{\mathbb{H}} \,.$$

Now simply let $S := \cup_{i=1}^{n} S_i$. Then $S \subseteq \mathscr{X}_v$, and $A = \cup_{x_v \in S} (X_v = x_v)_{\mathbb{H}}$. □

Consider now any $u \in \mathscr{U}^{\mathbb{H}}$. Then on the one hand, for any $A \in \mathscr{A}_u^{\mathbb{H}}$, it clearly holds that $A \in \mathscr{E}(\Omega_{\mathbb{H}})$. On the other hand, for any $x_u \in \mathscr{X}_u$, the event

$$(X_u = x_u)_{\mathbb{H}} := \{\omega \in \Omega_{\mathbb{H}} \,:\, \omega|_u = x_u\}$$

belongs to $\mathscr{E}(\Omega_{\mathbb{H}})_{\supset \emptyset}$, because it follows from Equation (2.8)$_{65}$ that this event is non-empty. Now, if for any $A \in \mathscr{A}_u^{\mathbb{H}}$ and $x_u \in \mathscr{X}_u$ we let $(A, X_u = x_u)_{\mathbb{H}} := \big(A, (X_u = x_u)_{\mathbb{H}}\big)$, then it follows that $(A, X_u = x_u)_{\mathbb{H}} \in \mathscr{E}(\Omega_{\mathbb{H}}) \times \mathscr{E}(\Omega_{\mathbb{H}})_{\supset \emptyset}$. It is also worth noting that if $u = \emptyset$, then $\omega|_u = x_u$ is vacuously true, which implies that in that case, $(X_u = x_u)_{\mathbb{H}} = \Omega_{\mathbb{H}}$.

We can now introduce the domains of the stochastic processes that we consider, as follows.

Definition 2.10 (Domain)**.** *We let*

$$\mathscr{C}^{\mathrm{SP}}_{\mathbb{R}_{\geq 0}} := \Big\{ (A, X_u = x_u)_{\mathbb{R}_{\geq 0}} : u \in \mathscr{U}^{\mathbb{R}_{\geq 0}},\, x_u \in \mathscr{X}_u,\, A \in \mathscr{A}_u^{\mathbb{R}_{\geq 0}} \Big\}$$

be the set of conditional events that will constitute the domain of continuous-time stochastic processes. Conversely, for any discrete time domain $\mathbb{D}$ with canonical time-index τ, we let

$$\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}} := \Big\{ (A, X_u = x_u)_{\mathbb{D}} \,:\, u \in \big\{ \{\tau_0, \dots, \tau_n\} \,|\, n \in \mathbb{Z}_{\geq 0} \big\} \cup \{\emptyset\},\, x_u \in \mathscr{X}_u,\, A \in \mathscr{A}_u^{\mathbb{D}} \Big\}$$

be the set of conditional events that will constitute the domain of discrete-time stochastic processes with time domain $\mathbb{D}$.

When we do not want or need to distinguish between continuous- or discrete time domains, we generically let $\mathbb{H} \in \{\mathbb{R}_{\geq 0}, \mathbb{D}\}$, for any discrete time domain $\mathbb{D}$, and then write $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ to denote the domain of stochastic processes with time domain $\mathbb{H}$; see Definition 2.11$_{\curvearrowright}$ below.

Due to our above considerations, it holds that $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ is a subset of $\mathscr{E}(\Omega_{\mathbb{H}}) \times \mathscr{E}(\Omega_{\mathbb{H}})_{\supset \emptyset}$, regardless of whether $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$ for some discrete time domain $\mathbb{D}$. However, it should be noted that the sets

$\mathscr{C}^{\mathrm{SP}}_{\mathbb{R}_{\geq 0}}$ and $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ are not entirely structurally similar. In particular, $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ contains only events $(A, X_u = x_u)_{\mathbb{D}}$ for which the conditioning event $(X_u = x_u)_{\mathbb{D}}$ depends on *all* time-points $u = \tau_0, \ldots, \tau_n$ up to τ_n (or is trivial). Conversely, the events $(A, X_u = x_u)_{\mathbb{R}_{\geq 0}}$ are such that u is either empty, or u is a finite subset of $\mathbb{R}_{\geq 0}$; thus there are always time-points between the time-points on which the conditioning event $(X_u = x_u)_{\mathbb{R}_{\geq 0}}$ depends. Consequently, we can slightly simplify the characterisation of $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$.

Lemma 2.19. *For any discrete time domain $\mathbb{D}$ with canonical time index τ,*

$$\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}} = \Big\{ (A, X_u = x_u)_{\mathbb{D}} : u \in \big\{ \{\tau_0, \ldots, \tau_n\} \,|\, n \in \mathbb{Z}_{\geq 0} \big\} \cup \{\emptyset\},\ x_u \in \mathscr{X}_u,\ A \in \mathscr{A}^{\mathbb{D}}_{\emptyset} \Big\}.$$

Proof. Clearly, it suffices to prove that, for any $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$, it holds that $\mathscr{A}^{\mathbb{D}}_u = \mathscr{A}^{\mathbb{D}}_{\emptyset}$. To this end, we first note that $\emptyset \cup \mathbb{D}_{>\emptyset} = \mathbb{D}$, and hence

$$\mathscr{E}^{\mathbb{D}}_{\emptyset} = \Big\{ (X_t = x)_{\mathbb{D}} : x \in \mathscr{X},\ t \in \mathbb{D} \Big\}.$$

Next we use that, by definition, τ is a strictly monotone bijection from $\mathbb{Z}_{\geq 0}$ to $\mathbb{D}$, to write

$$u \cup \mathbb{D}_{>u} = \{\tau_0, \ldots, \tau_n\} \cup \Big\{ \tau_m : m \in \mathbb{Z}_{\geq 0}, m > n \Big\} = \{\tau_m : m \in \mathbb{Z}_{\geq 0}\} = \mathbb{D}.$$

Hence it follows that $\mathscr{E}^{\mathbb{D}}_u = \mathscr{E}^{\mathbb{D}}_{\emptyset}$, whence $\mathscr{A}^{\mathbb{D}}_u = \langle \mathscr{E}^{\mathbb{D}}_u \rangle = \langle \mathscr{E}^{\mathbb{D}}_{\emptyset} \rangle = \mathscr{A}^{\mathbb{D}}_{\emptyset}$. □

Moving on, we can now finally formalise our definition of a stochastic process.

Definition 2.11 (Stochastic Process)**.** *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain. A* stochastic process with time domain $\mathbb{H}$ *is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$. We denote the set of all stochastic processes with time domain $\mathbb{H}$ by $\mathbb{P}^{\mathbb{H}}$.*

Corollary 2.20. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let P be a real-valued map from $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ to $\mathbb{R}$. Then P is a stochastic process if and only if it is the restriction of a full conditional probability on $\mathscr{E}(\Omega_{\mathbb{H}}) \times \mathscr{E}(\Omega_{\mathbb{H}})_{\supset\emptyset}$.*

Proof. Trivial consequence of Corollary 2.449. □

The following two definitions provide us with the particular stochastic processes that are of interest in this work.

Definition 2.12 (Continuous-Time Stochastic Process)**.** *Let P be a stochastic process with time domain $\mathbb{H} = \mathbb{R}_{\geq 0}$. Then P is called a* continuous-time stochastic process.

For notational brevity, when no confusion should arise, we drop the time domain $\mathbb{H}$ from our notation when talking about continuous-time stochastic processes: a continuous-time stochastic process is a coherent conditional probability on the domain $\mathscr{C}^{\mathrm{SP}}$; continuous-time paths are collected in the set Ω; continuous-time events are denoted $(X_t = x)$; and so forth. Moreover, in these cases we also often simply write "stochastic processes" to mean "continuous-time stochastic processes".

As mentioned before, the other type of stochastic process in which we are interested are the discrete-time ones. We give the general definition here, but will study these processes in more depth in Chapter 3_{83}.

Definition 2.13 (Discrete-Time Stochastic Process)**.** *Let $\mathbb{D}$ be a discrete time domain, and let P be a stochastic process with time domain $\mathbb{H} = \mathbb{D}$. Then P is called a* discrete-time stochastic process with time domain $\mathbb{D}$.

Let us now turn to the motivation for our definition(s) of stochastic processes. There are two reasons why we restrict ourselves to the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$. The most important reason is simply that all of the important results in this work can be expressed using only this domain, because *they are all concerned with events or functions that depend on a finite number of time points*.

The second reason is that this restriction will allow us to state uniqueness results that *do not necessarily extend to larger domains*; see for example Corollary 5.5_{186}. However, it is important to realise that our restriction of the domain does not impose any real limitations, because, as we know from Theorem 2.3_{49}, the domain of a coherent conditional probability—and hence also a stochastic process—can always be extended. Hence, if one is interested in results on larger domains, it suffices to work with these extensions. The cost that one pays for this, however, is potentially a loss of uniqueness; some results may only hold for *certain* extensions. We briefly touch on some of these questions in Chapter 8_{363}.

A third reason is that, in particular for the discrete-time setting, the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ allows us to connect our models to other results in the literature; we will rely on these connections for some of the core results in Chapter 3_{83}.

Finally, we would like to point out that it is also possible to use a different—yet equivalent—definition for stochastic processes. Indeed, due to Corollary 2.20, a stochastic process can also be defined as the restriction of a full conditional probability to $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$, regardless of whether $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is any discrete time domain. Of course, given this observation, one may start to wonder why we have gone through the trouble of introducing coherence, because this alternative definition would not require this concept. The reason why we nevertheless need coherence, is that it allows us to establish the existence of stochastic processes that have certain properties, which will often be necessary in the proofs of our results.

For instance, for the sake of providing an illustration, suppose that we are given an arbitrary function $p : \mathbb{R}_{>0} \to [0,1]$, and that we want to know if there is a continuous-time stochastic process P for which, for some $x, y \in \mathscr{X}$,

$$P(X_t = y \,|\, X_0 = x) = p(t) \text{ for all } t \in \mathbb{R}_{>0}. \tag{2.9}$$

Had we defined a continuous-time stochastic process as the restriction of a full conditional probability to $\mathscr{C}^{\mathrm{SP}}$, without introducing coherence, then answering this question would have been entirely non-trivial, because it would essentially require us to construct a full conditional probability that coincides with p on the relevant part of its domain.

In contrast, as illustrated by the following example, the introduction of coherence takes care of most of the heavy lifting in such an existence proof.

Example 2.3. Let $\mathscr{X}$ be a state space that contains at least two states, fix any two—possibly equal—states $x, y \in \mathscr{X}$, and consider any function $p : \mathbb{R}_{>0} \to [0,1]$. The aim of this example is to prove that there is a (continuous-time) stochastic process P that satisfies Equation (2.9).

The crucial step of the proof is to consider a smaller (than $\mathscr{C}^{\mathrm{SP}}$) domain

$$\mathscr{C} := \{(X_t = y, X_0 = x) \colon t \in \mathbb{R}_{>0}\},$$

and a function $\tilde{P}$ that is defined by

$$\tilde{P}(X_t = y | X_0 = x) := p(t) \text{ for all } (X_t = y, X_0 = x) \in \mathscr{C}, \tag{2.10}$$

and to prove that this function is a coherent conditional probability on $\mathscr{C}$, or equivalently, that it satisfies Definition 2.2$_{48}$.

So consider any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \dots, n\}$, some $(X_{t_i} = y, X_0 = x) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. According to Definition 2.2$_{48}$, we now have to show that

$$\max\left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_x(\omega(0)) \left(\tilde{P}(X_{t_i} = y | X_0 = x) - \mathbb{I}_y(\omega(t_i))\right) \,\middle|\, \omega \in C_0 \right\} \geq 0, \tag{2.11}$$

with $C_0 := \bigcup_{i=1}^{n} (X_0 = x) = (X_0 = x)$. In doing so, we can assume without loss of generality that $i \neq j$ implies $t_i \neq t_j$, because if $t_i = t_j$ for some $i \neq j$, then we can simply sum the corresponding two summands in Equation (2.11).

Let $z \in \mathscr{X}$ be any state such that $z \neq y$, let $u := (0, t_1, \dots, t_n) \in \mathscr{U}$, and let $x_u \in \mathscr{X}_u$ be the unique state assignment such that $x_0 := x$ and

$$x_{t_i} := \begin{cases} y & \text{if } \lambda_i < 0 \\ z & \text{if } \lambda_i \geq 0 \end{cases} \quad \text{for all } i \in \{1, \dots, n\}.$$

Furthermore, let $N_{<0} \coloneqq \{i \in \{1,\ldots,n\} \colon \lambda_i < 0\}$ and $N_{\geq 0} \coloneqq \{i \in \{1,\ldots,n\} \colon \lambda_i \geq 0\}$. Since n is finite, Equation $(2.8)_{65}$ now guarantees that there is some $\omega \in \Omega$ such that $\omega|_u = x_u$. Evaluating the sum in Equation (2.11) using this ω, we find that

$$\begin{aligned}
&\sum_{i=1}^{n} \lambda_i \mathbb{I}_x(\omega(0)) \left(\tilde{P}(X_{t_i} = y | X_0 = x) - \mathbb{I}_y(\omega(t_i))\right) \\
&= \sum_{i=1}^{n} \lambda_i \left(\tilde{P}(X_{t_i} = y | X_0 = x) - \mathbb{I}_y(\omega(t_i))\right) \\
&= \sum_{i \in N_{<0}} \lambda_i \left(\tilde{P}(X_{t_i} = y | X_0 = x) - 1\right) + \sum_{i \in N_{\geq 0}} \lambda_i \tilde{P}(X_{t_i} = y | X_0 = x) \\
&\geq \sum_{i \in N_{<0}} \lambda_i \left(\tilde{P}(X_{t_i} = y | X_0 = x) - 1\right) = \sum_{i \in N_{<0}} |\lambda_i| \left(1 - \tilde{P}(X_{t_i} = y | X_0 = x)\right) \geq 0,
\end{aligned}$$

where the two inequalities follow from the fact that $\tilde{P}(X_{t_i} = y | X_0 = x) = p(t_i) \in [0,1]$. Furthermore, because $\omega(0) = x$, we also have that $\omega \in (X_0 = x) = C_0$. Therefore, we find that Equation (2.11) indeed holds. Hence, we conclude that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$.

The rest of the proof is now straightforward. Since $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$, and because $\mathscr{C}$ is a subset of $\mathscr{C}^{\mathrm{SP}}$, it follows from Theorem 2.3_{49} that $\tilde{P}$ can be extended to a coherent conditional probability P on $\mathscr{C}^{\mathrm{SP}}$, or equivalently, to a stochastic process P. Since this stochastic process P is an extension of $\tilde{P}$, Equation (2.9) is now an immediate consequence of Equation (2.10). ◊

2.4 Inferences for Stochastic Processes

With the formalisms to represent stochastic processes in place, let us conclude this chapter by discussing how to use these models to reason about their behaviour. In the sequel, we will refer to this as making *inferences* about them. Broadly speaking, an inference is a quantified statement about the system of interest, taking into account both our knowledge of, and our uncertainty about, the system's behaviour. Formally, an inference will typically be a *conditional expectation* of a function of interest, taken with respect to the coherent conditional probabilities that constitute our uncertainty model.

As explained in Section $2.1.1_{51}$, our general definition of conditional expectations only covers functions whose corresponding coherent conditional prevision follows uniquely from the coherent conditional probabilities that are specified. It is therefore important to know for which functions this is the case. The vast majority of this dissertation focusses on a particular type of functions, which we call u-measurable functions. As we will see in Proposition 2.23_{73} further on, the conditional previsions (and thus expectations) of such functions are uniquely and easily determined, at least for suitably chosen conditioning events. Let us start with developing this result.

Definition 2.14. *Let $\mathbb{H}$ be a time domain. A real-valued function on $\Omega_{\mathbb{H}}$ is called u-*measurable*, for a specific $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$, if for all $\omega, \omega' \in \Omega_{\mathbb{H}}$ for which $\omega|_u = \omega'|_u$, it holds that $f(\omega) = f(\omega')$.*

So, a function is called u-measurable, if its value in $\omega \in \Omega_{\mathbb{H}}$ only depends on the value of $\omega|_u \in \mathscr{X}_u$. Such functions are bounded:

Proposition 2.21. *Let $\mathbb{H}$ be a time domain, and consider any $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$. Then any u-measurable function is bounded.*

Proof. Because any particular $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$ is finite, with $u = t_0, \dots, t_n$ and $n \in \mathbb{Z}_{\geq 0}$, say, a u-measurable function can take at most $|\mathscr{X}|^{n+1}$ different values, all of which are in $\mathbb{R}$. Hence this function obtains its extremal values in $\mathbb{R}$, and thus is bounded. □

There is an obvious correspondence between elements of the set $\mathscr{L}(\mathscr{X}_u)$ and u-measurable functions on $\Omega_{\mathbb{H}}$. The following definition introduces the requisite notation to obtain this correspondence.

Definition 2.15. *Let $\mathbb{H}$ be a time domain. For any $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$ and any $f \in \mathscr{L}(\mathscr{X}_u)$, we introduce the function $f(X_u) : \Omega_{\mathbb{H}} \to \mathbb{R}$ that is defined, for all $\omega \in \Omega_{\mathbb{H}}$, as $f(X_u)(\omega) := f(\omega|_u)$.*

Proposition 2.22. *Let $\mathbb{H}$ be a time domain. For any $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$, the map $f \mapsto f(X_u)$ is a bijection from $\mathscr{L}(\mathscr{X}_u)$ to the set of u-measurable functions on $\Omega_{\mathbb{H}}$.*

Proof. Fix any $u \in \mathscr{U}_{\supset\emptyset}^{\mathbb{H}}$, and first choose any $f \in \mathscr{L}(\mathscr{X}_u)$. Then it follows from Definition 2.15 that $f(X_u)$ is u-measurable, according to Definition 2.14. So the map is indeed from $\mathscr{L}(\mathscr{X}_u)$ to the set of u-measurable functions on $\Omega_{\mathbb{H}}$.

Now consider any $g \in \mathscr{L}(\mathscr{X}_u)$, and suppose that $f(X_u) = g(X_u)$. Fix any $x_u \in \mathscr{X}_u$ and any $\omega \in \Omega_{\mathbb{H}}$ such that $\omega|_u = x_u$; this ω exists due to Equation $(2.8)_{65}$. Then it holds that

$$f(x_u) = f(\omega|_u) = f(X_u)(\omega) = g(X_u)(\omega) = g(\omega|_u) = g(x_u),$$

from which it follows that $f = g$ because $x_u \in \mathscr{X}_u$ was arbitrary. Thus the map is injective (i.e., one-to-one).

For the other direction, fix any u-measurable function f on $\Omega_{\mathbb{H}}$. We now identify $g \in \mathscr{L}(\mathscr{X}_u)$ by setting, for all $x_u \in \mathscr{X}_u$, its value $g(x_u) := f(\omega)$, for any $\omega \in \Omega_{\mathbb{H}}$ such that $\omega|_u = x_u$; this ω exists due to Equation $(2.8)_{65}$. Note that $g(x_u)$ is then uniquely determined because

f is u-measurable, thus the specific choice of ω does not matter here. Specifically, because f is u-measurable it follows that, for any $\omega \in \Omega_{\mathbb{H}}$, it holds that $g(\omega|_u) = f(\omega)$. Let $g(X_u)$ be the u-measurable function identified by $g \in \mathscr{L}(\mathscr{X}_u)$. Then it follows that, for any $\omega \in \Omega_{\mathbb{H}}$, it holds that $g(X_u)(\omega) = g(\omega|_u) = f(\omega)$, whence $g(X_u) = f$, and thus the map is surjective (i.e., onto). $\square$

We can use this mapping between functions $f \in \mathscr{L}(\mathscr{X}_u)$ on a stochastic process' state space $\mathscr{X}_u$ at a finite number of time points, and functions $f(X_u)$ on this process' realisations, to talk about expectations of these functions f on the states at specific points in time. Conversely, given a u-measurable function on $\Omega_{\mathbb{H}}$, we can identify with it an element of $\mathscr{L}(\mathscr{X}_u)$ whose value $f(x_u)$ denotes the common value of this u-measurable function in all $\omega \in \Omega_{\mathbb{H}}$ for which $\omega|_u = x_u$. We then write $f(X_u)$ to emphasise this u-measurable function's dependence on the time points u. In the sequel, we will often make these identifications implicitly.

The following result provides an explicit and straightforward expression for the kind of conditional expectations with which we will be most concerned in the remainder of this work.

Proposition 2.23. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let $P \in \mathbb{P}^{\mathbb{H}}$ be a stochastic process. Then, for any $u, v \in \mathscr{U}^{\mathbb{H}}$ such that $v \neq \emptyset$ and $v \subset u \cup \mathbb{H}_{>u}$, any $f \in \mathscr{L}(\mathscr{X}_v)$, and any $x_u \in \mathscr{X}_u$, it holds that*

$$\mathbb{E}_P[f(X_v) \,|\, X_u = x_u] = \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u)\,,$$

provided that if $\mathbb{H} = \mathbb{D}$, it also holds that either $u = \emptyset$ or $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$, where τ is the canonical time index of $\mathbb{D}$.

Proof. Let us first prove that

$$(X_v = y_v, X_u = x_u)_{\mathbb{H}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{H}} \quad \text{for all } y_v \in \mathscr{X}_v. \tag{2.12}$$

Because $v \subset u \cup \mathbb{H}_{>u}$ it follows that $(X_v = y_v)_{\mathbb{H}} \in \mathscr{A}^{\mathbb{H}}_u$ for all $y_v \in \mathscr{X}_v$. Hence it follows from Definition 2.10$_{67}$ that Equation (2.12) is satisfied, where, if $\mathbb{H} = \mathbb{D}$ is a discrete time domain with canonical time index τ, we use the additional assumption that $u = \emptyset$ or $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$.

Now consider the function $F : \Omega_{\mathbb{H}} \to \mathbb{R}$, defined for all $\omega \in \Omega_{\mathbb{H}}$ as

$$F(\omega) := \sum_{y_v \in \mathscr{X}_v} f(y_v) \mathbb{I}_{(X_v = y_v)_{\mathbb{H}}}(\omega)\,. \tag{2.13}$$

Then, using Equation (2.12)$_{\curvearrowleft}$, it follows from Proposition 2.12$_{56}$ that

$$\mathbb{E}_P[F \,|\, X_u = x_u] = \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u)\,.$$

It remains to show that the function F in Equation (2.13)$_{\curvearrowleft}$ is exactly $f(X_v)$. To this end, fix any $\omega \in \Omega_{\mathbb{H}}$. It is clear that $\mathbb{I}_{(X_v = y_v)_{\mathbb{H}}}(\omega) = 1$ if and only if $\omega|_v = y_v$. Hence

$$F(\omega) = \sum_{y_v \in \mathscr{X}_v} f(y_v) \mathbb{I}_{(X_v = y_v)_{\mathbb{H}}}(\omega) = f(\omega|_v)\,,$$

and therefore $F = f(X_v)$ by Definition 2.15$_{72}$. □

Let us next state the following property, which is well-known to hold for expectations as they are typically defined, but which is worth verifying for our current formalism.

Proposition 2.24. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let $P \in \mathbb{P}^{\mathbb{H}}$ be a stochastic process. Then for any $u, v \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$ such that $v \subseteq u$, any $f \in \mathscr{L}(\mathscr{X}_v)$, and any $x_u \in \mathscr{X}_u$, it holds that*

$$\mathbb{E}_P[f(X_v) \,|\, X_u = x_u] = f(x_v)\,,$$

provided that if $\mathbb{H} = \mathbb{D}$, it also holds that either $u = \emptyset$ or $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$, where τ is the canonical time index of $\mathbb{D}$.

Proof. Because $v \subseteq u$ it follows that $v \subset u \cup \mathbb{H}_{>u}$. Therefore, and because $v \neq \emptyset$, it follows from Proposition 2.23$_{\curvearrowleft}$ that

$$\mathbb{E}_P[f(X_v) \,|\, X_u = x_u] = \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u)\,. \tag{2.14}$$

Because $v \subseteq u$ it holds that $(X_u = x_u)_{\mathbb{H}} \subseteq (X_v = x_v)_{\mathbb{H}}$, and hence it follows from Property F2$_{47}$ that $P(X_v = x_v \,|\, X_u = x_u) = 1$. Due to Properties F1$_{47}$, F3$_{47}$, and F8$_{47}$ this implies that $P(X_v = y_v \,|\, X_u = x_u) = 0$ for all $y_v \in \mathscr{X}_v$ with $y_v \neq x_v$. By combining this with Equation (2.14) we find that

$$\mathbb{E}_P[f(X_v) \,|\, X_u = x_u] = f(x_v) P(X_v = x_v \,|\, X_u = x_u) = f(x_v)\,,$$

which concludes the proof. □

The following notational convention will also be useful.

Definition 2.16. *Let $\mathbb{H}$ be a time domain. For any $u, v \in \mathscr{U}^{\mathbb{H}}$ such that $v \neq \emptyset$ and $u \cap v = \emptyset$, any $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$, and any $x_u \in \mathscr{X}_u$, we define $f(x_u, X_v)$ to be the v-measurable function whose value in $\omega \in \Omega_{\mathbb{H}}$ is $f(x_u, \omega|_v)$.*[7]

[7] In the special case that $u = \emptyset$, we will adopt the convention that $f(x_u, X_v) = f(x_\emptyset, X_v)$ should simply represent $f(X_{u \cup v}) = f(X_v)$ itself, so that its value in ω is $f(x_\emptyset, \omega|_v) := f(\omega|_v)$.

Now, Proposition 2.24 already gave us a simple expression for conditional expectations of functions that depend on the state at time points that are fully contained in the conditioning event. Definition 2.16 allows us to extend this result to the case where only part of these time points are in the conditioning event, as follows.

Proposition 2.25. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let $P \in \mathbb{P}^{\mathbb{H}}$ be a stochastic process. Then for any $u, v \in \mathscr{U}^{\mathbb{H}}$ such that $v \neq \emptyset$ and $u < v$, any $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$, and any $x_u \in \mathscr{X}_u$, it holds that*

$$\mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u] = \mathbb{E}_P[f(x_u, X_v) \,|\, X_u = x_u]\,,$$

provided that if $\mathbb{H} = \mathbb{D}$, it also holds that either $u = \emptyset$ or $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$, where τ is the canonical time index of $\mathbb{D}$.

Proof. Because $v \neq \emptyset$ it follows that $u \cup v \neq \emptyset$. Moreover, because $u < v$ it follows that $v \subset u \cup \mathbb{H}_{>u}$ and, hence, that $u \cup v \subset u \cup \mathbb{H}_{>u}$. Therefore, it follows from Proposition 2.23$_{73}$ that

$$\mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u] = \sum_{y_{u \cup v} \in \mathscr{X}_{u \cup v}} f(y_{u \cup v}) P(X_{u \cup v} = y_{u \cup v} \,|\, X_u = x_u)\,. \quad (2.15)$$

Now first suppose that $u = \emptyset$. Then $u \cup v = v$, whence it follows that

$$\begin{aligned} \sum_{y_{u \cup v} \in \mathscr{X}_{u \cup v}} f(y_{u \cup v}) P(X_{u \cup v} = y_{u \cup v} \,|\, X_u = x_u) &= \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u) \\ &= \mathbb{E}_P[f(X_v) \,|\, X_u = x_u] \\ &= \mathbb{E}_P[f(x_u, X_v) \,|\, X_u = x_u]\,, \end{aligned}$$

where for the second equality we used Proposition 2.23$_{73}$, which is valid since $v \neq \emptyset$ and because we already established that $v \subset u \cup \mathbb{H}_{>u}$, and where for the third equality we used the convention that $f(x_u, X_v) = f(x_\emptyset, X_v) = f(X_v)$ established in Definition 2.16. This concludes the proof for the case where $u = \emptyset$.

Before we consider the case $u \neq \emptyset$, we next prove that

$$(X_v = y_v, X_u = x_u)_{\mathbb{H}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{H}} \quad \text{for all } y_v \in \mathscr{X}_v. \quad (2.16)$$

Because we already established that $v \subset u \cup \mathbb{H}_{>u}$ it follows that $(X_v = y_v)_{\mathbb{H}} \in \mathscr{A}_u^{\mathbb{H}}$ for all $y_v \in \mathscr{X}_v$. Hence it follows from Definition 2.10$_{67}$ that Equation (2.16) is satisfied, where, if $\mathbb{H} = \mathbb{D}$ is a discrete time domain with canonical time index τ, we use the additional assumption that $u = \emptyset$ or $u = \tau_0, \ldots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$.

Let us now consider the case that $u \neq \emptyset$. Fix any $y_u \in \mathscr{X}_u$ and $y_v \in \mathscr{X}_v$. We consider two cases: $y_u = x_u$, and $y_u \neq x_u$. So first suppose that $y_u = x_u$; it then holds that

$$\begin{aligned} P(X_u = y_u, X_v = y_v \,|\, X_u = x_u) &= P(X_u = x_u, X_v = y_v \,|\, X_u = x_u) \\ &= P(X_v = y_v \,|\, X_u = x_u)\,, \end{aligned}$$

where we used Property F6$_{47}$ for the second equality, and where we use Equation (2.16)$_{\frown}$ to ensure that these quantities are well-defined.

For the other case, suppose that $y_u \neq x_u$. Then it holds that

$$(X_u = y_u, X_v = y_v)_{\mathbb{H}} \cap (X_u = x_u)_{\mathbb{H}} = \emptyset\,,$$

and hence by Property F6$_{47}$ we find that

$$P(X_u = y_u, X_v = y_v \,|\, X_u = x_u) = P(\emptyset \,|\, X_u = x_u) = 0\,,$$

using Property F7$_{47}$ for the final equality. In summary, we have found that

$$P(X_u = y_u, X_v = y_v \,|\, X_u = x_u) = \begin{cases} P(X_v = y_v \,|\, X_u = x_u) & \text{if } y_u = x_u, \text{ and} \\ 0 & \text{otherwise}. \end{cases}$$

By combining this with Equation (2.15)$_{\frown}$, we find that

$$\begin{aligned} \mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u] &= \sum_{y_v \in \mathscr{X}_v} \sum_{y_u \in \mathscr{X}_u} f(y_{u \cup v}) P(X_{u \cup v} = y_{u \cup v} \,|\, X_u = x_u) \\ &= \sum_{y_v \in \mathscr{X}_v} f(x_u, y_v) P(X_v = y_v \,|\, X_u = x_u)\,. \end{aligned}$$

Using Definition 2.16$_{74}$, and because we already established that $v \subset u \cup \mathbb{H}_{>u}$, it now follows from Proposition 2.23$_{73}$ that

$$\begin{aligned} \mathbb{E}_P[f(x_u, X_v) \,|\, X_u = x_u] &= \sum_{y_v \in \mathscr{X}_v} f(x_u, y_v) P(X_v = y_v \,|\, X_u = x_u) \\ &= \mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u]\,, \end{aligned}$$

which concludes the proof. □

We note that for any $u, v \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$ and any $f \in \mathscr{L}(\mathscr{X}_v)$, the conditional expectation $\mathbb{E}_P[f(X_v) \,|\, X_u = x_u]$—provided it exists—is a real-valued function of $x_u \in \mathscr{X}_u$. As such, we can associate with it the u-measurable function $\mathbb{E}_P[f(X_v) \,|\, X_u]$, whose value in $\omega \in \Omega_{\mathbb{H}}$ we define as $\mathbb{E}_P[f(X_v) \,|\, X_u](\omega) := \mathbb{E}_P[f(X_v) \,|\, X_u = \omega|_u]$. Using this notation, we can state the following result, which is known as the *law of iterated expectation*.

Proposition 2.26. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let $P \in \mathbb{P}^{\mathbb{H}}$ be a stochastic process. Then for any $u \in \mathscr{U}^{\mathbb{H}}$ and any $v, w \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$ such that $u < v < w$, any $f \in \mathscr{L}(\mathscr{X}_{u\cup v\cup w})$, and any $x_u \in \mathscr{X}_u$, it holds that*

$$\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_u = x_u] = \mathbb{E}_P\Big[\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_{u\cup v}] \,\Big|\, X_u = x_u\Big],$$

provided that if $\mathbb{H} = \mathbb{D}$, it also holds that either $u = \emptyset$ and $v = \tau_0, \ldots, \tau_m$ with $m \in \mathbb{Z}_{\geq 0}$, or $u = \tau_0, \ldots, \tau_n$ and $v = \tau_{(n+1)}, \ldots, \tau_m$ with $n, m \in \mathbb{Z}_{\geq 0}$ such that $n < m$, where τ is the canonical time index of $\mathbb{D}$.

Proof. First note that because $u < v < w$ it holds that $u \cup v \subset u \cup \mathbb{H}_{>u}$, that $u \cup v \cup w \subset u \cup \mathbb{H}_{>u}$, and that $u \cup v \cup w \subset (u \cup v) \cup \mathbb{H}_{>(u\cup v)}$, and hence it follows from Definition 2.1067 that

$$(X_{u\cup v} = y_{u\cup v}, X_u = x_u)_{\mathbb{H}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{H}} \quad \text{for all } y_{u\cup v} \in \mathscr{X}_{u\cup v}, \tag{2.17}$$

that

$$(X_{u\cup v\cup w} = y_{u\cup v\cup w}, X_u = x_u)_{\mathbb{H}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{H}} \quad \text{for all } y_{u\cup v\cup w} \in \mathscr{X}_{u\cup v\cup w}, \tag{2.18}$$

and that

$$(X_{u\cup v\cup w} = y_{u\cup v\cup w}, (X_u = x_u, X_v = y_v))_{\mathbb{H}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{H}} \quad \text{for all } y_{u\cup v\cup w} \in \mathscr{X}_{u\cup v\cup w}, \tag{2.19}$$

where if $\mathbb{H} = \mathbb{D}$, we use the additional assumption that either $u = \emptyset$ and $v = \tau_0, \ldots, \tau_m$ with $m \in \mathbb{Z}_{\geq 0}$, or $u = \tau_0, \ldots, \tau_n$ and $v = \tau_{(n+1)}, \ldots, \tau_m$ with $n, m \in \mathbb{Z}_{\geq 0}$ such that $n < m$, where τ is the canonical time index of $\mathbb{D}$.

Now because $u < v < w$ it follows that $w \subset (u \cup v) \cup \mathbb{H}_{>(u\cup v)}$. Hence, it follows from Propositions 2.2575 and 2.2373 that, for all $y_v \in \mathscr{X}_v$,

$$\begin{aligned}
&\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_u = x_u, X_v = y_v] \\
&\quad = \mathbb{E}_P[f(x_u, y_v, X_w) \,|\, X_u = x_u, X_v = y_v] \\
&\quad = \sum_{z_w \in \mathscr{X}_w} f(x_u, y_v, z_w) P(X_w = z_w \,|\, X_u = x_u, X_v = y_v) \\
&\quad = \sum_{z_w \in \mathscr{X}_w} f(x_u, y_v, z_w) P(X_u = x_u, X_v = y_v, X_w = z_w \,|\, X_u = x_u, X_v = y_v),
\end{aligned} \tag{2.20}$$

where we used Property F647 for the final equality, together with the fact that all relevant events are in the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ of P, due to Equation (2.19).

Next let $g \in \mathscr{L}(\mathscr{X}_{u\cup v})$ be defined, for all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$, as

$$g(y_{u\cup v}) := \mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_{u\cup v} = y_{u\cup v}]. \tag{2.21}$$

Because $u < v$ it follows that $v \subset u \cup \mathbb{H}_{>u}$. Hence, it follows from Propositions 2.25$_{75}$ and 2.23$_{73}$ that

$$\begin{aligned}\mathbb{E}_P[g(X_{u\cup v}) \,|\, X_u = x_u] &= \mathbb{E}_P[g(x_u, X_v) \,|\, X_u = x_u] \\ &= \sum_{y_v \in \mathscr{X}_v} g(x_u, y_v) P(X_v = y_v \,|\, X_u = x_u) \\ &= \sum_{y_v \in \mathscr{X}_v} g(x_u, y_v) P(X_u = x_u, X_v = y_v \,|\, X_u = x_u)\,, \quad (2.22)\end{aligned}$$

where we used Property F6$_{47}$ for the final equality, together with the fact that all relevant events are in the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ of P, due to Equation (2.17)$_{\curvearrowleft}$.

Combining Equations (2.21)$_{\curvearrowleft}$, (2.22), and (2.20)$_{\curvearrowleft}$, we obtain

$$\begin{aligned}&\mathbb{E}_P\Big[\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_{u\cup v}] \,\Big|\, X_u = x_u\Big] \\ &= \sum_{y_v \in \mathscr{X}_v} P(X_u = x_u, X_v = y_v \,|\, X_u = x_u) \mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_u = x_u, X_v = y_v] \\ &= \sum_{y_v \in \mathscr{X}_v} P(X_u = x_u, X_v = y_v \,|\, X_u = x_u) \\ &\qquad \sum_{z_w \in \mathscr{X}_w} f(x_u, y_v, z_w) P(X_u = x_u, X_v = y_v, X_w = z_w \,|\, X_u = x_u, X_v = y_v) \\ &= \sum_{y_v \in \mathscr{X}_v} \sum_{z_w \in \mathscr{X}_w} f(x_u, y_v, z_w) P(X_u = x_u, X_v = y_v, X_w = z_w \,|\, X_u = x_u) \\ &= \mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_u = x_u]\,,\end{aligned}$$

where we used Property F4$_{47}$ for the third equality, together with the fact that all relevant events are in the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{H}}$ of P, due to Equation (2.18)$_{\curvearrowleft}$; and where we used Proposition 2.23$_{73}$ for the final equality, which is valid because $u < v < w$ implies that $u \cup v \cup w \subset u \cup \mathbb{H}_{>u}$. □

Let us conclude this section by stating some generally useful properties of conditional expectations of u-measurable functions, that we will use repeatedly in the remainder of this work.

Proposition 2.27. *Let $\mathbb{H}$ be a time domain such that either $\mathbb{H} = \mathbb{R}_{\geq 0}$ or $\mathbb{H} = \mathbb{D}$, where $\mathbb{D}$ is a discrete time domain, and let $P \in \mathbb{P}^{\mathbb{H}}$ be a stochastic process. Then, for any $u, v \in \mathscr{U}^{\mathbb{H}}$ such that $v \neq \emptyset$ and $v \subset u \cup \mathbb{H}_{>u}$, any $f, g \in \mathscr{L}(\mathscr{X}_v)$, any $\lambda, \mu \in \mathbb{R}$, and any $x_u \in \mathscr{X}_u$, it holds that*

CE1: $\min_{y_v \in \mathscr{X}_v} f(y_v) \leq \mathbb{E}_P[f(X_v) \,|\, X_u = x_u] \leq \max_{y_v \in \mathscr{X}_v} f(y_v)$;

CE2: $\mathbb{E}_P[f(X_v) + g(X_v) \,|\, X_u = x_u] = \mathbb{E}_P[f(X_v) \,|\, X_u = x_u] + \mathbb{E}_P[g(X_v) \,|\, X_u = x_u]$;

CE3: $\mathbb{E}_P[\lambda f(X_v) \,|\, X_u = x_u] = \lambda \mathbb{E}_P[f(X_v) \,|\, X_u = x_u]$;

CE4: $f \leq g \Rightarrow \mathbb{E}_P[f(X_v) \,|\, X_u = x_u] \leq \mathbb{E}_P[g(X_v) \,|\, X_u = x_u]$;

CE5: $\mathbb{E}_P[\mu \,|\, X_u = x_u] = \mu$;

CE6: $\mathbb{E}_P[f(X_v) + \mu \,|\, X_u = x_u] = \mathbb{E}_P[f(X_v) \,|\, X_u = x_u] + \mu$;

provided that if $\mathbb{H} = \mathbb{D}$, it also holds that either $u = \emptyset$ or $u = \tau_0, \dots, \tau_n$, with $n \in \mathbb{Z}_{\geq 0}$, where τ is the canonical time index of $\mathbb{D}$.

Proof. Because $f, g, f+g, \lambda f$, and $f+\mu$ are all in $\mathscr{L}(\mathscr{X}_v)$, Proposition 2.23$_{73}$ implies that the expectations of $f(X_v), f(X_v)+g(X_v), \lambda f(X_v)$, and $f(X_v)+\lambda$, conditional on $X_u = x_u$, are well-defined, whence $\big(f(X_v), X_u = x_u\big)$, $\big(g(X_v), X_u = x_u\big)$, $\big((f+g)(X_v), X_u = x_u\big)$, $\big(\lambda f(X_v), X_u = x_u\big)$, and $\big(f(X_v)+\mu, X_u = x_u\big)$ are all necessarily elements of $\mathscr{D}_P$ due to Definition 2.5$_{54}$.

Therefore, and due to Proposition 2.10$_{54}$, Property CE1 follows from Property E1$_{52}$ and Definition 2.15$_{72}$; Property CE2 follows from Property E2$_{52}$; and Property CE3 follows from Property E3$_{52}$.

For Property CE4, assume that $f \leq g$. Then it follows from Proposition 2.23$_{73}$ that

$$\begin{aligned}\mathbb{E}_P[f(X_v) \,|\, X_u = x_u] &= \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u) \\ &\leq \sum_{y_v \in \mathscr{X}_v} g(y_v) P(X_v = y_v \,|\, X_u = x_u) = \mathbb{E}_P[g(X_v) \,|\, X_u = x_u],\end{aligned}$$

where the inequality used Property F1$_{47}$ and the assumption that $f \leq g$.

For Property CE5, note that $\mu \in \mathbb{R}$ is trivially identified with the constant function in $\mathscr{L}(\mathscr{X}_v)$ whose value in all $y_v \in \mathscr{X}_v$ equals μ. Hence it follows from Proposition 2.23$_{73}$ that

$$\mathbb{E}_P[\mu \,|\, X_u = x_u] = \sum_{y_v \in \mathscr{X}_v} \mu P(X_v = y_v \,|\, X_u = x_u) = \mu,$$

where the final equality used Properties F3$_{47}$ and F8$_{47}$.

Property CE6 now follows by combining Properties CE2 and CE5. □

Appendix

2.A A Gambling Interpretation of Coherence

In this short appendix we aim to provide a basic exposition of the gambling interpretation for coherent conditional probabilities. A more extensive discussion, which also provides some historical context, can be found, amongst others, in References [4, 5, 87, 109, 112, 117, 118].

Basically, the idea is to interpret P as a set of gambles on the actual—but unknown—value of X in Ω, which some bettor is willing to either buy or sell, and to impose a rationality criterion on this set of gambles.

Concretely, for every pair $(A,C) \in \mathscr{C}$, $P(A|C)$ is interpreted as a bettor's fair price for a ticket that yields a reward of one currency unit to its holder if the event A occurs, and zero if it does not; provided that C also happens. In other words, the bettor is willing to either sell or buy such a ticket at this price, provided that the bet is *called off* if C does not happen. If a bet is called off, the bettor is refunded and no reward is obtained. Furthermore, it is also assumed that the bettor's utility is linear, which implies that she is willing to vary the stakes of her bets arbitrarily, and that multiple bets can be combined through summation.

Suppose for example that the actual value of X ends up being ω. For each ticket that the bettor sold, she has then received $P(A|C)$ currency units in advance, but after the value of X is revealed, she loses one currency unit if A has happened, that is, she loses $\mathbb{I}_A(\omega)$ currency units. Because the bet is called off if C does not happen, her net profit is $\mathbb{I}_C(\omega)(P(A|C) - \mathbb{I}_A(\omega))$, with negative profit being loss. Note that if C does not happen, that is, if $\mathbb{I}_C(\omega) = 0$, the bet is called off and she neither gains nor loses anything. Since we also allow for arbitrary stakes, we conclude that for any $\lambda \in \mathbb{R}_{\geq 0}$, the bettor is willing to accept the uncertain net profit $\lambda \mathbb{I}_C(\omega)(P(A|C) - \mathbb{I}_A(\omega))$.

Similarly, for each ticket that the bettor buys, she first has to pay $P(A|C)$ to buy the ticket, but will then receive one unit of currency if A happens. Her profit is then $\mathbb{I}_A(\omega) - P(A|C)$ per ticket. However, she only receives this profit if event C also came to pass, and otherwise gets refunded. Hence, if we again take into account that the stake can be chosen arbitrarily, we find that for any $\lambda \in \mathbb{R}_{\geq 0}$ the bettor is willing to accept the uncertain net profit $\lambda \mathbb{I}_C(\omega)(\mathbb{I}_A(\omega) - P(A|C))$.

By combining the arguments for selling and buying, we conclude from the above that for any $\lambda \in \mathbb{R}$, the bettor is willing to accept a bet in which she receives the uncertain net profit $\lambda \mathbb{I}_C(\omega)(P(A|C) - \mathbb{I}_A(\omega))$, with negative profit being loss.

Because the bettor's utility was assumed to be linear, it follows that the bettor must be willing to combine any finite number of such transactions. That is, if we consider any $n \in \mathbb{Z}_{>0}$ and, for every $i \in \{1, \ldots, n\}$, some $\lambda_i \in \mathbb{R}$ and $(A_i, C_i) \in \mathscr{C}$, then the bettor is willing to accept a bet in which her net profit is equal to

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\left(P(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\right).$$

The coherence of P is now equivalent to requiring that any such bet *avoids sure loss*, in the sense that there should be at least one "non-trivial" outcome ω for which her total profit is non-negative, the trivial case being when none of the events C_i happen—$\omega \notin \cup_{i=1}^{n} C_i$—because then all bets are called off and she gets refunded completely.

Incidentally, the interpretation of coherent conditional previsions, which we discussed in Section 2.1.1$_{51}$, is similar: essentially, any bounded function $f : \Omega \to \mathbb{R}$ can be understood as a ticket that yields a reward of $f(\omega)$ currency units when the value of X turns out to be $\omega \in \Omega$ (again, with negative profit being loss). A conditional prevision $E[f|C]$ can then be interpreted as a bettor's fair price for f, provided that the bet is called off if C does not happen, in which case the bettor is again refunded and no reward is obtained. In relation to the conditional expectations that we discussed in Section 2.1.1$_{51}$, this fair price is therefore the expectation, or the *expected value of f*, conditional on the event C obtaining. As above, the coherence condition of the conditional prevision requires that the bettor cannot be forced into accepting a (finite) collection of such transactions, which jointly would lead to her incurring a sure loss regardless of the actual outcome ω.

3

Discrete-Time (Imprecise-)Markov Chains

"We are here and it is now.
The way I see it is, after that, everything tends towards guesswork."

Terry Pratchett, "Small Gods"

We discussed in Chapter 2[45] that stochastic processes can be used to represent and reason about dynamical systems that behave in a manner about which we are uncertain. *Markov chains* are a particular type of stochastic process: they satisfy the so-called *Markov property*. This is a particular independence condition, which essentially states that the future behaviour of the system is independent of its historical behaviour, given its current state. Symbolically, in a discrete-time setting the Markov property may be written as saying that

$$P(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} \,|\, X_n = x_n)\,.$$

It is this crucial independence assumption that leads to these models being relatively easy to parameterise and tractable from both a computational and analytical point of view. For those reasons, Markov chains have arguably become one of the most popular and successful types of stochastic process, both in theory and in applications.

Like general stochastic processes, Markov chains have both continuous-time and discrete-time variants. Although the continuous-time setting is of primary interest in this dissertation, this current chapter focusses solely on discrete-time Markov chains. We do this for several reasons. First, from a didactic point of view: we find discrete-time models to be simply easier to work with and conceptualise. Because there are many analogies between these two settings, we can introduce some concepts here that we will need later, without worrying about the technical details surrounding continuous time domains.

Secondly, the imprecise-probabilistic generalisation of discrete-time Markov chains has already been studied extensively in the literature. As such, the content of this chapter should be viewed mostly as a collection and summary of results from the discrete-time (imprecise-)Markov chain literature, with an emphasis on concepts that will be useful when we study continuous-time imprecise-Markov chains in later chapters. Indeed, although we have included many proofs for didactic reasons, and because we felt it easier to prove them in our notation than to provide the translation to external work, the content of this chapter should not really be understood as novel. As we will explain throughout, we base ourselves in large part on developments in References [20–22, 69], and the origin of the ideas presented in this chapter can be found there.

Our final reason for including the discussion in this chapter, is with the aim of providing the (hitherto) unpublished results from Chapter 7$_{335}$, which in many ways provide a unification of the theories of imprecise-Markov chains in discrete and continuous time. These results require a formal basis also for discrete-time Markov chains, and one that is slightly more general than what is typically encountered in the literature; specifically, we need to deal with arbitrary discrete time domains, rather than just the natural numbers as is usually done. This slight generalisation is fairly straightforward, as the crucial observation is only this: any discrete time domain $\mathbb{D}$ is simply a re-labelling of the prototypical discrete time domain $\mathbb{Z}_{\geq 0}$. Nevertheless, we will be careful throughout in stating our results and proofs under this minor generalisation.

In terms of content, we start in Section 3.1 by introducing and discussing some concepts that are specific to the discrete-time setting, and that we did not want to include in Chapter 2$_{45}$. We also make the connection there to the work in [69], which provides the formal basis for many of our technical results. In Section 3.2$_{89}$, we then finally define discrete-time Markov chains using our formalism of stochastic processes. We discuss how they can be parameterised using (families of) *transition matrices*: row-stochastic matrices that encode the *transition probabilities* of these Markov chains. We also discuss how the linear maps encoded by these transition matrices can be used to represent the conditional expectations of Markov chains.

We finally come to the generalisation to *imprecise*-Markov chains in Section 3.3_{101}. There we discuss and study *sets* of transition matrices and *sets* of stochastic processes, and investigate their corresponding *lower-* and *upper expectations*. In Section 3.4_{116}, we study *lower transition operators*. As we shall see, these are essentially non-linear generalisations of transition matrices; in particular, we discuss their relationship with lower envelopes of sets of transition matrices. Having studied these objects in the abstract, we relate them to imprecise-Markov chains in Section 3.5_{121}, where we show that lower transition operators can be used to represent lower expectations of imprecise-Markov chains, in a manner analogous to how transition matrices are used for (precise) Markov chains.

3.1 Some Properties of Discrete-Time Stochastic Processes

In this section we discuss some concepts that will set up the results in the remainder of this chapter. In particular, we describe some simplifying concepts and expressions for the domain of discrete-time stochastic processes, and introduce the machinery that connects our formalism and results to those in the literature. Let us start by introducing some shorthand notation that will be exceptionally helpful in the remainder of this chapter. For any discrete time domain $\mathbb{D}$ with canonical time index τ, we will write $\tau_{0:n} := \tau_0, \ldots, \tau_n$ for any $n \in \mathbb{Z}_{\geq 0}$. Moreover, we will adopt the convention that $\tau_{0:(-1)} := \emptyset$.

To relate some of our technical results to existing work in the literature, we will mostly base ourselves on the results in [69]. To make this connection, we first introduce the notion of a *situation*; this is an event of the form $(X_u = x_u)_{\mathbb{D}}$, with $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, and $x_u \in \mathscr{X}_u$, where τ is the canonical time index of $\mathbb{D}$. Because $\mathbb{D}$ is discrete, such a situation fully describes the realisation of a stochastic process up to time τ_n. For notational convenience, we also consider the "initial" situation, corresponding to $u = \emptyset$. In the sequel, we use the following definition:

Definition 3.1 (Situations). *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ. Then we define the set $\mathscr{S}_{\mathbb{D}}$ of* situations *(with time domain $\mathbb{D}$) as*

$$\mathscr{S}_{\mathbb{D}} := \left\{ (X_u = x_u)_{\mathbb{D}} \,:\, u \in \{\tau_{0:n} \,|\, n \in \mathbb{Z}_{\geq 0}\} \cup \{\emptyset\}, x_u \in \mathscr{X}_u \right\}.$$

The following two results establish the connection between the[1] do-

[1] In fact, Reference [69] considers many different domains for stochastic processes, but we primarily concern ourselves with the one defined in [69, Equation 3.14].

main of discrete-time stochastic processes as they are defined in [69], and as we have defined them in Chapter 2_{45}. For the remainder of this section, we have moved the proofs of some results to Appendix 3.A$_{125}$.

Lemma 3.1. *Let $\mathbb{D}$ be a discrete time domain, and let $\mathscr{S}_{\mathbb{D}}$ be the set of situations with time domain $\mathbb{D}$. Then $\mathscr{A}_{\emptyset}^{\mathbb{D}} = \langle \mathscr{S}_{\mathbb{D}} \rangle$.*

Lemma 3.2. *Let $\mathbb{D}$ be a discrete time domain, let $\mathscr{S}_{\mathbb{D}}$ be the set of situations with time domain $\mathbb{D}$, and consider the set*

$$\mathscr{C}_{\mathbb{D}}^{*} := \Big\{ (A, C)_{\mathbb{D}} \,:\, A \in \langle \mathscr{S}_{\mathbb{D}} \rangle, C \in \mathscr{S}_{\mathbb{D}} \Big\}. \tag{3.1}$$

Then $\mathscr{C}_{\mathbb{D}}^{*} = \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$.

Proof. Using Lemma 2.19$_{68}$ and Definition 3.1$_{\curvearrowleft}$, we see that

$$\mathscr{C}_{\mathbb{D}}^{\mathrm{SP}} = \Big\{ (A, C)_{\mathbb{D}} \,:\, A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}, C \in \mathscr{S}_{\mathbb{D}} \Big\}.$$

Hence, and because $\langle \mathscr{S}_{\mathbb{D}} \rangle = \mathscr{A}_{\emptyset}^{\mathbb{D}}$ by Lemma 3.1, it follows from Equation (3.1) that $\mathscr{C}_{\mathbb{D}}^{*} = \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$. □

Moreover, the result from Proposition 2.18$_{66}$ that events can be written in a kind of "normal form", has a convenient corollary that states that discrete-time events can always be expressed as a union of situations, as follows.

Lemma 3.3. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ and consider any $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$. Then there is some $n \in \mathbb{Z}_{\geq 0}$ and some $S \subseteq \mathscr{X}_{\tau_{0:n}}$, such that $A = \cup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}}$.*

Now, one other concept that we need from the literature is that of a *probability tree* [69, Section 3.3]. For a given discrete time domain $\mathbb{D}$, this is a map

$$p : \mathscr{X} \times \mathscr{S}_{\mathbb{D}} : \big(x, (X_u = x_u)_{\mathbb{D}}\big) \mapsto p(x \,|\, x_u)$$

such that, for all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, $p(x \,|\, x_u)$, as a function of $x \in \mathscr{X}$, is a probability mass function (on $\mathscr{X}$). It can be interpreted as describing a stochastic process with time domain $\mathbb{D}$, in that for every situation $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ with $u = \tau_{0:n}$, it gives the probability $p(x \,|\, x_u)$ that the process will be in state x at time τ_{n+1}. Formally, we use the following definition.

Definition 3.2. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, let $P \in \mathbb{P}^{\mathbb{D}}$ be a discrete-time stochastic process with time domain $\mathbb{D}$,* *and let* $p : \mathscr{X} \times \mathscr{S}_{\mathbb{D}} \to \mathbb{R}$ *be a probability tree with time domain* $\mathbb{D}$*. We say that* P corresponds to p *if, for all* $x \in \mathscr{X}$*,*

$$P(X_{\tau_0} = x) = p(x \,|\, x_\emptyset) \tag{3.2}$$

and, for all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ *for which* $u = \tau_{0:n}$*, with* $n \in \mathbb{Z}_{\geq 0}$*,*

$$P(X_{\tau_{n+1}} = x \,|\, X_u = x_u) = p(x \,|\, x_u)\,. \tag{3.3}$$

As the following result makes clear, any discrete-time stochastic process that corresponds to a given probability tree in the manner above, has a convenient expression for the conditional probabilities that it assigns to situations. Our proof of this result is based on, and conceptually essentially the same as, the second part of the proof of [69, Lemma 14]; we here only provide the argument explicitly to deal with our slightly more general setting.

Lemma 3.4. *Let* $\mathbb{D}$ *be a discrete time domain with canonical time index* τ*, and let* p *be a probability tree with time domain* $\mathbb{D}$*. Let* $P \in \mathbb{P}^{\mathbb{D}}$ *be a discrete-time stochastic process with time domain* $\mathbb{D}$ *that corresponds to* p*. Consider any* $n, m \in \mathbb{Z}_{\geq 0}$ *and let* $u := \tau_{0:(n-1)}$ *and* $v := \tau_{0:m}$*. Then for all* $x_u \in \mathscr{X}_u$ *and* $y_v \in \mathscr{X}_v$*, it holds that*

$$P(X_v = y_v \,|\, x_u = x_u) = \begin{cases} \prod_{i=n}^{m} p(y_{\tau_i} \,|\, y_{\tau_{0:(i-1)}}) & \text{if } n \leq m \text{ and } y_u = x_u \\ 1 & \text{if } m < n \text{ and } y_v = x_v \\ 0 & \text{otherwise.} \end{cases}$$

We note that it follows from Lemmas 3.1, 3.2 and 3.3 that for any conditional event $(A, C)_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$, there are some $n, m \in \mathbb{Z}_{\geq 0}$ and, with $u := \tau_{0:(n-1)}$ and $v := \tau_{0:m}$, some $x_u \in \mathscr{X}_u$ and $S \subseteq \mathscr{X}_v$, such that $C = (X_u = x_u)_{\mathbb{D}}$ and $A = \cup_{y_v \in S} (X_v = y_v)_{\mathbb{D}}$, and it follows from Lemma 3.2 that $(X_v = y_v, X_u = x_u)_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ for all $y_v \in S$. Therefore, and because any discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$ is a coherent conditional probability by Definition 2.13$_{69}$, it follows from Property F3$_{47}$ that

$$P(A \,|\, C) = \sum_{y_v \in S} P(X_v = y_v \,|\, X_u = x_u)\,,$$

which, together with Lemma 3.4, provides an explicit expression for the conditional probability assigned to any conditional event $(A, C)_{\mathbb{D}}$ by a discrete-time stochastic process P that corresponds to a given probability tree p.

We note that, so far, we have not yet shown that there *exist* discrete-time stochastic processes that correspond to a given probability tree. The following result confirms that this is indeed the case. In fact, as is perhaps unsurprising in light of the preceding discussion, such processes are *uniquely* determined (on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$) by the probability tree.

Theorem 3.5. *Let $\mathbb{D}$ be a discrete time domain, and let p be a probability tree with time domain $\mathbb{D}$. Then there is a unique discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$ that corresponds to p.*

Lemmas 3.2_{86} and $3.4_{\curvearrowleft}$ and Theorem 3.5 together complete the connection with the work in Reference [69]. In particular, [69, Lemma 14] states these results for the special case where the time domain $\mathbb{D} = \mathbb{Z}_{>0}$, and where $\Omega_{\mathbb{Z}_{>0}}$ is the set of *all* paths with time domain $\mathbb{Z}_{>0}$. This is reflective of the typical literature on discrete-time stochastic processes, where one typically considers either $\mathbb{Z}_{>0}$ or $\mathbb{Z}_{\geq 0}$ as the time domain, and where this choice for the outcome space $\Omega_{\mathbb{Z}_{>0}}$ is common. Contrariwise, in this work we want to consider general discrete time domains $\mathbb{D}$ and general outcome spaces $\Omega_{\mathbb{D}}$; in particular for the time domains, we need this level of generality to later make the connection with continuous-time stochastic processes where, essentially, we consider the embedding of $\mathbb{D}$ in $\mathbb{R}_{\geq 0}$. As for the choice of the outcome spaces, we simply have no need to impose a stronger assumption than that used in Definition 2.9_{65}, in order to obtain our results.

To conclude this section, we note that Theorem 3.5 implies that discrete-time stochastic processes are uniquely determined by their value on the events that form the domain of a probability tree:

Corollary 3.6. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and consider two discrete-time stochastic processes $P, P' \in \mathbb{P}^{\mathbb{D}}$ such that, for all $x \in \mathscr{X}$,*

$$P(X_{\tau_0} = x) = P'(X_{\tau_0} = x)\,, \tag{3.4}$$

and, for all $n \in \mathbb{Z}_{\geq 0}$, all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, and all $x \in \mathscr{X}$,

$$P(X_{\tau_{n+1}} = x \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) = P'(X_{\tau_{n+1}} = x \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}})\,. \tag{3.5}$$

Then $P = P'$.

Proof. Let $p : \mathscr{X} \times \mathscr{S}_{\mathbb{D}} \to \mathbb{R}$ be the probability tree that is defined, for all $x \in \mathscr{X}$ and $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ as

$$p(x\,|\,x_u) := \begin{cases} P(X_{\tau_0} = x) & \text{if } u = \emptyset \\ P(X_{\tau_{n+1}=x} \,|\, X_u = x_u) & \text{if } u = \tau_{0:n} \text{ with } n \in \mathbb{Z}_{\geq 0}. \end{cases}$$

Then it follows from Theorem 3.5 that P is the *unique* element of $\mathbb{P}^{\mathbb{D}}$ that corresponds to p. However, due to Equations (3.4) and (3.5) it holds, for all $x \in \mathscr{X}$ and $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, that

$$p(x\,|\,x_u) = \begin{cases} P'(X_{\tau_0} = x) & \text{if } u = \emptyset \\ P'(X_{\tau_{n+1}=x} \,|\, X_u = x_u) & \text{if } u = \tau_{0:n} \text{ with } n \in \mathbb{Z}_{\geq 0}. \end{cases}$$

It follows from Theorem 3.5 that P' is *also* the unique element of $\mathbb{P}^{\mathbb{D}}$ that corresponds to p, and therefore $P = P'$. □

3.2 Markov Chains and Transition Matrices

Let us now move on to the discussion of Markov chains. A *Markov chain* is simply a stochastic process that satisfies a particular independence condition; for the discrete-time case, this is given by Equation (3.6) below.

Definition 3.3 (Discrete-Time Markov Chain)**.** *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ. A stochastic process $P \in \mathbb{P}^{\mathbb{D}}$ satisfies the* Markov property *if for all $n \in \mathbb{Z}_{\geq 0}$ and all $x_{\tau_n}, x_{\tau_{n+1}} \in \mathscr{X}$, there is some real number $P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) \in \mathbb{R}$ such that, for all $x_{\tau_{0:(n-1)}} \in \mathscr{X}_{\tau_{0:(n-1)}}$,*

$$P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) = P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n})\,. \tag{3.6}$$

If P satisfies this property then it is called a (discrete-time) Markov chain*. We denote the set of all Markov chains with time domain $\mathbb{D}$ as $\mathbb{P}^{\mathbb{D},\mathrm{M}}$. For any Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$, we refer to the quantities on the right-hand side of Equation* (3.6) *as the* transition probabilities *of P.*

As Equation (3.6) makes clear, the behaviour of a Markov chain at time τ_{n+1} only depends on the state of the process at time τ_n. In other words, given the state X_{τ_n}, the future behaviour of the system is probabilistically independent of the history of the system before time τ_n. It is this crucial property that makes Markov chains tractable to work with; as we shall see later on this chapter, this property leads to straightforward expressions for their conditional expectations, and makes them relatively straightforward to parameterise.

Note that the phrasing of the above definition is perhaps a bit awkward, in the sense that on the right-hand side of Equation (3.6), the transition probability $P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n})$ appears to have no direct connection with the process P, despite notational appearances. Indeed, both notationally and conceptually, this quantity is read as "the probability that P assigns to the conditional event $(X_{\tau_{n+1}} = x_{\tau_{n+1}}, X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} \in \mathscr{E}(\Omega_{\mathbb{D}}) \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$". We cannot however use this interpretation formally, as this conditional event is not in the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ of P. Nevertheless, as the next result shows, this interpretation can be made exact: for any Markov chain P, its transition probability $P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n})$ is the unique value assigned to the conditional event $(X_{\tau_{n+1}} = x_{\tau_{n+1}}, X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} \in \mathscr{E}(\Omega_{\mathbb{D}}) \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ by *any* coherent conditional probability P^* that extends P and has this event in its domain.

Proposition 3.7. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain. Then for any $n \in \mathbb{Z}_{\geq 0}$, any $x_{\tau_n}, x_{\tau_{n+1}} \in \mathscr{X}$, and any coherent conditional probability P^* on $\mathscr{C}$ that extends P, with $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}} \subseteq \mathscr{C} \subseteq \mathscr{E}(\Omega_{\mathbb{D}}) \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ such that*

$(X_{\tau_{n+1}} = x_{\tau_{n+1}}, X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} \in \mathscr{C}$*, it holds that*

$$P^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) = P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}), \tag{3.7}$$

where the right-hand side denotes the transition probabilities of P.

Proof. If $n = 0$ then $(X_{\tau_1} = x_{\tau_1}, X_{\tau_0} = x_{\tau_0})_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ and therefore, since P^* extends P, Equation (3.7) follows immediately.

So let us suppose that $n > 0$, and let $u := \tau_{0:(n-1)}$. Let $\tilde{P}^*$ be any full conditional probability on $\mathscr{E}(\Omega_{\mathbb{D}}) \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ that extends P^*, which exists by Corollary 2.449. Then it holds that

$$\begin{aligned}
&P^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = \tilde{P}^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = \tilde{P}^*\left(\bigcup_{x_u \in \mathscr{X}_u} (X_{\tau_{n+1}} = x_{\tau_{n+1}}, X_u = x_u)_{\mathbb{D}} \,\middle|\, X_{\tau_n} = x_{\tau_n} \right) \\
&\quad = \sum_{x_u \in \mathscr{X}_u} \tilde{P}^*(X_{\tau_{n+1}} = x_{\tau_{n+1}}, X_u = x_u \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = \sum_{x_u \in \mathscr{X}_u} \tilde{P}^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}, X_u = x_u) \tilde{P}^*(X_u = x_u \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = \sum_{x_u \in \mathscr{X}_u} P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}, X_u = x_u) \tilde{P}^*(X_u = x_u \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = \sum_{x_u \in \mathscr{X}_u} P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) \tilde{P}^*(X_u = x_u \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) \sum_{x_u \in \mathscr{X}_u} \tilde{P}^*(X_u = x_u \,|\, X_{\tau_n} = x_{\tau_n}) \\
&\quad = P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}),
\end{aligned}$$

where we used that $\tilde{P}^*$ extends P^* for the first equality; Property F347 for the third equality; Property F447 for the fourth equality; the fact that $\tilde{P}^*$ extends P^*, that P^* extends P, and that $(X_{\tau_{n+1}} = x_{\tau_n+1}, (X_{\tau_n} = x_{\tau_n}, X_u = x_u))_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ for the fifth equality; the fact that P is a Markov chain together with the definition of its transition probabilities for the sixth equality; the fact that the transition probabilities do not depend on x_u for the seventh equality; and Properties F347 and F847 for the last equality. □

Note that Proposition 3.7↶ holds even if $P(X_{\tau_n} = x_{\tau_n}) = 0$; if this probability is strictly positive then the claim follows trivially from Bayes's rule—Property F447— which makes the identification

$$P^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n}) = \frac{P^*(X_{\tau_{n+1}} = x_{\tau_{n+1}},\, X_{\tau_n} = x_{\tau_n})}{P^*(X_{\tau_n} = x_{\tau_n})},$$

provided that $P^*(X_{\tau_n} = x_{\tau_n}) > 0$. Since the events in both the numerator and denominator in this right-hand side are in the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ of P, this then provides the unique way to identify $P^*(X_{\tau_{n+1}} = x_{\tau_{n+1}} \,|\, X_{\tau_n} = x_{\tau_n})$ for any P^* that extends P. However, our use of full and coherent conditional probabilities—which, as explained in Chapter 245, takes conditional probabilities as elementary, rather than derived, entities—allows us sometimes to still make this unique identification even when the conditioning event has probability zero; Proposition 3.789 establishes that P being a Markov chain is a sufficient condition to do this.

Moving on, let us next introduce another property that is often encountered when working with Markov chains: that of *time-homogeneity*.

Definition 3.4. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain.*

Then P is called (time-)homogeneous *if*

$$P\big(X_{\tau_{n+1}} = y \,\big|\, X_{\tau_n} = x\big) = P\big(X_{\tau_1} = y \,\big|\, X_{\tau_0} = x\big)\,, \tag{3.8}$$

for all $n \in \mathbb{Z}_{\geq 0}$ and all $x, y \in \mathscr{X}$. We use $\mathbb{P}^{\mathbb{D},\mathrm{HM}}$ to denote the set of all homogeneous Markov chains with time domain $\mathbb{D}$.

So, for a Markov chain that is time-homogeneous—that satisfies Equation (3.8)—the probability of moving from a state $x \in \mathscr{X}$ at time τ_n to a state $y \in \mathscr{X}$ at time τ_{n+1}, is the same for all $n \in \mathbb{Z}_{\geq 0}$. In other words, its transition probabilities do not depend on the point in time at which they are considered.

It will be useful to consider a different way to describe the transition probabilities of Markov chains. To this end, let us introduce the notion of a *transition matrix*. As the following definition makes explicit, a transition matrix T is simply a matrix that is row-stochastic, meaning that, for each $x \in \mathscr{X}$, the row $T(x, \cdot)$ is a probability mass function on $\mathscr{X}$.

Definition 3.5 (Transition Matrix)**.** *A real-valued matrix T is said to be a* transition matrix *if*

T1: $\sum_{y \in \mathscr{X}} T(x, y) = 1$ *for all* $x \in \mathscr{X}$;

T2: $T(x, y) \geq 0$ *for all* $x, y \in \mathscr{X}$.

We will use $\mathbb{T}$ to denote the set of all transition matrices.

The following property of transition matrices will be crucial in the remainder of this dissertation.

Proposition 3.8. *For any two transition matrices $T, S \in \mathbb{T}$, their product TS is also a transition matrix.*

Proof. The proof is elementary, but since the result is so crucial we will verify it here. First fix any $x \in \mathscr{X}$. Then, using the properties of matrix multiplication, we have that

$$
\begin{aligned}
\sum_{y\in\mathscr{X}} (TS)(x,y) &= \sum_{y\in\mathscr{X}} \sum_{z\in\mathscr{X}} T(x,z)S(z,y) \\
&= \sum_{z\in\mathscr{X}} \sum_{y\in\mathscr{X}} T(x,z)S(z,y) \\
&= \sum_{z\in\mathscr{X}} T(x,z) \sum_{y\in\mathscr{X}} S(z,y) = \sum_{z\in\mathscr{X}} T(x,z) = 1\,,
\end{aligned}
$$

where for the last two equalities we used that $\sum_{y\in\mathscr{X}} S(z,y) = 1$ for all $z \in \mathscr{X}$ and $\sum_{z\in\mathscr{X}} T(x,z) = 1$ due to Property T1$_{\curvearrowleft}$. Because $x \in \mathscr{X}$ is arbitrary, this means that TS satisfies Property T1$_{\curvearrowleft}$.

Now, fix any $x,y \in \mathscr{X}$. Then it holds that

$$
(TS)(x,y) = \sum_{z\in\mathscr{X}} T(x,z)S(z,y) \geq 0\,,
$$

where the equality used the properties of matrix multiplication, and where the inequality follows from the fact that $T(x,z) \geq 0$ and $S(z,y) \geq 0$ for all $z \in \mathscr{X}$ due to Property T2$_{\curvearrowleft}$. Because $x,y \in \mathscr{X}$ are arbitrary, this means that TS satisfies Property T2$_{\curvearrowleft}$.

Hence, because TS satisfies both Property T1$_{\curvearrowleft}$ and T2$_{\curvearrowleft}$, it is a transition matrix by Definition 3.5$_{\curvearrowleft}$. ☐

We will also need the following straightforward property.

Lemma 3.9. *For any transition matrix T it holds that $\|T\| = 1$.*

Proof. Using Equation (2.7)$_{63}$, we find that

$$
\|T\| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} T(x,y) = 1\,,
$$

where we used Property T2$_{\curvearrowleft}$ for the second equality and Property T1$_{\curvearrowleft}$ for the third equality. ☐

As another general result on transition matrices, it will be useful to note that $\mathbb{T}$ is a complete metric space:

Proposition 3.10. *$\mathbb{T}$ is complete under the metric induced by our norm $\|\cdot\|$.*

Proof. Let $\{T_i\}_{i\in\mathbb{Z}_{>0}}$ be any Cauchy sequence in $\mathbb{T}$; as discussed in Appendix A.3$_{383}$, the space $\mathbb{M}$ is a finite-dimensional normed vector space, which is complete due to Proposition A.7$_{374}$. Since $\mathbb{T} \subseteq \mathbb{M}$ the limit $T_* := \lim_{i\to+\infty} T_i$ therefore exists in $\mathbb{M}$. We need to show that also $T_* \in \mathbb{T}$. So fix any $x, y \in \mathscr{X}$, and suppose *ex absurdo* that $T_*(x,y) < 0$. Then because $T_* = \lim_{i\to+\infty} T_i$ there is some $n \in \mathbb{Z}_{>0}$ such that $\|T_n - T_*\| < -T_*(x,y)$. Because $T_n \in \mathbb{T}$ it follows from Property T2$_{91}$ that $T_n(x,y) \geq 0$, and therefore, because $T_*(x,y) < 0$, that $T_n(x,y) - T_*(x,y) = |T_n(x,y) - T_*(x,y)|$. Hence it follows that

$$\begin{aligned} T_n(x,y) - T_*(x,y) &= |T_n(x,y) - T_*(x,y)| \\ &\leq \sum_{z\in\mathscr{X}} |T_n(x,z) - T_*(x,z)| \leq \|T_n - T_*\| < -T_*(x,y)\,, \end{aligned}$$

using Equation (2.7)$_{63}$ for the second inequality. Adding $T_*(x,y)$ to both sides of this equation yields $T_n(x,y) < 0$, which contradicts Property T2$_{91}$ and the fact that $T_n \in \mathbb{T}$. Hence $T_*(x,y) \geq 0$ and, because the $x, y \in \mathscr{X}$ were arbitrary, it follows that T_* satisfies property T2$_{91}$.

Next, fix any $x \in \mathscr{X}$, and let $f \in \mathscr{L}(\mathscr{X})$ be such that $f(y) := 1$ for all $y \in \mathscr{X}$. Then $T_*f(x) = \sum_{y\in\mathscr{X}} T_*(x,y)f(y) = \sum_{y\in\mathscr{X}} T_*(x,y)$, so in order to obtain Property T1$_{91}$ it suffices to prove that $T_*f(x) = 1$. Note that, for all $i \in \mathbb{Z}_{>0}$, it holds that $T_if(x) = 1$ because T_i satisfies Property T1$_{91}$ since $T_i \in \mathbb{T}$. Because $T_* = \lim_{i\to+\infty} T_i$ it follows from Lemma A.34$_{390}$ that $T_*f = \lim_{i\to+\infty} T_if$ and therefore, since $T_if(x) = 1$ for all $i \in \mathbb{Z}_{>0}$, that $T_*f(x) = \lim_{i\to+\infty} T_if(x) = 1$. Because $x \in \mathscr{X}$ is arbitrary, this means that T_* also satisfies Property T1$_{91}$. □

Corollary 3.11. *$\mathbb{T}$ is a compact and convex subset of the Banach space $\mathbb{M}$.*

Proof. To show the compactness, first note that it follows from Lemma 3.9 that $\|\mathbb{T}\| = \sup_{T\in\mathbb{T}} \|T\| = 1$, and hence $\mathbb{T}$ is bounded by S5$_{376}$. Moreover, by Proposition 3.10, any Cauchy sequence in $\mathbb{T}$ converges to a limit in $\mathbb{T}$. Because, using Definition A.11$_{374}$, any convergent sequence of matrices is a Cauchy sequence, it follows from Proposition A.8$_{376}$ that $\mathbb{T}$ is closed. Hence it follows that $\mathbb{T}$ is compact by Corollary A.12$_{378}$.

To see that $\mathbb{T}$ is convex, fix any $T, S \in \mathbb{T}$ and any $\lambda \in [0,1]$; we need to show that $\lambda T + (1-\lambda)S \in \mathbb{T}$. So fix any $x \in \mathscr{X}$. Then because T and S are both transition matrices, it follows from Definition 3.5$_{91}$ that

$$\sum_{y\in\mathscr{X}} \lambda T(x,y) + (1-\lambda)S(x,y) = \lambda \sum_{y\in\mathscr{X}} T(x,y) + (1-\lambda) \sum_{y\in\mathscr{X}} S(x,y) = 1\,,$$

where for the final equality we used that T and S both satisfy Property T1$_{91}$. This implies that $\lambda T + (1-\lambda)S$ also satisfies Property T1$_{91}$. Similarly, for any $y \in \mathscr{X}$ it holds that

$$\lambda T(x,y) + (1-\lambda)S(x,y) \geq 0\,,$$

since $\lambda \in [0,1]$ and because T and S both satisfy Property T2$_{91}$. This implies that $\lambda T + (1-\lambda)S$ also satisfies Property T2$_{91}$ and therefore, by Definition 3.5$_{91}$, that $\lambda T + (1-\lambda)S \in \mathbb{T}$. □

So let us now move on to the connection between transition matrices, and the transition probabilities of Markov chains. The easiest case to consider is that of homogeneous Markov chains, for which we can consider a single corresponding transition matrix, as follows.

Definition 3.6 (Corresponding Transition Matrix)**.** *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{HM}}$ be a homogeneous discrete-time Markov chain. Then the* transition matrix corresponding to *P is a matrix T that is defined, for all $x, y \in \mathscr{X}$, as*

$$T(x,y) := P\big(X_{\tau_1} = y \,\big|\, X_{\tau_0} = x\big)\,.$$

Proposition 3.12. *Let $\mathbb{D}$ be a discrete time domain, let $P \in \mathbb{P}^{\mathbb{D},\mathrm{HM}}$ be a time-homogeneous discrete-time Markov chain, and let T be its corresponding transition matrix. Then T is a transition matrix.*

Proof. Simply check both of the properties. □

The connection between non-homogeneous Markov chains and transition matrices is a bit more subtle, in that their transition probabilities do depend on the point in time at which they are considered. As such, we need to introduce a similar kind of time-dependency when considering their corresponding transition matrices. We will use the following definition.

Definition 3.7. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain. Then the* family of transition matrices corresponding to *P is a family (T_n) of matrices T_n that are defined, for all $n \in \mathbb{Z}_{\geq 0}$ and all $x, y \in \mathscr{X}$, as*

$$T_n(x,y) := P\big(X_{\tau_{n+1}} = y \,\big|\, X_{\tau_n} = x\big)\,.$$

Note that in the above definition, the family of transition matrices (T_n) might be more explicitly written as $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$; we drop the reference to its index set for notational brevity, when no confusion should arise. This convention will be especially helpful when we later consider more complicated families, like those introduced in Definition 3.8_{101} further on.

In any case, the following result should not be surprising.

Proposition 3.13. *Let $\mathbb{D}$ be a discrete time domain, let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain, and let (T_n) be its corresponding family of transition matrices. Then, for all $n \in \mathbb{Z}_{\geq 0}$, T_n is a transition matrix.*

Proof. Simply check both of the properties. □

We note that a time-homogeneous discrete-time Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{HM}}$ has both a corresponding transition matrix T, and a corresponding family of transition matrices (T_n). It follows from Definition 3.491 that, for homogeneous Markov chains, these transition matrices satisfy $T_n = T$ for all $n \in \mathbb{Z}_{\geq 0}$.

Moving on, the above discussion tells us that every discrete-time Markov chain has a corresponding family of transition matrices, that captures the transition probabilities of the system. As the next result shows, the converse statement is also true: every family $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ of transition matrices has a corresponding discrete-time Markov chain associated with it. Moreover, this Markov chain P is uniquely determined by $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ up to the specification of its initial distribution $P(X_0)$.

Proposition 3.14. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, consider any family $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ of transition matrices, and any probability mass function q on $\mathscr{X}$. Then there is a unique discrete-time Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ that has $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ as its corresponding family of transition matrices, and that satisfies $P(X_{\tau_0} = x) = q(x)$ for all $x \in \mathscr{X}$.*

Proof. Let $\mathscr{S}_{\mathbb{D}}$ denote the set of situations with time domain $\mathbb{D}$. Define the function $p : \mathscr{X} \times \mathscr{S}_{\mathbb{D}}$, for all $x \in \mathscr{X}$ and all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}$, as

$$p(x\,|\,x_u) := \begin{cases} q(x) & \text{if } u = \emptyset, \text{ and} \\ T_n(x_{\tau_n}, x) & \text{if } u = \tau_{0:n} \text{ for some } n \in \mathbb{Z}_{\geq 0}. \end{cases} \tag{3.9}$$

Because q is a probability mass function on $\mathscr{X}$ and because, for all $n \in \mathbb{Z}_{\geq 0}$ and $x_{\tau_n} \in \mathscr{X}$, $T_n(x_{\tau_n}, x)$, as a function of $x \in \mathscr{X}$, is also a probability mass function on $\mathscr{X}$, it follows that p is a probability tree, as defined in Section 3.185. Therefore, by Theorem 3.588, there is a unique discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$ that corresponds to p, and which therefore satisfies Equations (3.2)87 and (3.3)87.

Because P satisfies Equation (3.2)87, and using the definition of p, it follows that for all $x \in \mathscr{X}$ it holds that

$$P(X_{\tau_0} = x) = p(x\,|\,x_\emptyset) = q(x)\,.$$

Moreover, because P satisfies Equation (3.3)87, and using the definition of p, it follows that for all $x \in \mathscr{X}$ and all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ with $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, it holds that

$$P(X_{\tau_{n+1}} = x\,|\,X_u = x_u) = p(x\,|\,x_u) = T_n(x_{\tau_n}, x)\,.$$

Hence, it follows from Definition 3.389 that P is a Markov chain, with transition probabilities given by

$$P(X_{\tau_{n+1}} = y\,|\,X_{\tau_n} = x) = T_n(x,y)\,,$$

for all $x,y \in \mathscr{X}$ and all $n \in \mathbb{Z}_{\geq 0}$. Hence, it has a corresponding family of transition matrices, (S_n), say, and, for all $n \in \mathbb{Z}_{\geq 0}$ and all $x,y \in \mathscr{X}$,

$$S_n(x,y) = P(X_{\tau_{n+1}} = y \,|\, X_{\tau_n} = x) = T_n(x,y)\,,$$

from which we conclude that $(T_n)_{n\in\mathbb{Z}_{\geq 0}} = (S_n)$, whence P has $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ as its corresponding family of transition matrices. In summary, we have shown the existence of a discrete-time Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ with corresponding family of transition matrices $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ and such that $P(X_{\tau_0} = x) = q(x)$ for all $x \in \mathscr{X}$. It remains to show that this Markov chain is unique.

To this end, consider any discrete-time Markov chain $P_* \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ whose corresponding family of transition matrices is given by $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$, and that satisfies $P_*(X_{\tau_0} = x) = q(x)$ for all $x \in \mathscr{X}$; we will show that $P_* = P$. First, for all $x \in \mathscr{X}$ it follows from the fact that $P_*(X_{\tau_0} = x) = q(x)$, together with Equation (3.9)$_{\curvearrowleft}$, that

$$P_*(X_{\tau_0} = x) = q(x) = p(x \,|\, x_\emptyset)\,,$$

whence P_* satisfies Equation (3.2)$_{87}$. Similarly, for all $x \in \mathscr{X}$ and $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ with $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, it follows from the fact that P_* is a Markov chain, that

$$P_*(X_{\tau_{n+1}} = x \,|\, X_u = x_u) = P_*(X_{\tau_{n+1}} = x \,|\, X_{\tau_n} = x_{\tau_n}) = T_n(x_{\tau_n}, x) = p(x \,|\, x_u)\,,$$

where we used the definition of the corresponding transition matrix T_n for the second equality, and Equation (3.9)$_{\curvearrowleft}$ for the third equality. Hence it follows that P_* also satisfies Equation (3.3)$_{87}$. This implies that P_* corresponds to p. Because, by Theorem 3.5$_{88}$, P is the unique element of $\mathbb{P}^{\mathbb{D}}$ that corresponds to p, and because $\mathbb{P}^{\mathbb{D},\mathrm{M}} \subseteq \mathbb{P}^{\mathbb{D}}$, we conclude that $P_* = P$. □

We note that the above result is well-known, and that we do not intend to present it as novel. In fact, the result is so well-known that some textbooks do not even bother with the existence proof. However, similar statements—usually based on other formalisms—can be found throughout the literature. The measure-theoretic construction is typically based on Kolmogorov's Extension Theorem, see e.g. [96], or [54] for a more abstract treatment. Reference [69] considers the problem using both coherent conditional probabilities and game-theoretic probabilities [97, 98] as the underlying formalism, but without introducing transition matrices explicitly. In any case, the above result should be interpreted as a mere translation of those well-known results to the formalisation of Markov chains that we employ here.

In summary, we have seen above that (families of) transition matrices can be used to, essentially, parameterise discrete-time Markov chains. However, this is not the only reason that transition matrices are a useful tool when working with Markov chains. Because matrices are linear algebraic objects, we can also use them as *computational* tools. We will present some of these results in what follows; again, these properties are well-known in other formalisms—c.f. [39, 54, 82, 96]—but we repeat them here to show that they also hold for our current definitions.

Proposition 3.15. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain with corresponding family of transition matrices (T_n). Then for any $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, any $x_u \in \mathscr{X}_u$, any $m \in \mathbb{Z}_{\geq 0}$ such that $m > n$, and any $y \in \mathscr{X}$, it holds that*

$$P(X_{\tau_m} = y \,|\, X_u = x_u) = \left(\prod_{k=n}^{m-1} T_k\right)(x_{\tau_n}, y)\,. \tag{3.10}$$

Proof. We give a proof by induction on m. For the induction base, if $m = n+1$ then the result is immediate from Definitions 3.3$_{89}$ and 3.7$_{94}$.

So for the induction step, let us consider that $m > n+1$, and as the induction hypothesis let us suppose that

$$P(X_{\tau_{m-1}} = x_{\tau_{m-1}} \,|\, X_u = x_u) = \left(\prod_{k=n}^{m-2} T_k\right)(x_{\tau_n}, x_{\tau_{m-1}}) \quad \text{for all } x_{\tau_{m-1}} \in \mathscr{X}.$$

Define $v := \tau_{(n+1):(m-1)}$ and $w := v \setminus \{\tau_{m-1}\}$. Then it follows that

$$\begin{aligned}
P(X_{\tau_m} = y \,|\, X_u = x_u) &= \sum_{x_v \in \mathscr{X}_v} P(X_{\tau_m} = y, X_v = x_v \,|\, X_u = x_u) \\
&= \sum_{x_v \in \mathscr{X}_v} P(X_{\tau_m} = y \,|\, X_{u \cup v} = x_{u \cup v}) P(X_v = x_v \,|\, X_u = x_u) \\
&= \sum_{x_v \in \mathscr{X}_v} T_{m-1}(x_{\tau_{m-1}}, y) P(X_v = x_v \,|\, X_u = x_u) \\
&= \sum_{x_w \in \mathscr{X}_w} \sum_{x_{\tau_{m-1}} \in \mathscr{X}_{\tau_{m-1}}} T_{m-1}(x_{\tau_{m-1}}, y) P(X_v = x_v \,|\, X_u = x_u) \\
&= \sum_{x_{\tau_{m-1}} \in \mathscr{X}_{\tau_{m-1}}} T_{m-1}(x_{\tau_{m-1}}, y) \sum_{x_w \in \mathscr{X}_w} P(X_v = x_v \,|\, X_u = x_u) \\
&= \sum_{x_{\tau_{m-1}} \in \mathscr{X}_{\tau_{m-1}}} T_{m-1}(x_{\tau_{m-1}}, y) P(X_{\tau_{m-1}} = x_{\tau_{m-1}} \,|\, X_u = x_u)\,,
\end{aligned}$$

where we used Property F3$_{47}$ for the first equality, Property F4$_{47}$ for the second equality, Definitions 3.3$_{89}$ and 3.7$_{94}$ for the third equality, and Property F3$_{47}$ for the sixth equality. Hence, using the induction hypothesis, we get

$$\begin{aligned}
P(X_{\tau_m} = y \,|\, X_u = x_u) &= \sum_{x_{\tau_{m-1}} \in \mathscr{X}_{\tau_{m-1}}} T_{m-1}(x_{\tau_{m-1}}, y) P(X_{\tau_{m-1}} = x_{\tau_{m-1}} \,|\, X_u = x_u) \\
&= \sum_{x_{\tau_{m-1}} \in \mathscr{X}_{\tau_{m-1}}} T_{m-1}(x_{\tau_{m-1}}, y) \left(\prod_{k=n}^{m-2} T_k \right) (x_{\tau_n}, x_{\tau_{m-1}}) \\
&= \left(\prod_{k=n}^{m-1} T_k \right) (x_{\tau_n}, y) \,,
\end{aligned}$$

using the properties of matrix multiplication. □

To put the previous result into words, consider that the transition matrix T_n of a Markov chain describes the probabilities that the system will move from any state at time τ_n, to any state at time τ_{n+1}. In other words, they contain the one-step transition probabilities. Equation (3.10)↶, then, tells us that if we are interested in the probabilities of moving between states in *multiple* steps, it suffices to look at products of these transition matrices. Equation (3.10)↶ is also known as the *Chapman-Kolmogorov equation* in the Markov chain literature. This result has a specialisation for homogeneous Markov chains that provides an even simpler expression.

Corollary 3.16. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{HM}}$ be a discrete-time homogeneous Markov chain with corresponding transition matrix T. Then for any $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, any $x_u \in \mathscr{X}_u$, any $m \in \mathbb{Z}_{\geq 0}$ such that $m > n$, and any $y \in \mathscr{X}$, it holds that*

$$P(X_{\tau_m} = y \,|\, X_u = x_u) = T^{m-n}(x_{\tau_n}, y) \,,$$

where the term T^{m-n} denotes the $(m-n)$-th matrix power of T.

Proof. Recall that for a homogeneous Markov chain, its corresponding family of transition matrices (T_n) satisfies $T_n = T$ for all $n \in \mathbb{Z}_{\geq 0}$. Now apply Proposition 3.15↶. □

The above results can further be generalised to yield expressions for conditional expectations with respect to Markov chains, in terms of products of their transition matrices. This elementary but fundamental result will be crucial further on.

Proposition 3.17. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain with corresponding family of transition matrices (T_n). Then for any $u = \tau_{0:n}$, $n \in \mathbb{Z}_{\geq 0}$, any*

$x_u \in \mathscr{X}_u$, any $m \in \mathbb{Z}_{\geq 0}$ such that $m > n$, and any $f \in \mathscr{L}(\mathscr{X})$, it holds that

$$\mathbb{E}_P[f(X_{\tau_m}) \,|\, X_u = x_u] = \left(\prod_{k=n}^{m-1} T_k \right) f(x_{\tau_n}) . \tag{3.11}$$

Proof. It holds that $\tau_m \in \mathbb{D} = \tau_{0:n} \cup \mathbb{D}_{>\tau_{0:n}}$ and therefore, it follows from Proposition 2.23$_{73}$—with $v = \{\tau_m\}$—that

$$\mathbb{E}_P[f(X_{\tau_m}) \,|\, X_u = x_u] = \sum_{y \in \mathscr{X}} f(y) P(X_{\tau_m} = y \,|\, X_u = x_u) .$$

Applying Proposition 3.15$_{97}$, we obtain

$$\begin{aligned}\mathbb{E}_P[f(X_{\tau_m}) \,|\, X_u = x_u] &= \sum_{y \in \mathscr{X}} f(y) P(X_{\tau_m} = y \,|\, X_u = x_u) \\ &= \sum_{y \in \mathscr{X}} f(y) \left(\prod_{k=n}^{m-1} T_k \right) (x_{\tau_n}, y) = \left(\prod_{k=n}^{m-1} T_k \right) f(x_{\tau_n}) ,\end{aligned}$$

using the properties of matrix-vector products for the last equality. □

A special case of the previous result (with $m = n+1$) states that

$$\mathbb{E}_P[f(X_{\tau_{n+1}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = T_n f(x_{\tau_n}) . \tag{3.12}$$

We see that the right-hand side does not depend on the full history $x_{\tau_{0:n}}$, but only on the state x_{τ_n} at time τ_n. This suggests that we can formulate a "Markov property" also in terms of conditional expectations. The following result provides the required setup.

Proposition 3.18. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ be a discrete-time Markov chain with corresponding family of transition matrices (T_n). Then for any $n \in \mathbb{Z}_{\geq 0}$, any $x_{\tau_n} \in \mathscr{X}_{\tau_n}$, and any $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\mathbb{E}_P[f(X_{\tau_{n+1}}) \,|\, X_{\tau_n} = x_{\tau_n}] = T_n f(x_{\tau_n}) .$$

Proof. Let E be any coherent conditional prevision corresponding to P that is defined on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}} \subseteq \mathscr{D}$ and $(f(X_{\tau_{n+1}}), (X_{\tau_n} = x_{\tau_n})_{\mathbb{D}}) \in \mathscr{D}$. Let E^* be any coherent extension of E to $\mathbb{B} \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$, which exists by Theorem 2.6$_{52}$. Let $\mathscr{C} := \mathscr{E}(\Omega_{\mathbb{D}}) \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$; then $\mathscr{D}_{\mathscr{C}} \subseteq \mathbb{B} \times \mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$. Let $\tilde{E}^*$ be the restriction of E^* to $\mathscr{D}_{\mathscr{C}}$; because E^* is a coherent conditional prevision it follows from Definition 2.3$_{52}$ that $\tilde{E}^*$ is a coherent conditional prevision on $\mathscr{D}_{\mathscr{C}}$. Let P^* be the real-valued map on $\mathscr{C}$ that is defined by $P^*(A \,|\, C) := \tilde{E}^*[\mathbb{I}_A \,|\, C]$ for all $(A, C)_{\mathbb{D}} \in \mathscr{C}$. Then because $\tilde{E}^*$ is a coherent conditional prevision, it follows from Proposition 2.7_{53} that P^* is a coherent conditional probability. Moreover, because E corresponds to P, because E^* extends E, and because $\tilde{E}^*$ is the restriction of E^* to $\mathscr{D}_{\mathscr{C}}$, it follows that for all $(A,C)_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}} \subseteq \mathscr{C}$ it holds that

$$P(A\,|\,C) = E[\mathbb{I}_A\,|\,C] = E^*[\mathbb{I}_A\,|\,C] = \tilde{E}^*[\mathbb{I}_A\,|\,C] = P^*(A\,|\,C)\,, \tag{3.13}$$

which implies that P^* extends P.

Using the same line of reasoning as used in the proof of Proposition 2.23_{73}, we represent the τ_{n+1}-measurable function $f(X_{\tau_{n+1}})$ as $\sum_{y\in\mathscr{X}} f(y)\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}$. It follows that

$$\begin{aligned} E[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}] &= E^*[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}] \\ &= E^*\left[\sum_{y\in\mathscr{X}} f(y)\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}\,\middle|\,X_{\tau_n}=x_{\tau_n}\right] \\ &= \sum_{y\in\mathscr{X}} f(y)E^*\left[\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}\,\middle|\,X_{\tau_n}=x_{\tau_n}\right] \\ &= \sum_{y\in\mathscr{X}} f(y)P^*(X_{\tau_{n+1}}=y\,|\,X_{\tau_n}=x_{\tau_n})\,, \end{aligned}$$

where for the first equality we used that E^* extends E and that $\big(f(X_{\tau_{n+1}}),(X_{\tau_n}=x\tau_n)_{\mathbb{D}}\big) \in \mathscr{D}$; for the third equality we used the linearity of E^*, i.e. Properties $\mathrm{E2}_{52}$ and $\mathrm{E3}_{52}$; and for the last equality we used Equation (3.13). Because P^* extends P and because $(X_{\tau_{n+1}}=y, X_{\tau_n}=x_{\tau_n})_{\mathbb{D}} \in \mathscr{C}$ for all $y\in\mathscr{X}$, it follows from Proposition 3.7_{89} that

$$\begin{aligned} E[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}] &= \sum_{y\in\mathscr{X}} f(y)P^*(X_{\tau_{n+1}}=y\,|\,X_{\tau_n}=x_{\tau_n}) \\ &= \sum_{y\in\mathscr{X}} f(y)P(X_{\tau_{n+1}}=y\,|\,X_{\tau_n}=x_{\tau_n}) \\ &= \sum_{y\in\mathscr{X}} f(y)T_n(x_{\tau_n},y) = T_nf(x_{\tau_n})\,, \end{aligned}$$

where for the third equality we used Definition 3.7_{94}.

Because the coherent conditional prevision E corresponding to P and its domain $\mathscr{D} \subseteq \mathbb{B}\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ with $\mathscr{D}_{\mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}} \subseteq \mathscr{D}$ and $(f(X_{\tau_{n+1}}),(X_{\tau_n}=x_{\tau_n})_{\mathbb{D}}) \in \mathscr{D}$ are arbitrary, it follows from Definition 2.5_{54} that $\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}] = T_nf(x_{\tau_n})$. □

By combining Equation $(3.12)_{\curvearrowleft}$ and Proposition $3.18_{\curvearrowleft}$, we obtain the equality

$$\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}] = \mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]\,, \tag{3.14}$$

which is the stated "Markov property" formulated in terms of conditional expectations (c.f. Definition 3.389).

Let us conclude this section with a final definition, that captures the transition matrices corresponding to general discrete-time stochastic processes. Because such processes do not necessarily satisfy the Markov property, we need to index this family also in terms of the process its historical behaviour. This does not really lead to any nice properties like the ones discussed above, but we need this definition to streamline our notation for future results.

Definition 3.8. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $P \in \mathbb{P}^{\mathbb{D}}$ be a discrete-time stochastic process. Then the* family of history-dependent transition matrices corresponding to *P is a multi-index family $\left(T_{n,x_u}\right)$ of matrices T_{n,x_u} with $u = \tau_{0:(n-1)}$ for all $n \in \mathbb{Z}_{\geq 0}$, that are defined, for all $n \in \mathbb{Z}_{\geq 0}$ and $x_u \in \mathscr{X}_u$, and all $x,y \in \mathscr{X}$, as*

$$T_{n,x_u}(x,y) := P\left(X_{\tau_{n+1}} = y \,\middle|\, X_{\tau_n} = x, X_u = x_u\right).$$

As before, it should be clear that each history-dependent transition matrix T_{n,x_u} is, indeed, a transition matrix. Moreover, since any Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ is a stochastic process, it has both a corresponding family of transition matrices (T_n), and a family of history-dependent transition matrices (T_{n,x_u}). It follows from Definition 3.389 that $T_n = T_{n,x_u}$ for all $n \in \mathbb{Z}_{\geq 0}$ and all $x_u \in \mathscr{X}_u$, with $u = \tau_{0:(n-1)}$. If this Markov chain is moreover homogeneous, then its corresponding transition matrix T will satisfy $T = T_n = T_{n,x_u}$ for all $n \in \mathbb{Z}_{\geq 0}$ and all $x_u \in \mathscr{X}_u$, with $u = \tau_{0:(n-1)}$.

3.3 Discrete-Time Imprecise-Markov Chains

Having formally introduced the concept of Markov chains using our current formalism for stochastic processes, we can now finally start with the generalisation of these models to *imprecise*-Markov chains, which are actually the objects that we aim to study in this work. We have already briefly explained in Chapter 129 that imprecise-probabilistic models can be viewed as, essentially, being *sets* of traditional ("precise") probabilistic models. So let us now consider how to apply these ideas when the precise models that we are working with are Markov chains, or more generally stochastic processes. We start in Section 3.3.1↷ by defining imprecise-Markov chains as sets of stochastic processes. We show how to parameterise these models using (families of) sets of transition matrices, and study some of their properties. In Section 3.3.2105 we introduce and study the *lower*- and *upper* expectations for these discrete-time imprecise-Markov chains, which, as we discussed in Chapter 129, are the inferences in which we are interested when working with imprecise probabilities.

3.3.1 *Sets of Processes*

The aim of this section is to introduce the definition and parameterisation of imprecise-Markov chains, and to study some of the properties of the resulting sets of models. We have seen in Section 3.289 that Markov chains can, essentially, be parameterised using (families of) transition matrices. Analogously, imprecise-Markov chains will be parameterised using *sets* of transition matrices, or, more generally, families of sets of transition matrices.

To this end, we introduce the notion of *consistency* of a process, with such a family of sets, as follows.

Definition 3.9. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and consider any family $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ of sets $\mathscr{T}_n$ of transition matrices. We say that a discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$ is* consistent *with $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ if its corresponding family of history-dependent transition matrices (T_{n,x_u}) satisfies $T_{n,x_u} \in \mathscr{T}_n$ for all $n \in \mathbb{Z}_{\geq 0}$ and all $x_u \in \mathscr{X}_u$, with $u = \tau_{0:(n-1)}$. If P is consistent with $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$, we write $P \sim (\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$.*

Before we can proceed, let us recall from Section 3.289 that (families of) transition matrices only identify discrete-time Markov chains P up to the specification of their initial distribution $P(X_{\tau_0})$. We need a similar construction for imprecise-Markov chains. To this end, for any set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, and for any discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$, we will say that P is *consistent* with $\mathscr{M}$, and write $P \sim \mathscr{M}$, if the map $p : \mathscr{X} \to \mathbb{R} : x \mapsto P(X_{\tau_0} = x)$ satisfies $p \in \mathscr{M}$, where τ is the canonical time index of $\mathbb{D}$.

We now have all the elements that we need to formally define discrete-time imprecise-Markov chains. Perhaps unsurprisingly, these are sets of discrete-time stochastic processes that are consistent with some given family of sets of transition matrices, and with some given set of probability mass functions. Let us first define the following.

Definition 3.10 (Set of Consistent Processes)**.** *Let $\mathbb{D}$ be a discrete time domain, and consider any family $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ of sets $\mathscr{T}_n$ of transition matrices, and any set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for any set $\mathscr{P} \subseteq \mathbb{P}^{\mathbb{D}}$ we define the subset of $\mathscr{P}$ that is consistent with both $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $\mathscr{M}$, which we denote as $\mathscr{P}_{(\mathscr{T}_n),\mathscr{M}}$, as*

$$\mathscr{P}_{(\mathscr{T}_n),\mathscr{M}} := \left\{ P \in \mathscr{P} \,\middle|\, P \sim (\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}} \text{ and } P \sim \mathscr{M} \right\}.$$

Note that for notational brevity, we have removed the explicit mention of the index set of the family $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ in the first subscript of $\mathscr{P}_{(\mathscr{T}_n),\mathscr{M}}$. Moreover, *if $\mathscr{M}$ is the set of* all *probability mass functions on $\mathscr{X}$, then we will simply write $\mathscr{P}_{(\mathscr{T}_n)}$ for the subset of $\mathscr{P}$ that is consistent with $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$.*[2]

The reason that we use this level of generality for $\mathscr{P}$ in Definition 3.10 is that the definition of imprecise-Markov chains in the literature is a bit ambiguous; there are different sets of discrete-time stochastic processes that have been studied in the literature, and that one might call (and have been called) imprecise-Markov chains.

For starters, we might consider the set $\mathbb{P}^{\mathbb{D},\mathrm{M}}_{(\mathscr{T}_n),\mathscr{M}}$, which is the set of all Markov chains (with time domain $\mathbb{D}$) that are consistent with both $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $\mathscr{M}$. This set contains Markov chains that are, in general, non-homogeneous. Essentially, it captures those Markov chains whose (time-dependent) transition matrix T_n is contained in $\mathscr{T}_n$, for all $n\in\mathbb{Z}_{\geq 0}$. Albeit under slightly different formalisations of the underlying processes, this model has been studied in, amongst others, [48,101]. It was first introduced by Hartfiel, who called them *Markov set chains* [44–46]. Readers who are familiar with the general theory of graphical models for imprecise probabilities (e.g. [13, 16]) might find it helpful to note that these are sometimes called imprecise-Markov chains under *complete independence* [14],[3] as is explained in, for example, [69].

A different model that one could consider is the set $\mathbb{P}^{\mathbb{D},\mathrm{HM}}_{(\mathscr{T}_n),\mathscr{M}}$, i.e. the set of all *homogeneous* Markov chains consistent with $\mathscr{M}$ and $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$. Because, as we have seen in Section 3.289, a homogeneous Markov chain P is identified by a single transition matrix T, we can only have $P\in\mathbb{P}^{\mathbb{D},\mathrm{HM}}_{(\mathscr{T}_n),\mathscr{M}}$ if $T\in\mathscr{T}_n$ for all $n\in\mathbb{Z}_{\geq 0}$, or in other words, if $T\in\cap_{n\in\mathbb{Z}_{\geq 0}}\mathscr{T}_n$. So, in this setting we can parameterise the transition matrices using a single set $\mathscr{T}$, and it is more convenient to focus on the model

$$\mathbb{P}^{\mathbb{D},\mathrm{HM}}_{\mathscr{T},\mathscr{M}} := \left\{ P\in\mathbb{P}^{\mathbb{D},\mathrm{HM}} \,:\, T_P\in\mathscr{T}, P\sim\mathscr{M} \right\},$$

where, for all $P\in\mathbb{P}^{\mathbb{D},\mathrm{HM}}$, T_P is the transition matrix corresponding to P. This model has been studied in, amongst others, [10, 57]. However, as noted in e.g. [69], this model is generally very difficult to perform inferences with. Nevertheless, for some classes of problems the model is tractable. This is in particular the case with inferences for which these different notions of imprecise-Markov chains coincide in terms of their lower expectations; see e.g. the results in [64, 69].

---

[2] Consistency with $\mathscr{M}$ is trivial if $\mathscr{M}$ contains all probability mass functions on $\mathscr{X}$.

[3] Reference [48] refers to this condition as *strong independence*. As explained in [14], strong independence requires the set of distributions to be convex. While convexity of the set $\mathbb{P}^{\mathbb{D},\mathrm{M}}_{(\mathscr{T}_n),\mathscr{M}}$ at a global level seems non-obvious, Reference [48] spends some effort in showing that its induced set of distributions for each separate X_{τ_n} is convex, under certain conditions on $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $\mathscr{M}$.

Finally, yet another definition of an imprecise-Markov chain—and the one that we will mainly use in the remainder of this work—is the set $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ of *all* stochastic processes consistent with $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $\mathscr{M}$. This type of imprecise-Markov chain has been studied in, amongst others, [20–22, 27, 48, 69]. Again from an imprecise graphical model perspective, it can be viewed as an imprecise-Markov chain under *epistemic irrelevance*: a conditional independence property for imprecise probabilities that is weaker than the notion of *complete independence* that we mentioned above [14]. We refer to [48] for further information on this perspective. From here on out, unless we explicitly mention otherwise, we will use the following definition when we refer to a discrete-time imprecise-Markov chain:

Definition 3.11 (Discrete-Time Imprecise-Markov Chain)**.** *Let $\mathbb{D}$ be a discrete time domain, consider any family $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ of non-empty sets of transition matrices, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then we define the corresponding* discrete-time imprecise-Markov chain *(*DTIMC*) $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ to be the set of all discrete-time stochastic processes with time domain $\mathbb{D}$, that are consistent with both $(\mathscr{T}_n)_{n\in\mathbb{Z}_{\geq 0}}$ and $\mathscr{M}$, following the notation from Definition 3.10*$_{102}$*.*

We would like to provide some intuition at this point to explain why we call this model an imprecise-"Markov" chain, while we explicitly do not impose Markovianity on its constituent processes. This is because, as we will show in, e.g., Proposition 3.25$_{110}$ further on, this model nevertheless satisfies an *imprecise-Markov property*, in the sense that inferences derived from it *are* history-independent in a manner analogous to the Markov property. Although we must postpone explaining the technical details until we get to these results further on, one intuitive way to see this is that the sets $\mathscr{T}_n$, $n \in \mathbb{Z}_{>0}$, that parameterise the DTIMC, *are* in fact history-independent.

Moving on, in Definition 3.11 we restrict attention to sets of transition matrices, and sets of probability mass functions, that are non-empty; this simply serves to prevent trivialities by ensuring that $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ is non-empty. Let us start with an auxiliary result before formalising this statement.

Lemma 3.19. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $p \in \mathscr{M}$, there is some $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ such that $P(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$.*

Proof. Fix any $p \in \mathscr{M}$ and, for all $k \in \mathbb{Z}_{\geq 0}$, any $T_k \in \mathscr{T}_k$; this is possible because $\mathscr{M}$ and $\mathscr{T}_k$, $k \in \mathbb{Z}_{\geq 0}$, are non-empty by Definition 3.11.

By Proposition 3.14$_{95}$, there is a discrete-time Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}}$ such that $P(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$, and that has $(T_k)_{k\in\mathbb{Z}_{\geq 0}}$ as its corresponding family of transition matrices. Because $p \in \mathscr{M}$ and $T_k \in \mathscr{T}_k$ for all $k \in \mathbb{Z}_{\geq 0}$, and since $\mathbb{P}^{\mathbb{D},\mathrm{M}} \subseteq \mathbb{P}^{\mathbb{D}}$, it follows from Definition 3.11 that $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$. ☐

Hence that $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ is non-empty now follows immediately:

Lemma 3.20. *Let $\mathbb{D}$ be a discrete time domain and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}} \neq \emptyset$.*

Proof. This is an immediate consequence of Lemma 3.19 together with the fact that $\mathscr{M}$ is non-empty by Definition 3.11. ☐

3.3.2 *Lower and Upper Expectations for DTIMCs*

We have already mentioned in Chapter 1₂₉ that computing inferences for imprecise-probabilistic models essentially involves computing the lower (or upper) envelopes over the inferences of the precise-probabilistic models that constitute these sets: these are the corresponding *lower*- and *upper* expectations. Let us formalise this here for discrete-time imprecise-Markov chains.

Definition 3.12. *Let $\mathbb{D}$ be a discrete time domain, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then we define its corresponding (conditional)* lower- *and* upper *expectations, respectively, as*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}[\cdot\,|\,\cdot] := \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}} \mathbb{E}_P[\cdot\,|\,\cdot] \quad \text{and} \quad \overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}[\cdot\,|\,\cdot] := \sup_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}} \mathbb{E}_P[\cdot\,|\,\cdot],$$

whose domain(s) we take to be the intersection of the domains $\mathscr{D}_P$ of the conditional expectations $\mathbb{E}_P$ corresponding to the elements $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$.

Note that the domain of the lower (and upper) expectation is simply such, that the precise expectations $\mathbb{E}_P$ are well-defined for all $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$. Following Definition 2.5₅₄, the domain of $\mathbb{E}_P$ depends strongly on P, despite all elements of $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$ having the same domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$. This makes the domain of the lower and upper expectation somewhat difficult to write explicitly. Nevertheless, it follows from Proposition 2.23₇₃ that all u-measurable functions, for $u \in \mathscr{U}^{\mathbb{D}}_{\supset\emptyset}$, are in this domain, provided that care is taken that the conditioning events are situations. The following result makes this explicit.

Lemma 3.21. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $v \in \mathscr{U}^{\mathbb{D}}_{\supset\emptyset}$, all $f \in \mathscr{L}(\mathscr{X}_v)$, and all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, the lower and upper* *expectations* $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u]$ *and* $\overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u]$ *are well-defined. In particular, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u] = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{x_v\in\mathscr{X}_v} f(x_v)P(X_v=x_v\,|\,X_u=x_u) \quad (3.15)$$

and

$$\overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u] = \sup_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{x_v\in\mathscr{X}_v} f(x_v)P(X_v=x_v\,|\,X_u=x_u)\,. \quad (3.16)$$

Proof. Because $(X_u=x_u)_{\mathbb{D}}$ is a situation, it follows that either $u=\emptyset$ or $u=\tau_{0:n}$ for some $n\in\mathbb{Z}_{\geq 0}$. Moreover, it holds that $v\subset\mathbb{D}=u\cup\mathbb{D}_{>u}$. Therefore, it follows from Proposition 2.2373 that, for all $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$,

$$\mathbb{E}_P[f(X_v)\,|\,X_u=x_u] = \sum_{x_v\in\mathscr{X}_v} f(x_v)P(X_v=x_v\,|\,X_u=x_u)\,. \quad (3.17)$$

Hence in particular $\mathbb{E}_P[f(X_v)\,|\,X_u=x_u]$ is well-defined for all $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, whence $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u]$ and $\overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u=x_u]$ are well-defined by Definition 3.12$_{\curvearrowleft}$. Equations (3.15) and (3.16) now follow immediately from Equation (3.17) and Definition 3.12$_{\curvearrowleft}$. □

We recall from Chapter 129 the important *conjugacy* relation between lower- and upper expectations, which in our current setting can be written as

$$\overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}[\cdot\,|\,\cdot] = -\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}[-\cdot\,|\,\cdot]\,.$$

Because this relation means that we can always translate results about lower expectations to results about upper expectations (and vice versa), this implies that we can mostly content ourselves with discussing either. Hence, in the sequel, we will mostly phrase results in terms of lower expectations, but these results can always be translated to also hold for upper expectations using the above conjugacy relation.

Moreover, we note that *lower*- and *upper probabilities* can always be expressed using the lower- and upper expectations of indicators of events; for any $(A,C)_{\mathbb{D}}\in\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ it holds that

$$\underline{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}(A\,|\,C) := \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}} P(A\,|\,C) = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}[\mathbb{I}_A\,|\,C]\,,$$

where we used Proposition 2.1256 for the equality. Upper probabilities are derived analogously using upper expectations, and can therefore be obtained from lower expectations using the above-mentioned conjugacy relation. Hence, because lower- and upper probabilities can always be derived using lower expectations,[4] we will in the sequel express our results mostly in terms of the latter.

We next translate Proposition 2.25_{75} to the imprecise setting.

Lemma 3.22. *Let $\mathbb{D}$ be a discrete time domain and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, all $v \in \mathscr{U}^{\mathbb{D}}_{\supset\emptyset}$ such that $u < v$, and all $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{u\cup v})\,|\,X_u = x_u] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_u,X_v)\,|\,X_u = x_u]\,,$$

with $f(x_u,X_v)$ as in Definition 2.16_{74}.

Proof. It follows from Lemma 3.21_{105} that $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{u\cup v})\,|\,X_u = x_u]$ and $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_u,X_v)\,|\,X_u = x_u]$ are both well-defined. Because $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, it follows from Definition 3.1_{85} that either $u = \emptyset$ or $u = \tau_{0:n}$ for some $n \in \mathbb{Z}_{\geq 0}$. Therefore, and because $v \neq \emptyset$ and $u < v$, it follows from Proposition 2.25_{75} that for all $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ it holds that

$$\mathbb{E}_P\big[f(X_{u\cup v})\,\big|\,X_u = x_u\big] = \mathbb{E}_P\big[f(x_u,X_v)\,\big|\,X_u = x_u\big]\,.$$

Hence it follows from Definition 3.12_{105} that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{u\cup v})\,|\,X_u = x_u] &= \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P\big[f(X_{u\cup v})\,\big|\,X_u = x_u\big] \\ &= \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P\big[f(x_u,X_v)\,\big|\,X_u = x_u\big] \\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_u,X_v)\,|\,X_u = x_u]\,,\end{aligned}$$

which concludes the proof. □

The following technical observation will be useful: it verifies that the conditional lower expectation of any u-measurable function is real-valued; and hence, in particular, that it is bounded.

Lemma 3.23. *Let $\mathbb{D}$ be a discrete time domain and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $v \in \mathscr{U}^{\mathbb{D}}_{\supset\emptyset}$, all $f \in \mathscr{L}(\mathscr{X}_v)$, and all $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, it holds that $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u = x_u] \in \mathbb{R}$. In particular,*

$$\min_{y_v\in\mathscr{X}_v} f(y_v) \leq \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u = x_u] \leq \max_{y_v\in\mathscr{X}_v} f(y_v)\,.$$

[4] In contrast with "precise" probabilistic models, for which the expectation operators are linear—see e.g. the statement and proof of Proposition 2.12_{56}—the converse is in general *not* true: one cannot always recover the lower- and upper expectation operators from the specification of the lower- and upper probabilities [114, Section 2.7]. On the other hand, there are known sufficient conditions under which this *is* possible (*ibid.*), but we will not consider the technical details here.

Proof. Because $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$, it follows from Lemma 3.286 that

$$(X_v = x_v, X_u = x_u)_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}} \quad \text{for all } x_v \in \mathscr{X}_v.$$

Hence, by Property CE178 we find that, for all $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, it holds that

$$\min_{y_v \in \mathscr{X}_v} f(y_v) \leq \mathbb{E}_P[f(X_v)\,|\,X_u = x_u] \leq \max_{y_v \in \mathscr{X}_v} f(y_v)\,. \tag{3.18}$$

Note that because f is v-measurable, it follows from Proposition 2.2172 that it is bounded, i.e. it obtains its extremal values in $\mathbb{R}$, whence the minimum and maximum operations in the inequalities above are well-defined. Because Equation (3.18) holds for all $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ it follows from Definition 3.12105 that

$$\min_{y_v \in \mathscr{X}_v} f(y_v) \leq \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u = x_u]\,.$$

Moreover, by Lemma 3.20105, $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ is non-empty. Hence there is some $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ such that, using Definition 3.12105,

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u = x_u] \leq \mathbb{E}_P[f(X_v)\,|\,X_u = x_u] \leq \max_{y_v \in \mathscr{X}_v} f(y_v)\,,$$

and hence $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_v)\,|\,X_u = x_u] \in \mathbb{R}$. □

The next result provides us with an expression for the conditional lower expectations of discrete-time Markov chains, in terms of the sets of transition matrices that parameterise them.

Proposition 3.24. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $f \in \mathscr{L}(\mathscr{X})$, all $n \in \mathbb{Z}_{\geq 0}$, and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \inf_{T \in \mathscr{T}_n} Tf(x_{\tau_n})\,. \tag{3.19}$$

Proof. We first note that, due to Lemma 3.21105, the lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}} = x_{\tau_{0:n}}]$ is well-defined. Now consider an arbitrary family $(T_k)_{k\in\mathbb{Z}_{\geq 0}}$ such that $T_k \in \mathscr{T}_k$ for all $k \in \mathbb{Z}_{\geq 0}$, and choose any $p \in \mathscr{M}$. Due to Proposition 3.1495, there is then a unique Markov chain $P \in \mathbb{P}^{\mathbb{D},\mathrm{M}} \subseteq \mathbb{P}^{\mathbb{D}}$ that has $(T_k)_{k\in\mathbb{Z}_{\geq 0}}$ as its corresponding family of transition matrices, and such that $P(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$. Moreover, it is immediately clear that $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, since $p \in \mathscr{M}$ and $T_k \in \mathscr{T}_k$ for all $k \in \mathbb{Z}_{\geq 0}$. Due to Proposition 3.1798, it holds that

$$\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}} = x_{\tau_{0:n}}] = T_n f(x_{\tau_n})\,,$$

so it follows from Definition 3.12$_{105}$ that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]\leq \mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]=T_nf(x_{\tau_n})\,.$$

Because $T_n\in\mathscr{T}_n$ was arbitrary, it follows that also

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]\leq \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})\,.$$

To get the inequality in the other direction, fix any $\varepsilon>0$. Then, because $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]$ is real-valued by Lemma 3.23$_{107}$, and because $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ is non-empty due to Lemma 3.20$_{105}$, and using Definition 3.12$_{105}$, there is some $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ with corresponding conditional expectation $\mathbb{E}_P$, such that

$$\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]<\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]+\varepsilon\,. \tag{3.20}$$

Because $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, its corresponding history-dependent transition matrix T_{n,x_u}, with $u=\tau_{0:(n-1)}$, satisfies $T_{n,x_u}\in\mathscr{T}_n$ due to Definitions 3.11$_{104}$ and 3.9$_{102}$ and, by Definition 3.8$_{101}$, it holds for all $y\in\mathscr{X}$ that

$$T_{n,x_u}(x_{\tau_n},y)=P(X_{\tau_{n+1}}=y\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}})\,.$$

Therefore, and using Proposition 2.23$_{73}$, it holds that

$$\begin{aligned}\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]&=\sum_{y\in\mathscr{X}}f(y)P(X_{\tau_{n+1}}=y\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}})\\&=\sum_{y\in\mathscr{X}}f(y)T_{n,x_u}(x_{\tau_n},y)=T_{n,x_u}f(x_{\tau_n})\,,\end{aligned}$$

using the properties of matrix-vector multiplication. Because $T_{n,x_u}\in\mathscr{T}_n$, it therefore follows that

$$\inf_{T\in\mathscr{T}_n}Tf(x_{\tau_n})\leq T_{n,x_u}f(x_{\tau_n})=\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]\,.$$

Hence it follows from Equation (3.20) that

$$\begin{aligned}\inf_{T\in\mathscr{T}_n}Tf(x_{\tau_n})&\leq\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]\\&<\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]+\varepsilon\,.\end{aligned}$$

Because $\varepsilon>0$ is arbitrary, this implies that also

$$\inf_{T\in\mathscr{T}_n}Tf(x_{\tau_n})\leq\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}}=x_{\tau_{0:n}}]\,,$$

which concludes the proof. □

Note that the right-hand side of Equation $(3.19)_{108}$ only depends on the state x_{τ_n} at time τ_n, and not on the entire history $x_{\tau_{0:n}}$ as the left-hand side of that equation does. In other words, this implies that this model satisfies an *imprecise-Markov property*, which motivates the terminology that this model is an imprecise-*Markov* chain. The next result makes this even more notationally explicit, in analogy with the (precise) Markov property formulated in terms of (precise) conditional expectations, as stated in Equation $(3.14)_{100}$. Because the proof is fairly long, we have deferred it to Appendix $3.\mathrm{B}_{130}$.

Proposition 3.25. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $f \in \mathscr{L}(\mathscr{X})$, all $n \in \mathbb{Z}_{\geq 0}$, and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}}) \,|\, X_{\tau_n} = x_{\tau_n}] \,,$$

whenever the lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}}) \,|\, X_{\tau_n} = x_{\tau_n}]$ is well-defined.

Let us now introduce the convention that for any $n \in \mathbb{Z}_{\geq 0}$, any $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}})$, and any $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, we can consider the function $f(x_{\tau_{0:n}}, \cdot) \in \mathscr{L}(\mathscr{X}_{\tau_{n+1}})$, which is a projection of f onto $\mathscr{L}(\mathscr{X}_{\tau_{n+1}})$, defined such that $f(x_{\tau_{0:n}}, \cdot)(x_{\tau_{n+1}}) \coloneqq f(x_{\tau_{0:(n+1)}})$ for all $x_{\tau_{n+1}} \in \mathscr{X}_{\tau_{n+1}}$. Comparing this with Definition 2.16_{74}, this essentially identifies the element of $\mathscr{L}(\mathscr{X}_{\tau_{n+1}})$ corresponding to the τ_{n+1}-measurable function $f(x_{\tau_{0:n}}, X_{\tau_{n+1}})$.

With this notation, we can generalise Proposition 3.24_{108} as follows.

Corollary 3.26. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $n \in \mathbb{Z}_{\geq 0}$, all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}})$, and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:(n+1)}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \inf_{T \in \mathscr{T}_n} \big[T f(x_{\tau_{0:n}}, \cdot)\big](x_{\tau_n}) \,.$$

Proof. By combining Lemma 3.22_{107} and Proposition 3.24_{108} we find that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:(n+1)}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_{\tau_{0:n}}, X_{\tau_{n+1}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] \\ &= \inf_{T \in \mathscr{T}_n} \big[T f(x_{\tau_{0:n}}, \cdot)\big](x_{\tau_n}) \,,\end{aligned}$$

using the notational convention for $f(x_{\tau_{0:n}}, \cdot)$ as established above. ☐

Let us now introduce some important terminology that we will need in the sequel; the definition below characterises a structural property of sets of (transition) matrices that is instrumental in obtaining certain decomposition properties (e.g., Corollary 3.28_{112}) of lower expectations for imprecise-Markov chains that are parameterised by these sets.

Definition 3.13. *Let $\mathbb{T}$ be the set of all transition matrices, and consider any set $\mathscr{T} \subseteq \mathbb{T}$. We then say that $\mathscr{T}$ has* separately specified rows *if it holds that*

$$\mathscr{T} = \left\{ T \in \mathbb{T} \,\middle|\, \forall x \in \mathscr{X} : T(x,\cdot) \in \mathscr{T}_x \right\},$$

where, for all $x \in \mathscr{X}$, $\mathscr{T}_x := \left\{ T(x,\cdot) \,|\, T \in \mathscr{T} \right\}$ is the set of x-rows of the elements of $\mathscr{T}$.

This property is often encountered in the literature on imprecise-Markov chains; see e.g. [48, 101]. It requires, effectively, that one can recombine the elements of $\mathscr{T}$ by their rows, and that $\mathscr{T}$ is closed under this recombination. Formally, if $\mathscr{T}$ has separately specified rows then if for all $x \in \mathscr{X}$ we select any matrix $T_x \in \mathscr{T}$, and if we define $S \in \mathbb{T}$ such that $S(x,\cdot) := T_x(x,\cdot)$ for all $x \in \mathscr{X}$, then it holds that also $S \in \mathscr{T}$. As we will see below, this property is crucial in obtaining some important properties of imprecise-Markov chains.[5]

To state these results, we note that, as in the discussion preceding Proposition 2.2677, for a fixed function f in its domain, the conditional lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}]$ can be viewed as a function of $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$. In other words, we can associate with this conditional lower expectation a $\tau_{0:n}$-measurable function, which we denote by $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f \,|\, X_{\tau_{0:n}}]$, and whose value in $\omega \in \Omega_{\mathbb{D}}$ is given by

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f \,|\, X_{\tau_{0:n}}](\omega) := \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f \,|\, X_{\tau_{0:n}} = \omega|_{\tau_{0:n}}].$$

The following important result serves as the crucial step to stating a version of Proposition 2.2677 in the imprecise setting; this will provide a *law of iterated lower expectations*. This property of imprecise-Markov chains is well-known in the literature, see e.g. [20–22, 48, 69, 107]. Our proof, which can be found in Appendix 3.B130, uses the terminology and notation from this current work, but is based on—and conceptually essentially the same as—the proof of [69, Theorem 21].

Lemma 3.27. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain such that $\mathscr{T}_k$ has separately specified rows for all $k \in \mathbb{Z}_{\geq 0}$. Then for all $n, m \in \mathbb{Z}_{\geq 0}$ such that*

[5]Incidentally, as noted in [101], the concept of *separately specified rows* has a counterpart in the theory of imprecise graphical models—to which, as noted in Section 3.3.1102, imprecise-Markov chains are related—in which sets of probabilities are said to be separately specified [15].

$n \geq m$, all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$, and all $x_{\tau_{0:(m-1)}} \in \mathscr{X}_{\tau_{0:(m-1)}}$, it holds that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}&[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\middle|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\right].\end{aligned} \tag{3.21}$$

The next statement can be understood as more fully describing this law of iterated lower expectations:

Corollary 3.28. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain such that $\mathscr{T}_k$ has separately specified rows for all $k \in \mathbb{Z}_{\geq 0}$. Then for all $n, \ell, m \in \mathbb{Z}_{\geq 0}$ such that $n > \ell \geq m$, all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$, and all $x_{\tau_{0:(m-1)}} \in \mathscr{X}_{\tau_{0:(m-1)}}$ it holds that*

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}&[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}]\,\middle|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\right].\end{aligned}$$

Proof. If $\ell = n-1$ then it holds that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}],$$

and the result is then an immediate consequence of Lemma 3.27$_{\curvearrowleft}$.

So let us assume that $\ell < n-1$, and fix any $y_{\tau_{0:\ell}} \in \mathscr{X}_{\tau_{0:\ell}}$. Applying Lemma 3.27$_{\curvearrowleft}$ we find that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}&[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}} = y_{\tau_{0:\ell}}] \\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\middle|\,X_{\tau_{0:\ell}} = y_{\tau_{0:\ell}}\right].\end{aligned} \tag{3.22}$$

Now if $\ell = n-2$ it holds that $n-1 = \ell+1$, and since $y_{\tau_{0:\ell}} \in \mathscr{X}_{\tau_{0:\ell}}$ is arbitrary, this implies that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(\ell+1)}}]\,\middle|\,X_{\tau_{0:\ell}}\right]. \tag{3.23}$$

Conversely, if $\ell < n-2$, we use the fact that in the right-hand side of Equation (3.22), the inner conditional lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]$ is a $\tau_{0:(n-1)}$-measurable function (which is real-valued due to Lemma 3.23$_{107}$), and because $\ell < n-2$, we can again apply Lemma 3.27$_{\curvearrowleft}$ to find

$$\begin{aligned}&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}} = y_{\tau_{0:\ell}}] \\ &\quad= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\middle|\,X_{\tau_{0:(n-2)}}\right]\,\middle|\,X_{\tau_{0:\ell}} = y_{\tau_{0:\ell}}\right].\end{aligned}$$

By continuing in this way, after a total of $n-\ell-1$ applications of Lemma 3.27$_{111}$ we find that

$$\begin{aligned}&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}=y_{\tau_{0:\ell}}]\\&=\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\cdots\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\cdots\,\middle|\,X_{\tau_{0:(\ell+1)}}\right]\,\middle|\,X_{\tau_{0:\ell}}=y_{\tau_{0:\ell}}\right],\end{aligned}$$

and because this is true for all $y_{\tau_{0:\ell}}\in\mathscr{X}_{\tau_{0:\ell}}$, it follows that

$$\begin{aligned}&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}]\\&\quad=\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\cdots\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\cdots\,\middle|\,X_{\tau_{0:(\ell+1)}}\right]\,\middle|\,X_{\tau_{0:\ell}}\right],\end{aligned}\tag{3.24}$$

which is essentially the more general version of Equation (3.23).

Now for notational brevity, let $u:=\tau_{0:(m-1)}$. We proceed similarly by repeatedly expanding the conditional lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u=x_u]$ using Lemma 3.27$_{111}$, until we obtain

$$\begin{aligned}&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u=x_u]\\&\quad=\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\cdots\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\cdots\,\middle|\,X_{\tau_{0:\ell}}\right]\,\middle|\,X_u=x_u\right].\end{aligned}\tag{3.25}$$

Substituting Equation (3.24) (or Equation (3.23) in case $\ell=n-2$) into Equation (3.25) yields

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u=x_u]=\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:\ell}}]\,\middle|\,X_u=x_u\right],$$

which, since $u=\tau_{0:(m-1)}$, concludes the proof. □

One particular reason that Corollary 3.28 is so useful, is that it allows us to, essentially, separately consider time points on which a function depends, and recursively deal with the resulting lower expectations. We will next present some results that illustrate the use of this machinery to separately deal with the initial distribution—that is, the state at time τ_0—which allows us to express unconditional lower expectations in a more convenient manner. We start by establishing the following crucial property:

Lemma 3.29. *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $f\in\mathscr{L}(\mathscr{X})$ it holds that*

$$\min_{x\in\mathscr{X}}f(x)\leq\inf_{p\in\mathscr{M}}\sum_{x\in\mathscr{X}}p(x)f(x)\leq\max_{x\in\mathscr{X}}f(x).$$

Proof. For any $p \in \mathscr{M}$ it holds that

$$\sum_{x\in\mathscr{X}} p(x)f(x) \geq \sum_{x\in\mathscr{X}} p(x) \min_{y\in\mathscr{X}} f(y) = \min_{y\in\mathscr{X}} f(y)\,,$$

where we used that p is a probability mass function on $\mathscr{X}$. Because this is true for all $p \in \mathscr{M}$ it follows that $\min_{x\in\mathscr{X}} f(x) \leq \inf_{p\in\mathscr{M}} \sum_{x\in\mathscr{X}} p(x)f(x)$. Similarly, because $\mathscr{M}$ is non-empty there is some $q \in \mathscr{M}$ such that $\inf_{p\in\mathscr{M}} \sum_{x\in\mathscr{X}} p(x)f(x) \leq \sum_{x\in\mathscr{X}} q(x)f(x)$, which implies that

$$\inf_{p\in\mathscr{M}} \sum_{x\in\mathscr{X}} p(x)f(x) \leq \sum_{x\in\mathscr{X}} q(x)f(x) \leq \sum_{x\in\mathscr{X}} q(x) \max_{y\in\mathscr{X}} f(y) = \max_{y\in\mathscr{X}} f(y)\,,$$

where we used that q is a probability mass function on $\mathscr{X}$. ☐

This allows us to provide the following definition, which in effect represents the lower expectation for the initial distribution only in terms of $\mathscr{M}$, without further reference to the set $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$. We note that Lemma 3.29↶ ensures that this map is indeed real-valued.

Definition 3.14. *For any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we define the map $\underline{\mathbb{E}}_{\mathscr{M}} : \mathscr{L}(\mathscr{X}) \to \mathbb{R} : f \mapsto \underline{\mathbb{E}}_{\mathscr{M}}[f]$ where, for all $f \in \mathscr{L}(\mathscr{X})$, we let*

$$\underline{\mathbb{E}}_{\mathscr{M}}[f] \coloneqq \inf_{p\in\mathscr{M}} \sum_{x\in\mathscr{X}} p(x)f(x)\,.$$

Moreover, for notational convenience, for any discrete time domain $\mathbb{D}$ with canonical time index τ, and any τ_0-measurable function $f(X_{\tau_0}) : \Omega_{\mathbb{D}} \to \mathbb{R}$, we let $\underline{\mathbb{E}}_{\mathscr{M}}[f(X_{\tau_0})] \coloneqq \underline{\mathbb{E}}_{\mathscr{M}}[f]$, where f is the element of $\mathscr{L}(\mathscr{X})$ corresponding to $f(X_{\tau_0})$, as described in Section 2.4$_{71}$.

The following result shows that the map $\underline{\mathbb{E}}_{\mathscr{M}}$ properly captures the lower expectation of functions that depend only on the state at time τ_0, for the discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$.

Proposition 3.30. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. Then for all $f \in \mathscr{L}(\mathscr{X})$ it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_0})] = \underline{\mathbb{E}}_{\mathscr{M}}[f(X_{\tau_0})]\,.$$

Proof. It follows from Lemma 3.21$_{105}$ (with $u = \emptyset$ and $v = \{\tau_0\}$) that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_0})] = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{x\in\mathscr{X}} f(x)P(X_{\tau_0} = x)\,.$$

Now, for any $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ it holds that $P \sim \mathscr{M}$ due to Definition 3.11$_{104}$ which, by the definition in Section 3.3.1$_{102}$, implies that there is some

$p \in \mathscr{M}$ such that $p(x) = P(X_{\tau_0} = x)$ for all $x \in \mathscr{X}$. Conversely, it follows from Lemma 3.19$_{104}$ that for all $p \in \mathscr{M}$, there is some $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ such that $P(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$. This implies that

$$\inf_{P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{x \in \mathscr{X}} f(x) P(X_{\tau_0} = x) = \inf_{p \in \mathscr{M}} \sum_{x \in \mathscr{X}} f(x) p(x) .$$

Hence it follows from Definition 3.14 that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_0})] &= \inf_{P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{x \in \mathscr{X}} f(x) P(X_{\tau_0} = x) \\ &= \inf_{p \in \mathscr{M}} \sum_{x \in \mathscr{X}} f(x) p(x) = \underline{\mathbb{E}}_{\mathscr{M}}[f] .\end{aligned}$$

Finally, again due to Definition 3.14, it holds that $\underline{\mathbb{E}}_{\mathscr{M}}[f(X_{\tau_0})] = \underline{\mathbb{E}}_{\mathscr{M}}[f]$ because $f(X_{\tau_0})$ is the τ_0-measurable function corresponding to f. $\square$

The next result now uses Corollary 3.28$_{112}$ to isolate the lower expectation for the initial model, when considering unconditional lower expectations of functions that depend on multiple time points.

Corollary 3.31. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain such that $\mathscr{T}_k$ has separately specified rows for all $k \in \mathbb{Z}_{\geq 0}$. Then for all $n \in \mathbb{Z}_{>0}$ and all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_0}]\right] .$$

Proof. Because, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k$ has separately specified rows, and because $n > 0$, it follows from Corollary 3.28$_{112}$ (with $\ell = m = 0$) that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_0}]\right] .$$

The inner conditional lower expectation on the right-hand side of this equality is clearly a τ_0-measurable function (which is real-valued due to Lemma 3.23$_{107}$), whence it follows from Proposition 3.30 that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_0}]\right] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_0}]\right] ,$$

which concludes the proof. $\square$

Effectively, Corollary 3.31 tells us that when working with unconditional lower expectations $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})]$, we can separately deal with the lower expectation at time τ_0—through $\underline{\mathbb{E}}_{\mathscr{M}}$—and with the conditional lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_0}]$. In the remainder of this chapter, we will introduce some machinery that allows us to further simplify the expressions for such conditional lower expectations.

3.4 Lower Transition Operators

We discussed in Section 3.2_{89} that transition matrices play an important role in the theory of Markov chains, and that they can be used both for parameterisation and as a computational tool. We also discussed in Section 3.3_{101} how to parameterise imprecise-Markov chains using *sets* of transition matrices. Some crucial results of that section—Propositions 3.24_{108} and 3.25_{110} in particular—showed that imprecise-Markov chains satisfy an imprecise-Markov property, in the sense that their conditional lower expectations are history-independent. Proposition 3.24_{108} derived an expression for these history-independent lower expectations of a function f as an infimum $\inf_{T\in\mathscr{T}} Tf(x)$, which is also called the *lower envelope* of the set $\mathscr{T}$ of transition matrices.

As we will discuss in this section, such lower envelopes are a particular type of *lower transition operators*; essentially, these are (non-linear) generalisations of the transition matrices that we discussed before [22, 48]. We will here study such objects on a relatively abstract level. Although we prove some properties explicitly, we do not aim to present the results in this section as novel. In Section 3.5_{121} we then connect back to discrete-time imprecise-Markov chains, where we will use these lower transition operators to provide expressions for lower expectations. Let us now start with the general definition. We follow [23, Definition 8][6] and [17, Definition 1] in providing the following, relatively abstract, definition.

Definition 3.15 ([23, Definition 8]). *A map $\underline{T}$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ is called a* lower transition operator *if, for all $f, g \in \mathscr{L}(\mathscr{X})$, all $\lambda \in \mathbb{R}_{\geq 0}$, and all $x \in \mathscr{X}$:*

LT1: $\underline{T}f(x) \geq \min_{y\in\mathscr{X}} f(y)$; *(lower bounds)*

LT2: $\underline{T}(f+g)(x) \geq \underline{T}f(x) + \underline{T}g(x)$; *(super-additivity)*

LT3: $\underline{T}(\lambda f)(x) = \lambda \underline{T}f(x)$. *(non-negative homogeneity)*

We will use $\underline{\mathbb{T}}$ to denote the set of all lower transition operators.

Such operators furthermore satisfy the following properties.

[6] The authors of [23] consider more generally what they call *coherence preserving maps*, but they note in [23, Section 7.2] that the lower transition operators in e.g. [22] are a special case of these. As far as we know this is the earliest reference that considers such maps in this context, whence we provide the reference here for historical context, even if the authors did not refer to them explicitly as (only) being lower transition operators.

Proposition 3.32 ([17]). *For any lower transition operator $\underline{T}$, any $f,g \in \mathscr{L}(\mathscr{X})$, any $\mu \in \mathbb{R}$ and any two non-negatively homogeneous operators A,B from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$:*

LT4: $\|\underline{T}\| \leq 1$; *(at most unit norm)*

LT5: $f \leq g \Rightarrow \underline{T}f \leq \underline{T}g$; *(monotonicity)*

LT6: $\underline{T}(f+\mu) = \underline{T}f + \mu$; *(constant additivity)*

LT7: $\|\underline{T}A - \underline{T}B\| \leq \|A-B\|$.

Moreover, as the next result shows, compositions of lower transition operators are also, themselves, lower transition operators. Although the result is well-known—see e.g. [23, Proposition 2]—we provide an explicit proof below because we believe it has some didactic value.

Proposition 3.33 ([23, Proposition 2]). *For any two lower transition operators $\underline{T},\underline{S} \in \underline{\mathbb{T}}$, their composition $\underline{T}\,\underline{S}$ is again a lower transition operator.*

Proof. Consider any $f,g \in \mathscr{L}(\mathscr{X})$, $\lambda \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$. Since $\underline{T}$ and $\underline{S}$ are lower transition operators, they both satisfy LT1, and therefore, we find that

$$\underline{T}\,\underline{S}f(x) \geq \min_{y\in\mathscr{X}} \underline{S}f(y) \geq \min_{y\in\mathscr{X}} \min_{z\in\mathscr{X}} f(z) = \min_{z\in\mathscr{X}} f(z),$$

which implies that $\underline{T}\,\underline{S}$ satisfies LT1 as well. Similarly, $\underline{T}\,\underline{S}$ satisfies LT2 because

$$\underline{T}\,\underline{S}(f+g)(x) \geq \underline{T}(\underline{S}f + \underline{S}g)(x) \geq \underline{T}\,\underline{S}f(x) + \underline{T}\,\underline{S}g(x),$$

where the first inequality follows from LT5 and the fact that $\underline{S}$ satisfies LT2, and where the second inequality follows from the fact that $\underline{T}$ satisfies LT2. Finally, since $\underline{T}$ and $\underline{S}$ both satisfy LT3, it follows that $\underline{T}\,\underline{S}$ also satisfies LT3, because

$$\underline{T}\,\underline{S}(\lambda f)(x) = \underline{T}(\lambda \underline{S}f)(x) = \lambda \underline{T}\,\underline{S}f(x).$$

We conclude that $\underline{T}\,\underline{S}$ satisfies LT1–LT3, and therefore, because of Definition 3.15, it is a lower transition operator. □

It is important to note that any transition matrix T is also a lower transition operator. To see this, observe that T is a linear map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, and hence satisfies both property LT2 (with equality) and property LT3$_{116}$ (even if $\lambda < 0$). To see that it also satisfies property LT1$_{116}$, note that it follows from Definition 3.5$_{91}$ that, for any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$,

$$Tf(x) = \sum_{y\in\mathscr{X}} T(x,y)f(y) \geq \sum_{y\in\mathscr{X}} T(x,y)\min_{z\in\mathscr{X}} f(z) = \min_{z\in\mathscr{X}} f(z)\,,$$

where the inequality used property T2$_{91}$ and the final equality used property T1$_{91}$. Hence, it is therefore clear that lower transition operators are a generalisation of transition matrices.

In order to motivate this specific generalisation, let us now introduce the *lower envelope* of a given set of transition matrices. We need the following result to introduce it.

Lemma 3.34. *For any non-empty set $\mathscr{T}$ of transition matrices, any $f \in \mathscr{L}(\mathscr{X})$, and any $x \in \mathscr{X}$, it holds that $\inf_{T\in\mathscr{T}} Tf(x) \in \mathbb{R}$. In particular, it holds that $\min_{y\in\mathscr{X}} f(y) \leq \inf_{T\in\mathscr{T}} Tf(x) \leq \max_{y\in\mathscr{X}} f(y)$.*

Proof. Fix any $T \in \mathscr{T}$. Then we know from the discussion above that T is a lower transition operator, and hence it follows from Property LT1$_{116}$ that $Tf(x) \geq \min_{y\in\mathscr{X}} f(y)$. Moreover, because T is a linear operator on $\mathscr{L}(\mathscr{X})$ it holds that $-Tf = T(-f)$, and hence it follows from Property LT1$_{116}$ that

$$-Tf(x) = T(-f)(x) \geq \min_{y\in\mathscr{X}} -f(y) = -\max_{y\in\mathscr{X}} f(y)\,,$$

using the conjugacy property $\max\{\cdot\} = -\min\{-\cdot\}$. Hence, reordering terms, we find that $Tf(x) \leq \max_{y\in\mathscr{X}} f(y)$.

Because this is true for all $T \in \mathscr{T}$, it follows that $\min_{y\in\mathscr{X}} f(y) \leq \inf_{T\in\mathscr{T}} Tf(x)$. Moreover, and because $\mathscr{T}$ is non-empty, there is some $S \in \mathscr{T}$ such that $\inf_{T\in\mathscr{T}} Tf(x) \leq Sf(x) \leq \max_{y\in\mathscr{X}} f(y)$.

In summary, we found that $\min_{y\in\mathscr{X}} \leq \inf_{T\in\mathscr{T}} Tf(x) \leq \max_{y\in\mathscr{X}} f(y)$. Since $f \in \mathscr{L}(\mathscr{X})$, and because $\mathscr{X}$ is finite, it holds that both $\min_{y\in\mathscr{X}} f(y)$ and $\max_{y\in\mathscr{X}} f(y)$ are real-valued, and hence $\inf_{T\in\mathscr{T}} Tf(x) \in \mathbb{R}$. □

Let us now give the definition of the lower envelope of $\mathscr{T}$; note that Lemma 3.34 ensures that the codomain of this map is indeed $\mathscr{L}(\mathscr{X})$.

Definition 3.16 (Lower Envelope)**.** *For any non-empty set $\mathscr{T}$ of transition matrices, we define its* lower envelope $\underline{T} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{T}f$ *where, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$, we let $\underline{T}f(x) := \inf_{T\in\mathscr{T}} Tf(x)$.*

The next result is well-known in the imprecise-Markov chain literature, but because it is so crucial we will prove it here for completeness.

Proposition 3.35 ([23, Proposition 4])**.** *For any non-empty set $\mathscr{T}$ of transition matrices, its lower envelope $\underline{T}$ is a lower transition operator.*

Proof. Fix any $x \in \mathscr{X}$ and any $f \in \mathscr{L}(\mathscr{X})$, and consider any $T \in \mathscr{T}$. Because, as we have seen above, T is a lower transition operator, it satisfies property LT1$_{116}$, and hence $Tf(x) \geq \min_{y\in\mathscr{X}} f(y)$. Because this is true for all $T \in \mathscr{T}$, it follows that $\underline{T}f(x) = \inf_{T\in\mathscr{T}} Tf(x) \geq \min_{y\in\mathscr{X}} f(y)$. Because $x \in \mathscr{X}$ and $f \in \mathscr{L}(\mathscr{X})$ are arbitrary, this means that $\underline{T}$ satisfies property LT1$_{116}$.

Properties LT2$_{116}$ and LT3$_{116}$ follow directly from the properties of the infimum and the fact that the elements $T \in \mathscr{T}$ are linear maps. That is, for any $f, g \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$,

$$\begin{aligned}\big[\underline{T}(f+g)\big](x) &= \inf_{T\in\mathscr{T}} \big[T(f+g)\big](x)\\ &= \inf_{T\in\mathscr{T}} \big(Tf(x) + Tg(x)\big)\\ &\geq \inf_{T\in\mathscr{T}} Tf(x) + \inf_{T\in\mathscr{T}} Tg(x) = \underline{T}f(x) + \underline{T}g(x)\,,\end{aligned}$$

whence $\underline{T}$ satisfies property LT2$_{116}$.

Similarly, for any $f \in \mathscr{L}(\mathscr{X})$, $\lambda \in \mathbb{R}_{\geq 0}$, and $x \in \mathscr{X}$ it holds that

$$\begin{aligned}\big[\underline{T}(\lambda f)\big](x) = \inf_{T\in\mathscr{T}} \big[T(\lambda f)\big](x) &= \inf_{T\in\mathscr{T}} \lambda Tf(x)\\ &= \lambda \inf_{T\in\mathscr{T}} Tf(x) = \lambda \underline{T}f(x)\,,\end{aligned}$$

and hence $\underline{T}$ satisfies property LT3$_{116}$.

Because $\underline{T}$ satisfies properties LT1$_{116}$–LT3$_{116}$, it is a lower transition operator by Definition 3.15$_{116}$. □

Due to Proposition 3.35, we also refer to the lower envelope $\underline{T}$ of a given set $\mathscr{T}$ of transition matrices, as the *lower transition operator corresponding to $\mathscr{T}$*.

The following result provides sufficient conditions on the set $\mathscr{T}$ for the value of $\underline{T}f$ to be reached by Tf, for some $T \in \mathscr{T}$, where $\underline{T}$ is the lower transition operator corresponding to $\mathscr{T}$. In other words, under those conditions the lower envelope is actually a minimum, rather than an infimum; and in particular, this minimum is achieved uniformly over all elements of $\mathscr{X}$.

Proposition 3.36. *Let $\mathscr{T}$ be a non-empty and closed set of transition matrices that has separately specified rows, and let $\underline{T}$ be its corresponding lower transition operator. Then for all $f \in \mathscr{L}(\mathscr{X})$, there is some $T \in \mathscr{T}$ such that $Tf(x) = \underline{T}f(x)$ for all $x \in \mathscr{X}$.*

Proof. Fix any $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$, and consider a sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{\geq 0}$ such that $\lim_{i\to+\infty} \varepsilon_i = 0$. Then, for all $i \in \mathbb{Z}_{>0}$, because of Definition 3.16, there is some $T_x^{(i)} \in \mathscr{T}$ such that $\underline{T}f(x) \leq T_x^{(i)} f(x) < \underline{T}f(x) + \varepsilon_i$.

Note that the set $\mathscr{T}$ is bounded because, using Lemma 3.9$_{92}$, it holds that $\|\mathscr{T}\| = \sup_{T\in\mathscr{T}} \|T\| = 1$. Since $\mathscr{T}$ is also closed by assumption, it follows from Corollary A.12$_{378}$ that it is sequentially compact. Hence, the sequence $\{T_x^{(i)}\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{T}$ has a convergent subsequence, $\{T_x^{(i_k)}\}_{k\in\mathbb{Z}_{>0}}$, say, and $\lim_{k\to+\infty} T_x^{(i_k)} =: T_x^* \in \mathscr{T}$.

Moreover, for all $k \in \mathbb{Z}_{>0}$ it holds that

$$\underline{T}\,f(x) \leq T_x^{(i_k)} f(x) < \underline{T}\,f(x) + \varepsilon_{i_k}\,,$$

and hence $T_x^* f(x) = \underline{T}\,f(x)$ because also $\lim_{k\to+\infty} \varepsilon_{i_k} = 0$.

Now, because $\mathscr{T}$ has separately specified rows, there is some $T \in \mathscr{T}$ such that $T(x,\cdot) = T_x^*(x,\cdot)$ for all $x \in \mathscr{X}$. Because of the above, it therefore holds that $Tf(x) = \underline{T}\,f(x)$ for all $x \in \mathscr{X}$. □

So, we have seen above that any set $\mathscr{T}$ has a corresponding lower transition operator. We will now reason in the opposite direction; given an arbitrary lower transition operator $\underline{T}$, is there a set $\mathscr{T}$ of transition matrices that corresponds to it? To this end, we consider the set of transition matrices that *dominate* this lower transition operator, as follows.

Definition 3.17. *For any lower transition operator $\underline{T}$, we define its* dominating set of transition matrices *$\mathscr{T}_{\underline{T}}$ as*

$$\mathscr{T}_{\underline{T}} := \left\{ T \in \mathbb{T} \,\middle|\, Tf \geq \underline{T}\,f \text{ for all } f \in \mathscr{L}(\mathscr{X}) \right\}.$$

It turns out that this set of dominating transition matrices satisfies a number of convenient properties. This result is also well-known in the literature, although it is not often proved explicitly. Hence, we here provide a proof for the sake of completeness; because it is a bit involved, however, it is deferred to Appendix 3.C$_{137}$.

Proposition 3.37. *For any lower transition operator $\underline{T}$, its dominating set of transition matrices $\mathscr{T}_{\underline{T}}$ is a non-empty, closed, and convex set of transition matrices that has separately specified rows, and that has $\underline{T}$ as its corresponding lower transition operator.*

These properties characterise $\mathscr{T}_{\underline{T}}$ completely, in the sense that no other set satisfies them.

Corollary 3.38. *Let $\mathscr{T}$ be a non-empty, closed, and convex set of transition matrices that has separately specified rows, and that has $\underline{T}$ as its corresponding lower transition operator. Then $\mathscr{T} = \mathscr{T}_{\underline{T}}$.*

We conclude this section with some results about the set $\underline{\mathbb{T}}$ of all lower transition operators. In particular, we will later be interested in sequences, and limits of sequences, of lower transition operators. The following technical results will therefore be helpful.

Lemma 3.39. *[17, Proposition 1] Consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{Z}_{>0}}$ of lower transition operators such that $\underline{T}f = \lim_{i\to+\infty}\underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$. Then $\underline{T}$ is a lower transition operator.*

Lemma 3.40. *[17, Proposition 2] Let $\underline{T}$ be a lower transition operator, and consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{Z}_{>0}}$ of lower transition operators. Then, $\underline{T} = \lim_{i\to+\infty}\underline{T}_i$ if and only if $\underline{T}f = \lim_{i\to+\infty}\underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$.*

Proposition 3.41. *$\underline{\mathbb{T}}$ is complete under the metric induced by our norm $\|\cdot\|$.*

Proof. Consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{Z}_{>0}}$ of lower transition operators that is Cauchy with respect to the operator norm $\|\cdot\|$. We will prove that $\{\underline{T}_i\}_{i\in\mathbb{Z}_{>0}}$ converges to a limit $\underline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ that is itself a lower transition operator.

Consider any $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$. For any $k, \ell \in \mathbb{Z}_{>0}$, (N11)$_{64}$ then implies that

$$|\underline{T}_k f(x) - \underline{T}_\ell f(x)| \le \|\underline{T}_k f - \underline{T}_\ell f\| = \|(\underline{T}_k - \underline{T}_\ell)f\| \le \|\underline{T}_k - \underline{T}_\ell\| \, \|f\| .$$

Therefore, and because $\{\underline{T}_i\}_{i\in\mathbb{Z}_{>0}}$ is Cauchy with respect to the norm $\|\cdot\|$, it follows that $\{\underline{T}_i f(x)\}_{i\in\mathbb{Z}_{>0}}$ is Cauchy with respect to the norm $|\cdot|$. Hence, since $\mathbb{R}$ is (well known to be) complete with respect to the topology that is induced by $|\cdot|$, we find that $\{\underline{T}_i f(x)\}_{i\in\mathbb{Z}_{>0}}$ converges to a limit in $\mathbb{R}$, which we will denote by $\underline{T}f(x)$. Let $\underline{T}f$ be the unique function in $\mathscr{L}(\mathscr{X})$ that has $\underline{T}f(x)$, $x \in \mathscr{X}$, as its components. Then clearly, $\underline{T}f = \lim_{i\to+\infty}\underline{T}_i f$.

Let $\underline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ be the unique operator that maps any $f \in \mathscr{L}(\mathscr{X})$ to $\underline{T}f$. It then follows from Lemma 3.39 that $\underline{T}$ is a lower transition operator. Therefore, and because we already know that $\underline{T}f = \lim_{i\to+\infty}\underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$, it now follows from Lemma 3.40 that $\lim_{i\to+\infty}\underline{T}_i = \underline{T}$. □

3.5 Lower Expectations using Lower Transition Operators

Let us now connect the lower transition operators that we discussed in the previous section, to the discrete-time imprecise-Markov chains that we introduced earlier. Again, the results in this section are well-known in the imprecise-Markov chain literature. Let us start with the following result, which shows that we can write the one-step conditional lower expectations of an imprecise-Markov chain, using the corresponding lower transition operators.[7]

Proposition 3.42. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain. For all $k \in \mathbb{Z}_{\geq 0}$, let $\underline{T}_k$ denote the lower transition operator corresponding to $\mathscr{T}_k$. Then for all $f \in \mathscr{L}(\mathscr{X})$, all $n \in \mathbb{Z}_{\geq 0}$, and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \underline{T}_n f(x_{\tau_n})\,. \tag{3.26}$$

Proof. By Proposition 3.24_{108} and Definition 3.16_{118}, it holds that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \inf_{T \in \mathscr{T}_n} Tf(x_{\tau_n}) = \underline{T}_n f(x_{\tau_n})\,,$$

which concludes the proof. □

We now recall from the discussion preceding Corollary 3.26_{110} the notational convention that, for any $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}})$ and any $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, we write $f(x_{\tau_{0:n}},\cdot)$ for the projection of f onto $\mathscr{L}(\mathscr{X}_{\tau_{n+1}})$ obtained by fixing $x_{\tau_{0:n}}$, i.e. the element of $\mathscr{L}(\mathscr{X}_{\tau_{n+1}})$ corresponding to the τ_{n+1}-measurable function $f(x_{\tau_{0:n}},X_{\tau_{n+1}})$. With a slight abuse of notation, let us now introduce the following notational trick, which will be very convenient in the sequel: for any lower transition operator $\underline{T} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$, and for any discrete time domain $\mathbb{D}$ with canonical time index τ, we introduce for any $n \in \mathbb{Z}_{\geq 0}$ an associated operator $\underline{T} : \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}}) \to \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$. We define this map, for all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}})$ and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, as

$$\underline{T} f(x_{\tau_{0:n}}) := \big[\underline{T} f(x_{\tau_{0:n}},\cdot)\big](x_{\tau_n})\,,$$

where, on the right-hand side, we have applied the (original) operator $\underline{T} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ to $f(x_{\tau_{0:n}},\cdot)$. All of this is just a formal way of saying that we allow the operator $\underline{T}$ to be applied to f, by applying it a specific projection of f onto $\mathscr{L}(\mathscr{X}_{\tau_{n+1}})$. Because this projection depends on the value of $x_{\tau_{0:n}}$, the resulting function $\underline{T} f$ is an element of $\mathscr{L}(\mathscr{X}_{\tau_{0:n}})$.

Using this notation, we can reformulate Corollary 3.26_{110} using lower transition operators, as follows.

[7] Although the result here is derived, Reference [22] actually took the equality (3.26) as the definition of the lower transition operator corresponding to an imprecise-Markov chain, without reference to any set of transition matrices. However, because of the correspondence between lower transition operators and sets of transition matrices, which we discussed in Section 3.4_{116}, this turns out to be equivalent.

Corollary 3.43. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain such that $\mathscr{T}_k$ has corresponding lower transition operator $\underline{T}_k$, for all $k \in \mathbb{Z}_{\geq 0}$. Then for all $n \in \mathbb{Z}_{\geq 0}$, all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:(n+1)}})$, and all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:(n+1)}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \underline{T}_n f(x_{\tau_{0:n}})\,.$$

Proof. By Lemma3.22$_{107}$ it holds that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:(n+1)}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_{\tau_{0:n}}, X_{\tau_{n+1}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}]\,,$$

and it follows from Proposition 3.42 that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(x_{\tau_{0:n}}, X_{\tau_{n+1}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \left[\underline{T}_n f(x_{\tau_{0:n}}, \cdot)\right](x_{\tau_n})\,.$$

Using the notation that we introduced above that allows us to apply $\underline{T}_n$ directly to f, we have by definition that

$$\underline{T}_n f(x_{\tau_{0:n}}) = \left[\underline{T}_n f(x_{\tau_{0:n}}, \cdot)\right](x_{\tau_n})\,,$$

and hence

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:(n+1)}}) \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}] = \underline{T}_n f(x_{\tau_{0:n}})\,,$$

which concludes the proof. ☐

Finally, we can use the above notational convention to formulate the law of iterated lower expectations using lower transition operators. This result is also well-known in the literature; see e.g. [22, Theorem 3.1] or [48, Theorem 11.2] for analogous results. Although, as noted above, these references use a slightly differ manner to derive the corresponding lower transition operators, the result itself is conceptually essentially the same.

Proposition 3.44. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ be a discrete-time imprecise-Markov chain such that $\mathscr{T}_k$ has separately specified rows and corresponding lower transition operator $\underline{T}_k$, for all $k \in \mathbb{Z}_{\geq 0}$. Then for all $n, m \in \mathbb{Z}_{\geq 0}$ such that $n > m$, all $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$, and all $x_{\tau_{0:m}} \in \mathscr{X}_{\tau_{0:m}}$, it holds that*

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_{0:m}} = x_{\tau_{0:m}}] = \underline{T}_m \underline{T}_{m+1} \cdots \underline{T}_{n-1} f(x_{\tau_{0:m}})\,.$$

Proof. Because, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k$ has separately specified rows, using Lemma 3.27$_{111}$ it holds that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}&[f(X_{\tau_{0:n}}) \,|\, X_{\tau_{0:m}} = x_{\tau_{0:m}}] \\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_{0:(n-1)}}] \,\middle|\, X_{\tau_{0:m}} = x_{\tau_{0:m}}\right]. \end{aligned} \tag{3.27}$$

Using Corollary 3.43$_\frown$, we find that for any $y_{\tau_{0:(n-1)}} \in \mathscr{X}_{\tau_{0:(n-1)}}$ it holds that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}} = y_{\tau_{0:(n-1)}}] = \underline{T}_{n-1} f(y_{\tau_{0:(n-1)}})\,.$$

Because this is true for all $y_{\tau_{0:(n-1)}} \in \mathscr{X}_{\tau_{0:(n-1)}}$, it follows that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}] = \underline{T}_{n-1} f(X_{\tau_{0:(n-1)}})\,,$$

where the right-hand side denotes the $\tau_{0:(n-1)}$-measurable function corresponding to the element $\underline{T}_{n-1}f$ of $\mathscr{L}(\mathscr{X}_{\tau_{0:(n-1)}})$. Substituting this into Equation (3.27)$_\frown$ we obtain

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:m}} = x_{\tau_{0:m}}] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{T}_{n-1} f(X_{\tau_{0:(n-1)}})\,\Big|\,X_{\tau_{0:m}} = x_{\tau_{0:m}}\Big]\,.$$

Here, the right-hand side is a conditional lower expectation of the $\tau_{0:(n-1)}$ measurable function $\underline{T}_{n-1}f(X_{\tau_{0:(n-1)}})$. We can now proceed by repeatedly using the above argumentation to expand these functions, until we find that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}&[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:m}} = x_{\tau_{0:m}}]\\ &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{T}_{m+1}\cdots\underline{T}_{n-1} f(X_{\tau_{0:(m+1)}})\,\Big|\,X_{\tau_{0:m}} = x_{\tau_{0:m}}\Big]\,.\end{aligned}$$

Here, the right-hand side is taken over the $\tau_{0:(m+1)}$-measurable function $\underline{T}_{m+1}\cdots\underline{T}_{n-1}f(X_{\tau_{0:(m+1)}})$, so one last use of Corollary 3.43$_\frown$ yields

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:m}} = x_{\tau_{0:m}}] = \underline{T}_m\underline{T}_{m+1}\cdots\underline{T}_{n-1} f(x_{\tau_{0:m}})\,,$$

which concludes the proof. □

Proposition 3.44$_\frown$ essentially tells us that we can use compositions of lower transition operators, to compute lower expectations of functions that depend on the state of an imprecise-Markov chain at multiple time points. This observation is especially important, since it forms the basis of many efficient inference algorithms that have been published in the literature, see e.g. [107] for a general result that captures a wide variety of inference problems, or [18] for an application to *hidden* imprecise-Markov chains.

Finally, by combining this result with Corollary 3.31$_{115}$, we find that for any $n \in \mathbb{Z}_{>0}$ we can also express the unconditional lower expectation of any $f \in \mathscr{L}(\mathscr{X}_{\tau_{0:n}})$ as

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\big[f(X_{\tau_{0:n}})\big] = \underline{\mathbb{E}}_{\mathscr{M}}\big[\underline{T}_0\cdots\underline{T}_n f(X_{\tau_0})\big]\,.$$

Using this identity, it is possible to leverage the efficient inference algorithms mentioned above, also when computing unconditional lower expectations for the imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$.

Appendix

3.A Proofs of Results in Section 3.1

*Proof of Lemma 3.1*86. Let us recall from Section 2.364 that $\mathscr{A}_{\emptyset}^{\mathbb{D}} = \langle \mathscr{E}_{\emptyset}^{\mathbb{D}} \rangle$. Because two generated algebras are equal when they contain each other's generators, it therefore suffices to show that $\mathscr{E}_{\emptyset}^{\mathbb{D}} \subseteq \langle \mathscr{S}_{\mathbb{D}} \rangle$ and $\mathscr{S}_{\mathbb{D}} \subseteq \mathscr{A}_{\emptyset}^{\mathbb{D}}$.

So, for the first inclusion, fix any $(X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} \in \mathscr{E}_{\emptyset}^{\mathbb{D}}$, with $n \in \mathbb{Z}_{\geq 0}$ and $x_{\tau_n} \in \mathscr{X}$. Then, for all $x_{\tau_{0:(n-1)}} \in \mathscr{X}_{\tau_{0:(n-1)}}$, it follows from Definition 3.185 that $(X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$. Moreover, we have that

$$(X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} = \bigcup_{x_{\tau_{0:(n-1)}} \in \mathscr{X}_{\tau_{0:(n-1)}}} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}}\,.$$

Because $\mathscr{X}_{\tau_{0:(n-1)}}$ is finite since $n \in \mathbb{Z}_{\geq 0}$, and because $\langle \mathscr{S}_{\mathbb{D}} \rangle$ is closed under finite unions, it follows that $(X_{\tau_n} = x_{\tau_n})_{\mathbb{D}} \in \langle \mathscr{S}_{\mathbb{D}} \rangle$. This concludes the argument in the first direction.

For the other direction, fix any $(X_u = x_u)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$. Then if $u = \emptyset$ it holds that $(X_u = x_u)_{\mathbb{D}} = \Omega_{\mathbb{D}}$, in which case we trivially have $(X_u = x_u)_{\mathbb{D}} \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ since $\mathscr{A}_{\emptyset}^{\mathbb{D}}$ is an algebra. So, let us suppose that $u = \tau_{0:n}$, with $n \in \mathbb{Z}_{\geq 0}$. Then, for all $i \in \{0, \ldots, n\}$ it holds that $(X_{\tau_i} = x_{\tau_i})_{\mathbb{D}} \in \mathscr{E}_{\emptyset}^{\mathbb{D}}$. Moreover, we have

$$(X_u = x_u)_{\mathbb{D}} = \bigcap_{i=0}^{n} (X_{\tau_i} = x_{\tau_i})_{\mathbb{D}}\,,$$

from which it follows that $(X_u = x_u)_{\mathbb{D}} \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ since n is finite and $\mathscr{A}_{\emptyset}^{\mathbb{D}}$ is closed under finite intersections. This concludes the argument in the second direction. □

*Proof of Lemma 3.3*86. Because $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ it follows from Proposition 2.1866 that there are $v \in \mathscr{U}^{\mathbb{D}}$ and $S \subseteq \mathscr{X}_v$ such that $A = \cup_{x_v \in S}(X_v = x_v)_{\mathbb{D}}$. Now let $n \in \mathbb{Z}_{\geq 0}$ be such that $\tau_n = \max v$—this is always possible since $v \subset \mathbb{D}$—and let

$$S' := \left\{ x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}} : x_v \in S \right\},$$

i.e. we collect in S' all elements of $\mathscr{X}_{\tau_{0:n}}$ whose value in the time points in v belongs to S. Then, clearly,

$$\bigcup_{x_{\tau_{0:n}} \in S'} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}} = \bigcup_{x_v \in S} (X_v = x_v)_{\mathbb{D}} = A\,,$$

which concludes the proof. □

Proof of Lemma 3.4$_{87}$. First suppose that $m < n$ and $y_v = x_v$. Then $v \subseteq u$, and hence

$$(X_u = x_u)_{\mathbb{D}} = (X_v = x_v)_{\mathbb{D}} \cap (X_{u\setminus v} = x_{u\setminus v})_{\mathbb{D}}$$
$$= (X_v = y_v)_{\mathbb{D}} \cap (X_{u\setminus v} = x_{u\setminus v})_{\mathbb{D}} \subseteq (X_v = y_v)_{\mathbb{D}}\,.$$

Therefore, and because $P \in \mathbb{P}^{\mathbb{D}}$ is a coherent conditional probability due to Definition 2.13$_{69}$, it follows that $P(X_v = y_v \,|\, x_u = x_u) = 1$ due to Property F2$_{47}$.

Next, suppose that $n \leq m$ and $y_u = x_u$. Then for all $i \in \{n, \ldots, m\}$ it holds that

$$\begin{aligned} &P(X_{\tau_{0:i}} = y_{\tau_{0:i}} \,|\, X_u = y_u) \\ &\quad= P(X_{\tau_i} = y_{\tau_i} \,|\, X_{\tau_{0:(i-1)}} = y_{\tau_{0:(i-1)}}) P(X_{\tau_{0:(i-1)}} = y_{\tau_{0:(i-1)}} \,|\, X_u = y_u) \\ &\quad= p(y_{\tau_i} \,|\, y_{\tau_{0:(i-1)}}) P(X_{\tau_{0:(i-1)}} = y_{\tau_{0:(i-1)}} \,|\, X_u = y_u)\,, \end{aligned}$$

where for the first equality we use Property F4$_{47}$ and that P is a coherent conditional probability, and for the second equality we use that P corresponds to p, as in Definition 3.2$_{86}$. Because this is true for all $i \in \{n, \ldots, m\}$, it follows that

$$\begin{aligned} P(X_{\tau_{0:m}} = y_{\tau_{0:m}} \,|\, X_u = y_u) &= P(X_u = y_u \,|\, X_u = y_u) \prod_{i=n}^{m} p(y_{\tau_i} \,|\, y_{\tau_{0:(i-1)}}) \\ &= \prod_{i=n}^{m} p(y_{\tau_i} \,|\, y_{\tau_{0:(i-1)}})\,, \end{aligned}$$

where for the second equality we use Property F2$_{47}$. Because $y_u = x_u$ and $v = \tau_{0:m}$ we therefore conclude that

$$P(X_v = y_v \,|\, X_u = x_u) = \prod_{i=n}^{m} p(y_{\tau_i} \,|\, y_{\tau_{0:(i-1)}})\,.$$

Finally, suppose that either $n \leq m$ and $y_u \neq x_u$, or $m < n$ and $y_v \neq x_v$. In either case we have that $(X_u = x_u)_{\mathbb{D}} \cap (X_v = y_v)_{\mathbb{D}} = \emptyset$, and hence it follows from the fact that P is a coherent conditional probability, together with Properties F6$_{47}$ and F7$_{47}$, that

$$P(X_v = y_v \,|\, X_u = x_u) = P(\emptyset \,|\, X_u = x_u) = 0\,,$$

which concludes the proof. □

The following result is expressed for general time domains $\mathbb{H}$ rather than just discrete time domains; this is with the aim of re-using it later to prove some results in Chapter 5$_{181}$.

Lemma 3.45. *Let $\mathbb{H}$ be a time domain, choose any $m \in \mathbb{Z}_{\geq 0}$ and let $w = \{w_0, w_1, \dots, w_m\} \subset \mathbb{H}$ be a finite set of time points such that $w_0 < w_1 < \dots < w_m$. Let P_w be a real-valued function on*

$$\mathscr{C}_w :=$$
$$\big\{(X_{w_j} = y, X_u = x_u)_{\mathbb{H}} \colon\, j \in \{0, \dots, m\},\ u = \{w_0, \dots, w_{j-1}\}, y \in \mathscr{X},\ x_u \in \mathscr{X}_u\big\}$$

such that, for any $j \in \{0, \dots, m\}$, $u = \{w_0, \dots, w_{j-1}\}$ and $x_u \in \mathscr{X}_u$, $P_w(X_{w_j} = x \,|\, X_u = x_u)$, as a function of $x \in \mathscr{X}$, is a probability mass function on $\mathscr{X}$. Then P_w is a coherent conditional probability.

Proof. We provide a proof by induction on m. So choose any $0 \leq m$ and suppose that the statement is true for all $m' \in \mathbb{Z}_{\geq 0}$ with $m' < m$; this is trivially true for $m = 0$, which provides the induction base for the argument. We will show that this implies that the statement is also true for m.

Consider any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \dots, n\}$, choose $(A_i, C_i)_{\mathbb{H}} \in \mathscr{C}_w$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max\left\{\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C_0\right\} \geq 0, \tag{3.28}$$

with $C_0 := \cup_{i=1}^{n} C_i$.

For any $i \in \{1, \dots, n\}$, since $(A_i, C_i)_{\mathbb{H}} \in \mathscr{C}_w$, there is some $j_i \in \{0, \dots, m\}$ and, for all $\ell \in \{0, \dots, j_i\}$, some $z_{\ell,i} \in \mathscr{X}$ such that

$$A_i = (X_{w_{j_i}} = z_{j_i,i})_{\mathbb{H}} \;\text{ and }\; C_i = (X_{w_0} = z_{0,i}, \dots, X_{w_{j_i-1}} = z_{j_i-1,i})_{\mathbb{H}}.$$

Let $S = \{i \in \{1, \dots, n\} \colon\, j_i < m\}$. If $S \neq \emptyset$, then by the induction hypothesis, we know that

$$\max\left\{\sum_{i \in S} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C_0^*\right\} \geq 0,$$

with $C_0^* := \cup_{i \in S} C_i$. It follows that there is some $\omega^* \in C_0^* \subseteq C_0$ such that

$$\sum_{i \in S} \lambda_i \mathbb{I}_{C_i}(\omega^*)\big(P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega^*)\big) \geq 0. \tag{3.29}$$

If $S = \emptyset$, then let ω^* be any element of C_0. Equation (3.29) is then trivially satisfied. Hence, in all cases, we have found some $\omega^* \in C_0$ that satisfies Equation (3.29).

Let $C^* := \cap_{0 \leq \ell < m}\big(X_{w_\ell} = \omega^*(w_\ell)\big)_{\mathbb{H}}$ and $S^* := \{i \in \{1, \dots, n\} \colon C_i = C^*\}$. Then by the assumptions of this lemma, there is some probability mass function p on $\mathscr{X}$ such that, for all $x \in \mathscr{X}$, $P_w(X_{w_m} = x|C^*) = p(x)$. For all

$x \in \mathscr{X}$, let $\lambda_x := \sum_{\{i \in S^* : \, z_{m,i}=x\}} \lambda_i$. Now let y^* be any element of $\mathscr{X}$ such that $\min_{y \in \mathscr{X}} \lambda_y = \lambda_{y^*}$ (since $\mathscr{X}$ is finite, this is always possible). Since p is a probability mass function, it then follows that

$$\sum_{i \in S^*} \lambda_i P_w(A_i|C^*) = \sum_{x \in \mathscr{X}} \lambda_x p(x) \geq \sum_{x \in \mathscr{X}} \lambda_{y^*} p(x) = \lambda_{y^*}.$$

Let ω^{**} be any path in $\Omega_{\mathbb{H}}$ such that $\omega^{**} \in C^*$ and $\omega^{**}(w_m) = y^*$; Equation $(2.8)_{65}$ guarantees that this $\omega^{**} \in \Omega_{\mathbb{H}}$ exists. Then

$$\sum_{i \in S^*} \lambda_i \big(P_w(A_i|C^*) - \mathbb{I}_{A_i}(\omega^{**})\big) \geq \min_{y \in \mathscr{X}} \lambda_y - \sum_{i \in S^*} \lambda_i \mathbb{I}_{A_i}(\omega^{**}) = \lambda_{y^*} - \lambda_{y^*} = 0,$$

where the first equality holds because $A_i = (X_{w_m} = z_{m,i})_{\mathbb{H}}$ for all $i \in S^*$.

Let $S^{**} := \{1, \ldots, n\} \setminus (S \cup S^*)$. Since $\omega^{**} \in C^*$, we find that $\mathbb{I}_{C_i}(\omega^{**}) = \mathbb{I}_{C_i}(\omega^*)$ and $\mathbb{I}_{A_i}(\omega^{**}) = \mathbb{I}_{A_i}(\omega^*)$ for all $i \in S$, that $\mathbb{I}_{C_i}(\omega^{**}) = 1$ for all $i \in S^*$, and that $\mathbb{I}_{C_i}(\omega^{**}) = 0$ for all $i \in S^{**}$. Hence, it follows from Equation $(3.29)_{\curvearrowleft}$ that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega^{**}) \big(P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{**})\big) \geq \sum_{i \in S^*} \lambda_i \big(P_w(A_i|C^*) - \mathbb{I}_{A_i}(\omega^{**})\big).$$

By combining this inequality with the previous one, we find that in order to show that Equation $(3.28)_{\curvearrowleft}$ holds, it suffices to prove that $\omega^{**} \in C_0$.

In order to prove this, it suffices to notice that the question of whether or not a path $\omega \in \Omega_{\mathbb{H}}$ belongs to C_0, only depends on the values $\omega(t)$ of ω at time points $t \in \{w_0, \ldots, w_{m-1}\}$. Indeed, since we infer from $\omega^{**} \in C^*$ that the value of ω^* and ω^{**} at these time points is the same, and because $\omega^* \in C_0$, this implies that $\omega^{**} \in C_0$. $\square$

Proof of Theorem 3.5_{88}. Consider the set

$$\mathscr{C} := \left\{ (X_{\tau_n} = x_{\tau_n}, X_{\tau_{0:(n-1)}} = x_{\tau_{0:(n-1)}})_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}} \,\middle|\, n \in \mathbb{Z}_{\geq 0},\, x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}} \right\}, \tag{3.30}$$

and define $P_* : \mathscr{C} \to \mathbb{R}$, for all $(X_{\tau_n} = x_{\tau_n}, X_{\tau_{0:(n-1)}} = x_{\tau_{0:(n-1)}})_{\mathbb{D}} \in \mathscr{C}$, as

$$P_*\big(X_{\tau_n} = x_{\tau_n} \,\big|\, X_{\tau_{0:(n-1)}} = x_{\tau_{0:(n-1)}}\big) := p\big(x_{\tau_n} \,\big|\, x_{\tau_{0:(n-1)}}\big). \tag{3.31}$$

We will first show that P_* is a coherent conditional probability on $\mathscr{C}$. So fix any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \ldots, n\}$, any $(A_i, C_i)_{\mathbb{D}}$ in $\mathscr{C}$ and $\lambda_i \in \mathbb{R}$. According to Definition 2.2_{48}, we need to show that

$$\max \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega) \big(P_*(A_i \,|\, C_i) - \mathbb{I}_{A_i}(\omega)\big) \,\middle|\, \omega \in C_0 \right\} \geq 0, \tag{3.32}$$

with $C_0 := \cup_{i=1}^n C_i$. Now for any $i \in \{1,\dots,n\}$, because $(A_i,C_i)_{\mathbb{D}} \in \mathscr{C}$, it follows from Equation (3.30) that there is some $n_i \in \mathbb{Z}_{\geq 0}$ such that $A_i = (X_{\tau_{n_i}} = x_{\tau_{n_i}})_{\mathbb{D}}$ and $C_i = (X_{\tau_{0:(n_i-1)}} = x_{\tau_{0:(n_i-1)}})_{\mathbb{D}}$ for some $x_{\tau_{0:n_i}} \in \mathscr{X}_{x_{\tau_{0:n_i}}}$. Let $m := \max_{i\in\{1,\dots,n\}} n_i$, define $w := \tau_{0:m}$, and consider the set

$$\mathscr{C}_w := \left\{ (X_{\tau_j} = x_{\tau_j}, X_{\tau_{0:(j-1)}} = x_{\tau_{0:(j-1)}})_{\mathbb{D}} \,:\, j \in \{0,\dots,m\},\, x_{\tau_{0:j}} \in \mathscr{X}_{\tau_{0:j}} \right\}.$$

Then it follows from Equation (3.30) that $\mathscr{C}_w \subseteq \mathscr{C}$ and, in particular, for all $i \in \{1,\dots,n\}$, that $(A_i,C_i)_{\mathbb{D}} \in \mathscr{C}_w$ because $m \geq n_i$.

Let P'_* be the restriction of P_* to $\mathscr{C}_w$. Then since p is a probability tree, and because of Equation (3.31) and the fact that P'_* is the restriction of P_*, it follows that P'_* satisfies the conditions of Lemma 3.45_{127} and, hence, by Lemma 3.45_{127}, that P'_* is a coherent conditional probability on $\mathscr{C}_w$. Because $(A_i,C_i)_{\mathbb{D}} \in \mathscr{C}_w$ for all $i \in \{1,\dots,n\}$, it therefore follows from Definition 2.2_{48} that

$$\max\left\{ \sum_{i=1}^n \lambda_i \mathbb{I}_{C_i}(\omega)\big(P'_*(A_i\,|\,C_i) - \mathbb{I}_{A_i}(\omega)\big) \,\middle|\, \omega \in C_0 \right\} \geq 0,$$

which, because $P_*(A_i\,|\,C_i) = P'_*(A_i\,|\,C_i)$ for all $i \in \{1,\dots,n\}$ since P'_* is the restriction of P_*, implies that Equation (3.32) is also satisfied. This implies that P_* is a coherent conditional probability on $\mathscr{C}$.

Because P_* is a coherent conditional probability on $\mathscr{C}$, and because $\mathscr{C} \subset \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$, it follows from Theorem 2.3_{49} that P_* can be extended to a coherent conditional probability P on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ which, by Definition 2.13_{69}, implies that $P \in \mathbb{P}^{\mathbb{D}}$ is a discrete-time stochastic process with time domain $\mathbb{D}$. Moreover, it follows from Equation (3.31) and the fact that P extends P_*, that P satisfies Equations $(3.2)_{87}$ and $(3.3)_{87}$ and therefore, by Definition 3.2_{86}, that P corresponds to p.

It remains to prove that P is the unique element of $\mathbb{P}^{\mathbb{D}}$ that corresponds to p. To this end, consider any $P' \in \mathbb{P}^{\mathbb{D}}$ that corresponds to p; we will show that $P = P'$. So, consider any $(A,C)_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$. Then it follows from Lemma 3.2_{86} that $C \in \mathscr{S}_{\mathbb{D}}$, which implies that there is some $n \in \mathbb{Z}_{\geq 0}$ and, with $u := \tau_{0:(n-1)}$, some $x_u \in \mathscr{X}_u$, such that $C = (X_u = x_u)_{\mathbb{D}}$. Moreover, it follows from Lemma 2.19_{68} that $A \in \mathscr{A}^{\mathbb{D}}_{\emptyset}$, which by Lemma 3.3_{86} implies that there is some $m \in \mathbb{Z}_{\geq 0}$ and, with $v := \tau_{0:m}$, some $S \subseteq \mathscr{X}_v$, such that $A = \cup_{y_v \in S}(X_v = y_v)_{\mathbb{D}}$.

Because P and P' both correspond to p, it follows from Lemma 3.4_{87} that for all $y_v \in S$ it holds that

$$P(X_v = y_v\,|\,X_u = x_u) = P'(X_v = y_v\,|\,X_u = x_u).$$

Therefore, and because P and P' are both coherent conditional probabilities, it follows from Property $\mathrm{F3}_{47}$ and the fact that $A = \cup_{y_v\in S}(X_v = y_v)_{\mathbb{D}}$, that $P(A\,|\,C) = P'(A\,|\,C)$. Because this is true for all $(A,C)_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$, we conclude that $P = P'$. □

3.B Proofs of Results in Section 3.3

Proof of Proposition 3.25$_{110}$. Suppose that the lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]$ is well-defined. Due to Definition 3.12$_{105}$, this means that $\big(f(X_{\tau_{n+1}}),(X_{\tau_n}=x_{\tau_n})_{\mathbb{D}}\big)\in\mathscr{D}_P$ for all $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$.

So fix any $P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, and let E be any coherent conditional prevision on $\mathbb{B}\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ that corresponds to P; this E exists by Theorem 2.6$_{52}$. Because E corresponds to P, and because it trivially holds that $\mathscr{D}_{\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}}\subseteq\mathbb{B}\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$ and $\big(f(X_{\tau_{n+1}}),(X_{\tau_n}=x_{\tau_n})_{\mathbb{D}}\big)\in\mathbb{B}\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$, it follows from Definition 2.5$_{54}$ that

$$\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]=E[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]\,. \tag{3.33}$$

Let $\mathscr{C}:=\mathscr{E}(\Omega_{\mathbb{D}})\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$; then $\mathscr{D}_{\mathscr{C}}\subseteq\mathbb{B}\times\mathscr{E}(\Omega_{\mathbb{D}})_{\supset\emptyset}$. Let $\tilde{E}$ be the restriction of E to $\mathscr{D}_{\mathscr{C}}$. Because E is a coherent conditional prevision, it follows from Definition 2.3$_{52}$ that $\tilde{E}$ is a coherent conditional prevision on $\mathscr{D}_{\mathscr{C}}$. Let P^* be the map on $\mathscr{C}$ that is defined as $P^*(A\,|\,C):=\tilde{E}[\mathbb{I}_A\,|\,C]$ for all $(A,C)_{\mathbb{D}}\in\mathscr{C}$. Then it follows from Proposition 2.7$_{53}$ that P^* is a coherent conditional probability. Moreover, because E corresponds to P, because $\tilde{E}$ is the restriction of E to $\mathscr{D}_{\mathscr{C}}$, and because $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}\subseteq\mathscr{C}$, it follows that for all $(A,C)_{\mathbb{D}}\in\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ it holds that

$$P(A\,|\,C)=E[\mathbb{I}_A\,|\,C]=\tilde{E}[\mathbb{I}_A\,|\,C]=P^*(A\,|\,C)\,, \tag{3.34}$$

and hence P^* extends P.

Using the same line of reasoning as used in the proof of Proposition 2.23$_{73}$, we represent the τ_{n+1}-measurable function $f(X_{\tau_{n+1}})$ as $\sum_{y\in\mathscr{X}}f(y)\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}$. It then follows that

$$\begin{aligned}\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]&=E[f(X_{\tau_{n+1}})\,|\,X_{\tau_n}=x_{\tau_n}]\\&=E\left[\sum_{y\in\mathscr{X}}f(y)\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}\,\middle|\,X_{\tau_n}=x_{\tau_n}\right]\\&=\sum_{y\in\mathscr{X}}f(y)E\big[\mathbb{I}_{(X_{\tau_{n+1}}=y)_{\mathbb{D}}}\,\big|\,X_{\tau_n}=x_{\tau_n}\big]\\&=\sum_{y\in\mathscr{X}}f(y)P^*(X_{\tau_{n+1}}=y\,|\,X_{\tau_n}=x_{\tau_n})\,,\end{aligned} \tag{3.35}$$

where for the first equality we used Equation (3.33); for the second equality we used the linearity of E, i.e. Properties E2$_{52}$ and E3$_{52}$; and for the last equality we used Equation (3.34).

Now let $u := \tau_{0:(n-1)}$ and, for all $x_u \in \mathscr{X}_u$, let ${}^P T_{n,x_u}$ denote the history-dependent transition matrix corresponding to P. Then it follows that

$$
\begin{aligned}
&\mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] \\
&= \sum_{y\in\mathscr{X}} f(y) P^*(X_{\tau_{n+1}} = y\,|\,X_{\tau_n} = x_{\tau_n}) \\
&= \sum_{y\in\mathscr{X}} f(y) \sum_{x_u\in\mathscr{X}_u} P^*(X_{\tau_{n+1}} = y, X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \\
&= \sum_{y\in\mathscr{X}} f(y) \sum_{x_u\in\mathscr{X}_u} P^*(X_{\tau_{n+1}} = y\,|\,X_{\tau_n} = x_{\tau_n}, X_u = x_u) P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \\
&= \sum_{y\in\mathscr{X}} f(y) \sum_{x_u\in\mathscr{X}_u} P(X_{\tau_{n+1}} = y\,|\,X_{\tau_n} = x_{\tau_n}, X_u = x_u) P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \\
&= \sum_{x_u\in\mathscr{X}_u} P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \sum_{y\in\mathscr{X}} f(y) P(X_{\tau_{n+1}} = y\,|\,X_{\tau_n} = x_{\tau_n}, X_u = x_u) \\
&= \sum_{x_u\in\mathscr{X}_u} P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \sum_{y\in\mathscr{X}} f(y)\, {}^P T_{n,x_u}(x_{\tau_n}, y) \\
&= \sum_{x_u\in\mathscr{X}_u} P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n})\, {}^P T_{n,x_u} f(x_{\tau_n}) \\
&\geq \sum_{x_u\in\mathscr{X}_u} P^*(X_u = x_u\,|\,X_{\tau_n} = x_{\tau_n}) \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n}) \\
&= \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})\,,
\end{aligned}
$$

where we used Equation (3.35) for the first equality; Property F3$_{47}$ for the second equality; Property F4$_{47}$ for the third equality; the fact that P^* extends P and that $((X_{\tau_{n+1}} = y)_{\mathbb{D}}, (X_{\tau_n = x_{\tau_n}}, X_u = x_u)_{\mathbb{D}}) \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$ for all $x_u \in \mathscr{X}_u$ for the fourth equality; the definition of ${}^P T_{n,x_u}$ for the sixth equality; the fact that ${}^P T_{n,x_u} \in \mathscr{T}_n$ by Definition 3.11$_{104}$ for the inequality; and Properties F3$_{47}$ and F8$_{47}$ for the final equality.

Because $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ is arbitrary, it follows that

$$
\inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] \geq \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})\,. \tag{3.36}
$$

Next, we will show that $\inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})$ is real-valued. In particular, for any $T \in \mathscr{T}_n$ and any $x \in \mathscr{X}$ it holds that

$$
Tf(x) = \sum_{y\in\mathscr{X}} T(x,y) f(y) \geq \sum_{y\in\mathscr{X}} T(x,y) \min_{z\in\mathscr{X}} f(z) = \min_{z\in\mathscr{X}} f(z) \in \mathbb{R}\,,
$$

where we used the properties of matrix-vector products for the first equality, Property T2$_{91}$ for the inequality, Property T1$_{91}$ for the second equality, and the fact that $f \in \mathscr{L}(\mathscr{X})$ for the inclusion. Similarly, it holds that

$$
Tf(x) = \sum_{y\in\mathscr{X}} T(x,y) f(y) \leq \sum_{y\in\mathscr{X}} T(x,y) \max_{z\in\mathscr{X}} f(z) = \max_{z\in\mathscr{X}} f(z) \in \mathbb{R}\,,
$$

and, because this is true for all $T \in \mathscr{T}_n$, and because $\mathscr{T}_n$ is non-empty, it follows that

$$\min_{z\in\mathscr{X}} f(z) \leq \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n}) \leq \max_{z\in\mathscr{X}} f(z)\,.$$

Now fix any $\varepsilon > 0$. Then because $\inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})$ is real-valued, and since $\mathscr{T}_n$ is non-empty, there is some $T_n \in \mathscr{T}_n$ such that

$$T_n f(x_{\tau_n}) < \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n}) + \varepsilon\,. \tag{3.37}$$

Now take an arbitrary $p \in \mathscr{M}$ and, for all $k \in \mathbb{Z}_{\geq 0}$ such that $k \neq n$, an arbitrary $T_k \in \mathscr{T}_k$. Due to Proposition 3.14₉₅ there is then a unique Markov chain $P_* \in \mathbb{P}^{\mathbb{D},\mathrm{M}} \subseteq \mathbb{P}^{\mathbb{D}}$ that has $(T_k)_{k\in\mathbb{Z}_{\geq 0}}$ as its corresponding family of transition matrices, and that satisfies $P_*(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$. Moreover, it is immediately clear that in fact $P_* \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, because $p \in \mathscr{M}$ and $T_k \in \mathscr{T}_k$ for all $k \in \mathbb{Z}_{\geq 0}$. Due to Proposition 3.18₉₉, this Markov chain satisfies

$$\mathbb{E}_{P_*}[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] = T_n f(x_{\tau_n})\,. \tag{3.38}$$

By combining Equations (3.37), (3.36)↶, and (3.38) (in that order), we find that

$$\begin{aligned}
T_n f(x_{\tau_n}) &< \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n}) + \varepsilon \\
&\leq \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] + \varepsilon \\
&\leq \mathbb{E}_{P_*}[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] + \varepsilon \\
&= T_n f(x_{\tau_n}) + \varepsilon\,,
\end{aligned}$$

where we used that $P_* \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ for the third inequality. Because $\varepsilon > 0$ is arbitrary, this implies that

$$\inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P[f(X_{\tau_{n+1}})\,|\,X_{\tau_n} = x_{\tau_n}] = \inf_{T\in\mathscr{T}_n} Tf(x_{\tau_n})\,,$$

which, using Definition 3.12₁₀₅ and Proposition 3.24₁₀₈, concludes the proof. □

Proof of Lemma 3.27₁₁₁. We start by handling the case $n = m$ separately; we want to show that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\big|\,X_{\tau_{0:(n-1)}} = x_{\tau_{0:(n-1)}}\big]\,.
\end{aligned}$$

For notational convenience, let $u \coloneqq \tau_{0:(n-1)}$. It follows from Lemma 3.21[105] that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\big|\,X_u = x_u\big] \\
&\qquad = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \sum_{y_u\in\mathscr{X}_u} \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u = y_u] P(X_u = y_u\,|\,X_u = x_u) \\
&\qquad = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u = x_u] = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u = x_u]\,,
\end{aligned}$$

where for the second equality we used that, for all $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, it holds that $P(X_u = y_u\,|\,X_u = x_u) = 1$ if $y_u = x_u$ (due to Property F2[47]), and $P(X_u = y_u\,|\,X_u = x_u) = 0$ otherwise (due to Properties F6[47] and F7[47]). Since $u = \tau_{0:(n-1)}$ and $n = m$, it holds that $u = \tau_{0:(m-1)}$, whence this implies that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad\qquad = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\big]\,,
\end{aligned}$$

which concludes the proof for the case where $n = m$.

So, for the remainder of this proof, let us suppose that $n > m$, and let $u \coloneqq \tau_{0:(n-1)}$. Since $n > m$ and $m \in \mathbb{Z}_{\geq 0}$ this implies that $n > 0$ and that $u \neq \emptyset$. We now first show that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad\qquad \geq \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big]\,.
\end{aligned}$$

Following Definition 3.12[105] and Proposition 2.26[77] we find that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad = \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P\Big[\mathbb{E}_P[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] \\
&\qquad \geq \inf_{P\in\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}} \mathbb{E}_P\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] \\
&\qquad = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big]\,,
\end{aligned}$$

where we used Definition 3.12[105] together with Property CE4[78] for the inequality.

For the other direction, fix any $\varepsilon > 0$. It follows from Corollary 3.26[110] and the fact that $\mathscr{T}_{n-1}$ is non-empty that for all $y_u \in \mathscr{X}_u$

there is some $T_{y_u} \in \mathscr{T}_{n-1}$ such that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u = y_u] = \inf_{T\in\mathscr{T}_{n-1}} \big[Tf(y_u,\cdot)\big](y_{\tau_{n-1}}) \\ > \big[T_{y_u}f(y_u,\cdot)\big](y_{\tau_{n-1}}) - \frac{\varepsilon}{2}\,. \tag{3.39}$$

Moreover, using Definition 3.12_{105}, Lemma 3.20_{105}, and Lemma 3.23_{107}, there is some $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ such that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] \\ > \mathbb{E}_P\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] - \frac{\varepsilon}{2}\,. \tag{3.40}$$

Now let $p : \mathscr{X} \times \mathscr{S}_{\mathbb{D}}$ be a probability tree such that, for all $(X_v = y_v)_{\mathbb{D}} \in \mathscr{S}_{\mathbb{D}}$ and all $x \in \mathscr{X}$,

$$p(x\,|\,y_v) := \begin{cases} P(X_{\tau_0} = x) & \text{if } v = \emptyset, \\ P(X_{\tau_{k+1}} = x\,|\,X_v = y_v) & \text{if } v = \tau_{0:k},\ k \in \mathbb{Z}_{\geq 0}, \text{ and } k \neq (n-1), \\ T_{y_v}(y_{\tau_{n-1}}, x) & \text{otherwise (if } v = u)\,. \end{cases}$$

In words, this probability tree p agrees with the process P on all situations *except* those that depend on the time points $u = \tau_{0:(n-1)}$, for which it agrees with the selections from Equation (3.39).

By Theorem 3.5_{88}, there is a unique stochastic process $P_* \in \mathbb{P}^{\mathbb{D}}$ that corresponds to p. Let us prove that $P_* \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$. First, it follows from Equation $(3.2)_{87}$ that $P_*(X_{\tau_0} = x) = p(x\,|\,x_\emptyset) = P(X_{\tau_0} = x)$ for all $x \in \mathscr{X}$ and therefore, since $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, that P_* is consistent with $\mathscr{M}$.

Next, let $({}^{P_*}T_{k,y_v})$ denote the family of history-dependent transition matrices corresponding to P_*. To show that $P_* \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$ it remains to show that, for all $k \in \mathbb{Z}_{\geq 0}$, it holds that ${}^{P_*}T_{k,y_v} \in \mathscr{T}_k$ for all $y_v \in \mathscr{X}_v$, with $v = \tau_{0:(k-1)}$. So fix any $k \in \mathbb{Z}_{\geq 0}$, let $v := \tau_{0:(k-1)}$, and choose any $y_v \in \mathscr{X}_v$. We consider two different cases. If $k = n-1$ then $\tau_{0:k} = \tau_{0:(n-1)} = u$, and then it follows from Definition 3.8_{101}, Equation $(3.3)_{87}$, and the definition of p that, for all $y_{\tau_k}, x \in \mathscr{X}$,

$$\begin{aligned} {}^{P_*}T_{k,y_v}(y_{\tau_k}, x) &= P_*(X_{\tau_{k+1}} = x\,|\,X_{\tau_k} = y_{\tau_k}, X_v = y_v) \\ &= p(x\,|\,y_u) \\ &= T_{y_u}(y_{\tau_k}, x)\,, \end{aligned}$$

with T_{y_u} as in Equation (3.39). Because $T_{y_u} \in \mathscr{T}_k$, and because $\mathscr{T}_k$ has separately specified rows, it follows from Definition 3.13_{111} that there is some $T \in \mathscr{T}_k$ such that, for all $y_{\tau_k}, x \in \mathscr{X}$,

$$T(y_{\tau_k}, x) = T_{y_u}(y_{\tau_k}, x) = {}^{P_*}T_{k,y_v}(y_{\tau_k}, x)\,,$$

which implies that $T = {}^{P_*}T_{k,y_v}$. This means that ${}^{P_*}T_{k,y_v} \in \mathscr{T}_k$.

For the other case, suppose that $k \neq n-1$. Then it follows from Definition 3.8$_{101}$, Equation (3.3)$_{87}$, and the definition of p that, for all $y_{\tau_k}, x \in \mathscr{X}$,

$$\begin{aligned}
{}^{P_*}T_{k,y_v}(y_{\tau_k}, x) &= P_*(X_{\tau_{k+1}} = x \,|\, X_{\tau_k} = y_{\tau_k}, X_v = y_v) \\
&= p(x \,|\, y_{\tau_{0:k}}) \\
&= P(X_{\tau_{k+1}} = x \,|\, X_{\tau_k} = y_{\tau_k}, X_v = y_v) \\
&= {}^{P}T_{k,y_v}(y_{\tau_k}, x),
\end{aligned}$$

where ${}^{P}T_{k,y_v}$ denotes the history-dependent transition matrix corresponding to P. Hence it follows that ${}^{P_*}T_{k,y_v} = {}^{P}T_{k,y_v}$ and therefore, since $P \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$, that ${}^{P_*}T_{k,y_v} \in \mathscr{T}_k$. Because this covers all cases, we conclude that, indeed, $P_* \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}$.

Moving on, let now $v := \tau_{m:(n-1)}$; then $v \neq \emptyset$ since $n > m$. Moreover, for any $g \in \mathscr{L}(\mathscr{X}_u)$, it holds that

$$\begin{aligned}
&\mathbb{E}_{P_*}[g(X_u) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\quad = \mathbb{E}_{P_*}[g(x_{\tau_{0:(m-1)}}, X_v) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\quad = \sum_{x_v \in \mathscr{X}_v} g(x_{\tau_{0:(m-1)}}, x_v) P_*(X_v = x_v \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}) \\
&\quad = \sum_{x_v \in \mathscr{X}_v} g(x_{\tau_{0:(m-1)}}, x_v) \prod_{i=m}^{n-1} P_*(X_{\tau_i} = x_{\tau_i} \,|\, X_{\tau_{0:(i-1)}} = x_{\tau_{0:(i-1)}}) \\
&\quad = \sum_{x_v \in \mathscr{X}_v} g(x_{\tau_{0:(m-1)}}, x_v) \prod_{i=m}^{n-1} p(x_{\tau_i} \,|\, x_{\tau_{0:(i-1)}}) \\
&\quad = \sum_{x_v \in \mathscr{X}_v} g(x_{\tau_{0:(m-1)}}, x_v) \prod_{i=m}^{n-1} P(X_{\tau_i} = x_{\tau_i} \,|\, X_{\tau_{0:(i-1)}} = x_{\tau_{0:(i-1)}}) \\
&\quad = \sum_{x_v \in \mathscr{X}_v} g(x_{\tau_{0:(m-1)}}, x_v) P(X_v = x_v \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}) \\
&\quad = \mathbb{E}_{P}[g(x_{\tau_{0:(m-1)}}, X_v) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\quad = \mathbb{E}_{P}[g(X_u) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}],
\end{aligned} \tag{3.41}$$

where we used Proposition 2.25$_{75}$ and that $v \neq \emptyset$ and $\tau_{0:(m-1)} < v$ for the first equality; Proposition 2.23$_{73}$ for the second equality; Property F4$_{47}$ (repeatedly) for the third equality, using that $v = \tau_{m:(n-1)}$ and hence that $(X_{\tau_v} = x_{\tau_v})_{\mathbb{D}} = \cap_{i=m}^{n-1}(X_{\tau_i} = x_{\tau_i})_{\mathbb{D}}$; the correspondence of P_* with p together with Equation (3.3)$_{87}$ for the fourth equality; the definition of p for the fifth equality; Property F4$_{47}$ (repeatedly) for the sixth equality, again using that $v = \tau_{m:(n-1)}$ and hence that $(X_{\tau_v} = x_{\tau_v})_{\mathbb{D}} = \cap_{i=m}^{n-1}(X_{\tau_i} = x_{\tau_i})_{\mathbb{D}}$;

Proposition 2.23$_{73}$ for the seventh equality; and Proposition 2.25$_{75}$ and that $v \neq \emptyset$ and $\tau_{0:(m-1)} < v$, for the final equality.

Moreover, for all $y_u \in \mathscr{X}_u$ it holds that

$$\begin{aligned}
\mathbb{E}_{P_*}[f(X_{\tau_{0:n}}) \,|\, X_u = y_u] &= \mathbb{E}_{P_*}[f(y_u, X_{\tau_n}) \,|\, X_u = y_u] \\
&= \sum_{x_{\tau_n} \in \mathscr{X}_{\tau_n}} f(y_u, x_{\tau_n}) P_*(X_{\tau_n} = x_{\tau_n} \,|\, X_u = y_u) \\
&= \sum_{x_{\tau_n} \in \mathscr{X}_{\tau_n}} f(y_u, x_{\tau_n}) p(x_{\tau_n} \,|\, y_u) \\
&= \sum_{x_{\tau_n} \in \mathscr{X}_{\tau_n}} f(y_u, x_{\tau_n}) T_{y_u}(y_{\tau_{n-1}}, x_{\tau_n}) \\
&= \big[T_{y_u} f(y_u, \cdot)\big](y_{\tau_{n-1}}) \\
&< \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_u = y_u] + \frac{\varepsilon}{2}\,,
\end{aligned} \tag{3.42}$$

where we used Proposition 2.25$_{75}$ and that $u = \tau_{0:(n-1)}$ and $u < \tau_n$ for the first equality, Proposition 2.23$_{73}$ for the second equality, the correspondence of P_* with p together with Equation (3.3)$_{87}$ for the third equality, the definition of p for the fourth equality, the properties of matrix-vector products together with the notational convention $f(y_u, \cdot) \in \mathscr{L}(\mathscr{X}_{\tau_n})$ for the fifth equality, and Equation (3.39)$_{134}$ for the inequality.

Finally, due to Proposition 2.26$_{77}$ it holds that

$$\mathbb{E}_{P_*}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] = \mathbb{E}_{P_*}\Big[\mathbb{E}_{P_*}[f(X_{\tau_{0:n}}) \,|\, X_u] \,\Big|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big]\,. \tag{3.43}$$

Now, by combining Equations (3.43), (3.42), (3.41)$_{\frown}$, and (3.40)$_{134}$, we find that

$$\begin{aligned}
&\mathbb{E}_{P_*}[f(X_{\tau_{0:n}}) \,|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}] \\
&\qquad = \mathbb{E}_{P_*}\Big[\mathbb{E}_{P_*}[f(X_{\tau_{0:n}}) \,|\, X_u] \,\Big|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] \\
&\qquad \leq \mathbb{E}_{P_*}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_u] + \frac{\varepsilon}{2} \,\Big|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] \\
&\qquad = \mathbb{E}_{P_*}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_u] \,\Big|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] + \frac{\varepsilon}{2} \\
&\qquad < \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}}) \,|\, X_u] \,\Big|\, X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\Big] + \varepsilon\,,
\end{aligned}$$

where we used Property CE4$_{78}$ for the first inequality and Property CE6$_{79}$ for the second equality.

Due to Definition 3.12$_{105}$, this implies that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}] \\
&\qquad \leq \mathbb{E}_{P_*}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}] \\
&\qquad < \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}\Big]+\varepsilon\,,
\end{aligned}$$

and, because $\varepsilon > 0$ is arbitrary, we conclude that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}] \\
&\qquad \leq \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_u]\,\Big|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}\Big]\,.
\end{aligned}$$

Because we already proved the inequality in the other direction, and using that $u=\tau_{0:(n-1)}$, we therefore find that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}] \\
&\qquad = \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(n-1)}}]\,\Big|\,X_{\tau_{0:(m-1)}}=x_{\tau_{0:(m-1)}}\Big]\,,
\end{aligned}$$

which concludes the proof. □

3.C Proofs of Results in Section 3.4

The following lemma uses the interpretation of the rows of matrices as elements of the dual space $\mathscr{L}(\mathscr{X})^\top$ of $\mathscr{L}(\mathscr{X})$; see Appendices A.2$_{380}$ and A.3$_{383}$ for details.

Lemma 3.46. *Let $\mathscr{T}$ be a set of transition matrices and, for all $x \in \mathscr{X}$, let $\mathscr{T}_x := \{T(x,\cdot)\,|\,T \in \mathscr{T}\}$ denote the set of x-rows of elements of $\mathscr{T}$, which we interpret as elements of $\mathscr{L}(\mathscr{X})^\top$. Then, for all $x \in \mathscr{X}$, $\mathscr{T}_x$ is convex if $\mathscr{T}$ is convex, and $\mathscr{T}_x$ is closed if $\mathscr{T}$ is closed. Moreover, if $\mathscr{T}$ has separately specified rows, then $\mathscr{T}$ is closed if $\mathscr{T}_x$ is closed for all $x \in \mathscr{X}$, and $\mathscr{T}$ is convex if $\mathscr{T}_x$ is convex for all $x \in \mathscr{X}$.*

Proof. First assume that $\mathscr{T}$ is convex, and fix any $x \in \mathscr{X}$, any $T(x,\cdot), S(x,\cdot) \in \mathscr{T}_x$, and any $\lambda \in [0,1]$. Then there are $T, S \in \mathscr{T}$ such that $T(x,\cdot)$ is the x-row of T, and $S(x,\cdot)$ is the x-row of S. Because $\mathscr{T}$ is convex, it holds that $\lambda T + (1-\lambda)S \in \mathscr{T}$, whence also $\lambda T(x,\cdot) + (1-\lambda)S(x,\cdot) \in \mathscr{T}_x$, which implies that $\mathscr{T}_x$ is convex.

Next assume that $\mathscr{T}$ is closed, fix any $x \in \mathscr{X}$, and consider any convergent sequence $\{T_i(x,\cdot)\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{T}_x$. To show that $\mathscr{T}_x$ is closed, by Proposition A.8$_{376}$, we need to show that the limit of this sequence is an element of $\mathscr{T}_x$. For any $i \in \mathbb{Z}_{>0}$, because $T_i(x,\cdot) \in \mathscr{T}_x$, there is some $T_i \in \mathscr{T}$ such that $T_i(x,\cdot)$ is the x-row of T_i. It follows from Lemma 3.9$_{92}$

that $\|\mathscr{T}\| = 1$, and hence that $\mathscr{T}$ is bounded. Because $\mathscr{T}$ is also closed (by assumption), it follows from Corollary A.12$_{378}$ that $\mathscr{T}$ is sequentially compact. This implies the existence of a convergent subsequence $\{T_{i_j}\}_{j\in\mathbb{Z}_{>0}}$ with limit $T_* := \lim_{j\to+\infty} T_{i_j}$ that satisfies $T_* \in \mathscr{T}$, which implies that also $T_*(x,\cdot) \in \mathscr{T}_x$. Because the sequence $\{T_i(x,\cdot)\}_{i\in\mathbb{Z}_{>0}}$ was convergent, the subsequence $\{T_{i_j}(x,\cdot)\}$ has the same limit, say $T'_*(x,\cdot) := \lim_{j\to+\infty} T_{i_j}(x,\cdot)$. Because we already know that $T_*(x,\cdot) \in \mathscr{T}_x$, it now suffices to show that $T_*(x,\cdot) = T'_*(x,\cdot)$. To this end, fix any $\varepsilon > 0$. Then there is some $n \in \mathbb{Z}_{>0}$ such that for all $j > n$ it holds that

$$\left\|T'_*(x,\cdot) - T_{i_j}(x,\cdot)\right\|_* < \frac{\varepsilon}{2} \quad \text{and} \quad \left\|T_{i_j} - T_*\right\| < \frac{\varepsilon}{2},$$

which implies that

$$\begin{aligned}\left\|T'_*(x,\cdot) - T_*(x,\cdot)\right\|_* &\leq \left\|T'_*(x,\cdot) - T_{i_j}(x,\cdot)\right\|_* + \left\|T_{i_j}(x,\cdot) - T_*(x,\cdot)\right\|_* \\ &< \frac{\varepsilon}{2} + \left\|T_{i_j} - T_*\right\| < \varepsilon,\end{aligned}$$

where we used Proposition A.33$_{390}$ for the second inequality. Because $\varepsilon > 0$ is arbitrary this implies that $\|T'_*(x,\cdot) - T_*(x,\cdot)\|_* = 0$, or equivalently, that $T'_*(x,\cdot) = T_*(x,\cdot)$. This concludes the proof that $\mathscr{T}_x$ is closed whenever $\mathscr{T}$ is closed.

We now prove the implications in the other direction. Let us start by assuming that $\mathscr{T}_x$ is closed for all $x \in \mathscr{X}$. Fix any convergent sequence $\{T_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{T}$ with $T = \lim_{i\to+\infty} T_i$; we need to show that $T \in \mathscr{T}$. Now fix any $x \in \mathscr{X}$. Then, for all $i \in \mathbb{Z}_{>0}$, it holds that $T_i(x,\cdot) \in \mathscr{T}_x$, and

$$\|T_i(x,\cdot) - T(x,\cdot)\|_* \leq \|T_i - T\|,$$

due to Proposition A.33$_{390}$, whence $\lim_{i\to+\infty} T_i(x,\cdot) = T(x,\cdot)$. Because $\mathscr{T}_x$ is closed, this implies that $T(x,\cdot) \in \mathscr{T}_x$. Because $\mathscr{T}$ has separately specified rows, this means that $T \in \mathscr{T}$, which concludes the proof that $\mathscr{T}$ is closed.

To establish the convexity, assume that $\mathscr{T}_x$ is convex for all $x \in \mathscr{X}$. Fix any $T, S \in \mathscr{T}$ and any $\lambda \in [0,1]$. We want to show that $T_\lambda := \lambda T + (1-\lambda)S$ is an element of $\mathscr{T}$.

Now, for any $x \in \mathscr{X}$, it holds that $T(x,\cdot), S(x,\cdot) \in \mathscr{T}_x$. Moreover, because $\mathscr{T}_x$ is convex, it holds that $T_\lambda(x,\cdot) = \lambda T(x,\cdot) + (1-\lambda)S(x,\cdot) \in \mathscr{T}_x$. Thus $T_\lambda \in \mathscr{T}$ because $\mathscr{T}$ has separately specified rows. □

Proof of Proposition 3.37$_{120}$. For all $x \in \mathscr{X}$, we define the map $\underline{T}_x : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ such that, for all $f \in \mathscr{L}(\mathscr{X})$, $\underline{T}_x f := \underline{T} f(x)$. Then, because $\underline{T}$ is a lower transition operator, it follows from Definition 3.15$_{116}$ that, for all $x \in \mathscr{X}$, all $f, g \in \mathscr{L}(\mathscr{X})$, and all $\lambda \in \mathbb{R}_{\geq 0}$, it holds that

CLP1: $\underline{T}_x f \geq \min_{y\in\mathscr{X}} f(y)$; (lower bounds)

CLP2: $\underline{T}_x(f+g) \geq \underline{T}_x f + \underline{T}_x g$; (super-additivity)

CLP3: $\underline{T}_x(\lambda f) = \lambda \underline{T}_x f$. (non-negative homogeneity)

Therefore, it follows from [114, Definition 2.3.3] that, for all $x \in \mathscr{X}$, the map $\underline{T}_x$ is a *coherent lower prevision*[8] on $\mathscr{L}(\mathscr{X})$.

Next, a *coherent linear prevision* on $\mathscr{L}(\mathscr{X})$ is a *linear* functional $p^\top : \mathscr{L}(\mathscr{X}) \to \mathbb{R} : f \mapsto p^\top f$—i.e. an element of the dual space $\mathscr{L}(\mathscr{X})^\top$ of $\mathscr{L}(\mathscr{X})$ that we introduce in Appendix A.2$_{380}$—that satisfies $p^\top f \geq \min_{x\in\mathscr{X}} f(x)$ for all $f \in \mathscr{L}(\mathscr{X})$ [114, Theorem 2.8.4]. Because these are linear maps, they are trivially super-additive (because they are additive) and non-negatively homogeneous (because they are homogeneous). Hence, any coherent linear prevision is also a coherent lower prevision, by the above definition.

Moreover, for any $x \in \mathscr{X}$, let $\mathbb{I}_x \in \mathscr{L}(\mathscr{X})$ denote the indicator of x, defined for all $y \in \mathscr{X}$ such that $\mathbb{I}_x(y) := 1$ if $x = y$, and $\mathbb{I}_x(y) := 0$, otherwise. Fix any coherent linear prevision $p^\top$, and let $p \in \mathscr{L}(\mathscr{X})$ be defined such that $p(x) := p^\top \mathbb{I}_x$ for all $x \in \mathscr{X}$. Then for any $x \in \mathscr{X}$, it holds that $p(x) = p^\top \mathbb{I}_x \geq \min_{y\in\mathscr{X}} \mathbb{I}_x(y) = 0$. Moreover, it holds that

$$\sum_{x\in\mathscr{X}} p(x) = \sum_{x\in\mathscr{X}} p^\top \mathbb{I}_x = p^\top \sum_{x\in\mathscr{X}} \mathbb{I}_x = p^\top 1\,,$$

using the linear character of $p^\top$ for the second equality, and where 1 denotes the constant function in $\mathscr{L}(\mathscr{X})$ whose value in every $x \in \mathscr{X}$ is one. Because $p^\top$ is a coherent linear prevision, it holds that $p^\top 1 \geq \min_{x\in\mathscr{X}} 1 = 1$, and because it is linear, that

$$-p^\top 1 = p^\top(-1) \geq \min_{x\in\mathscr{X}} -1 = -1\,,$$

and hence $p^\top 1 \leq 1$. It follows that $\sum_{x\in\mathscr{X}} p(x) = 1$. Because we have already seen that $p(x) \geq 0$ for all $x \in \mathscr{X}$, we conclude that p is a probability mass function on $\mathscr{L}(\mathscr{X})$.

Now, for any $x \in \mathscr{X}$, by [114, Theorem 3.6.1], there is a unique set $\mathscr{T}_x \subseteq \mathscr{L}(\mathscr{X})^\top$ of coherent linear previsions on $\mathscr{L}(\mathscr{X})$ that is non-empty and convex (see Proposition A.18$_{380}$ and Definition A.12$_{376}$), that *dominates* $\underline{T}_x$ in the sense that $\underline{T}_x f \leq p^\top f$ for all $p^\top \in \mathscr{T}_x$ and all $f \in \mathscr{L}(\mathscr{X})$,

[8] As the terminology indicates, these objects are strongly related to the coherent previsions that we discussed in Chapter 2$_{45}$. Essentially, where we used coherent previsions to come up with a notion of expectations, coherent *lower* previsions can be used to derive a notion of *lower* expectations. In particular, this can be done axiomatically, e.g. by imposing properties CLP1–CLP3, without the explicit reference to sets of probabilities used in this work. We refer to [109, 114] for further information.

that satisfies $\underline{T}_x f = \inf_{p^\top \in \mathscr{T}_x} p^\top f$ for all $f \in \mathscr{L}(\mathscr{X})$, and that is compact in the weak$\star$ topology (on $\mathscr{L}(\mathscr{X})^\top$). By Corollary A.23$_{383}$, this $\mathscr{T}_x$ is then also compact in the metric topology induced by the dual norm $\|\cdot\|_*$ on $\mathscr{L}(\mathscr{X})^\top$. Hence $\mathscr{T}_x$ is closed and bounded by Corollary A.12$_{378}$.

Following the discussion in Appendix A.2$_{380}$, for any matrix $T \in \mathbb{M}$ and any $x \in \mathscr{X}$, we can interpret the x-row $T(x,\cdot)$ of T as an element of $\mathscr{L}(\mathscr{X})^\top$. Moreover, it then holds that $T(x,y) = T(x,\cdot)\mathbb{I}_y$, so it follows from the above discussion that if $T(x,\cdot)$ is a coherent linear prevision, that then $T(x,y)$, as a function of y, is a probability mass function on $\mathscr{L}(\mathscr{X})$. In other words, if we consider the set

$$\mathscr{T} := \left\{ T \in \mathbb{M} \,\middle|\, \forall x \in \mathscr{X} : T(x,\cdot) \in \mathscr{T}_x \right\},$$

then it follows from Definition 3.5$_{91}$ that $\mathscr{T}$ is a set of transition matrices. Because each $\mathscr{T}_x$ is non-empty, also $\mathscr{T}$ is non-empty. Moreover, it is clear that, by Definition 3.13$_{111}$, $\mathscr{T}$ has separately specified rows, since each $\mathscr{T}_x$ represents the set of x-rows of the elements of $\mathscr{T}$. It follows from Lemma 3.46$_{137}$ that $\mathscr{T}$ is closed and convex because $\mathscr{T}$ has separately specified rows and $\mathscr{T}_x$ is closed and convex for all $x \in \mathscr{X}$.

Moreover, for any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$, it holds that

$$\inf_{T \in \mathscr{T}} Tf(x) = \inf_{T(x,\cdot) \in \mathscr{T}_x} T(x,\cdot) f = \underline{T}_x f = \underline{T} f(x)\,.$$

Because this is true for all $x \in \mathscr{X}$ and all $f \in \mathscr{L}(\mathscr{X})$, it follows that $\underline{T}$ is the lower transition operator corresponding to $\mathscr{T}$.

In summary, we have shown the existence of a set $\mathscr{T}$ that satisfies all of the properties that we claimed in the statement of this proposition. It remains to show that $\mathscr{T} = \mathscr{T}_{\underline{T}}$. To this end, note that because $\underline{T}$ is the lower transition operator corresponding to $\mathscr{T}$, it holds for all $f \in \mathscr{L}(\mathscr{X})$ and all $T \in \mathscr{T}$ that $Tf \geq \underline{T} f$, and hence $\mathscr{T} \subseteq \mathscr{T}_{\underline{T}}$.

To prove the inclusion in the other direction, fix any $T \in \mathscr{T}_{\underline{T}}$. Then it holds that $Tf \geq \underline{T} f$ for all $f \in \mathscr{L}(\mathscr{X})$ and hence in particular, for any $x \in \mathscr{X}$, this means that $Tf(x) \geq \underline{T} f(x) = \underline{T}_x f$ for all $f \in \mathscr{L}(\mathscr{X})$. In other words, for all $x \in \mathscr{X}$, the x-row $T(x,\cdot)$ of T, which as we know, is a coherent linear prevision on $\mathscr{L}(\mathscr{X})$, dominates $\underline{T}_x$. This means that $T(x,\cdot) \in \mathscr{T}_x$ by [114, Theorem 3.6.1]. Because $\mathscr{T}$ has separately specified rows, this means that $T \in \mathscr{T}$, and hence $\mathscr{T}_{\underline{T}} \subseteq \mathscr{T}$. □

Proof of Corollary 3.38$_{120}$. For all $x \in \mathscr{X}$, let $\underline{T}_x : \mathscr{L} \to \mathbb{R}$ be as in the proof of Proposition 3.37$_{120}$ and, following the proof of Proposition 3.37$_{120}$, due to [114, Theorem 3.6.1] there is a *unique* set $\mathscr{S}_x$ of coherent linear previsions on $\mathscr{L}(\mathscr{X})$ that is non-empty, convex, closed, that dominates $\underline{T}_x$, and that satisfies $\underline{T}_x f = \inf_{p^\top \in \mathscr{S}_x} p^\top f$ for all

$f \in \mathscr{L}(\mathscr{X})$. Moreover, we know from the proof of Proposition 3.37$_{120}$ that $\mathscr{S}_x = \{T(x,\cdot) : T \in \mathscr{T}_{\underline{T}}\}$ is the set of x-rows of the elements of $\mathscr{T}_{\underline{T}}$.

Now let $\mathscr{T}_x := \{T(x,\cdot) : T \in \mathscr{T}\}$ be the set of x-rows of the elements of $\mathscr{T}$. Then $\mathscr{T}_x$ is non-empty because $\mathscr{T}$ is non-empty, closed and convex due to Lemma 3.46$_{137}$, and satisfies $\inf_{p^\top \in \mathscr{T}_x} p^\top f = \underline{T}f(x) = \underline{T}_x f$ because $\mathscr{T}$ has $\underline{T}$ as its corresponding lower transition operator, which also implies that $\mathscr{T}_x$ dominates $\underline{T}_x$. Moreover, because $\mathscr{T}$ is a set of transition matrices, and following the proof of Proposition 3.37$_{120}$, each $p^\top \in \mathscr{T}_x$ is a coherent linear prevision on $\mathscr{L}(\mathscr{X})$. Therefore, it follows that $\mathscr{T}_x = \mathscr{S}_x$ because, as we already know, $\mathscr{S}_x$ is the unique set of coherent linear previsions that satisfies these properties.

Because this is true for all $x \in \mathscr{X}$, and because $\mathscr{T}$ and $\mathscr{T}_{\underline{T}}$ both have separately specified rows, it therefore follows that $\mathscr{T} = \mathscr{T}_{\underline{T}}$. □

4

Dynamics of Continuous-Time Stochastic Processes

"It was called dub,
a sensuous mosaic cooked from vast libraries of digitalized pop;
it was worship, Molly said, and a sense of community."

William Gibson, "Neuromancer"

In this chapter we introduce the technical machinery that we require to describe the behaviour of continuous-time stochastic processes. We discussed in Chapter 3_{83} that the behaviour of *discrete*-time stochastic processes—and in particular Markov chains—can be described using transition matrices. These transition matrices effectively describe the transition probabilities of the system for a single step in time. As such, and because of the discrete nature of the time domain considered there, it suffices to work with a single transition matrix (in the case of homogeneous Markov chains), or a countable family of transition matrices (in the case of non-homogeneous Markov chains). In contrast, in a continuous-time setting there is no notion of "single step in time", because one could always consider a smaller step. Hence, we require some additional machinery to describe these processes.

We start in Section 4.1↷ by introducing a particular type of continuous-time stochastic processes, that we call *well-behaved*. Essentially, these are processes whose behaviour is not too pathological, in a specific sense; the majority of this dissertation will focus on these types of processes. In Section 4.2_{148} we introduce the notion of corresponding (history-dependent) transition matrices for continuous-time processes, in analogy to our developments in Chapter 3_{83}.

In Section 4.3_{150} we then introduce *transition rate matrices* which, as we shall see, form the core of the methods to parameterise and perform inference with continuous-time processes. We discuss their relation with transition matrices, and introduce the notion of the corresponding *matrix exponential*. In Section 4.4_{156} we introduce and study *transition matrix systems*: these are families of transition matrices satisfying some specific properties that will allow us to use them as a parameterisation of continuous-time Markov chains in Chapter 5_{181}. We also study restrictions of these transition matrix systems, in Section 4.5_{158}, and investigate ways to, essentially, combine two or more of them into a new transition matrix system.

We conclude with Section 4.6_{166}, where we introduce the *outer partial derivatives* of the transition matrices of continuous-time stochastic processes. These are set-valued generalisations of (as their name suggests) the partial derivatives of transition matrices. We prove that, for well-behaved stochastic processes, these always are non-empty and compact sets of transition rate matrices. These outer partial derivatives are crucial to our definition of continuous-time *imprecise*-Markov chains, in Chapter 5_{181}.

As an aside before we start, recall from Chapter 2_{45} that we take the continuous-time setting as the implicit default in our notation. As such, in contrast with the discussion in Chapter 3_{83}, we will (usually) no longer be mentioning the time domain $\mathbb{H} = \mathbb{R}_{\geq 0}$ explicitly in our discussion and notation. Moreover, we have deferred the technical details of some running examples to Appendices $4.\mathrm{A}_{173}$ and $4.\mathrm{B}_{177}$. A few of our proofs rely on technical norm inequalities that can be found in Appendix B_{391}.

4.1 Well-Behaved Stochastic Processes

Let us begin by introducing the notion of *well-behaved* stochastic processes. We have already mentioned in the introduction to this chapter that these are processes that are not "too" pathological. To make this explicit, these are essentially processes for which the conditional probabilities of the elementary events $(X_t = x)$ do not change instantaneously as a function of $t \in \mathbb{R}_{\geq 0}$. To see the need for this condition, and to illustrate the kind of extreme behaviour we would otherwise have to deal with, consider that Example 2.3_{70} tells us in particular that for any two states $x, y \in \mathscr{X}$, there is a stochastic process $P \in \mathbb{P}$ such that

$P(X_t = y|X_0 = x) = \mathbb{I}_{\mathbb{Q}_{>0}}(t)$, where $\mathbb{I}_{\mathbb{Q}_{>0}}$ is the indicator of the positive rational numbers. Such processes are somewhat difficult to work with analytically, and are arguably too unrealistic for applications. Therefore, we simply exclude them from our analysis. What we require, then, is that the rate of change of a stochastic process should remain bounded. We formalise this requirement through the notion of *well-behavedness*.

Definition 4.1 (Well-Behaved Stochastic Process). *A stochastic process $P \in \mathbb{P}$ is said to be* well-behaved *if, for any—possibly empty—time sequence $u \in \mathscr{U}$, any $x_u \in \mathscr{X}_u$, any $x, y \in \mathscr{X}$ and any $t \in \mathbb{R}_{\geq 0}$ such that $t > u$:*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left| P(X_{t+\Delta} = y | X_t = x, X_u = x_u) - \mathbb{I}_x(y) \right| < +\infty \tag{4.1}$$

and, if $t \neq 0$,

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left| P(X_t = y | X_{t-\Delta} = x, X_u = x_u) - \mathbb{I}_x(y) \right| < +\infty. \tag{4.2}$$

The set of all well-behaved stochastic processes is denoted by $\mathbb{P}^{\mathrm{W}}$.

It should hopefully be clear that the specific pathological behaviour mentioned above is prevented by imposing this condition of well-behavedness, since clearly $\limsup_{\Delta \to 0^+} {}^1\!/\!_\Delta \left| \mathbb{I}_{\mathbb{Q}_{>0}}(\Delta) \right| = +\infty$; this implies that a process that satisfies $P(X_t = y\,|\,X_0 = x) = \mathbb{I}_{\mathbb{Q}_{>0}}(t)$ for all $t \in \mathbb{R}_{>0}$, with $x \neq y$, fails to satisfy Equation (4.1).

Moreover, to illustrate the interpretation of Equations (4.1) and (4.2) as bounds on the rate of change of the process, observe that it follows from Properties F2$_{47}$, F6$_{47}$, and F7$_{47}$ that

$$P(X_t = y | X_t = x, X_u = x_u) = \mathbb{I}_x(y) \quad \text{for all } x, y \in \mathscr{X},$$

and therefore in particular, that these inequalities bound the rate of change for Δ around zero. This interpretation might also be understood as imposing a kind of "local" Lipschitz continuity, in the sense that there is a real number—a Lipschitz constant, if you want—that bounds the rate of change. However, this bound is only local in the sense that it can depend on both the value of t, and on the time-points u and state assignment x_u. Hence, the rate of change need not be *uniformly* bounded, whence the property is not truly Lipschitzian.

In light of the above, we note that the definition of well-behavedness is related to continuity and differentiability, but stronger than the former and weaker than the latter. Let us first establish that the relevant probabilities described by any well-behaved process are indeed continuous in this sense.

Proposition 4.1. *Let $P \in \mathbb{P}^{\mathrm{W}}$ be a well-behaved stochastic process. Then for all $t \in \mathbb{R}_{\geq 0}$, all $u \in \mathscr{U}_{<t}$, all $x, y \in \mathscr{X}$, and all $x_u \in \mathscr{X}_u$, the map $p_t : \mathbb{R} \to \mathbb{R}$, defined as*

$$p_t(\Delta) := \begin{cases} P(X_{t+\Delta} = y \,|\, X_t = x, X_u = x_u) & \text{if } \Delta \geq 0 \\ P(X_t = y \,|\, X_{t-|\Delta|} = x, X_u = x_u) & \text{if } \Delta < 0 \text{ and } t - |\Delta| > u \\ 0 & \text{otherwise,} \end{cases}$$

is continuous in 0 *if $t > 0$; if $t = 0$ then it is right-continuous in* 0.

Proof. Fix any $\varepsilon > 0$. To show the continuity (or right-continuity, if $t = 0$) we need to establish that there are $\delta_-, \delta_+ > 0$ such that for all $\Delta \in (-\delta_-, \delta_+)$ (with $\Delta \geq 0$ if $t = 0$), it holds that $|p_t(\Delta) - p_t(0)| < \varepsilon$.

Because P is well-behaved, and by Equation (4.1)$_{\curvearrowleft}$, there is some $\delta_+ \in \mathbb{R}_{>0}$ and some $B_+ \in \mathbb{R}_{>0}$ such that for all $\Delta \in \mathbb{R}_{>0}$ with $\Delta < \delta_+$, it holds that

$$\frac{1}{\Delta} \left| P(X_{t+\Delta} = y | X_t = x, X_u = x_u) - \mathbb{I}_x(y) \right| < B_+ \,,$$

which implies that then also

$$\left| P(X_{t+\Delta} = y | X_t = x, X_u = x_u) - \mathbb{I}_x(y) \right| < \Delta B_+ \,. \tag{4.3}$$

Now let $\delta_+^* \in \mathbb{R}_{>0}$ be such that $\delta_+^* \leq \delta_+$ and such that $\delta_+^* B_+ \leq \varepsilon$; this is clearly always possible since $B_+ \in \mathbb{R}_{>0}$.

Moreover, if $t > 0$ then it follows from Equation (4.2)$_{\curvearrowleft}$ that there is some $\delta_- \in \mathbb{R}_{>0}$ and some $B_- \in \mathbb{R}_{>0}$ such that for all $\Delta \in \mathbb{R}_{>0}$ with $\Delta < \delta_-$, it holds that

$$\frac{1}{\Delta} \left| P(X_t = y | X_{t-\Delta} = x, X_u = x_u) - \mathbb{I}_x(y) \right| < B_- \,,$$

which implies that then also

$$\left| P(X_t = y | X_{t-\Delta} = x, X_u = x_u) - \mathbb{I}_x(y) \right| < \Delta B_- \,. \tag{4.4}$$

Now let $\delta_-^* \in \mathbb{R}_{>0}$ be such that $\delta_-^* \leq \delta_-$, such that $t - \delta_-^* > u$, and such that $\delta_-^* B_- \leq \varepsilon$; this is clearly always possible since $t > u$ and $B_- \in \mathbb{R}_{>0}$. Conversely, if $t = 0$ then let $\delta_-^* \in \mathbb{R}_{>0}$ be arbitrary.

We note, as already mentioned above, that it follows from Properties F2$_{47}$, F6$_{47}$, and F7$_{47}$ that

$$p_t(0) = P(X_t = y | X_t = x, X_u = x_u) = \mathbb{I}_x(y) \,. \tag{4.5}$$

Now fix any $\Delta \in (-\delta_-^*, \delta_+^*)$. Then if $\Delta \geq 0$ it follows from the definition of p_t and Equation (4.5) that

$$|p_t(\Delta) - p_t(0)| = \left| P(X_{t+\Delta} = y \,|\, X_t = x, X_u = x_u) - \mathbb{I}_x(y) \right| < \Delta B_+ < \delta_+^* B_+ \leq \varepsilon \,,$$

where we used Equation (4.3) and the fact that $\Delta < \delta_+^* \leq \delta_+$ for the first inequality. If $t = 0$ then we only have to consider $\Delta \geq 0$, so in that case we are done, having established the right-continuity of p_t in 0.

On the other hand, if $t > 0$ we also have to consider the case $\Delta < 0$; suppose that this holds. Then $|\Delta| < \delta_-^*$, which implies that $t - |\Delta| > u$ since $t - \delta_-^* > u$. Therefore, it follows from the definition of p_t and Equation (4.5) that

$$\begin{aligned} |p_t(\Delta) - p_t(0)| &= \left|P(X_t = y \,|\, X_{t-|\Delta|} = x, X_u = x_u) - \mathbb{I}_x(y)\right| \\ &< |\Delta| B_- < \delta_-^* B_- \leq \varepsilon, \end{aligned}$$

where we used Equation (4.3) and the fact that $|\Delta| < \delta_-^* \leq \delta_-$ for the first inequality. Hence if $t > 0$ we have shown that $|p_t(\Delta) - p_t(0)| < \varepsilon$ for all $\Delta \in (-\delta_-^*, \delta_+^*)$, which establishes the continuity of p_t in 0. ☐

The next example provides further intuition, and suggests that the converse need not be true: there are processes where these probabilities are continuous at certain time points, but which do not satisfy the well-behavedness condition there; and, as claimed above, there are processes that satisfy the well-behavedness condition at certain time points, but for which these probabilities are not differentiable there.

Example 4.1. Let $\mathscr{X}$ be a state space that contains at least two states, fix two states $x, y \in \mathscr{X}$ such that $x \neq y$, and consider any function $p : \mathbb{R}_{>0} \to [0,1]$. Then as we know from Example 2.3$_{70}$, there is a stochastic process P such that

$$P(X_\Delta = y | X_0 = x) = p(\Delta) \text{ for all } \Delta \in \mathbb{R}_{>0}.$$

Furthermore, since $x \neq y$, it follows from F6$_{47}$ and F7$_{47}$ that $P(X_0 = y | X_0 = x) = 0$. We now consider two choices for p.

If we let $p(\Delta) := \sqrt{\Delta}$ for $\Delta \in (0,1]$ and $p(\Delta) := 1$ for $\Delta \geq 1$, then $P(X_\Delta = y | X_0 = x)$ is continuous on $\mathbb{R}_{\geq 0}$ because $P(X_0 = y | X_0 = x) = 0$. However, we also find that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_\Delta = y | X_0 = x) - \mathbb{I}_x(y)| = \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \sqrt{\Delta} = +\infty$$

and therefore, it follows from Equation (4.1)$_{145}$—with $t = 0$ and $u = \emptyset$—that P is not well-behaved.

On the other hand, if we let $p(\Delta) := \Delta |\sin(1/\Delta)|$ for $\Delta \in \mathbb{R}_{>0}$, we find that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_\Delta = y | X_0 = x) - \mathbb{I}_x(y)| = \limsup_{\Delta \to 0^+} |\sin(1/\Delta)| = 1.$$

In this case—at least for $t = 0$ and $u = \emptyset$—P does exhibit the behaviour of a well-behaved stochastic process. Furthermore, $P(X_\Delta = y|X_0 = x)$ is again continuous on $\mathbb{R}_{\geq 0}$.[1] However, $P(X_\Delta = y|X_0 = x)$ is not differentiable in $\Delta = 0$, because ${}^1\!/\!_\Delta P(X_\Delta = y|X_0 = x) = |\sin({}^1\!/\!_\Delta)|$ oscillates ever more wildly as Δ approaches zero, so has no limit there. ◊

4.2 Corresponding Transition Matrices

Let us now introduce the notion of transition matrices corresponding to continuous-time stochastic processes. This is largely analogous to the concept of transition matrices corresponding to discrete-time stochastic processes, as discussed in Chapter 3_{83}, but with some important distinctions. We start by considering corresponding *history-dependent* transition matrices. For a given stochastic process, these matrices form a multi-index family (T_{t,x_u}^s) of transition matrices, with indices t, s, and x_u, defined as follows.

Definition 4.2 (Corresponding History-Dependent Transition Matrix)**.** *Let $P \in \mathbb{P}$ be a stochastic process. Then the* family of history-dependent transition matrices corresponding to P *is a multi-index family (T_{t,x_u}^s) of matrices T_{t,x_u}^s with $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$, that are defined, for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, all $u \in \mathscr{U}_{<t}$, and all $x_u \in \mathscr{X}_u$, as*

$$T_{t,x_u}^s(x_t, x_s) := P(X_s = x_s|X_t = x_t, X_u = x_u) \quad \textit{for all } x_t, x_s \in \mathscr{X} .$$

For notational convenience, if u is empty we write $T_t^s = T_{t,x_\emptyset}^s$.

Let us compare this to the family (T_{n,x_v}) of history-dependent transition matrices corresponding to a discrete-time stochastic process $P \in \mathbb{P}^{\mathbb{D}}$, with τ the canonical time index of $\mathbb{D}$, as in Definition 3.8_{101}. Two straightforward differences are that $n \in \mathbb{Z}_{\geq 0}$ is an index in a countably infinite index set, whereas $t \in \mathbb{R}_{\geq 0}$ has an uncountably infinite domain; and $u \in \mathscr{U}_{<t}$ contains any finite set of time points that precedes t, whereas $v = \tau_0, \ldots, \tau_{n-1}$ contains all time points up to time n.

The more striking difference is the additional index s, which has no counterpart in the discrete-time setting. We need this due to the continuous nature of the time domain considered here; whereas T_{n,x_u} contains the probabilities for the system to move from the state X_{τ_n} at time τ_n to the state $X_{\tau_{n+1}}$ at the *next* time point τ_{n+1}, there is no proper

[1] Note that $p(\Delta) = \Delta|\sin({}^1\!/\!_\Delta)|$ is *not* continuous (or even defined) at $\Delta = 0$, but this function is only used to define $P(X_\Delta = y|X_0 = x)$ for $\Delta > 0$, and hence this is not a problem.

notion of "next" in a continuous-time setting. Hence, we need to explicitly account for the time point s for which the matrix T_{t,x_u}^s contains these probabilities.

The following proposition establishes some simple properties of these corresponding (history-dependent) transition matrices.

Proposition 4.2. *Let $P \in \mathbb{P}$ be a stochastic process with corresponding family of history-dependent transition matrices (T_{t,x_u}^s). Then, for any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the corresponding (history-dependent) transition matrix T_{t,x_u}^s is—as its name suggests—a transition matrix, and $T_{t,x_u}^t = I$. Furthermore, P is well-behaved if and only if, for every—possibly empty—time sequence $u \in \mathscr{U}$, any $x_u \in \mathscr{X}_u$ and any $t \in \mathbb{R}_{\geq 0}$ such that $t > u$:*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t,x_u}^{t+\Delta} - I \right\| < +\infty \tag{4.6}$$

and, if $t \neq 0$,

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta,x_u}^{t} - I \right\| < +\infty. \tag{4.7}$$

Proof. The first part of the statement follows from Corollary 2.20_{68} and Definitions 3.5_{91} and 2.1_{47}. The second part is a consequence of Definition 4.1_{145} and Equation $(2.7)_{63}$. □

Recall also from Chapter 3_{83} that the transition matrices corresponding to discrete-time processes, could be used to represent conditional expectations for these processes. It is perhaps not surprising that an analogous property holds in the continuous-time setting.

Proposition 4.3. *Let $P \in \mathbb{P}$ be a stochastic process with corresponding family of history-dependent transition matrices (T_{t,x_u}^s). Then for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, all $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$, $x_t \in \mathscr{X}$, and $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\mathbb{E}_P[f(X_s) \,|\, X_t = x_t, X_u = x_u] = T_{t,x_u}^s f(x_t)\,.$$

Proof. Because $u < t \leq s$, it holds that $\{s\} \in \{t\} \cup u \cup \mathbb{R}_{>(\{t\} \cup u)}$, and hence, by Proposition 2.23_{73}—with $v = \{s\}$—it holds that

$$\begin{aligned}
\mathbb{E}_P[f(X_s) \,|\, X_t = x_t, X_u = x_u] &= \sum_{y \in \mathscr{X}} f(y) P(X_s = y \,|\, X_t = x_t, X_u = x_u) \\
&= \sum_{y \in \mathscr{X}} f(y) T_{t,x_u}^s(x_t, y) \\
&= T_{t,x_u}^s f(x_t)\,,
\end{aligned}$$

where we used Definition 4.2 for the second equality, and the properties of matrix-vector multiplication for the last equality. □

We end this section by considering history-*independent* transition matrices corresponding to a given stochastic process, which we also simply call *corresponding transition matrices*. These will later be useful to describe continuous-time Markov chains, but because we have not yet formally defined those, for now we simply give this definition for general stochastic processes.

Definition 4.3 (Corresponding Transition Matrix). *Let $P \in \mathbb{P}$ be a stochastic process. Then the* family of transition matrices corresponding to P *is a multi-index family (T_t^s) of matrices T_t^s with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, that are defined, for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, as*

$$T_t^s(x_t,x_s) := P(X_s = x_s | X_t = x_t) \quad \textit{for all } x_t,x_s \in \mathscr{X} .$$

Note that the family (T_t^s) of all transition matrices corresponding to a stochastic process P, contains in particular the elements T_{t,x_u}^s of the family (T_{t,x_u}^s) of history-dependent transition matrices corresponding to P, for the choice $u = \emptyset$. Hence, the following result should not be surprising.

Corollary 4.4. *Let $P \in \mathbb{P}$ be a stochastic process with corresponding family of transition matrices (T_t^s). Then for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, the matrix T_t^s is a transition matrix, and $T_t^t = I$.*

Moreover, for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, all $x_t \in \mathscr{X}$, and all $f \in \mathscr{L}(\mathscr{X})$, it holds that

$$\mathbb{E}_P[f(X_s)\,|\,X_t = x_t] = T_t^s f(x_t) .$$

Proof. This follows from Propositions 4.2 and 4.3. □

4.3 Transition Rate Matrices

We now turn our attention to the concept of *transition rate matrices* [82]. As we shall discuss in this and following chapters, these matrices are closely related to the corresponding (history-dependent) transition matrices of stochastic processes, and therefore serve as an alternative parameterisation for them. Let us start with the general definition.

Definition 4.4 (Transition Rate Matrix). *A real-valued matrix Q is said to be a* transition rate matrix, *or sometimes simply* rate matrix, *if*

- R1: $\sum_{y \in \mathscr{X}} Q(x,y) = 0$ *for all $x \in \mathscr{X}$;*
- R2: $Q(x,y) \geq 0$ *for all $x,y \in \mathscr{X}$ such that $x \neq y$.*

We use $\mathscr{R}$ to denote the set of all transition rate matrices.

As the next result shows, the set $\mathscr{R}$ forms a convex cone in the space of all real-valued matrices; that is, $\mathscr{R}$ is closed under addition and multiplication with non-negative scalars. Hence in particular, the convex combination $\lambda Q_1 + (1-\lambda)Q_2$ of two rate matrices $Q_1, Q_2 \in \mathscr{R}$, with $\lambda \in [0,1]$, is a rate matrix.

Proposition 4.5. *For any $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ and any $Q_1, Q_2 \in \mathscr{R}$, it holds that $\lambda_1 Q_1 + \lambda_2 Q_2 \in \mathscr{R}$.*

Proof. First note that for all $x \in \mathscr{X}$, it holds that

$$\sum_{y\in\mathscr{X}} \big(\lambda_1 Q_1 + \lambda_2 Q_2\big)(x,y) = \sum_{y\in\mathscr{X}} \lambda_1 Q_1(x,y) + \lambda_2 Q_2(x,y) = 0\,,$$

because both Q_1 and Q_2 satisfy Property R1. Hence $\lambda_1 Q_1 + \lambda_2 Q_2$ also satisfies R1. Moreover, for all $x, y \in \mathscr{X}$ it holds that

$$\big(\lambda_1 Q_1 + \lambda_2 Q_2\big)(x,y) = \lambda_1 Q_1(x,y) + \lambda_2 Q_2(x,y) \geq 0\,,$$

because both Q_1 and Q_2 satisfy Property R2, and $\lambda_1, \lambda_2 \geq 0$. Hence $\lambda_1 Q_1 + \lambda_2 Q_2$ also satisfies R2, whence $\lambda_1 Q_1 + \lambda_2 Q_2$ is a rate matrix by Definition 4.4. □

Moreover, it is useful to observe that $\mathscr{R}$ is a complete metric space:

Proposition 4.6. *$\mathscr{R}$ is complete under the metric induced by our norm $\|\cdot\|$.*

Proof. Let $\{Q_i\}_{i\in\mathbb{Z}_{>0}}$ be any Cauchy sequence in $\mathscr{R}$; as discussed in Appendix A369, the space of all matrices $\mathbb{M}$ is complete, so the limit $Q_* := \lim_{i\to+\infty} Q_i$ exists in $\mathbb{M}$. We need to show that also $Q_* \in \mathscr{R}$.

First fix any $x, y \in \mathscr{X}$ with $x \neq y$, and suppose *ex absurdo* that $Q_*(x,y) < 0$. Then because $Q_* = \lim_{i\to+\infty} Q_i$ there is some $n \in \mathbb{Z}_{>0}$ such that $\|Q_n - Q_*\| < -Q_*(x,y)$. Because $Q_n \in \mathscr{R}$ it follows from Property R2 that $Q_n(x,y) \geq 0$, and therefore, because $Q_*(x,y) < 0$, that $Q_n(x,y) - Q_*(x,y) = |Q_n(x,y) - Q_*(x,y)|$. Hence it follows that

$$\begin{aligned} Q_n(x,y) - Q_*(x,y) &= |Q_n(x,y) - Q_*(x,y)| \\ &\leq \sum_{z\in\mathscr{X}} |Q_n(x,z) - Q_*(x,z)| \leq \|Q_n - Q_*\| < -Q_*(x,y)\,, \end{aligned}$$

using Equation (2.7)63 for the second inequality. Adding $Q_*(x,y)$ to both sides of this equation yields $Q_n(x,y) < 0$, which contradicts Property R2 and the fact that $Q_n \in \mathscr{R}$. Hence $Q_*(x,y) \geq 0$ and, because $x, y \in \mathscr{X}$ with $x \neq y$ are arbitrary, it follows that Q_* satisfies property R2.

Next, fix any $x \in \mathscr{X}$, and let $f \in \mathscr{L}(\mathscr{X})$ be such that $f(y) := 1$ for all $y \in \mathscr{X}$. Then $Q_* f(x) = \sum_{y\in\mathscr{X}} Q_*(x,y) f(y) = \sum_{y\in\mathscr{X}} Q_*(x,y)$, so in order to show that Q_* satisfies Property R1 it suffices to show that $Q_* f(x) = 0$.

Note that, for all $i \in \mathbb{Z}_{>0}$, it holds that $Q_i f(x) = 0$ because Q_i satisfies Property R1150 since $Q_i \in \mathscr{R}$. Because $Q_* = \lim_{i\to+\infty} Q_i$ it follows from Lemma A.34390 that $Q_* f = \lim_{i\to+\infty} Q_i f$ and therefore, since $Q_i f(x) = 0$ for all $i \in \mathbb{Z}_{>0}$, that $Q_* f(x) = \lim_{i\to+\infty} Q_i f(x) = 0$. Because $x \in \mathscr{X}$ is arbitrary, this means that Q_* also satisfies Property R1150. □

The following result provides an easy expression for the norm of any rate matrix:

Lemma 4.7. *For any $Q \in \mathscr{R}$ it holds that $\|Q\| = -2\min_{x\in\mathscr{X}} Q(x,x)$.*

Proof. First fix any $x \in \mathscr{X}$. Then because $Q \in \mathscr{R}$, it follows from Property R1150 that $\sum_{y\in\mathscr{X}\setminus\{x\}} Q(x,y) = -Q(x,x)$ and therefore, due to Property R2150, that $Q(x,x) \leq 0$ and that $|Q(x,x)| = \sum_{y\in\mathscr{X}\setminus\{x\}} Q(x,y)$, which implies that $2|Q(x,x)| = \sum_{y\in\mathscr{X}} |Q(x,y)|$. From Equation (2.7)63 we obtain

$$\|Q\| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |Q(x,y)| = 2\max_{x\in\mathscr{X}} |Q(x,x)| = -2\min_{x\in\mathscr{X}} Q(x,x)\,,$$

where for the final equality we used the conjugacy property $\min\{\cdot\} = -\max\{-\cdot\}$ and the fact that $Q(x,x) \leq 0$. □

Now consider any set $\mathscr{Q} \subseteq \mathscr{R}$ of rate matrices. We note from Definition A.12376 that $\mathscr{Q}$ is said to be *bounded* if $\|\mathscr{Q}\| := \sup_{Q\in\mathscr{Q}} \|Q\| < +\infty$. The following proposition provides a simple alternative characterisation of boundedness.

Proposition 4.8. *A set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ is bounded if and only if*

$$\inf_{Q\in\mathscr{Q}} Q(x,x) > -\infty \text{ for all } x \in \mathscr{X}. \tag{4.8}$$

Proof. We start by proving that Equation (4.8) implies $\|\mathscr{Q}\| < +\infty$; so assume that Equation (4.8) holds. Then it follows from Lemma 4.7 that

$$\|\mathscr{Q}\| = \sup_{Q\in\mathscr{Q}} \|Q\| = \sup_{Q\in\mathscr{Q}} -2\min_{x\in\mathscr{X}} Q(x,x) = -2\min_{x\in\mathscr{X}} \inf_{Q\in\mathscr{Q}} Q(x,x)\,,$$

where in the final equality we used that $\mathscr{X}$ is finite together with the conjugacy property $\inf\{\cdot\} = -\sup\{-\cdot\}$. This implies that there is some $x \in \mathscr{X}$ such that $\|\mathscr{Q}\| = -2\inf_{Q\in\mathscr{Q}} Q(x,x)$ which, using Equation (4.8), implies that $\|\mathscr{Q}\| < +\infty$. This concludes the proof for the first direction.

We next show that $\|\mathscr{Q}\| < +\infty$ implies Equation (4.8). So suppose that $\|\mathscr{Q}\| < +\infty$, and assume *ex absurdo* that Equation (4.8) does not hold. Then there is some $x \in \mathscr{X}$ and $Q \in \mathscr{Q}$ such that $Q(x,x) < -\|\mathscr{Q}\|$. However, this implies that $\|\mathscr{Q}\| \geq \|Q\| \geq |Q(x,x)| > \|\mathscr{Q}\|$, which is a contradiction. Hence, it follows that Equation (4.8) must be true, which concludes the proof for the second direction. □

The connection between transition matrices and rate matrices is perhaps best illustrated intuitively, as follows. Suppose that at some time point t, we want to describe for any state x the probability of ending up in state y at some time $s \geq t$. Let us collect all these probabilities in a family (T_t^s) of transition matrices T_t^s, with $t, s \in \mathbb{R}_{\geq 0}$ and $t \leq s$. Note first of all that it is reasonable to assume that at any point t in time, the system can only be in one state. That is, if we are in state x at time t, then the probability of still being in state x at time $s = t$, should be one. Hence, we should have $T_t^t = I$, with I the identity matrix, for all $t \in \mathbb{R}_{\geq 0}$. Put differently, this essentially says that the system cannot change between states without time moving forward.

A rate matrix Q is then used to describe the transition matrix $T_t^{t+\Delta}$ after a small period of time, $\Delta \in \mathbb{R}_{\geq 0}$, has elapsed. Specifically, the scaled matrix ΔQ serves as a linear approximation of the change from T_t^t to $T_t^{t+\Delta}$. The following proposition states that, for small enough Δ, this linear approximation is also a transition matrix.

Proposition 4.9. *Consider any rate matrix $Q \in \mathscr{R}$, and any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta \|Q\| \leq 1$. Then $(I + \Delta Q)$ is a transition matrix.*

Proof. T1$_{91}$ follows from R1$_{150}$: for all $x \in \mathscr{X}$, R1$_{150}$ implies that

$$\sum_{y \in \mathscr{X}} \big(I + \Delta Q\big)(x,y) = \sum_{y \in \mathscr{X}} I(x,y) + \Delta \sum_{y \in \mathscr{X}} Q(x,y) = 1 + \Delta 0 = 1.$$

T2$_{91}$ follows from R2$_{150}$ and that $0 \leq \Delta \|Q\| \leq 1$: for all $x, y \in \mathscr{X}$ such that $x \neq y$, $0 \leq \Delta \|Q\| \leq 1$ implies that

$$\big(I + \Delta Q\big)(x,x) = 1 + \Delta Q(x,x) \geq 1 - \Delta \|Q\| \geq 0\,,$$

and R2$_{150}$ and $\Delta \geq 0$ imply that $\big(I + \Delta Q\big)(x,y) = \Delta Q(x,y) \geq 0$. □

This also explains the terminology used; a rate matrix describes the "rate of change" of a (continuously) time-dependent transition matrix over a small enough period of time. Of course, this notion can also be reversed; given a transition matrix $T_t^{t+\Delta}$, what is the change that it underwent compared to $T_t^t = I$? The following proposition states that such a change can always be described using a rate matrix.

Proposition 4.10. *Consider any transition matrix T, and any $\Delta \in \mathbb{R}_{>0}$. Then $^1/_\Delta(T - I)$ is a rate matrix.*

Proof. This proof is analogous to that of Proposition 4.9; simply verify both of the properties in Definition 4.4$_{150}$. □

Note that Proposition 4.10↶ essentially states that the finite-difference $1/\Delta(T_t^{t+\Delta} - T_t^t)$ is a rate matrix. Intuitively, if we now take the limit as this Δ goes to zero, this states that the derivative of a continuously time-dependent transition matrix will be given by a rate matrix $Q \in \mathscr{R}$—assuming that this limit exists, of course. We will make this connection more explicit in Section 4.6166.

We next introduce a function that often appears in the context of continuous-time Markov chains, and that will play an important role in the remainder of this work: the *matrix exponential* e^{Qt} of Qt, with Q a rate matrix and $t \in \mathbb{R}_{\geq 0}$. There are various equivalent ways in which this matrix exponential can be defined. Some notable ones are given below.

Definition 4.5 ([111, Section 4]). *Consider any rate matrix $Q \in \mathscr{R}$ and any $t \in \mathbb{R}_{\geq 0}$. Then the matrix exponential e^{Qt} of the matrix Qt can be defined in the following equivalent ways:*

ME1: $e^{Qt} := \sum_{k=0}^{+\infty} \frac{Q^k t^k}{k!}$;

ME2: $e^{Qt} := T_t$, *where* $T : \mathbb{R} \to \mathbb{M} : s \mapsto T_s$ *is such that* $T_0 := I$ *and*

$$\frac{\mathrm{d}}{\mathrm{d}s} T_s := \lim_{\Delta \to 0} \frac{1}{\Delta}\big(T_{s+\Delta} - T_s\big) = QT_s \quad \textit{for all } s \in \mathbb{R};^{2}$$

ME3: $e^{Qt} := \lim_{k \to +\infty}(I + t/k\, Q)^k$.

Before investigating any of these definitions, the following well-known result suggests why we are interested in matrix exponentials in the first place.

Proposition 4.11 ([82, Theorem 2.1.2]). *Consider any rate matrix $Q \in \mathscr{R}$ and any $t \in \mathbb{R}_{\geq 0}$. Then e^{Qt} is a transition matrix.*

Although conceptually similar, the properties of this matrix exponential differ a bit from those of the (scalar) exponential function. The following results highlight an important such difference—that stems from the fact that, in contrast to scalars, matrices do not always commute—and which will be relevant to us further on.

Proposition 4.12 ([111, Theorem 5]). *For any $Q_1, Q_2 \in \mathscr{R}$ it holds that $e^{(Q_1+Q_2)t} = e^{Q_1 t} e^{Q_2 t}$ for all $t \in \mathbb{R}_{\geq 0}$, if and only if Q_1 and Q_2 commute, i.e. when it holds that $Q_1 Q_2 = Q_2 Q_1$.*

[2] Following reference [111], T_s is defined for all $s \in \mathbb{R}$, even if in this work we only consider e^{Qt} with $t \in \mathbb{R}_{\geq 0}$.

The following result casts the above into a form that will be more directly useful to us.

Lemma 4.13. *Let $Q_1, Q_2 \in \mathscr{R}$ be two commuting rate matrices, i.e. such that $Q_1Q_2 = Q_2Q_1$. Then $e^{Q_1t+Q_2s} = e^{Q_1t}e^{Q_2s} = e^{Q_2s}e^{Q_1t}$ for all $t,s \in \mathbb{R}_{\geq 0}$.*

Proof. Let $Q_1' := Q_1t$ and $Q_2' := Q_2s$. Then, by Proposition 4.5$_{151}$, it holds that $Q_1', Q_2' \in \mathscr{R}$. Moreover, because Q_1 and Q_2 commute, it follows that also Q_1' and Q_2' commute, since

$$Q_1'Q_2' = tsQ_1Q_2 = tsQ_2Q_1 = Q_2'Q_1'\,.$$

It now follows from Proposition 4.12 that

$$e^{Q_1t+Q_2s} = e^{Q_1'+Q_2'} = e^{Q_1'}e^{Q_2'} = e^{Q_1t}e^{Q_2s}\,.$$

Completely analogously, we see that also $e^{Q_1t+Q_2s} = e^{Q_2s}e^{Q_1t}$, whence it follows that $e^{Q_1t}e^{Q_2s} = e^{Q_2s}e^{Q_1t}$. □

Note that Lemma 4.13 does not hold vacuously; for any $Q_1 \in \mathscr{R}$ and $\lambda \in \mathbb{R}_{\geq 0}$, it follows from Proposition 4.5$_{151}$ that $Q_2 := \lambda Q_1$ is a rate matrix, and $Q_1Q_2 = \lambda Q_1Q_1 = Q_2Q_1$, whence Q_1 and Q_2 are two commuting rate matrices.

Let us now consider the various forms by which Definition 4.5 defines the matrix exponential e^{Qt}. First, the power series expression in ME1 is also the form in which e^{at} is often defined for scalar $a \in \mathbb{R}$.

The expression in ME2 provides the well-known connection of the exponential function to differential forms. In particular, this gives the definition of the exponential function as the solution T_t of the homogeneous matrix-valued initial value problem $\frac{\mathrm{d}}{\mathrm{d}s}T_s = QT_s$ with initial (boundary) condition $T_0 = I$. Note that this implies that the k-th derivative of $e^{Qt} = T_t$ in t is given by

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}T_t = Q^kT_t\,,$$

from which we see that the power series in ME1 is exactly the Taylor expansion of e^{Qt} around $t = 0$. Moreover, this also means that Q is the derivative of e^{Qt} at $t = 0$. This result will be crucial in the sequel, so we formalise it below.

Lemma 4.14 ([82, Theorem 2.1.1])**.** *For any $Q \in \mathscr{R}$, we have that*

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{Qt}\Big|_{t=0} = \lim_{\Delta\to 0}\frac{1}{\Delta}(e^{Q\Delta} - I) = Q.$$

Finally, the expression in ME3$_{154}$ can be understood as the Euler solution of the ordinary differential equation expressed in ME2$_{154}$. One way to understand this is that, for large enough k, it follows from Proposition 4.9$_{153}$ that $(I + {}^{t}/_{k}Q)$ is a transition matrix. Following the discussion around that result, we can understand this matrix as representing the transition probabilities over the (small) time step ${}^{t}/_{k}$. Since we know from Proposition 3.8$_{91}$ that products of transition matrices are, themselves, transition matrices, it follows that $(I + {}^{t}/_{k}Q)^k$ is a transition matrix. This compound matrix can be understood as containing the transition probabilities over k steps, where for each step, the probabilities are given by $(I + {}^{t}/_{k}Q)$. In the limit as k goes to infinity, this composition becomes the matrix exponential e^{Qt} which, as we know from Proposition 4.11$_{154}$, is a transition matrix. In light of this interpretation, it should be intuitively clear how matrix exponentials of rate matrices play an important role in the theory of continuous-time Markov chains; we shall formalise this in Chapter 5$_{181}$.

4.4 (Well-Behaved) Transition Matrix Systems

In the previous section, we discussed the relationship between transition rate matrices, and transition matrices. In particular, we motivated this connection in terms of the (differential) behaviour of continuously time-dependent transition matrices T_t (or T_t^s). Moreover, in Section 4.2$_{148}$ we introduced the families $(T_{t\,x_u}^s)$ and (T_t^s) of (history-dependent) transition matrices corresponding to continuous-time stochastic processes. Indeed, these families can be understood as continuously time-dependent transition matrices. It is the goal of this section to study such families in an abstract sense. In Chapter 5$_{181}$ we will make the explicit connection to continuous-time stochastic processes, and in particular to continuous-time Markov chains.

In this section we focus on two-parameter families (T_t^s) with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, where each T_t^s is a transition matrix. We have already noted in the previous section that, due to the underlying interpretation we want to work with, it is reasonable to assume that $T_t^t = I$. If the transition matrices of a family (T_t^s) satisfy this property, and if they furthermore satisfy the *semigroup* property—see Equation 4.9 below—we call this family a *transition matrix system*.

Definition 4.6 (Transition Matrix System)**.** *A* transition matrix system *(T_t^s) is a two-parameter family of transition matrices T_t^s, defined for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, such that for all $t,r,s \in \mathbb{R}_{\geq 0}$ with $t \leq r \leq s$, it holds that*

$$T_t^s = T_t^r T_r^s \,, \tag{4.9}$$

and for all $t \in \mathbb{R}_{\geq 0}$, that $T_t^t = I$.

We use $\boldsymbol{\mathcal{T}}$ to denote the set of all transition matrix systems.

In the previous section, we have seen that for any transition matrix T and any $\Delta \in \mathbb{R}_{>0}$, the matrix $1/\Delta(T - I)$ is a rate matrix, and therefore, in particular, that the finite difference $1/\Delta(T_t^{t+\Delta} - I)$ is a rate matrix. We here note that this is also true for $1/\Delta(T_{t-\Delta}^t - I)$ whenever $(t - \Delta) > 0$.

We now consider this property in the context of a transition matrix system (T_t^s). For all $t \in \mathbb{R}_{\geq 0}$ and all $\Delta \in \mathbb{R}_{>0}$, such a transition matrix system specifies a transition matrix $T_t^{t+\Delta}$ and—if $(t - \Delta) \geq 0$—a transition matrix $T_{t-\Delta}^t$. We now consider the behaviour of these matrices for various values of Δ. In particular, we look what happens to these finite differences if we take Δ to be increasingly smaller.

For each $\Delta \in \mathbb{R}_{>0}$, due to the property that we have just recalled, there will be a rate matrix that corresponds to these finite differences. If the norm of these rate matrices never diverges to $+\infty$ as we take Δ to zero, we call the family (T_t^s) *well-behaved*.

Definition 4.7 (Well-Behaved Transition Matrix System)**.** *A transition matrix system (T_t^s) is called* well-behaved *if*

$$(\forall t \in \mathbb{R}_{\geq 0}) \ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| < +\infty, \tag{4.10}$$

and

$$(\forall t \in \mathbb{R}_{>0}) \ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta}^t - I \right\| < +\infty. \tag{4.11}$$

We stress that the interpretation is analogous to the definition of well-behavedness that we introduced in Section 4.1$_{144}$, and that we encountered in terms of families of transition matrices in Proposition 4.2$_{149}$.

We now consider an important special type of transition matrix systems. We have seen in the previous section that for any $Q \in \mathscr{R}$ and any $t \in \mathbb{R}_{\geq 0}$, the matrix exponential e^{Qt} is a transition matrix. Hence we can consider the one-parameter family (e^{Qt}), with $t \in \mathbb{R}_{\geq 0}$, of transition matrices e^{Qt}. This family is known as the *semigroup generated by Q*, and Q is known as the *generator* of this semigroup [94, Chapter 13]. This family satisfies the semigroup property, analogous to Equation (4.9), in the following sense.

Proposition 4.15. *Consider any rate matrix $Q \in \mathscr{R}$, and consider the family (e^{Qt}), with $t \in \mathbb{R}_{\geq 0}$, of transition matrices e^{Qt}. Then for all $t, s \in \mathbb{R}_{\geq 0}$ it holds that $e^{Q(t+s)} = e^{Qt}e^{Qs} = e^{Qs}e^{Qt}$.*

Proof. This follows from Lemma 4.13$_{155}$ and the fact that Q trivially commutes with itself. □

Although the family (e^{Qt}) has only a single index $t \in \mathbb{R}_{\geq 0}$, we can identify with it a (two-parameter) family $(e^{Q(s-t)})$, with parameters $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$; although this second index is superfluous, this allows us to streamline our notation for what follows. Hence, we introduce the following definition.

Definition 4.8. *For any rate matrix $Q \in \mathscr{R}$, we use $(e^{Q(s-t)})$ to denote the two-parameter family of transition matrices $e^{Q(s-t)}$, defined for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$. We call this family $(e^{Q(s-t)})$ the* exponential transition matrix system *corresponding to Q.*

The next result motivates the terminology.

Proposition 4.16. *For any $Q \in \mathscr{R}$, $(e^{Q(s-t)})$ is a well-behaved transition matrix system.*

Proof. For ease of notation, let $T_t^s := e^{Q(s-t)}$ for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$. We start by showing that $(e^{Q(s-t)})$ is a transition matrix system. Because of Proposition 4.11$_{154}$, $(e^{Q(s-t)})$ is clearly a family of transition matrices. Consider now any $t,r,s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$. It then follows from the definition of $(e^{Q(s-t)})$ and Proposition 4.15$_{\curvearrowleft}$ that

$$T_t^s = e^{Q(s-t)} = e^{Q(s-r+r-t)} = e^{Q(r-t)}e^{Q(s-r)} = T_t^r T_r^s\,,$$

and $T_t^t = I$. Because the t,r,s are arbitrary, it follows from Definition 4.6$_{156}$ that $(e^{Q(s-t)})$ is a transition matrix system.

To prove that $(e^{Q(s-t)})$ is well-behaved, note that for any $t \in \mathbb{R}_{\geq 0}$, because of Definition 4.8,

$$\begin{aligned}\limsup_{\Delta\to 0^+} \frac{1}{\Delta}\left\|T_t^{t+\Delta} - I\right\| &= \limsup_{\Delta\to 0^+}\left\|\frac{1}{\Delta}(T_t^{t+\Delta} - I) - Q + Q\right\| \\ &\leq \limsup_{\Delta\to 0^+}\left\|\frac{1}{\Delta}(T_t^{t+\Delta} - I) - Q\right\| + \|Q\| = \|Q\| < \infty\,,\end{aligned}$$

where the first inequality follows from Proposition 2.16$_{63}$, the second equality follows from Lemma 4.14$_{155}$ and the final inequality uses that Q is real-valued. Because this holds for any $t \in \mathbb{R}_{\geq 0}$, the condition in Equation (4.10)$_{\curvearrowleft}$ is satisfied. A similar argument shows that Equation (4.11)$_{\curvearrowleft}$ is also satisfied, whence $(e^{Q(s-t)})$ is well-behaved. □

4.5 Restricted Transition Matrix Systems

We will now introduce some machinery that allows us to combine two (or more) transition matrix systems into another, new, transition matrix system. The fundamental concept that we need is the restriction of a transition matrix system (T_t^s) to a closed interval $\mathbf{I}$ in its index set $\mathbb{R}_{\geq 0}$. By a *closed interval* $\mathbf{I}$, we here mean a non-empty closed subset $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$ that is connected, in the sense that for any $t, s \in \mathbf{I}$ such that $t \leq s$, and any $r \in [t, s]$, it holds that $r \in \mathbf{I}$. Note that for any $c \in \mathbb{R}_{\geq 0}$, $[c, +\infty)$ is such a closed interval.

For any transition matrix system (T_t^s) and any such closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, we use $(T_t^s)_{\mathbf{I}}$ to denote the restriction of (T_t^s) to $\mathbf{I}$. This restriction is a family of transition matrices T_t^s that is defined for all $t, s \in \mathbf{I}$ such that $t \leq s$. We call such a family $(T_t^s)_{\mathbf{I}}$ a *restricted transition matrix system* on $\mathbf{I}$. The set of all restricted transition matrix systems on $\mathbf{I}$ is denoted by $\mathcal{T}_{\mathbf{I}}$.

Proposition 4.17. *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $(T_t^s)_{\mathbf{I}}$ be a family of transition matrices T_t^s that is defined for all $t, s \in \mathbf{I}$ with $t \leq s$. Then $(T_t^s)_{\mathbf{I}}$ is a restricted transition matrix system on $\mathbf{I}$ if and only if, for all $t, r, s \in \mathbf{I}$ with $t \leq r \leq s$, it holds that $T_t^s = T_t^r T_r^s$ and $T_t^t = I$.*

Proof. If $(T_t^s)_{\mathbf{I}}$ is a restricted transition matrix system, then, by definition, it is the restriction to $\mathbf{I}$ of some transition matrix system (T_t^s). Therefore, the 'only if' part of this result follows trivially from Definition 4.6156.

For the 'if' part, we need to prove that for any family of transition matrices $(T_t^s)_{\mathbf{I}}$ such that, for all $t, r, s \in \mathbf{I}$ with $t \leq r \leq s$, it holds that $T_t^s = T_t^r T_r^s$ and $T_t^t = I$, there is a transition matrix system (T_t^s) that coincides with $(T_t^s)_{\mathbf{I}}$ on $\mathbf{I}$. In order to prove this, it suffices to show that the unique family of transition matrices (T_t^s) that coincides with $(T_t^s)_{\mathbf{I}}$ on $\mathbf{I}$ and that is otherwise defined, for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$ and $[t, s] \not\subseteq \mathbf{I}$, by

$$T_t^s := \begin{cases} I & \text{if } s < \min \mathbf{I} \\ T_{\min \mathbf{I}}^s & \text{if } t < \min \mathbf{I} \text{ and } s \in \mathbf{I} \\ T_{\min \mathbf{I}}^{\sup \mathbf{I}} & \text{if } t < \min \mathbf{I} \text{ and } \sup \mathbf{I} < s \\ T_t^{\sup \mathbf{I}} & \text{if } t \in \mathbf{I} \text{ and } \sup \mathbf{I} < s \\ I & \text{if } \sup \mathbf{I} < t \end{cases} \tag{4.12}$$

is a transition matrix system. This is a matter of straightforward verification. In the sequel, we will also refer to this transition matrix system (T_t^s) as the *canonical extension* of $(T_t^s)_{\mathbf{I}}$. □

We call a restricted transition matrix system $(T_t^s)_{\mathbf{I}}$ well-behaved if it is the restriction to $\mathbf{I}$ of a well-behaved transition matrix system.

Proposition 4.18. *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $(T_t^s)_{\mathbf{I}}$ be a restricted transition matrix system on $\mathbf{I}$. Then $(T_t^s)_{\mathbf{I}}$ is well-behaved if and* *only if*

$$(\forall t \in \mathbf{I}^+)\ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| < +\infty, \tag{4.13}$$

and

$$(\forall t \in \mathbf{I}^-)\ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta}^{t} - I \right\| < +\infty, \tag{4.14}$$

where $\mathbf{I}^+ := \mathbf{I} \setminus \{\sup \mathbf{I}\}$ *and* $\mathbf{I}^- := \mathbf{I} \setminus \{\min \mathbf{I}\}$.

Proof. If $(T_t^s)_{\mathbf{I}}$ is well-behaved, then, by definition, it is the restriction to $\mathbf{I}$ of some well-behaved transition matrix system (T_t^s). Therefore, the 'only if' part of this result follows trivially from Definition 4.7157.

For the 'if' part, we need to show that for any restricted transition matrix system $(T_t^s)_{\mathbf{I}}$ on $\mathbf{I}$ that satisfies Equations (4.13) and (4.14), there is a well-behaved transition matrix system (T_t^s) that coincides with $(T_t^s)_{\mathbf{I}}$ on $\mathbf{I}$. Let (T_t^s) be the canonical extension of $(T_t^s)_{\mathbf{I}}$, as constructed in the proof of Proposition 4.17↶. Then, as explained in that proof, (T_t^s) is a transition matrix system that coincides with $(T_t^s)_{\mathbf{I}}$ on $\mathbf{I}$. Therefore, it suffices to prove that (T_t^s) is well-behaved. We start by proving that (T_t^s) satisfies Equation (4.10)157. So consider any $t \in \mathbb{R}_{\geq 0}$. If $t \in \mathbf{I}^+$, then the desired inequality follows from Equation (4.13). If $t \notin \mathbf{I}^+$, then either $t < \min \mathbf{I}$ or $t \geq \sup \mathbf{I}$, and therefore, for sufficiently small $\Delta > 0$, it follows from Equation (4.12)↶ that $T_t^{t+\Delta} = I$, thereby making the desired inequality trivially true. That (T_t^s) satisfies Equation (4.11)157 can be proved similarly. □

In order to combine multiple such restricted transition matrix systems into a single, new, (restricted) transition matrix system, we introduce a concatenation operator, as follows.

Definition 4.9 (Concatenation Operator)**.** *For any two closed intervals* $\mathbf{I}, \mathbf{J} \subseteq \mathbb{R}_{\geq 0}$ *such that* $\max \mathbf{I}$ *exists and equals* $\min \mathbf{J}$*, and any two restricted transition matrix systems* $(T_t^s)_{\mathbf{I}}$ *and* $(S_t^s)_{\mathbf{J}}$*, the* concatenation $(T_t^s)_{\mathbf{I}} \otimes (S_t^s)_{\mathbf{J}}$ *of* $(T_t^s)_{\mathbf{I}}$ *and* $(S_t^s)_{\mathbf{J}}$ *is defined as the family* $(T_t^s)_{\mathbf{I}} \otimes (S_t^s)_{\mathbf{J}} := (R_t^s)_{\mathbf{I} \cup \mathbf{J}}$ *of transition matrices* R_t^s *that is given by*

$$R_t^s := \begin{cases} T_t^s & \text{if } t, s \in \mathbf{I} \\ S_t^s & \text{if } t, s \in \mathbf{J} \\ T_t^r S_r^s & \text{if } t \in \mathbf{I} \text{ and } s \in \mathbf{J} \end{cases} \qquad \text{for all } t, s \in \mathbf{I} \cup \mathbf{J} \text{ such that } t \leq s,$$

where $r := \max \mathbf{I} = \min \mathbf{J}$.

This concatenation operator satisfies the following intuitive property.

Proposition 4.19. *Consider two closed intervals* $\mathbf{I},\mathbf{J} \subseteq \mathbb{R}_{\geq 0}$ *such that* $\max \mathbf{I}$ *exists and equals* $\min \mathbf{J}$*, and any two restricted transition matrix systems* $(T_t^s)_{\mathbf{I}}$ *and* $(S_t^s)_{\mathbf{J}}$*. Then their concatenation* $(R_t^s)_{\mathbf{I}\cup\mathbf{J}} := (T_t^s)_{\mathbf{I}} \otimes (S_t^s)_{\mathbf{J}}$ *is a restricted transition matrix system on* $\mathbf{I}\cup\mathbf{J}$*. Furthermore, if both* $(T_t^s)_{\mathbf{I}}$ *and* $(S_t^s)_{\mathbf{J}}$ *are well-behaved, then* $(R_t^s)_{\mathbf{I}\cup\mathbf{J}}$ *is also well-behaved.*

Proof. For all $t,s \in \mathbf{I}\cup\mathbf{J}$ such that $t \leq s$, it follows from Proposition 3.8$_{91}$ that the matrix R_t^s is a transition matrix. Furthermore, for all $t \in \mathbf{I}\cup\mathbf{J}$, we have that either $t \in \mathbf{I}$ or $t \in \mathbf{J}$. In either case, we have that $R_t^t = I$, because either $R_t^t = T_t^t = I$ or $R_t^t = S_t^t = I$. Next, we show that for all $t,q,s \in \mathbf{I}\cup\mathbf{J}$ with $t \leq q \leq s$, it holds that

$$R_t^s = R_t^q R_q^s .$$

If both $t,s \in \mathbf{I}$ or if both $t,s \in \mathbf{J}$, this clearly holds. Therefore, we may assume that $t \in \mathbf{I}$ and $s \in \mathbf{J}$. Suppose furthermore that $q \in \mathbf{I}$. Then, from the definition of the concatenation operator $\otimes$, we have that $R_q^s = T_q^r S_r^s$, with $r = \max \mathbf{I} = \min \mathbf{J}$. Because $t,q,r \in \mathbf{I}$, we know that $T_t^q T_q^r = T_t^r$, and hence, by the definition of the concatenation operator,

$$R_t^q R_q^s = T_t^q T_q^r S_r^s = T_t^r S_r^s = R_t^s .$$

An exactly analogous argument proves the case for $q \in \mathbf{J}$. Therefore, it follows from Proposition 4.17$_{159}$ that $(R_t^s)_{\mathbf{I}\cup\mathbf{J}}$ is a restricted transition matrix system.

It remains to prove that if $(T_t^s)_{\mathbf{I}}$ and $(S_t^s)_{\mathbf{J}}$ are both well-behaved, that then $(R_t^s)_{\mathbf{I}\cup\mathbf{J}}$ is also well-behaved. Due to Proposition 4.18$_{159}$, it suffices to prove that $(R_t^s)_{\mathbf{I}\cup\mathbf{J}}$ satisfies Equations (4.13) and (4.14). We only prove that it satisfies Equation (4.13), that is, that

$$\left(\forall t \in (\mathbf{I}\cup\mathbf{J})^+\right) \; \limsup_{\Delta\to 0^+} \left\| \frac{1}{\Delta}\left(R_t^{t+\Delta} - I\right)\right\| < +\infty.$$

The proof for Equation (4.14) is completely analogous. So consider any $t \in (\mathbf{I}\cup\mathbf{J})^+$. Since $\sup \mathbf{I} = \max \mathbf{I} = \min \mathbf{J}$, it follows that

$$\begin{aligned}(\mathbf{I}\cup\mathbf{J})^+ &:= (\mathbf{I}\cup\mathbf{J}) \setminus \{\sup(\mathbf{I}\cup\mathbf{J})\} \\ &= (\mathbf{I}\cup\mathbf{J}) \setminus \{\sup \mathbf{J}\} = (\mathbf{I}\setminus \sup \mathbf{I}) \cup (\mathbf{J}\setminus \sup \mathbf{J}) = \mathbf{I}^+ \cup \mathbf{J}^+.\end{aligned}$$

Therefore, without loss of generality, we may assume that $t \in \mathbf{I}^+$. The desired result now follows by applying Proposition 4.18$_{159}$ to the well-behaved restricted transition matrix system $(T_t^s)_{\mathbf{I}}$. □

The following example illustrates how we can compose two (restricted) transition matrix systems into a new (unrestricted) transition matrix system.

Example 4.2. Consider any two rate matrices $Q_1, Q_2 \in \mathscr{R}$ such that $Q_1 \neq Q_2$, and let $\left(e^{Q_1(s-t)}\right)$ and $\left(e^{Q_2(s-t)}\right)$ be their exponential transition matrix systems, which, as we know from Proposition 4.16158, are well-behaved. Now choose any $r \in \mathbb{R}_{\geq 0}$ and define

$$(T_t^s) := \left(e^{Q_1(s-t)}\right)_{[0,r]} \otimes \left(e^{Q_2(s-t)}\right)_{[r,+\infty)}.$$

It then follows from Proposition 4.19↶ that (T_t^s) is a well-behaved transition matrix system. For any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$, the transition matrix T_t^s satisfies $T_t^s = T_t^r T_r^s = e^{Q_1(r-t)} e^{Q_2(s-r)}$. ◊

The last technical result of this section is one that will allow us to consider limits of sequences of (restricted) transition matrix systems. To this end, we first introduce a metric d between restricted transition matrix systems that are defined on the same interval $\mathbf{I}$. For any two such restricted transition matrix systems $(T_t^s)_\mathbf{I}$ and $(S_t^s)_\mathbf{I}$, we let

$$d\big((T_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}\big) := \sup\left\{\|T_t^s - S_t^s\| \,:\, t, s \in \mathbf{I}, t \leq s\right\}. \tag{4.15}$$

Note that the set $\boldsymbol{\mathcal{T}}$ of all (unrestricted) transition matrix systems is equal to the set $\boldsymbol{\mathcal{T}}_{\mathbb{R}_{\geq 0}}$ of all transition matrix systems that are "restricted" to $\mathbb{R}_{\geq 0}$. Hence the following results also extend to the set $\boldsymbol{\mathcal{T}}$.

Proposition 4.20. *For any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, the map d defined in Equation* (4.15) *is a metric on $\boldsymbol{\mathcal{T}}_\mathbf{I}$.*

Proof. Let us show that d satisfies all the defining properties of a metric, as in Definition A.6372. Fix any $(T_t^s)_\mathbf{I}, (R_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}$ in $\mathscr{T}_\mathbf{I}$. From the definition of the norm $\|\cdot\|$ it follows that $\|T_t^s - S_t^s\| \geq 0$ for all $t, s \in \mathbf{I}$ with $t \leq s$, which by Equation (4.15) implies that $d\big((T_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}\big) \geq 0$.

Next, we note that, using the definition of the norm $\|\cdot\|$, it holds that

$$d\big((T_t^s)_\mathbf{I}, (T_t^s)_\mathbf{I}\big) = \sup\left\{\|T_t^s - T_t^s\| \,:\, t, s \in \mathbf{I}, t \leq s\right\} = 0.$$

Conversely, suppose that $d\big((T_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}\big) = 0$. Then it clearly follows from Equation (4.15) that $\|T_t^s - S_t^s\| = 0$ for all $t, s \in \mathbf{I}$ with $t \leq s$. This implies that $T_t^s = S_t^s$ for all $t, s \in \mathbf{I}$ with $t \leq s$, which means that $(T_t^s)_\mathbf{I} = (S_t^s)_\mathbf{I}$. Hence we have found that $d\big((T_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}\big) = 0$ if and only if $(T_t^s)_\mathbf{I} = (S_t^s)_\mathbf{I}$.

Next, again by the definition of the norm $\|\cdot\|$, we have that

$$\begin{aligned} d\big((T_t^s)_\mathbf{I}, (S_t^s)_\mathbf{I}\big) &= \sup\left\{\|T_t^s - S_t^s\| \,:\, t, s \in \mathbf{I}, t \leq s\right\} \\ &= \sup\left\{\|S_t^s - T_t^s\| \,:\, t, s \in \mathbf{I}, t \leq s\right\} = d\big((S_t^s)_\mathbf{I}, (T_t^s)_\mathbf{I}\big). \end{aligned}$$

Finally, once more using the definition of the norm $\|\cdot\|$, we have that

$$
\begin{aligned}
&d\big((T_t^s)_{\mathbf{I}},(S_t^s)_{\mathbf{I}}\big)\\
&\qquad= \sup\{\|T_t^s - S_t^s\| : t,s\in\mathbf{I}, t\leq s\}\\
&\qquad\leq \sup\{\|T_t^s - R_t^s\| + \|R_t^s - S_t^s\| : t,s\in\mathbf{I}, t\leq s\}\\
&\qquad\leq \sup\{\|T_t^s - R_t^s\| : t,s\in\mathbf{I}, t\leq s\} + \sup\{\|R_t^s - S_t^s\| : t,s\in\mathbf{I}, t\leq s\}\\
&\qquad= d\big((T_t^s)_{\mathbf{I}},(R_t^s)_{\mathbf{I}}\big) + d\big((R_t^s)_{\mathbf{I}},(S_t^s)_{\mathbf{I}}\big),
\end{aligned}
$$

which concludes the proof. □

Proposition 4.21. *Consider any interval* $\mathbf{I}\subseteq\mathbb{R}_{\geq 0}$ *and let* d *be the metric that is defined in Equation* (4.15)*. The metric space* $(\mathcal{T}_{\mathbf{I}}, d)$ *is then complete.*

Proof. Consider any sequence $\{({}^iT_t^s)_{\mathbf{I}}\}_{i\in\mathbb{Z}_{>0}}$ of restricted transition matrix systems $({}^iT_t^s)_{\mathbf{I}}$ in $\mathcal{T}_{\mathbf{I}}$ that is Cauchy. We will prove that this sequence converges to a limit that belongs to $\mathcal{T}_{\mathbf{I}}$.

Since $\{({}^iT_t^s)_{\mathbf{I}}\}_{i\in\mathbb{Z}_{>0}}$ is Cauchy, it follows from Equation (4.15) that

$$(\forall \varepsilon\in\mathbb{R}_{>0})\,(\exists n_\varepsilon\in\mathbb{Z}_{>0})\,(\forall k,\ell>n_\varepsilon)\,(\forall t,s\in\mathbf{I}\colon t\leq s)\ \left\|{}^kT_t^s - {}^\ell T_t^s\right\| < \varepsilon. \tag{4.16}$$

Clearly, for any $t,s\in\mathbf{I}$ such that $t\leq s$, this implies that the sequence $\{{}^iT_t^s\}_{i\in\mathbb{Z}_{>0}}$ of transition matrices is Cauchy. Since the set $\mathbb{T}$ of all transition matrices is complete by Proposition 3.10_{92}, this implies that the sequence $\{{}^iT_t^s\}_{i\in\mathbb{Z}_{>0}}$ has a limit T_t^s, and that this limit is a transition matrix. We use $(T_t^s)_{\mathbf{I}}$ to denote the family of transition matrices that consists of these limits T_t^s, with $t,s\in\mathbf{I}$ and $t\leq s$.

Fix any $t,r,s\in\mathbf{I}$ such that $t\leq r\leq s$. Then for any $i\in\mathbb{Z}_{>0}$, because $({}^iT_t^s)_{\mathbf{I}}$ is a restricted transition matrix system, we know that ${}^iT_t^t = I$ and ${}^iT_t^s = {}^iT_t^r\,{}^iT_r^s$, which implies that $\|T_t^t - I\| = \left\|T_t^t - {}^iT_t^t\right\|$ and, due to Lemma $\mathrm{B.5}_{393}$, that

$$
\begin{aligned}
\|T_t^s - T_t^rT_r^s\| &\leq \left\|T_t^s - {}^iT_t^s\right\| + \left\|{}^iT_t^r\,{}^iT_r^s - T_t^rT_r^s\right\|\\
&\leq \left\|T_t^s - {}^iT_t^s\right\| + \left\|{}^iT_t^r - T_t^r\right\| + \left\|{}^iT_r^s - T_r^s\right\|.
\end{aligned}
$$

Since we know that $\lim_{i\to+\infty}{}^iT_t^t = T_t^t$, $\lim_{i\to+\infty}{}^iT_t^s = T_t^s$, $\lim_{i\to+\infty}{}^iT_t^r = T_t^r$ and $\lim_{i\to+\infty}{}^iT_r^s = T_r^s$, this implies that $\|T_t^t - I\| = 0$ and $\|T_t^s - T_t^rT_r^s\| = 0$, or equivalently, that $T_t^t = I$ and $T_t^s = T_t^rT_r^s$. Since this is true for any $t,r,s\in\mathbf{I}$ such that $t\leq r\leq s$, and because we already know that the family $(T_t^s)_{\mathbf{I}}$ consists of transition matrices, it follows from Proposition 4.17_{159} that $(T_t^s)_{\mathbf{I}}$ is a restricted transition matrix system. In the remainder of this proof, we will show that $(T_t^s)_{\mathbf{I}} = \lim_{i\to\infty}({}^iT_t^s)_{\mathbf{I}}$.

Fix any $\varepsilon>0$ and consider the corresponding $n_\varepsilon\in\mathbb{Z}_{>0}$ whose existence is guaranteed by Equation (4.16). Fix any $k>n_\varepsilon$. For any $t,s\in\mathbf{I}$

such that $t \leq s$, it then follows from Equation $(4.16)_{\curvearrowleft}$ that, for all $\ell > n_\varepsilon$:

$$\left\| {}^kT_t^s - T_t^s \right\| \leq \left\| {}^kT_t^s - {}^\ell T_t^s \right\| + \left\| {}^\ell T_t^s - T_t^s \right\| < \varepsilon + \left\| {}^\ell T_t^s - T_t^s \right\|.$$

Since $\lim_{\ell \to +\infty} {}^\ell T_t^s = T_t^s$, this implies that $\left\| {}^kT_t^s - T_t^s \right\| \leq \varepsilon$. Since this is true for all $t, s \in \mathbf{I}$ such that $t \leq s$, it follows from Equation $(4.15)_{162}$ that $d\big(({}^kT_t^s)_{\mathbf{I}}, (T_t^s)_{\mathbf{I}}\big) \leq \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we conclude that

$$(\forall \varepsilon \in \mathbb{R}_{>0})\,(\exists n_\varepsilon \in \mathbb{Z}_{>0})\,(\forall k > n_\varepsilon)\; d\big(({}^kT_t^s)_{\mathbf{I}}, (T_t^s)_{\mathbf{I}}\big) \leq \varepsilon,$$

which implies that $(T_t^s)_{\mathbf{I}} = \lim_{i \to \infty} ({}^iT_t^s)_{\mathbf{I}}$. □

We now conclude this section with some examples that illustrate how this result can be used. Moreover, these examples will serve as the formal basis of some examples in future chapters. Because the results are somewhat technical, we only present the basic ideas here and have moved the bulk of the formal effort of proving their correctness to Appendix $4.\mathrm{A}_{173}$.

Example 4.3. Consider some positive constant $c \in \mathbb{R}_{>0}$ and let $\{Q_i\}_{i \in \mathbb{Z}_{\geq 0}}$ be a sequence of rate matrices such that, for all $i \in \mathbb{Z}_{>0}$, $\|Q_i - Q_{i-1}\| \leq c$. We can then construct the following sequence of transition matrix systems. For $i = 0$, we let $({}^0T_t^s) \coloneqq \big(e^{Q_0(s-t)}\big)$ and, for all $i \in \mathbb{Z}_{>0}$, we let

$$({}^iT_t^s) \coloneqq \big(e^{Q_i(s-t)}\big)_{[0,\delta_i]} \otimes \big({}^{i-1}T_t^s\big)_{[\delta_i,+\infty)} \tag{4.17}$$

where, for all $i \in \mathbb{Z}_{\geq 0}$, $\delta_i \coloneqq 2^{-i}$. The resulting sequence $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{\geq 0}}$ is then in $\mathcal{T}$ and, because of Propositions 4.16_{158} and 4.19_{161}, every transition matrix system in this sequence is well-behaved. Furthermore, as is proved in Appendix $4.\mathrm{A}_{173}$, $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{\geq 0}}$ is a Cauchy sequence, which basically means that its elements become arbitrarily close to each other as the sequence progresses.

The reason why this is of interest to us is that in a complete metric space, every Cauchy sequence converges to a limit that belongs to the same space. Hence, since $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{\geq 0}}$ is Cauchy, Proposition $4.21_{\curvearrowleft}$ tells us that $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{\geq 0}}$ converges to a limit $(T_t^s) \coloneqq \lim_{i \to \infty} ({}^iT_t^s)$ in $\mathcal{T}$. ◊

As this example illustrates, Proposition $4.21_{\curvearrowleft}$ allows us to (i) establish the existence of limits of sequences of (restricted) transition matrix systems and (ii) prove that these limits are restricted transition matrix systems themselves. In order to make this concept of a limit of transition matrix systems less abstract, we now provide, for a particular case of the sequence in Example 4.3, closed-form expressions for some of the transition matrices that correspond to its limit.

Example 4.4. Let $Q_1, Q_2 \in \mathscr{R}$ be two commuting rate matrices. For example, fix $Q_1 \in \mathscr{R}$ and let $Q_2 := \alpha Q_1$, with $\alpha \in \mathbb{R}_{\geq 0}$.

Now let $\{Q_i\}_{i \in \mathbb{Z}_{\geq 0}}$ be defined by $Q_i := Q_1$ if i is odd and $Q_i := Q_2$ if i is even, let $\delta_i := 2^{-i}$ for all $i \in \mathbb{Z}_{\geq 0}$, and consider the corresponding sequence of transition matrix systems $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{>0}}$ that was defined in Example 4.3. Since $\|Q_i - Q_{i-1}\| = \|Q_1 - Q_2\|$ for all $i \in \mathbb{Z}_{>0}$, the sequence $\{Q_i\}_{i \in \mathbb{Z}_{\geq 0}}$ clearly satisfies the conditions in Example 4.3—just choose $c = \|Q_1 - Q_2\|$—and therefore, as we have seen, $\{({}^iT_t^s)\}_{i \in \mathbb{Z}_{>0}}$ converges to a limit $(T_t^s) := \lim_{i \to \infty}({}^iT_t^s)$ in $\mathcal{T}$.

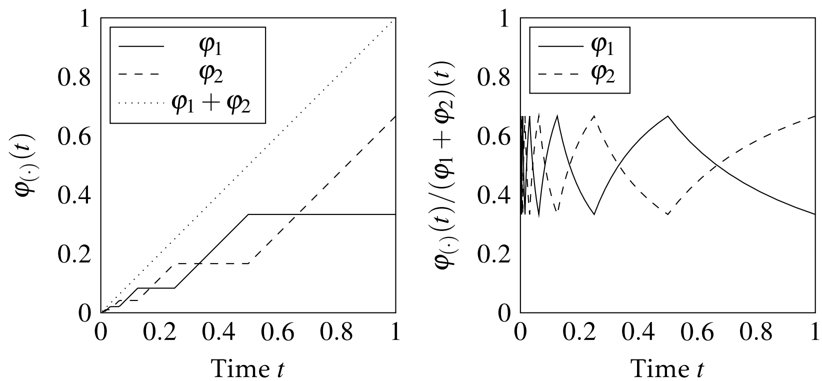

Figure 4.1: Left: Graphical illustration of the functions φ_1 and φ_2 from Equations (4.19) and (4.20), respectively. Note that $\phi_1(t) + \phi_2(t) = t$ for all $t \in (0,1]$. Right: Relative proportion of φ_1 and φ_2 as mixing coefficients of the rate matrices Q_1 and Q_2, as used in Equation (4.18). Note that these proportions keep oscillating as t goes to zero.

As proved in Appendix 4.A$_{173}$, it then holds that for any $t \in (0,1]$, the transition matrix T_0^t of this limiting transition matrix system is equal to

$$T_0^t = e^{Q_1\varphi_1(t) + Q_2\varphi_2(t)}, \tag{4.18}$$

with

$$\varphi_1(t) := \begin{cases} t - 2/3\,\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ odd} \\ 2/3\,\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ even} \end{cases} \tag{4.19}$$

and

$$\varphi_2(t) := \begin{cases} 2/3\,\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ odd} \\ t - 2/3\,\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ even.} \end{cases} \tag{4.20}$$

These somewhat abstract looking functions are illustrated in Figure 4.1↶. We note that $\varphi_1(t) + \varphi_2(t) = t$ for all $t \in (0,1]$, so Equation (4.18)↶ can also be understood as showing the matrix exponential $e^{Q_t t}$, where $Q_t := Q_1 \varphi_1(t)/t + Q_2 \varphi_2(t)/t$. Which is to say, φ_1 and φ_2 essentially give mixing coefficients that together determine a rate matrix Q_t at time t. As the right side of Figure 4.1↶ shows, these mixing coefficients oscillate ever more wildly as t goes to zero.

It can be shown that the transition matrix system (T_t^s) is well-behaved—again, see Appendix 4.A$_{173}$ for a proof. Moreover, the particular character that this transition matrix system exhibits will be used to show, in Example 4.6 further on, that the transition matrix T_0^t corresponding to this transition matrix system is not differentiable (directionally; from the right) in $t = 0$; this highlights the difference between well-behavedness and differentiability. ◊

The transition matrix system (T_t^s) in our previous example was well-behaved, and was constructed as a limit of well-behaved transition matrix systems. Therefore, one might think that the former is implied by the latter. However, as our next example illustrates, this is not the case: a limit of well-behaved transition matrix systems need not be well-behaved itself.

Example 4.5. Consider any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$ and, for all $i \in \mathbb{Z}_{\geq 0}$, define $Q_i := iQ$ and let $({}^iT_t^s)$ and δ_i be defined as in Example 4.3$_{164}$. Then since $\{Q_i\}_{i\in\mathbb{Z}_{\geq 0}}$ satisfies the conditions of Example 4.3$_{164}$ with $c = 1$, the sequence $\{({}^iT_t^s)\}_{i\in\mathbb{Z}_{\geq 0}}$ has a limit $(T_t^s) := \lim_{i\to\infty}({}^iT_t^s)$ in $\boldsymbol{\mathcal{T}}$.

However, despite the fact that we know from Example 4.3$_{164}$ that each of the transition matrix systems $({}^iT_t^s)$, $i \in \mathbb{Z}_{\geq 0}$, is well-behaved, the limit (T_t^s) itself is not well-behaved; see Appendix 4.A$_{173}$ for a proof. ◊

4.6 Outer Partial Derivatives

We conclude this chapter by considering the connection between rate matrices, and the families of transition matrices corresponding to continuous-time stochastic processes. We have already discussed in Section 4.3$_{150}$ that rate matrices can be understood as representing the rate of change of continuously time-dependent transition matrices. Here, we consider this interpretation in the context of the dynamics of stochastic processes.

One seemingly obvious way of describing these dynamics is to use the derivatives of the transition matrices that correspond to stochastic processes. However, because we do not impose differentiability assumptions on these processes, such derivatives may not exist. We will therefore instead introduce *outer partial derivatives* below. It will be instructive, however, to first consider ordinary *directional partial derivatives.*

Definition 4.10 (Directional Partial Derivatives). *For any stochastic process $P \in \mathbb{P}$ with corresponding family of history-dependent transition matrices (T^s_{t,x_u}), any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the* right-sided partial derivative *of T^t_{t,x_u} is defined by*

$$\partial_+ T^t_{t,x_u} := \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T^{t+\Delta}_{t,x_u} - T^t_{t,x_u}) = \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T^{t+\Delta}_{t,x_u} - I)$$

and, if $t \neq 0$, the left-sided partial derivative *of T^t_{t,x_u} is defined by*

$$\partial_- T^t_{t,x_u} := \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T^t_{t-\Delta,x_u} - T^t_{t,x_u}) = \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T^t_{t-\Delta,x_u} - I).$$

If these partial derivatives both exist and coincide, we write $\partial T^t_{t,x_u}$ to denote their common value. For $t = 0$, we let $\partial T^t_{t,x_u} := \partial_+ T^t_{t,x_u}$.

We note that whenever these (directional) partial derivatives exist, then because of Propositions 4.6_{151} and 4.10_{153}, they are guaranteed to belong to the set of all rate matrices $\mathscr{R}$.

The following example establishes that such directional partial derivatives need not always exist. In particular, they need not exist even for all well-behaved processes.

Example 4.6. Let $Q_1, Q_2 \in \mathscr{R}$ be two commuting rate matrices such that $Q_1 \neq Q_2$—for example, let $Q_1 \neq 0$ be an arbitrary rate matrix and let $Q_2 := \alpha Q_1$, with $\alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}$—and consider a well-behaved stochastic process $P \in \mathbb{P}^{\mathrm{W}}$ of which, for all $t \in (0, 1]$, the corresponding transition matrix T^t_0 is given by Equation $(4.18)_{165}$ in Example 4.4_{165}.

For now, we simply assume that this is possible. A formal proof for the existence of such a process requires some additional machinery, and we therefore postpone it to Example 5.1_{184}, where we construct a well-behaved continuous-time Markov chain that is compatible with Equation $(4.18)_{165}$.

The aim of the present example is to show that for any such process, the right-sided partial derivative $\partial_+ T^0_0$—which corresponds to choosing $t = 0$ and $u = \emptyset$ in Definition 4.10—does not exist. The reason for this is that—as is proved in Appendix $4.\mathrm{B}_{177}$—for any $\lambda \in [1/3, 2/3]$, there is a sequence $\{\Delta_i\}_{i \in \mathbb{Z}_{>0}} \to 0^+$ such that

$$\lim_{i \to +\infty} \frac{1}{\Delta_i}(T^{\Delta_i}_0 - I) = Q_\lambda \tag{4.21}$$

with $Q_\lambda := \lambda Q_1 + (1-\lambda)Q_2$. The reason why this indeed implies that $\partial_+ T_0^0$ does not exist, is that if it would exist, then Equation (4.21)↶ would imply that $\partial_+ T_0^0 = Q_\lambda$ for all $\lambda \in [1/3, 2/3]$. The only way for this to be possible is that $Q_1 = Q_2$, but this was excluded in the beginning of this example.

This property is also explained graphically in the right side of Figure 4.1$_{165}$: because the mixing proportions keep oscillating as we take t (or in our current notation, Δ) to zero, the sequence $Q_\Delta := Q_1 \varphi_1(\Delta)/\Delta + Q_2 \varphi_2(\Delta)/\Delta$ has multiple accumulation points depending on the sequence $\{\Delta_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+$ that we choose. Every such accumulation point is given by some Q_λ, $\lambda \in [1/3, 2/3]$. Our proof of this example—again, see Appendix 4.B$_{177}$—works out the technical details of this observation. ♢

Observe, therefore, that the problem is essentially that the finite-difference expressions $1/\Delta(T_{t,x_u}^{t+\Delta} - I)$ and $1/\Delta(T_{t-\Delta,x_u}^{t} - I)$, parameterised in Δ, can have multiple accumulation points as we take Δ to 0. Therefore, it will be more convenient to instead work with what we call *outer partial derivatives*. These can be seen as a kind of set-valued derivatives, containing all these accumulation points obtained as Δ goes to zero.

Definition 4.11 (Directional Outer Partial Derivatives)**.** *For any stochastic process $P \in \mathbb{P}$ with corresponding family of history-dependent transition matrices (T_{t,x_u}^{s}), any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the* right-sided outer partial derivative *of T_{t,x_u}^{t} is defined by*

$$\overline{\partial}_+ T_{t,x_u}^{t} := \left\{ Q \in \mathscr{R} \,\middle|\, \exists \{\Delta_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+ : \lim_{i\to+\infty} \frac{1}{\Delta_i}(T_{t,x_u}^{t+\Delta_i} - I) = Q \right\} \quad (4.22)$$

and, if $t \neq 0$, the left-sided outer partial derivative *of T_{t,x_u}^{t} is defined by*

$$\overline{\partial}_- T_{t,x_u}^{t} := \left\{ Q \in \mathscr{R} \,\middle|\, \exists \{\Delta_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+ : \lim_{i\to+\infty} \frac{1}{\Delta_i}(T_{t-\Delta_i,x_u}^{t} - I) = Q \right\}. \quad (4.23)$$

Furthermore, the outer partial derivative *of T_{t,x_u}^{t} is defined as*

$$\overline{\partial} T_{t,x_u}^{t} := \overline{\partial}_+ T_{t,x_u}^{t} \cup \overline{\partial}_- T_{t,x_u}^{t} \text{ if } t > 0 \text{ and } \overline{\partial} T_{t,x_u}^{t} := \overline{\partial}_+ T_{t,x_u}^{t} \text{ if } t = 0.$$

For a given stochastic process $P \in \mathbb{P}$, we collect these outer partial derivatives in the multi-index family $(\overline{\partial} T_{t,x_u}^{t})$, with $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$, which we call the *family of outer partial derivatives corresponding to P*. The so-called *directional outer partial derivatives*—i.e. the right- and left-sided outer partial derivatives in Definition 4.11—are similarly collected in corresponding families $(\overline{\partial}_+ T_{t,x_u}^{t})$ and $(\overline{\partial}_- T_{t,x_u}^{t})$. For

well-behaved processes $P \in \mathbb{P}^{\mathrm{W}}$, as our next result shows, these corresponding outer partial derivatives are always non-empty and compact.

Proposition 4.22. *Consider any $P \in \mathbb{P}^{\mathrm{W}}$ with corresponding families of (directional) outer partial derivatives $(\overline{\partial} T_{t,x_u}^t)$, $(\overline{\partial}_+ T_{t,x_u}^t)$, and $(\overline{\partial}_- T_{t,x_u}^t)$. Then, for all $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$, $\overline{\partial}_+ T_{t,x_u}^t$, $\overline{\partial}_- T_{t,x_u}^t$ and $\overline{\partial} T_{t,x_u}^t$ are non-empty and compact subsets of $\mathscr{R}$.*

Proof. We only give the proof for $\overline{\partial}_+ T_{t,x_u}^t$. The proof for $\overline{\partial}_- T_{t,x_u}^t$ is completely analogous. The proof for $\overline{\partial} T_{t,x_u}^t$ then follows trivially because a union of two non-empty and compact sets is always non-empty and compact itself.

We start by establishing the boundedness of $\overline{\partial}_+ T_{t,x_u}^t$. Since P is well-behaved, it follows from Proposition 4.2_{149} that there are some $B > 0$ and $\delta > 0$ such that

$$(\forall 0 < \Delta < \delta) \ \left\| \frac{1}{\Delta}(T_{t,x_u}^{t+\Delta} - I) \right\| = \frac{1}{\Delta} \left\| (T_{t,x_u}^{t+\Delta} - I) \right\| \leq B. \tag{4.24}$$

Consider now any $Q \in \overline{\partial}_+ T_{t,x_u}^t$. Because of Equation (4.22), Q is the limit of a sequence of matrices $\{Q_k\}_{k \in \mathbb{Z}_{>0}}$, defined by

$$Q_k := \frac{1}{\Delta_k}(T_{t,x_u}^{t+\Delta_k} - I) \text{ for all } k \in \mathbb{Z}_{>0}. \tag{4.25}$$

Because of Equation (4.24), the norms $\|Q_k\|$ of these matrices are eventually (for large enough k) bounded above by B, and then also

$$\|Q\| = \|Q - Q_k + Q_k\| \leq \|Q - Q_k\| + \|Q_k\| \leq \|Q - Q_k\| + B,$$

from which it follows that also $\|Q\| \leq B$ since $\lim_{k \to +\infty} \|Q_k - Q\| = 0$. Since this is true for any $Q \in \overline{\partial}_+ T_{t,x_u}^t$, we find that $\overline{\partial}_+ T_{t,x_u}^t$ is bounded.

In order to prove that $\overline{\partial}_+ T_{t,x_u}^t$ is non-empty, we consider any sequence $\{\Delta_k\}_{k \in \mathbb{Z}_{>0}} \to 0^+$. The corresponding sequence of matrices $\{Q_k\}_{k \in \mathbb{Z}_{>0}}$, as defined by Equation (4.25), is then bounded because P is well-behaved—see Proposition 4.2_{149}—and therefore, it follows from Corollary $A.14_{378}$ that it has a convergent subsequence $\{Q_{k_i}\}_{i \in \mathbb{Z}_{>0}}$ whose limit we denote by Q^*. Hence, we have found a sequence $\{\Delta_{k_i}\}_{i \in \mathbb{Z}_{>0}} \to 0^+$ such that $\{Q_{k_i}\}_{i \in \mathbb{Z}_{>0}} \to Q^*$. Since we know from Lemma 4.10_{153} that each of the matrices in $\{Q_{k_i}\}_{i \in \mathbb{Z}_{>0}}$ is a rate matrix, by Proposition 4.6_{151}, the limit Q^* is also a rate matrix, which therefore belongs to $\overline{\partial}_+ T_{t,x_u}^t$.

We next show that $\overline{\partial}_+ T_{t,x_u}^t$ is closed, or equivalently, by Proposition $A.8_{376}$, that for any converging sequence $\{Q_k^*\}_{k \in \mathbb{Z}_{>0}}$ of rate matrices in $\overline{\partial}_+ T_{t,x_u}^t$, the limit $Q^* := \lim_{k \to +\infty} Q_k^*$ is again an element of $\overline{\partial}_+ T_{t,x_u}^t$. First, since each of the Q_k^* is a rate matrix, their limit Q^* is also a rate matrix by Proposition 4.6_{151}. Next, for any $k \in \mathbb{Z}_{>0}$, since $Q_k^* \in \overline{\partial}_+ T_{t,x_u}^t$, it follows from Equation $(4.22)_{168}$ that there is some $0 < \Delta_k < 1/k$ such that $\left\|Q_k - Q_k^*\right\| \leq 1/k$, with Q_k defined as in Equation $(4.25)_{\curvearrowleft}$. Consider now the sequences $\{Q_k\}_{k\in\mathbb{Z}_{>0}}$ and $\{\Delta_k\}_{k\in\mathbb{Z}_{>0}}$. Then on the one hand, we find that

$$\begin{aligned}0 \leq \limsup_{k\to+\infty} \|Q^* - Q_k\| &\leq \limsup_{k\to+\infty} \|Q^* - Q_k^*\| + \limsup_{k\to+\infty} \|Q_k^* - Q_k\| \\ &\leq \limsup_{k\to+\infty} \|Q^* - Q_k^*\| + \lim_{k\to+\infty} 1/k \\ &= \limsup_{k\to+\infty} \|Q^* - Q_k^*\| = 0,\end{aligned}$$

which implies that the sequence $\{Q_k\}_{k\in\mathbb{Z}_{>0}}$ converges to Q^*. On the other hand, we have that $\lim_{k\to+\infty} \Delta_k = 0$. Hence, because of Definition 4.11_{168}, it follows that $Q^* \in \overline{\partial}_+ T_{t,x_u}^t$. This implies that $\overline{\partial}_+ T_{t,x_u}^t$ is closed and, because we have already shown that it is bounded, it follows from Corollary A.12_{378} that $\overline{\partial}_+ T_{t,x_u}^t$ is compact. □

The following two examples provide this result with some intuition. Example 4.7 illustrates the validity of the result, while Example 4.8 shows that the requirement that P must be well-behaved is essential for the result to be true.

Example 4.7. Consider again the well-behaved stochastic process $P \in \mathbb{P}^{\mathrm{W}}$ from Example 4.6_{167} of which, for all $t \in (0,1]$, the corresponding transition matrix T_0^t is given by Equation $(4.18)_{165}$. As proved in Appendix 4.B_{177}, it holds for this particular process that $\overline{\partial}_+ T_0^0 = \left\{Q_\lambda : \lambda \in [1/3, 2/3]\right\}$ where, for every $\lambda \in [1/3, 2/3]$, $Q_\lambda := \lambda Q_1 + (1-\lambda) Q_2$ as in Example 4.6_{167}. ♢

Example 4.8. Fix any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$, let (T_t^s) be the transition matrix system of Example 4.5_{166}, and consider any stochastic process $P \in \mathbb{P}$ of which (T_t^s) is the corresponding family of transition matrices.

For now, we simply assume that such a process exists. A formal proof again requires some additional machinery—as in Example 4.6_{167}—and we therefore postpone it to Example 5.1_{184}, where we construct a continuous-time Markov chain whose corresponding family of transition matrices is equal to the transition matrix system (T_t^s).

As we prove in Appendix 4.B_{177}, for such a stochastic process P, the right-sided outer partial derivative $\overline{\partial}_+ T_0^0$ is empty. ♢

We end this section with two additional properties of the outer partial derivatives of well-behaved stochastic processes. First, as we establish in our next result, they satisfy an $\varepsilon - \delta$ expression that is similar to the limit expression of a partial derivative.

Proposition 4.23. *Consider any well-behaved stochastic process $P \in \mathbb{P}^{\mathrm{W}}$ with corresponding family of history-dependent transition matrices (T^s_{t,x_u}), and corresponding families of directional outer partial derivatives $(\overline{\partial}_+ T^t_{t,x_u})$ and $(\overline{\partial}_- T^t_{t,x_u})$. Then, for any $t \in \mathbb{R}_{\geq 0}$, any $u \in \mathscr{U}_{<t}$, any $x_u \in \mathscr{X}_u$, and any $\varepsilon > 0$, there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$:*

$$(\exists Q \in \overline{\partial}_+ T^t_{t,x_u}) \left\| \frac{1}{\Delta}(T^{t+\Delta}_{t,x_u} - I) - Q \right\| < \varepsilon \tag{4.26}$$

and, if $t \neq 0$,

$$(\exists Q \in \overline{\partial}_- T^t_{t,x_u}) \left\| \frac{1}{\Delta}(T^{t}_{t-\Delta,x_u} - I) - Q \right\| < \varepsilon. \tag{4.27}$$

Proof. Fix any $\varepsilon > 0$. Assume *ex absurdo* that

$$(\forall \delta > 0)(\exists \Delta \in (0,\delta))(\forall Q \in \overline{\partial}_+ T^t_{t,x_u}) \left\| \frac{1}{\Delta}(T^{t+\Delta}_{t,x_u} - I) - Q \right\| \geq \varepsilon.$$

Clearly, this implies the existence of a sequence $\{\Delta_k\}_{k\in\mathbb{Z}_{>0}} \to 0^+$ such that

$$\|Q_k - Q\| \geq \varepsilon \text{ for all } k \in \mathbb{Z}_{>0} \text{ and all } Q \in \overline{\partial}_+ T^t_{t,x_u}, \tag{4.28}$$

with Q_k defined as in Equation (4.25)$_{169}$. As we know from the proof of Proposition 4.22$_{169}$, the sequence $\{Q_k\}_{k\in\mathbb{Z}_{>0}}$ has a convergent subsequence $\{Q_{k_i}\}_{i\in\mathbb{Z}_{>0}}$ of which the limit Q^* belongs to $\overline{\partial}_+ T^t_{t,x_u}$. On the one hand, since $\lim_{i\to+\infty} Q_{k_i} = Q^*$, we now have that $\lim_{i\to+\infty} \|Q_{k_i} - Q^*\| = 0$. On the other hand, since $Q^* \in \overline{\partial}_+ T^t_{t,x_u}$, it follows from Equation (4.28) that $\lim_{i\to+\infty} \|Q_{k_i} - Q^*\| \geq \varepsilon > 0$. From this contradiction, it follows that there is some $\delta_1 > 0$ such that Equation (4.26) holds for all $0 < \Delta < \delta_1$. Similarly, using a completely analogous argument, we infer that if $t \neq 0$, there must be some $\delta_2 > 0$ such that Equation (4.27) holds for all $0 < \Delta < \delta_2$. Now let $\delta := \min\{\delta_1, \delta_2\}$ if $t \neq 0$ and let $\delta := \delta_1$ if $t = 0$. □

Secondly, these outer partial derivatives are a proper generalisation of directional partial derivatives. In particular, if the latter exist, their values correspond exactly to the single element of the former.

Corollary 4.24. *Consider any $P \in \mathbb{P}^{\mathrm{W}}$ with corresponding families of (directional) outer partial derivatives $(\overline{\partial} T^t_{t,x_u})$, $(\overline{\partial}_+ T^t_{t,x_u})$, and $(\overline{\partial}_- T^t_{t,x_u})$. Then, for all $t \in \mathbb{R}_{\geq 0}$, all $u \in \mathscr{U}_{<t}$, and all $x_u \in \mathscr{X}_u$, $\overline{\partial}_+ T^t_{t,x_u}$ is a singleton if and only if $\partial_+ T^t_{t,x_u}$ exists and, in that case, $\overline{\partial}_+ T^t_{t,x_u} = \{\partial_+ T^t_{t,x_u}\}$. Analogous results hold for $\overline{\partial}_- T^t_{t,x_u}$ and $\partial_- T^t_{t,x_u}$, and for $\overline{\partial} T^t_{t,x_u}$ and $\partial T^t_{t,x_u}$.*

Proof. We only give the proof for $\overline{\partial}_+ T^t_{t,x_u}$; the other claims follow completely analogously.

So, for the first direction, suppose that $\overline{\partial}_+ T_{t,x_u}^t = \{Q\}$ is a singleton, and fix any $\epsilon > 0$. Due to Proposition 4.23$_{\curvearrowleft}$, there is then some $\delta > 0$ such that, for all $0 < \Delta < \delta$, there is some $Q_\Delta \in \overline{\partial}_+ T_{t,x_u}^t$ such that $\left\| {}^1\!/\!_\Delta (T_{t,x_u}^{t+\Delta} - I) - Q_\Delta \right\| < \epsilon$. Because $\overline{\partial}_+ T_{t,x_u}^t$ is a singleton, this means that $Q_\Delta = Q$ for every such Δ, and hence it follows that $\lim_{\Delta \to 0^+} \left\| {}^1\!/\!_\Delta (T_{t,x_u}^{t+\Delta} - I) - Q \right\| = 0$. This means that $\lim_{\Delta \to 0^+} {}^1\!/\!_\Delta (T_{t,x_u}^{t+\Delta} - I)$ exists and equals Q which, by Definition 4.10$_{167}$, means that $\partial_+ T_{t,x_u}^t = Q$. This concludes the proof in the first direction.

For the other direction, suppose that $\partial_+ T_{t,x_u}^t$ exists, and for notational brevity write $Q := \partial_+ T_{t,x_u}^t$. Then, as in Definition 4.10$_{167}$, Q is a rate matrix and $\lim_{\Delta \to 0^+} {}^1\!/\!_\Delta (T_{t,x_u}^{t+\Delta} - I) = Q$. Now fix any sequence $\{\Delta_k\}_{k \in \mathbb{Z}_{>0}} \to 0^+$. Then it holds that $\lim_{k \to +\infty} {}^1\!/\!_{\Delta_k} (T_{t,x_u}^{t+\Delta_k} - I) = Q$ which implies that $Q \in \overline{\partial}_+ T_{t,x_u}^t$ by Definition 4.11$_{168}$. Because any such sequence $\{\Delta_k\}_{k \in \mathbb{Z}_{>0}} \to 0^+$ yields the same limit Q, it follows that $\overline{\partial}_+ T_{t,x_u}^t$ is a singleton, and in particular, that $\overline{\partial}_+ T_{t,x_u}^t = \{Q\}$. $\square$

Appendix

4.A Proofs of Examples in Section 4.5

Proof of Example 4.3$_{164}$. We will show that the sequence $\{({}^{i}T_t^s)\}_{i\in\mathbb{Z}_{\geq 0}}$ defined in Equation (4.17)$_{164}$ is Cauchy, or in other words, that

$$(\forall \varepsilon \in \mathbb{R}_{>0})\,(\exists n_\varepsilon \in \mathbb{Z}_{>0})\,(\forall k,\ell > n_\varepsilon)\; d\big(({}^{k}T_t^s),({}^{\ell}T_t^s)\big) < \varepsilon. \tag{4.29}$$

In order to prove this, the first step is to notice that for any $i \in \mathbb{Z}_{>0}$, the difference between $({}^{i}T_t^s)$ and $({}^{i-1}T_t^s)$ is essentially situated on the interval $[0,\delta_i]$. It should therefore be intuitively clear that $d\big(({}^{i}T_t^s),({}^{i-1}T_t^s)\big)$ is proportional to δ_i.

In fact, it holds that $d\big(({}^{i}T_t^s),({}^{i-1}T_t^s)\big) \leq \delta_i \|Q_i - Q_{i-1}\|$; we will start by proving this inequality. So, fix any $i \in \mathbb{Z}_{>0}$ and any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$. We now consider three cases. The first case is $t \geq \delta_i$. It then follows from Equation (4.17)$_{164}$ that $\big\|{}^{i}T_t^s - {}^{i-1}T_t^s\big\| = 0$. The second case is $s \leq \delta_i$. It then follows from Equation (4.17)$_{164}$ and Lemma B.11$_{395}$ that

$$\big\|{}^{i}T_t^s - {}^{i-1}T_t^s\big\| = \Big\|e^{Q_i(s-t)} - e^{Q_{i-1}(s-t)}\Big\| \leq (s-t)\|Q_i - Q_{i-1}\| \leq \delta_i \|Q_i - Q_{i-1}\|.$$

The third case is $t \leq \delta_i \leq s$. We then find that

$$\begin{aligned}\big\|{}^{i}T_t^s - {}^{i-1}T_t^s\big\| &= \Big\|{}^{i}T_t^{\delta_i}\,{}^{i}T_{\delta_i}^s - {}^{i-1}T_t^{\delta_i}\,{}^{i-1}T_{\delta_i}^s\Big\| \\ &\leq \Big\|{}^{i}T_t^{\delta_i} - {}^{i-1}T_t^{\delta_i}\Big\| + \Big\|{}^{i}T_{\delta_i}^s - {}^{i-1}T_{\delta_i}^s\Big\| \\ &= \Big\|{}^{i}T_t^{\delta_i} - {}^{i-1}T_t^{\delta_i}\Big\| \leq \delta_i \|Q_i - Q_{i-1}\|,\end{aligned}$$

where the first inequality follows from Lemma B.5$_{393}$, the second equality follows from the first case above, and the last inequality follows from the second case above. Hence, in all three cases, we find that $\big\|{}^{i}T_t^s - {}^{i-1}T_t^s\big\| \leq \delta_i \|Q_i - Q_{i-1}\|$. Since this inequality holds for any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, it now follows from Equation (4.15)$_{162}$ that $d\big(({}^{i}T_t^s),({}^{i-1}T_t^s)\big) \leq \delta_i \|Q_i - Q_{i-1}\|$.

Using this inequality, and because d is a metric, Equation (4.29) can now easily be proven. It suffices to choose n_ε in such a way that $c2^{-n_\varepsilon} < \varepsilon$. Indeed, in that case, for any $k,\ell > n_\varepsilon$, if we assume—without loss of generality—that $k \leq \ell$, it follows that

$$d\big(({}^{k}T_t^s),({}^{\ell}T_t^s)\big) \leq \sum_{i=k+1}^{\ell} d\big(({}^{i-1}T_t^s),({}^{i}T_t^s)\big) \leq \sum_{i=k+1}^{\ell} \delta_i \|Q_i - Q_{i-1}\| \leq c \sum_{i=k+1}^{\ell} \delta_i$$

and therefore, since we also know that

$$\sum_{i=k+1}^{\ell} \delta_i \leq \sum_{i=k+1}^{+\infty} \delta_i = \sum_{i=k+1}^{+\infty} 2^{-i} = 2^{-k},$$

it follows that $d\big(({}^kT_t^s),({}^\ell T_t^s)\big) \leq c2^{-k} \leq c2^{-n_\varepsilon} < \varepsilon$, as required.

As a side result, we also obtain a similar bound on the distance between $({}^kT_t^s)$ and (T_t^s). For any $\ell \geq k$, we know that

$$d\big(({}^kT_t^s),(T_t^s)\big) \leq d\big(({}^kT_t^s),({}^\ell T_t^s)\big) + d\big(({}^\ell T_t^s),(T_t^s)\big) \leq c2^{-k} + d\big(({}^\ell T_t^s),(T_t^s)\big),$$

and hence, since $(T_t^s) = \lim_{\ell\to+\infty}({}^\ell T_t^s)$, we find $d\big(({}^kT_t^s),(T_t^s)\big) \leq c2^{-k}$. □

Proof of Example 4.4$_{165}$. Let $(S_0^t)_{t\in(0,1]}$ be the family of transition matrices S_0^t that satisfy Equation (4.18)$_{165}$, i.e. let

$$S_0^t := e^{Q_1\varphi_1(t)+Q_2\varphi_2(t)} \quad \text{for all } t \in (0,1]. \tag{4.30}$$

We will first show that these matrices coincide with the matrices T_0^t corresponding to the limiting transition matrix system (T_t^s). To this end, we first establish some properties. Consider any $t \in (0,1]$ and let j be the unique element of $\mathbb{Z}_{\geq 0}$ such that $\delta_{j+1} < t \leq \delta_j$. If j is odd, it follows from Equations (4.19)$_{165}$ and (4.20)$_{165}$ that

$$\varphi_1(t) = t - {}^2\!/\!{}_3\delta_{j+1} = (t-\delta_{j+1}) + {}^1\!/\!{}_3\delta_{j+1} = (t-\delta_{j+1}) + \varphi_1(\delta_{j+1})$$

and $\varphi_2(t) = {}^2\!/\!{}_3\delta_{j+1} = \varphi_2(\delta_{j+1})$. Similarly, if j is even, it follows that $\varphi_1(t) = \varphi_1(\delta_{j+1})$ and $\varphi_2(t) = \varphi_2(\delta_{j+1}) + (t-\delta_{j+1})$. Hence, in both cases, we have that

$$Q_1\varphi_1(t) + Q_2\varphi_2(t) = Q_1\varphi_1(\delta_{j+1}) + Q_2\varphi_2(\delta_{j+1}) + Q_j(t-\delta_{j+1}).$$

Therefore, and since Q_1 and Q_2 commute, it now follows from Lemma 4.13$_{155}$ and Equation (4.30) that

$$\begin{aligned} S_0^t = e^{Q_1\varphi_1(t)+Q_2\varphi_2(t)} &= e^{Q_1\varphi_1(\delta_{j+1})+Q_2\varphi_2(\delta_{j+1})+Q_j(t-\delta_{j+1})} \\ &= e^{Q_1\varphi_1(\delta_{j+1})+Q_2\varphi_2(\delta_{j+1})}e^{Q_j(t-\delta_{j+1})} = S_0^{\delta_{j+1}}e^{Q_j(t-\delta_{j+1})}. \end{aligned} \tag{4.31}$$

For large enough $k \in \mathbb{Z}_{>0}$, a similar statement holds for the transition matrix ${}^kT_0^t$ that corresponds to $({}^kT_t^s)$. In particular, Lemma 4.13$_{155}$ together with Equation (4.17)$_{164}$ imply that

$${}^kT_0^t = {}^kT_0^{\delta_{j+1}}e^{Q_j(t-\delta_{j+1})} \quad \text{for all } k \geq j. \tag{4.32}$$

Hence, for all $j \in \mathbb{Z}_{\geq 0}$, by choosing $t = \delta_j$, and because $\delta_j - \delta_{j+1} = \delta_{j+1}$, it follows that

$$S_0^{\delta_j} = S_0^{\delta_{j+1}}e^{Q_j\delta_{j+1}} \;\text{ and }\; {}^kT_0^{\delta_j} = {}^kT_0^{\delta_{j+1}}e^{Q_j\delta_{j+1}} \;\text{ for all } k \geq j. \tag{4.33}$$

Finally, for all $k \in \mathbb{Z}_{\geq 0}$, it follows from Equations $(4.19)_{165}$ and $(4.20)_{165}$, and Equations (4.30) and $(4.17)_{164}$ that

$$S_0^{\delta_k} = e^{Q_k 2/3 \delta_k + Q_{k+1} 1/3 \delta_k} \text{ and } {}^kT_0^{\delta_k} = e^{Q_k \delta_k} \text{ for all } k \in \mathbb{Z}_{\geq 0}. \tag{4.34}$$

Using these properties, the remainder of the proof is now relatively easy. Consider any $t \in (0,1]$, let i be the unique element of $\mathbb{Z}_{\geq 0}$ such that $\delta_{i+1} < t \leq \delta_i$, and fix any $k > i$. Then on the one hand, we find that

$$\begin{aligned}
\left\| S_0^t - {}^kT_0^t \right\| &= \left\| S_0^{\delta_{i+1}} e^{Q_i(t-\delta_{i+1})} - {}^kT_0^{\delta_{i+1}} e^{Q_i(t-\delta_{i+1})} \right\| \\
&\leq \left\| S_0^{\delta_{i+1}} - {}^kT_0^{\delta_{i+1}} \right\| + \left\| e^{Q_i(t-\delta_{i+1})} - e^{Q_i(t-\delta_{i+1})} \right\| \\
&= \left\| S_0^{\delta_{i+1}} - {}^kT_0^{\delta_{i+1}} \right\| = \left\| S_0^{\delta_{i+2}} e^{Q_{i+1}\delta_{i+2}} - {}^kT_0^{\delta_{i+2}} e^{Q_{i+1}\delta_{i+2}} \right\| \\
&\leq \left\| S_0^{\delta_{i+2}} - {}^kT_0^{\delta_{i+2}} \right\| + \left\| e^{Q_{i+1}\delta_{i+2}} - e^{Q_{i+1}\delta_{i+2}} \right\| \\
&= \left\| S_0^{\delta_{i+2}} - {}^kT_0^{\delta_{i+2}} \right\| \leq \cdots \leq \left\| S_0^{\delta_k} - {}^kT_0^{\delta_k} \right\|,
\end{aligned}$$

where the first equality follows from Equations (4.31) and (4.32), the first inequality follows from Lemma B.5_{393} in Appendix B_{391}, the third equality follows from Equation (4.33), the second equality is again due to Lemma B.5_{393}, and the remaining steps consist in repeating the last steps over and over again. On the other hand, we also know that

$$\begin{aligned}
\left\| S_0^{\delta_k} - {}^kT_0^{\delta_k} \right\| &= \left\| e^{Q_k 2/3\delta_k + Q_{k+1} 1/3\delta_k} - e^{Q_k\delta_k} \right\| \\
&= \left\| e^{Q_k 2/3\delta_k} e^{Q_{k+1} 1/3\delta_k} - e^{Q_k 2/3\delta_k} e^{Q_k 1/3\delta_k} \right\| \\
&\leq \left\| e^{Q_k 2/3\delta_k} - e^{Q_k 2/3\delta_k} \right\| + \left\| e^{Q_{k+1} 1/3\delta_k} - e^{Q_k 1/3\delta_k} \right\| \\
&= \left\| e^{Q_{k+1} 1/3\delta_k} - e^{Q_k 1/3\delta_k} \right\| \leq \frac{\delta_k}{3} \left\| Q_k - Q_{k+1} \right\| = \frac{\delta_k}{3} \left\| Q_1 - Q_2 \right\|,
\end{aligned}$$

where the first equality follows from Equation (4.34), the second equality follows from Lemma 4.13_{155} because Q_1 and Q_2 commute, and the two inequalities follow from Lemmas B.5_{393} and B.11_{395} in Appendix B_{391}. Hence, we find that $\left\| S_0^t - {}^kT_0^t \right\| \leq \delta_k/3 \left\| Q_1 - Q_2 \right\|$. Since this is true for any $k > i$, it follows that $\lim_{k \to +\infty} {}^kT_0^t = S_0^t$. Therefore, and because $(T_t^s) := \lim_{i \to \infty} ({}^iT_t^s)$, we can conclude that $S_0^t = T_0^t$ for all $t \in (0,t]$.

We end this proof by showing that the transition matrix system (T_t^s) is well-behaved. Let $M := \max\{\|Q_1\|, \|Q_2\|\}$. We will prove that

$$\frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| \leq M \text{ for all } t, \Delta \in \mathbb{R}_{\geq 0}.$$

According to Definition 4.7_{157}, this clearly implies that (T_t^s) is well-behaved.

So fix any $t, \Delta \in \mathbb{R}_{\geq 0}$ and consider any $\varepsilon > 0$. Then since $\lim_{i \to +\infty} {}^iT_t^{t+\Delta} = T_t^{t+\Delta}$, there is some $j \in \mathbb{Z}_{>0}$ such that $\left\| T_t^{t+\Delta} - {}^jT_t^{t+\Delta} \right\| \leq \varepsilon$. Furthermore, since Q_1 and Q_2 commute, it follows from Equation $(4.17)_{164}$ that there are $\Delta_1, \Delta_2 \in \mathbb{R}_{\geq 0}$ such that $\Delta_1 + \Delta_2 = \Delta$ and ${}^jT_t^{t+\Delta} = e^{Q_1\Delta_1} e^{Q_2\Delta_2}$. Therefore, we find that

$$\begin{aligned}
\left\| T_t^{t+\Delta} - I \right\| &\leq \left\| T_t^{t+\Delta} - {}^jT_t^{t+\Delta} \right\| + \left\| e^{Q_1\Delta_1} e^{Q_2\Delta_2} - I \right\| \\
&\leq \varepsilon + \left\| e^{Q_1\Delta_1} - I \right\| + \left\| e^{Q_2\Delta_2} - I \right\| \\
&\leq \varepsilon + \Delta_1 \left\| Q_1 \right\| + \Delta_2 \left\| Q_2 \right\| \leq \Delta M,
\end{aligned}$$

where the second and third inequalities follow from Lemmas B.5_{393} and B.10_{394} in Appendix B_{391}, respectively. □

Proof of Example 4.5_{166}. We will prove that the transition matrix system (T_t^s) from Example 4.5_{166} is not well-behaved. In order to do this, we first fix any $n \in \mathbb{Z}_{\geq 0}$ and any $\Delta \in (0, \delta_n]$, and we let $i \geq n$ be the unique element of $\mathbb{Z}_{\geq 0}$ such that $\delta_{i+1} < \Delta \leq \delta_i$. It then follows from Equation $(4.17)_{164}$ that ${}^iT_0^\Delta = e^{Q_i\Delta}$, and therefore, we find that

$$\begin{aligned}
\left\| \Delta Q_i \right\| &\leq \left\| e^{Q_i\Delta} - (I + \Delta Q_i) \right\| + \left\| T_0^\Delta - {}^iT_0^\Delta \right\| + \left\| T_0^\Delta - I \right\| \\
&\leq \Delta^2 \left\| Q_i \right\|^2 + 2^{-i} + \left\| T_0^\Delta - I \right\|,
\end{aligned}$$

where the second inequality holds because of Lemma B.8_{394} in Appendix B_{391} and because—as proved at the end of the proof of Example 4.3_{164}—$d\big(({}^iT_t^s), (T_t^s)\big) \leq c2^{-i} = 2^{-i}$. Hence, since $\left\| Q_i \right\| = \left\| iQ \right\| = i \left\| Q \right\| = i$, we find that

$$\frac{1}{\Delta} \left\| T_0^\Delta - I \right\| \geq \left\| Q_i \right\| - \Delta \left\| Q_i \right\|^2 - \frac{1}{\Delta} 2^{-i} = i - \Delta i^2 - \frac{1}{\Delta} 2^{-i} \geq i - \delta_i i^2 - \frac{1}{\delta_{i+1}} 2^{-i}$$

and therefore, because $\delta_i = 2^{-i}$, $\delta_{i+1} = 2^{-i-1}$, $2^i \geq i$ and $i \geq n$, it follows that

$$\frac{1}{\Delta} \left\| T_0^\Delta - I \right\| \geq i - 2^{-i} i^2 - 2 \geq i - \frac{1}{2} i - 2 = \frac{i}{2} - 2 \geq \frac{n}{2} - 2. \tag{4.35}$$

Since $\Delta \in (0, \delta_n]$ is arbitrary, this inequality immediately implies that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_0^\Delta - I \right\| \geq \frac{n}{2} - 2.$$

and therefore, since this is true for every $n \in \mathbb{Z}_{\geq 0}$, we infer from Definition 4.7_{157} that (T_t^s) is not well-behaved, because the definition clearly fails for $t = 0$. □

4.B Proofs of Examples in Section 4.6

Proof of Example 4.6$_{167}$. We will prove that Equation (4.21)$_{167}$ indeed holds. So fix any $\lambda \in [1/3, 2/3]$ and consider the sequence $\{\Delta_i\}_{i\in\mathbb{Z}_{\geq 0}} \to 0^+$ whose elements are defined by $\Delta_i := (2\delta_{2i+1})/(3\lambda)$. For all $i \in \mathbb{Z}_{\geq 0}$, we then find that

$$\begin{aligned}\varphi_1(\Delta_i)Q_1 + \varphi_2(\Delta_i)Q_2 &= \frac{2}{3}\delta_{2i+1}Q_1 + (\Delta_i - \frac{2}{3}\delta_{2i+1})Q_2 \\ &= \lambda\Delta_i Q_1 + (1-\lambda)\Delta_i Q_2 = Q_\lambda \Delta_i,\end{aligned}$$

where the first equality follows from Equations (4.19)$_{165}$ and (4.20)$_{165}$, because $1/(3\lambda) \in [1/2, 1]$ implies that $\delta_{2i+1} \leq \Delta_i \leq \delta_{2i}$. Hence, for all $i \in \mathbb{Z}_{\geq 0}$, Equation (4.18)$_{165}$ now tells us that $T_0^{\Delta_i} = e^{Q_\lambda \Delta_i}$. Therefore, and because $\{\Delta_i\}_{i\in\mathbb{Z}_{\geq 0}} \to 0^+$, we find that indeed, as required,

$$\lim_{i\to+\infty} \frac{1}{\Delta_i}(T_0^{\Delta_i} - I) = \lim_{i\to+\infty} \frac{1}{\Delta_i}(e^{Q_\lambda\Delta_i} - I) = \frac{\mathrm{d}}{\mathrm{d}t} e^{Q_\lambda t}\Big|_{t=0} = Q_\lambda, \tag{4.36}$$

where we use Lemma 4.14$_{155}$ to establish the last equality. ☐

Proof of Example 4.7$_{170}$. Showing that $\{Q_\lambda : \lambda \in [1/3, 2/3]\}$ is a subset of $\overline{\partial}_+ T_0^0$ was, essentially, already done in Example 4.6$_{167}$, because for every $\lambda \in [1/3, 2/3]$, it follows from Equation (4.36) and Definition 4.10$_{167}$ that $Q_\lambda \in \overline{\partial}_+ T_0^0$. Therefore, we only need to show that $\overline{\partial}_+ T_0^0$ is a subset of $\{Q_\lambda : \lambda \in [1/3, 2/3]\}$, or equivalently, we need to show that for every $Q \in \overline{\partial}_+ T_0^0$, there is some $\lambda \in [1/3, 2/3]$ such that $Q = Q_\lambda$.

So consider any $Q \in \overline{\partial}_+ T_0^0$. Definition 4.11$_{168}$ then implies the existence of a sequence $\{\Delta_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+$ such that

$$\lim_{i\to+\infty} \frac{1}{\Delta_i}(T_0^{\Delta_i} - I) = Q. \tag{4.37}$$

The first step of the proof is to observe that, as we already noted in Example 4.4$_{165}$, for every $t \in (0, 1]$ it holds that $\varphi_1(t) + \varphi_2(t) = t$. In particular, it follows from Equations (4.19)$_{165}$ and (4.20)$_{165}$ that

$$\varphi_1(t) + \varphi_2(t) = t - 2/3\delta_{i+1} + 2/3\delta_{i+1} = t,$$

with $i \in \mathbb{Z}_{\geq 0}$ such that $\delta_{i+1} \leq t \leq \delta_i$.

Next, let us show that $t1/3 \leq \varphi_1(t) \leq t2/3$ for all $t \in (0, 1]$. To this end, first note that $\delta_i = 2\delta_{i+1}$ for all $i \in \mathbb{Z}_{\geq 0}$, and hence it follows that if $\delta_{i+1} \leq t \leq \delta_i$ then

$$t1/3 \leq 2/3\delta_{i+1} \leq t2/3. \tag{4.38}$$

Hence, if $\delta_{i+1} \leq t \leq \delta_i$ with i even, then it follows from Equation (4.19)$_{165}$ that $\varphi_1(t) = 2/3\delta_{i+1}$, and the desired inequalities follow from Equation $(4.38)_{\curvearrowleft}$. On the other hand, if i is odd then it follows from Equation $(4.20)_{165}$ that $\varphi_2(t) = {}^2/_3\delta_{i+1}$, and because we know that $\varphi_1(t) + \varphi_2(t) = t$, the inequalities again follow from Equation $(4.38)_{\curvearrowleft}$.

Given these observations, it follows that for all $t \in (0,1]$, there is some $\lambda_t \in [{}^1/_3, {}^2/_3]$ such that $\varphi_1(t) = \lambda_t t$ and $\varphi_2(t) = (1-\lambda_t)t$. Hence in particular, for large enough $i \in \mathbb{Z}_{>0}$ it holds that $\Delta_i \in (0,1]$ because $\{\Delta_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+$, and then there is some $\lambda_i \in [{}^1/_3, {}^2/_3]$ such that $\varphi_1(\Delta_i) = \lambda_i \Delta_i$ and $\varphi_2(\Delta_i) = (1-\lambda_i)\Delta_i$ and therefore also, due to Equation $(4.18)_{165}$, $T_0^{\Delta_i} = e^{Q_{\lambda_i}\Delta_i}$. In this way we obtain a sequence $\{\lambda_i\}_{i\in\mathbb{Z}_{>0}}$ in $[{}^1/_3, {}^2/_3]$; to obtain the elements λ_i for which $\Delta_i > 1$, simply set $\lambda_i := {}^1/_3$.

Furthermore, because of the Bolzano-Weierstrass theorem, the sequence $\{\lambda_i\}_{i\in\mathbb{Z}_{>0}}$ has a convergent subsequence $\{\lambda_{i_k}\}_{k\in\mathbb{Z}_{>0}}$ whose limit $\lambda := \lim_{k\to+\infty} \lambda_{i_k}$ clearly belongs to $[{}^1/_3, {}^2/_3]$.

In the remainder of this proof, we will show that $Q = Q_\lambda$. To this end, let us fix any $\varepsilon > 0$ and prove that $\|Q - Q_\lambda\| < \varepsilon$. First of all, since $\lambda = \lim_{k\to+\infty} \lambda_{i_k}$, there is some $n_1 \in \mathbb{Z}_{>0}$ such that, for all $k \geq n_1$, $\left|\lambda - \lambda_{i_k}\right| \|Q_1 - Q_2\| < {}^\varepsilon/_3$ and therefore also

$$\left\|Q_\lambda - Q_{\lambda_{i_k}}\right\| = \left\|(\lambda - \lambda_{i_k})(Q_1 - Q_2)\right\| = \left|\lambda - \lambda_{i_k}\right| \|Q_1 - Q_2\| < \frac{\varepsilon}{3}. \tag{4.39}$$

Secondly, Equation $(4.37)_{\curvearrowleft}$ implies that there is some $n_2 \in \mathbb{Z}_{>0}$ such that

$$\left\|\frac{1}{\Delta_{i_k}}\left(T_0^{\Delta_{i_k}} - I\right) - Q\right\| < \frac{\varepsilon}{3} \text{ for all } k \geq n_2. \tag{4.40}$$

Thirdly, Lemma B.8$_{394}$ in Appendix B$_{391}$ implies that for all $k \in \mathbb{Z}_{>0}$ it holds that

$$\left\|\frac{1}{\Delta_{i_k}}\left(e^{Q_{\lambda_{i_k}}\Delta_{i_k}} - I\right) - Q_{\lambda_{i_k}}\right\| \leq \Delta_{i_k}\left\|Q_{\lambda_{i_k}}\right\|^2. \tag{4.41}$$

Moreover, for any $k \in \mathbb{Z}_{>0}$, because $\lambda_{i_k} \in [{}^1/_3, {}^2/_3]$, it holds that

$$\begin{aligned}\left\|Q_{\lambda_{i_k}}\right\| &= \left\|\lambda_{i_k} Q_1 + (1-\lambda_{i_k})Q_2\right\| \\ &\leq \lambda_{i_k}\|Q_1\| + (1-\lambda_{i_k})\|Q_2\| \leq \max\{\|Q\|_1, \|Q_2\|\}.\end{aligned}$$

Because $\lim_{k\to+\infty} \Delta_{i_k} = 0$, this implies that there is some $n_3 \in \mathbb{Z}_{>0}$ such that, for all $k \in \mathbb{Z}_{>0}$ with $k \geq n_3$ it holds that $\Delta_{i_k}\|Q_{\lambda_{i_k}}\|^2 < {}^\varepsilon/_3$, and therefore also, using Equation (4.41), that

$$\left\|\frac{1}{\Delta_{i_k}}\left(e^{Q_{\lambda_{i_k}}\Delta_{i_k}} - I\right) - Q_{\lambda_{i_k}}\right\| < \frac{\varepsilon}{3} \text{ for all } k \geq n_3. \tag{4.42}$$

Finally, because $\lim_{k\to+\infty} \Delta_{i_k} = 0$, there is some $n_4 \in \mathbb{Z}_{>0}$ such that $\Delta_{i_k} \in (0,1]$ for all $k \geq n_4$.

Now consider any $k \geq \max\{n_1, n_2, n_3, n_4\}$. Then because $\Delta_{i_k} \in (0,1]$ it holds that $T_0^{\Delta_{i_k}} = e^{Q_{\lambda_{i_k}} \Delta_{i_k}}$, and hence it follows from Equations (4.39), (4.40), and (4.42) that

$$\begin{aligned}\|Q - Q_\lambda\| \leq &\left\| Q - \frac{1}{\Delta_{i_k}} \left(T_0^{\Delta_{i_k}} - I \right) \right\| + \left\| \frac{1}{\Delta_{i_k}} \left(e^{Q_{\lambda_{i_k}} \Delta_{i_k}} - I \right) - Q_{\lambda_{i_k}} \right\| \\ &\qquad\qquad + \left\| Q_{\lambda_{i_k}} - Q_\lambda \right\| \\ < &\ \varepsilon.\end{aligned}$$

Since this is true for every $\varepsilon > 0$, it follows that $\|Q - Q_\lambda\| = 0$, and therefore also, that $Q = Q_\lambda$, as desired. □

Proof of Example 4.8$_{170}$. We will prove that the right-sided outer partial derivative $\overline{\partial}_+ T_0^0$ is empty. To this end, assume *ex absurdo* that it is not empty, and consider any $Q \in \overline{\partial}_+ T_0^0$. It then follows from Equation (4.22)$_{168}$ that there is a sequence $\{\Delta_i\}_{i \in \mathbb{Z}_{>0}} \to 0^+$ such that $\lim_{i \to +\infty} {}^1/_{\Delta_i}(T_0^{\Delta_i} - I) = Q$. Consider now any $\varepsilon > 0$, any $n \in \mathbb{Z}_{\geq 0}$ such that $n \geq 6 + 2\varepsilon$, let $\delta_n := 2^{-n}$ as in Example 4.5$_{166}$, and consider any $i^* \in \mathbb{Z}_{>0}$ such that, for all $i \geq i^*$, $\Delta_i \in (0, \delta_n)$—such an i^* always exists because $\{\Delta_i\}_{i \in \mathbb{Z}_{>0}} \to 0^+$. For all $i \geq i^*$, using Equation (4.35)$_{176}$ from the proof of Example 4.5$_{166}$, and noting that $\|Q\| = 1$, we then find that

$$\left\| \frac{1}{\Delta_i}(T_0^{\Delta_i} - I) - Q \right\| \geq \left\| \frac{1}{\Delta_i}(T_0^{\Delta_i} - I) \right\| - \|Q\| \geq \frac{n}{2} - 3 \geq \varepsilon,$$

which, since $\varepsilon > 0$, contradicts the fact that $\lim_{i \to +\infty} {}^1/_{\Delta_i}(T_0^{\Delta_i} - I) = Q$. Hence, our assumption must be wrong, and it follows that $\overline{\partial}_+ T_0^0$ is indeed empty. □

5

Continuous-Time (Imprecise-)Markov Chains

"I am a leaf on the wind – watch how I soar."

Hoban 'Wash' Washburn, Serenity

In this chapter we finally come to the eponymous continuous-time imprecise-Markov chains that are, ultimately, the subject of this dissertation. We use the formalisms developed in Chapter 2$_{45}$, and the machinery from Chapter 4$_{143}$, to construct and analyse these models.

We start in Section 5.1$_{\curvearrowright}$ by formally defining continuous-time (precise) Markov chains, which, in analogy with discrete-time Markov chains, are stochastic processes whose future behaviour is independent of their past behaviour, given their current state. We study both homogeneous and non-homogeneous Markov chains, and show that some well-known properties from the literature also hold in our formalism.

Continuous-time imprecise-Markov chains are introduced formally in Section 5.2$_{188}$, as sets of stochastic processes whose (generalised) derivatives are contained in a given set of transition rate matrices. We consider three different types of such models, with increasing generality, which we all consider to be imprecise-Markov chains. We study some particular closure properties of these sets of models.

In Section 5.3$_{194}$, we consider the sets of transition matrices and sets of transition matrix systems that are induced by a given continuous-time imprecise-Markov chain. We characterise structural properties of these sets in terms of the structural properties of the set of transition rate matrices that define a continuous-time imprecise-Markov chain.

We conclude this chapter with Section 5.4198, where we introduce and analyse the lower and upper expectations corresponding to continuous-time imprecise-Markov chains, which, as we know from Chapter 129, are the inferences we are interested in. We derive characterisations of these lower (and upper) expectations in terms of the induced sets of transition matrices, and show that all three types of continuous-time imprecise-Markov chains introduced in Section 5.2188 satisfy imprecise-Markov properties, which motivates our terminology. We conclude by proving a *law of iterated lower expectations* for the most imprecise of our continuous-time imprecise-Markov chains, which is a crucial property for the development of efficient algorithms in Chapters 6259 and 7335.

5.1 Continuous-Time Markov Chains

Let us start by giving the formal definition of a (continuous-time) Markov chain.

Definition 5.1 (Markov Property, Markov Chain). *A stochastic process $P \in \mathbb{P}$ satisfies the* Markov property *if for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any time sequence $u \in \mathscr{U}_{<t}$, any $x_u \in \mathscr{X}_u$, and any states $x, y \in \mathscr{X}$:*

$$P(X_s = y \,|\, X_t = x, X_u = x_u) = P(X_s = y \,|\, X_t = x)\,.$$

A stochastic process that satisfies this property is called a Markov chain. *We denote the set of all Markov chains by $\mathbb{P}^{\mathrm{M}}$ and use $\mathbb{P}^{\mathrm{WM}}$ to refer to the subset that only contains the well-behaved Markov chains.*

We note that in the continuous-time setting, the *transition probabilities $P(X_s = y \,|\, X_t = x)$* are actually values that P assigns to the conditional event $(X_s = y, X_t = x) \in \mathscr{C}^{\mathrm{SP}}$. This is different from what we encountered in the definition of discrete-time Markov chains in Definition 3.389, where the analogous conditional events were not part of the domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ of discrete-time Markov chains. Other than that, the interpretation in the continuous-time setting is analogous to the discrete-time setting: a Markov chain is a stochastic process whose future behaviour is independent of its past behaviour, given the current state.

Moving on, we already know from Corollary 4.4150 that the transition matrices of a stochastic process—and therefore also, in particular, of a Markov chain—satisfy some simple properties. For the specific case of a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$, the family of transition matrices (T_t^s)

also satisfies an additional property. In particular, for any $t, r, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$, these transition matrices satisfy

$$T_t^s = T_t^r T_r^s. \tag{5.1}$$

In this context, this property is known as the *Chapman-Kolmogorov equation* or the semigroup property [68]. Indeed, this is the same semigroup property that we defined in Section 4.4_{156} to hold for transition matrix systems. The following result is therefore not surprising.

Proposition 5.1. *Consider a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ and let (T_t^s) be its corresponding family of transition matrices. Then (T_t^s) is a transition matrix system. Furthermore, (T_t^s) is well-behaved if and only if P is well-behaved.*

Proof. Consider any Markov chain $P \in \mathbb{P}^{\mathrm{M}}$, with (T_t^s) its corresponding family of transition matrices. Then, because P is a stochastic process, it follows from Corollary 4.4_{150} that $T_t^t = I$ for all $t \in \mathbb{R}_{\geq 0}$.

Consider now any $t, r, s \in \mathbb{R}_{\geq 0}$ with $t \leq r \leq s$. We need to show that $T_t^s = T_t^r T_r^s$. If $t = r$, we have that $T_t^s = T_t^r T_r^s = I T_r^s = T_t^s$, and hence the result follows trivially. Similarly, the claim is trivial for $r = s$. Hence, it remains to show that the claim holds for $t < r < s$. It follows from Definition 5.1 that for all $x_t, x_r, x_s \in \mathscr{X}$,

$$P(X_s = x_s \,|\, X_r = x_r, X_t = x_t) = P(X_s = x_s \,|\, X_r = x_r).$$

Furthermore, because P is a stochastic process, it follows from Corollary 2.20_{68} that P satisfies $\mathrm{F4}_{47}$ and $\mathrm{F3}_{47}$ on its domain. From $\mathrm{F4}_{47}$, we infer that

$$\begin{aligned} P(X_s = x_s, X_r = x_r \,|\, X_t = x_t) &= P(X_s = x_s \,|\, X_r = x_r, X_t = x_t) P(X_r = x_r \,|\, X_t = x_t) \\ &= P(X_s = x_s \,|\, X_r = x_r) P(X_r = x_r \,|\, X_t = x_t), \end{aligned}$$

where the second equality used the Markov property. From $\mathrm{F3}_{47}$, we infer that

$$P(X_s = x_s \,|\, X_t = x_t) = \sum_{x_r \in \mathscr{X}} P(X_s = x_s, X_r = x_r \,|\, X_t = x_t).$$

From Definition 4.3_{150} it now follows that, for any $x_t, x_s \in \mathscr{X}$,

$$\begin{aligned} T_t^s(x_t, x_s) &= P(X_s = x_s \,|\, X_t = x_t) \\ &= \sum_{x_r \in \mathscr{X}} P(X_s = x_s, X_r = x_r \,|\, X_t = x_t) \\ &= \sum_{x_r \in \mathscr{X}} P(X_s = x_s \,|\, X_r = x_r) P(X_r = x_r \,|\, X_t = x_t) \\ &= \sum_{x_r \in \mathscr{X}} T_t^r(x_t, x_r) T_r^s(x_r, x_s), \end{aligned}$$

and hence, by the properties of matrix multiplication, we find that $T_t^s = T_t^r T_r^s$. Therefore, and because the $t,r,s \in \mathbb{R}_{\geq 0}$ are arbitrary, (T_t^s) is a transition matrix system.

That (T_t^s) is well-behaved if and only if P is well-behaved follows immediately from Definition 4.7$_{157}$ and Proposition 4.2$_{149}$ because the Markov property implies that $T_{t,x_u}^{t+\Delta} = T_t^{t+\Delta}$ and $T_{t-\Delta,x_u}^t = T_{t-\Delta}^t$. □

At this point we know that every (well-behaved) Markov chain has a corresponding (well-behaved) transition matrix system. Our next result establishes that the converse is true as well: every (well-behaved) transition matrix system has a corresponding (well-behaved) Markov chain, and for a given initial distribution, this Markov chain is even unique. The proof of this statement is highly technical (and fairly cumbersome), so we have deferred it to Appendix 5.A$_{210}$.

Theorem 5.2. *Let p be any probability mass function on $\mathscr{X}$ and let (T_t^s) be a transition matrix system. Then there is a unique Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ with corresponding family of transition matrices (T_t^s) that satisfies $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$. Furthermore, P is well-behaved if and only if (T_t^s) is well-behaved.*

Hence, Markov chains—and well-behaved Markov chains in particular—are completely characterised by their transition matrices and their initial distributions. Our next example uses this result to formally establish the existence of the Markov chains that were used in Examples 4.6$_{167}$ and 4.8$_{170}$. Furthermore, it also illustrates that not every Markov chain is well-behaved.

Example 5.1. For any transition matrix system (T_t^s), it follows from Theorem 5.2—with p chosen arbitrarily—that there is a continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{M}} \subseteq \mathbb{P}$ with corresponding family of transition matrices (T_t^s) and, furthermore, that P is well-behaved if and only if (T_t^s) is well-behaved.

For example, for any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$, if we let (T_t^s) be the transition matrix system of Example 4.5$_{166}$, we find—as already claimed in Example 4.8$_{170}$—that there is a continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{M}} \subseteq \mathbb{P}$ with corresponding family of transition matrices (T_t^s). Furthermore, since we know from Example 4.5$_{166}$ that (T_t^s) is not well-behaved, it follows that P is not well-behaved either.

As another example, for any two commuting rate matrices $Q_1, Q_2 \in \mathscr{R}$, if we let (T_t^s) be the well-behaved transition matrix system of Example 4.4$_{165}$, we find—as already claimed in Example 4.6$_{167}$—that there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{WM}} \subseteq \mathbb{P}^{\mathrm{M}} \subseteq \mathbb{P}$ such that, for all $t \in (0,1]$, its transition matrix T_0^t is given by Equation (4.18)$_{165}$ in Example 4.4$_{165}$. ◇

The caveats surrounding the analogous Proposition 3.14₉₅ also apply here; this result is essentially well-known in other formalisms, and Theorem 5.2 should be understood as a translation of those results to our formalisation of continuous-time Markov chains using coherent conditional probabilities.

Let us conclude this section by noting that the Markov property not only simplifies the transition probabilities of a process, but also simplifies its dynamics.

Lemma 5.3. *Consider a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ with corresponding families of (history-dependent) transition matrices (T_{t,x_u}^s) and (T_t^s). Then, for all $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$, it holds that $T_{t,x_u}^s = T_t^s$.*

Proof. This is immediate from Definitions 4.2_{148}, 4.3_{150}, and 5.1_{182}. □

Proposition 5.4. *Consider a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$, with corresponding families of (directional) outer partial derivatives $(\overline{\partial} T_{t,x_u}^t)$, $(\overline{\partial}_+ T_{t,x_u}^t)$, and $(\overline{\partial}_- T_{t,x_u}^t)$. Then, for all $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$, it holds that $\overline{\partial}_+ T_{t,x_u}^t = \overline{\partial}_+ T_t^t$, $\overline{\partial}_- T_{t,x_u}^t = \overline{\partial}_- T_t^t$, and $\overline{\partial} T_{t,x_u}^t = \overline{\partial} T_t^t$.*

Proof. We only give the proof for $\overline{\partial}_+ T_{t,x_u}^t$; the proof for $\overline{\partial}_- T_{t,x_u}^t$ is completely analogous, and the proof for $\overline{\partial} T_{t,x_u}^t$ then follows immediately. We will show that $\overline{\partial}_+ T_{t,x_u}^t \subseteq \overline{\partial}_+ T_t^t$ and $\overline{\partial}_+ T_t^t \subseteq \overline{\partial}_+ T_{t,x_u}^t$.

So, first fix any $Q \in \overline{\partial}_+ T_{t,x_u}^t$. By Definition 4.11_{168}, there is then some sequence $\{\Delta_i\}_{i \in \mathbb{Z}_{>0}} \to 0^+$ such that $\lim_{i \to +\infty} {}^1\!/_{\Delta_i}(T_{t,x_u}^{t+\Delta_i} - I) = Q$. By Lemma 5.3 it holds that $T_{t,x_u}^{t+\Delta_i} = T_t^{t+\Delta_i}$ for all $i \in \mathbb{Z}_{>0}$, which means that also $\lim_{i \to +\infty} {}^1\!/_{\Delta_i}(T_t^{t+\Delta_i} - I) = Q$. By Definition 4.11_{168}, this means that $Q \in \overline{\partial}_+ T_t^t$. Since $Q \in \overline{\partial}_+ T_{t,x_u}^t$ is arbitrary it follows that $\overline{\partial}_+ T_{t,x_u}^t \subseteq \overline{\partial}_+ T_t^t$. The proof for the other direction is completely analogous. □

5.1.1 *Homogeneous Markov Chains*

We encountered *homogeneous* Markov chains in discrete time in Chapter 3_{83}. They were Markov chains for which the transition probabilities do not depend on the absolute moment in time that they are considered. The following definition introduces the analogous concept for continuous-time Markov chains.

Definition 5.2 (Homogeneous Markov chain)**.** *A Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ is called* time-homogeneous, *or simply* homogeneous, *if for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, and all $x, y \in \mathscr{X}$, it holds that*

$$P(X_s = y \,|\, X_t = x) = P(X_{s-t} = y \,|\, X_0 = x)\,.$$

We denote the set of all homogeneous Markov chains by $\mathbb{P}^{\mathrm{HM}}$ *and use* $\mathbb{P}^{\mathrm{WHM}}$ *to refer to the subset that consists of the well-behaved homogeneous Markov chains.*

Another way to put this is that a Markov chain is homogeneous if its transition matrices T_t^s only depend on the values of t and s through their difference $s-t$, i.e. if for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, it holds that

$$T_t^s = T_0^{s-t}\,. \tag{5.2}$$

Recall now from Chapter 4$_{143}$ the exponential transition matrix system $(e^{Q(s-t)})$ corresponding to some rate matrix $Q \in \mathscr{R}$. This family of transition matrices $(T_t^s) := (e^{Q(s-t)})$ therefore clearly satisfies Equation (5.2). Furthermore, by Proposition 4.16$_{158}$, $(e^{Q(s-t)})$ is well-behaved. Hence, we have the following result.

Corollary 5.5. *Consider any rate matrix* $Q \in \mathscr{R}$ *and let* p *be an arbitrary probability mass function on* $\mathscr{X}$*. Then there is a unique Markov chain* $P \in \mathbb{P}^{\mathrm{M}}$ *with corresponding family of transition matrices* (T_t^s) *such that* $(T_t^s) = (e^{Q(s-t)})$ *and, for all* $y \in \mathscr{X}$*,* $P(X_0 = y) = p(y)$*. Furthermore, this unique Markov chain is well-behaved and homogeneous.*

Proof. Since we know from Proposition 4.16$_{158}$ that $(e^{Q(s-t)})$ is a well-behaved transition matrix system, it follows from Theorem 5.2$_{184}$ that there is a unique Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ with corresponding family of transition matrices (T_t^s) such that $(T_t^s) = (e^{Q(s-t)})$ and, for all $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$, and that this Markov chain is furthermore well-behaved. Since it—trivially—follows from Definition 4.8$_{158}$ that $(T_t^s) = (e^{Q(s-t)})$ satisfies Equation (5.2), Definition 5.2$_{\curvearrowleft}$ implies that P is homogeneous. □

Our next result strengthens this connection between well-behaved homogeneous Markov chains and exponential transition matrix systems.[1] The proof can be found in Appendix 5.A$_{210}$.

Theorem 5.6. *For any well-behaved homogeneous Markov chain* $P \in \mathbb{P}^{\mathrm{WHM}}$ *with corresponding family of transition matrices* (T_t^s)*, there is a unique rate matrix* $Q \in \mathscr{R}$ *such that* $(T_t^s) = (e^{Q(s-t)})$*.*

[1] As before, although our proof for this result starts from scratch, this result is essentially well known. Our version of it should be regarded as a (re)formulation that is adapted to our terminology and notation and, in particular, to our use of coherent conditional probabilities.

By combining Corollary 5.5 and Theorem 5.6, we see that any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$ is completely characterised by its initial distribution and its rate matrix $Q \in \mathscr{R}$. We will denote this rate matrix by Q_P.

The dynamic behaviour of well-behaved homogeneous Markov chains is furthermore particularly easy to describe, as shown by the next result.

Proposition 5.7. *Consider any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$ and let $Q_P \in \mathscr{R}$ be its corresponding rate matrix. Then for all $t \in \mathbb{R}_{\geq 0}$ its corresponding (directional and/or outer) partial derivatives satisfy*

$$\partial T_t^t = \partial_+ T_t^t = Q_P \quad \text{and, if } t > 0, \quad \partial_- T_t^t = Q_P\,,$$

and

$$\overline{\partial} T_t^t = \overline{\partial}_+ T_t^t = \{Q_P\} \quad \text{and, if } t > 0, \quad \overline{\partial}_- T_t^t = \{Q_P\}\,.$$

Proof. The result about the partial derivatives is an immediate consequence of Lemma 4.14_{155} and the fact that $T_t^{t+\Delta} = e^{Q\Delta}$ and, if $t - \Delta \geq 0$, $T_{t-\Delta}^t = e^{Q\Delta}$. The result about the outer partial derivatives then follows from Corollary 4.24_{171}. □

5.1.2 *Non-Homogeneous Markov Chains*

In contrast with homogeneous Markov chains, a Markov chain for which Equation (5.2) does not hold is called—rather obviously—*non-homogeneous*. While we know from Theorem 5.6 that well-behaved homogeneous Markov chains can be characterised (up to an initial distribution) by a fixed rate matrix $Q \in \mathscr{R}$, this is not the case for well-behaved non-homogeneous Markov chains.

Instead, such systems are typically described by a function Q_t that gives for each time point $t \in \mathbb{R}_{\geq 0}$ a rate matrix $Q_t \in \mathscr{R}$. For any such function Q_t, the existence and uniqueness of a corresponding non-homogeneous Markov chain then depend on the specific properties of Q_t. Rather than attempt to treat all these different cases here, we instead refer to some examples from the literature.

Typically, some kind of continuity of Q_t in terms of t is assumed. The specifics of these assumptions may then depend on the intended generality of the results, computational considerations, the domain of application, and so forth. For example, Reference [1] assumes that Q_t is left-continuous and has bounded right-hand limits. As a stronger restriction, Reference [53] uses a collection $Q_1, \ldots, Q_n$ of commuting rate matrices, and defines Q_t as a weighted linear combination of these component rate matrices wherein the weights vary continuously with t. In Reference [88], a right-continuous and piecewise-constant Q_t is used, meaning that Q_t takes different values on various (half-open) intervals of $\mathbb{R}_{\geq 0}$, but fixed values within those intervals.

This idea of using a time-dependent rate matrix Q_t has the advantage of being rather intuitive, but it is rather difficult to formalise. Essentially, the problem with this approach is that it does not allow us to distinguish between left and right derivatives. Intuitively, Q_t is supposed to be "the" derivative. However, this is impossible if Q_t is discontinuous—for example in the piecewise constant case. Therefore, in this work, instead of explicitly using a function Q_t, we will characterise non-homogeneous Markov chains by means of their transition matrix system and their initial distribution, making use of the results in Proposition 5.1$_{183}$ and Theorem 5.2$_{184}$.

One technique for constructing transition matrix systems that is particularly important for our work, and especially in our proofs, is to combine restrictions of exponential transition matrix systems to form a new transition matrix system that is, loosely speaking, piecewise constant. Example 4.2$_{162}$ provided a simple illustration of this technique. More generally, these transition matrix systems will be of the form

$$\left(e^{Q_0(s-t)}\right)_{[0,t_0]} \otimes \left(e^{Q_1(s-t)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{Q_n(s-t)}\right)_{[t_{n-1},t_n]} \otimes \left(e^{Q_{n+1}(s-t)}\right)\big)_{[t_n,+\infty)} . \tag{5.3}$$

For example, the transition matrix systems $({}^i\mathcal{T}_t^s)$, $i \in \mathbb{Z}_{\geq 0}$, that we defined in Equation (4.17)$_{164}$ all have this form. As we know from Propositions 4.16$_{158}$ and 4.19$_{161}$, transition matrix systems that have this form are always well-behaved. The following is therefore a trivial consequence of Theorem 5.2$_{184}$.

Proposition 5.8. *Let p be a probability mass function on $\mathscr{X}$, let $u = t_0, \ldots, t_n$ be a finite sequence of time points in $\mathscr{U}_{\supset\emptyset}$, and let $Q_0, \ldots, Q_{n+1} \in \mathscr{Q}$ be a collection of rate matrices. Then there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{WM}}$ such that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$, and whose corresponding family of transition matrices is given by Equation* (5.3).

5.2 Continuous-Time Imprecise-Markov Chains

We are now finally ready to introduce the eponymous continuous-time imprecise-Markov chains that are the focus of this dissertation. In analogy with the concept of a discrete-time imprecise-Markov chain, as discussed in Chapter 3$_{83}$, they are sets of stochastic processes that are in a specific sense consistent with a given set of parameters.

Recall from Chapter 4$_{143}$ that for a given stochastic process $P \in \mathbb{P}$, its dynamics can be described by means of the outer partial derivatives

$\overline{\partial}T^{t}_{t,x_u}$ of its transition matrices, which can depend both on the time $t \in \mathbb{R}_{\geq 0}$ and on the history $x_u \in \mathscr{X}_u$. Furthermore, we also found that—at least for well-behaved processes—these outer partial derivatives are non-empty and compact sets of rate matrices. If all the outer partial derivatives of a process belong to the same non-empty set of rate matrices $\mathscr{Q}$, we say that this process is consistent with $\mathscr{Q}$.

Definition 5.3 (Consistency with a set of rate matrices)**.** *Consider a non-empty set of rate matrices $\mathscr{Q}$ and a stochastic process $P \in \mathbb{P}$ with corresponding family of outer partial derivatives $(\overline{\partial}T^{t}_{t,x_u})$. Then P is said to be* consistent *with $\mathscr{Q}$ if*

$$(\forall t \in \mathbb{R}_{\geq 0})(\forall u \in \mathscr{U}_{<t})(\forall x_u \in \mathscr{X}_u) : \overline{\partial}T^{t}_{t,x_u} \subseteq \mathscr{Q}.$$

If P is consistent with $\mathscr{Q}$, we will write $P \sim \mathscr{Q}$.

Hence, when a process is consistent with a set of rate matrices $\mathscr{Q}$, we know that its dynamics can always be described using rate matrices in that set. However, we do not know which of these rate matrices $Q \in \mathscr{Q}$ describe the dynamics at any given time $t \in \mathbb{R}_{\geq 0}$ or for any given history $x_u \in \mathscr{X}_u$. Furthermore, consistency of a process with a set of rate matrices $\mathscr{Q}$ does not tell us anything about the initial distribution of the process. Therefore, we also introduce the concept of consistency with a set of initial distributions $\mathscr{M}$.

Definition 5.4 (Consistency with a set of initial distributions)**.** *Consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$ and any stochastic process $P \in \mathbb{P}$. We then say that P is* consistent *with $\mathscr{M}$, and write $P \sim \mathscr{M}$, if the map $p : \mathscr{X} \to \mathbb{R} : x \mapsto P(X_0 = x)$ is in $\mathscr{M}$.*

In analogy with our developments in Chapter 383, we first introduce the notion of a consistent *subset* of a given set of processes. This will (again) allow us to consider various different types of continuous-time imprecise-Markov chains, using consistent notation and definitions.

Definition 5.5 (Consistent subset of processes)**.** *Consider a non-empty set of rate matrices $\mathscr{Q}$, a non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, and a set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$. Then, the* subset of $\mathscr{P}$ consistent with $\mathscr{Q}$ and $\mathscr{M}$ *is denoted by $\mathscr{P}_{\mathscr{Q},\mathscr{M}}$, and defined as*

$$\mathscr{P}_{\mathscr{Q},\mathscr{M}} := \left\{ P \in \mathscr{P} \,\middle|\, P \sim \mathscr{Q} \text{ and } P \sim \mathscr{M} \right\}.$$

When $\mathscr{M}$ is the set of all *probability mass functions on $\mathscr{X}$, we will write $\mathscr{P}_{\mathscr{Q}}$ for the sake of brevity.*

For some fixed $\mathscr{Q}$ and $\mathscr{M}$, different choices for $\mathscr{P}$ will result in different sets of consistent processes $\mathscr{P}_{\mathscr{Q},\mathscr{M}}$. Three specific choices of $\mathscr{P}$ will be particularly important in this work because, as we will now show, they lead to three different types of imprecise continuous-time Markov chains.

In particular, we have introduced three sets $\mathbb{P}^{\mathrm{W}}$, $\mathbb{P}^{\mathrm{WM}}$ and $\mathbb{P}^{\mathrm{WHM}}$ of well-behaved stochastic processes with different qualitative properties. $\mathbb{P}^{\mathrm{W}}$ is the set of all well-behaved stochastic processes, $\mathbb{P}^{\mathrm{WM}}$ consists of the processes in $\mathbb{P}^{\mathrm{W}}$ that are Markov chains, and $\mathbb{P}^{\mathrm{WHM}}$ is the set of all homogeneous Markov chains that are well-behaved, which is therefore a subset of $\mathbb{P}^{\mathrm{WM}}$. We now use Definition 5.5$_{\curvearrowleft}$ to define three sets of consistent processes that have these respective qualitative properties.

Definition 5.6 (Continuous-time imprecise-Markov chain)**.** *For any non-empty set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we define the following three sets of stochastic processes that are jointly consistent with $\mathscr{Q}$ and $\mathscr{M}$:*

- $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ *is the consistent set of all well-behaved stochastic processes;*
- $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ *is the consistent set of all well-behaved Markov chains;*
- $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ *is the consistent set of all well-behaved homogeneous Markov chains.*

We call each of these three sets a continuous-time imprecise-Markov chain, *and abbreviate this as CTIMC.*[2] *Following Definition 5.5$_{\curvearrowleft}$, we will write $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ when we take $\mathscr{M}$ to be the set of* all *probability mass functions on $\mathscr{X}$, and similarly for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q}}$.*

Since the sets $\mathbb{P}^{\mathrm{WHM}}$, $\mathbb{P}^{\mathrm{WM}}$ and $\mathbb{P}^{\mathrm{W}}$ are nested, it should be clear that this also true for the corresponding types of continuous-time imprecise-Markov chains.

Proposition 5.9. *For any non-empty set $\mathscr{Q} \subseteq \mathscr{R}$ and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, it holds that $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$.*

Proof. Use Definitions 5.5$_{\curvearrowleft}$ and 5.6 and that $\mathbb{P}^{\mathrm{WHM}} \subseteq \mathbb{P}^{\mathrm{WM}} \subseteq \mathbb{P}^{\mathrm{W}}$. □

[2] As in Chapter 3$_{83}$, this terminology derives from the *imprecise-Markov property* that these sets satisfy. For $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ we show in Proposition 5.26$_{202}$ that this is always true. For $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, we show in Proposition 5.28$_{203}$ and Equation (6.13)$_{281}$ that this is true under some structural assumptions on $\mathscr{Q}$.

Observe furthermore that for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, the additional properties of their elements allow us to simplify the notion of consistency in Definition 5.3$_{189}$, which leads to the following alternative characterisations:

$$\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} = \left\{P \in \mathbb{P}^{\mathrm{WM}} \,\middle|\, (\forall t \in \mathbb{R}_{\geq 0})\ \overline{\partial} T_t^t \subseteq \mathscr{Q}, P \sim \mathscr{M}\right\}, \tag{5.4}$$

and

$$\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}} = \left\{P \in \mathbb{P}^{\mathrm{WHM}} \,\middle|\, Q_P \in \mathscr{Q}, P \sim \mathscr{M}\right\}. \tag{5.5}$$

This first equality follows from the Markov property of the elements of $\mathbb{P}^{\mathrm{WM}}$, which, by Proposition 5.4$_{185}$, ensures that $\overline{\partial} T^t_{t,x_u} = \overline{\partial} T^t_t$. The second equality follows from the Markov property and the homogeneity of the processes $P \in \mathbb{P}^{\mathrm{WHM}}$, which, by Proposition 5.7$_{187}$, ensures that $\overline{\partial} T_t^t = \{Q_P\}$ for all $t \in \mathbb{R}_{\geq 0}$.

The following example further illustrates the difference between the three types of models that we consider.

Example 5.2. Let Q_* be an arbitrary transition rate matrix, fix any $\rho > 0$, and let $\mathscr{Q} := \{Q \in \mathscr{R} : \|Q - Q_*\| \leq \rho\}$ be the closed metric ball (in $\mathscr{R}$) of radius ρ around Q_*. This constructs $\mathscr{Q}$ as essentially a perturbation model around the rate matrix Q_*, which may arise naturally in contexts like for example sensitivity analysis. Although we do not currently aim to dwell on specifics, it follows from Proposition 6.20$_{278}$ further on that this set $\mathscr{Q}$ has a number of convenient structural properties, like boundedness and convexity. To finish setting up the example, let furthermore $\mathscr{M} := \{p\}$, with p an arbitrary but fixed probability mass function on $\mathscr{X}$.

The set $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ then contains an infinity of stochastic processes P, each of which is a well-behaved homogeneous Markov chain. They all have p as their initial distribution, in the sense that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$, but their transition rate matrices Q_P—whose existence is guaranteed by Theorem 5.6$_{186}$—are different: there is exactly one process P with $Q_P = Q$ for every $Q \in \mathscr{Q}$. The transition matrix systems of these Markov chains P are given by the exponential transition matrix systems $\left(e^{Q_P(s-t)}\right)$.

Since $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ is a subset of $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, all these homogeneous Markov chains belong to $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ as well. However, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ may also contain additional processes, which are not homogeneous and therefore do not belong to $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. For instance, for any $Q_1, Q_2 \in \mathscr{Q}$ such that $Q_1 \neq Q_2$, and any $r > 0$, it follows from Proposition 5.10$_{\curvearrowright}$ further on that there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ that has $\left(e^{Q_1(s-t)}\right)_{[0,r]} \otimes \left(e^{Q_2(s-t)}\right)_{[r,+\infty)}$ as its transition matrix system. Provided that $T_0^r = e^{Q_1 r}$ is different from $T_r^{2r} = e^{Q_2 r}$, it follows that this Markov chain is not homogeneous, and it therefore does not belong to $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$.

Since $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ is a subset of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, all of these processes belong to the set $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ as well. However the latter set may contain more complicated processes still. For instance, due to the properties of $\mathscr{Q}$ established by Proposition 6.20_{278}, and using Lemma 5.35_{215}—which is a more technical and slightly stronger version of Theorem 5.11 further on—it is possible to show that for any $u \in \mathscr{U}_{\supset\emptyset}$ and any $x_u, y_u \in \mathscr{X}_u$ such that $x_u \neq y_u$, $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ contains a stochastic process P such that, for all $t > u$, $\overline{\partial} T^t_{t,x_u} = \{Q_1\}$ and $\overline{\partial} T^t_{t,y_u} = \{Q_2\}$. For all $s > t > u$, the history-dependent transition matrices T^s_{t,x_u} and T^s_{t,y_u} of this process P are given by $T^s_{t,x_u} = e^{Q_1(s-t)}$ and $T^s_{t,y_u} = e^{Q_2(s-t)}$, which—again provided that $e^{Q_1(s-t)} \neq e^{Q_2(s-t)}$—implies that this process P does not satisfy the Markov property, and therefore, that it does not belong to $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. ◊

Proposition 5.10. *Consider any non-empty set of rate matrices $\mathscr{Q}$ and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for any $p \in \mathscr{M}$, any ordered finite sequence of time points $u = t_0, \ldots, t_n$ in $\mathscr{U}_{\supset\emptyset}$ and any collection of rate matrices $Q_0, \ldots, Q_{n+1} \in \mathscr{Q}$, there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ such that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$, and whose transition matrix system is given by Equation* $(5.3)_{188}$.

Proof. Proposition 5.8_{188} implies the existence of a process $P \in \mathbb{P}^{\mathrm{WM}}$ such that $P(X_0 = y)$ for all $y \in \mathscr{X}$ and such that its corresponding family of transition matrices (T^s_t) is given by Equation $(5.3)_{188}$. It remains to show that $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. Because of Equation $(5.4)_{\curvearrowleft}$, and since we already know that $p \in \mathscr{M}$, this means that we have to show that $\overline{\partial} T^r_r \subseteq \mathscr{Q}$ for all $r \in \mathbb{R}_{\geq 0}$. So, fix any $r \in \mathbb{R}_{\geq 0}$.

We consider several cases. If $r < t_0$, then $\overline{\partial} T^r_r$ corresponds to $\left(e^{Q_0(s-t)}\right)_{[0,t_0]}$, and it then follows from Lemma 4.14_{155} and Corollary 4.24_{171} that $\overline{\partial} T^r_r = \{Q_0\} \subseteq \mathscr{Q}$. If $r > t_n$, then $\overline{\partial} T^r_r$ corresponds to $\left(e^{Q_{n+1}(s-t)}\right)_{[t_n,+\infty)}$, whence $\overline{\partial} T^r_r = \{Q_{n+1}\} \subseteq \mathscr{Q}$. Similarly, if there is some $i \in \{1, \ldots, n\}$ such that $r \in (t_{i-1}, t_i)$, then $\overline{\partial} T^r_r$ corresponds to $\left(e^{Q_i(s-t)}\right)_{[t_{i-1},t_i]}$, and therefore $\overline{\partial} T^r_r = \{Q_i\} \subseteq \mathscr{Q}$. The only remaining case is when $r = t_i$ for some $i \in \{0, \ldots, n\}$. In this case, we have that $\overline{\partial}_+ T^r_r = \{Q_{i+1}\}$ and, if $r \neq 0$, that $\overline{\partial}_- T^r_r = \{Q_i\}$, and therefore, it follows from Definition 4.10_{167} that $\overline{\partial} T^r_r \subseteq \mathscr{Q}$. □

We conclude this section with some notes about closure properties of the different types of models that we consider, which are particularly useful for existence proofs. In particular, we focus on closure properties under recombination of known elements—colloquially, the "piecing together" of two or more processes in order to construct a new process that belongs to the same continuous-time imprecise-Markov chain.

Example $5.2_{\curvearrowleft}$ already suggested how to do this for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, by combining two well-behaved homogeneous Markov chains $P_1, P_2 \in \mathbb{P}^{\mathrm{WHM}}$

to form a new process $P \in \mathbb{P}^{\mathrm{WM}}$. Similarly, but more generally, for any two processes $P_1, P_2 \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, we can combine their transition matrix systems $\left({}^1T_t^s\right)$ and $\left({}^2T_t^s\right)$ to construct a new transition matrix system $(T_t^s) \coloneqq \left({}^1T_t^s\right)_{[0,r]} \otimes \left({}^2T_t^s\right)_{[r,+\infty)}$, with $r > 0$ chosen arbitrarily. Theorem 5.2$_{184}$ then guarantees that there is a Markov chain P whose corresponding family of transition matrices is (T_t^s), and that satisfies $P(X_0 = y) = P_1(X_0 = y)$ for all $y \in \mathscr{X}$. It is straightforward to verify that $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ because $P_1, P_2 \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$.

Clearly, a similar procedure is impossible for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, because the combination of two processes would make the resulting one lose the homogeneity property, which is required to be an element of $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$.

However, for our most imprecise type of CTIMC, which is $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it turns out that, as already suggested in Example 5.2$_{191}$, it is possible to recombine elements in an even more general, history-dependent way. That is, under some conditions on $\mathscr{Q}$ it is possible, for fixed time points $u \in \mathscr{U}$, to choose for every history $x_u \in \mathscr{X}_u$ a different process $P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, and to recombine these into a process P that agrees with these P_{x_u} conditional on the specific history x_u. Furthermore, the distribution on the time points u can be chosen to agree with any element $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. The crucial observation is that this new process P will again belong to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, provided that the conditions on $\mathscr{Q}$ are satisfied.

One important such condition on $\mathscr{Q}$ is that it has *separately specified rows*, in a manner analogous to how we defined this property for sets of transition matrices (c.f. Definition 3.13$_{111}$). We will require this condition for many results further on as well, so let us explicitly rephrase this definition for completeness:

Definition 5.7. *Let $\mathscr{R}$ be the set of all rate matrices, and consider any set $\mathscr{Q} \subseteq \mathscr{R}$. We then say that $\mathscr{Q}$ has* separately specified rows *if it holds that*

$$\mathscr{Q} = \left\{ Q \in \mathscr{R} \,\middle|\, \forall x \in \mathscr{X} : Q(x,\cdot) \in \mathscr{Q}_x \right\},$$

where, for all $x \in \mathscr{X}$, $\mathscr{Q}_x \coloneqq \left\{ Q(x,\cdot) \,\middle|\, Q \in \mathscr{Q} \right\}$ is the set of x-rows of the elements of $\mathscr{Q}$.

Again, this quality of having separately specified rows is completely analogous to the same property for sets of transition matrices: $\mathscr{Q}$ has separately specified rows if it is closed under the row-wise recombination of its elements.

The following theorem formalises the claim made above that, provided that $\mathscr{Q}$ is convex and has separately specified rows, we can recombine the elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ in a history-dependent manner. The (highly technical) proof can be found in Appendix 5.B$_{215}$.

Theorem 5.11. *Consider a non-empty and convex set of rate matrices $\mathscr{Q}$ that has separately specified rows, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Fix a finite sequence of time points $u \in \mathscr{U}$. Choose any $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and, for all $x_u \in \mathscr{X}_u$, choose some $P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then there is a stochastic process $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that, for all $u_1, u_2 \subseteq u$ such that $u_2 \neq \emptyset$ and $u_1 < u_2$, all $x_{u_1} \in \mathscr{X}_{u_1}$ and all $x_{u_2} \in \mathscr{X}_{u_2}$,*

$$P(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) = P_\emptyset(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) \tag{5.6}$$

and, for all $x_u \in \mathscr{X}_u$ and all $A \in \mathscr{A}_u$,

$$P(A | X_u = x_u) = P_{x_u}(A | X_u = x_u). \tag{5.7}$$

5.3 Sets of Transition Matrices for CTIMCs

Because transition matrices play such an important role in the theory of stochastic processes, and of Markov chains in particular, we find it interesting to also investigate the sets of transition matrices that are induced by a given CTIMC which, as we know from the previous section, is a set of stochastic processes.

In particular, we will investigate in this (mostly technical) section the structural properties of these induced sets of transition matrices, mostly in terms of the structural properties of the set $\mathscr{Q}$ that is used to parameterise the CTIMC. The structural properties of these sets are mostly related to the notions of non-emptiness, boundedness, closure, convexity, and having separately specified rows; we refer to Appendix A369, and to Definition A.12376 in particular, for the definition of all these properties but the last one—the quality of having separately specified rows was introduced in Definition 5.7↶ above.

As to the other structural properties that we require, we note that Proposition 4.5151 guarantees that convex combinations of rate matrices are themselves rate matrices, so it makes sense to talk about convex sets of them. The notion of boundedness is explicitly mentioned because, unlike when working with transition matrices, rate matrices can have an arbitrarily high norm; again, this follows from Proposition 4.5151. We also note that, due to Corollary A.12378, closure and boundedness of $\mathscr{Q}$ together imply its compactness (and vice versa).

The proofs of the results appearing in this section are unfortunately somewhat involved, so we have moved almost all of them to Appendix 5.D232. Moreover, these results rely on some general technical properties that we discuss in Appendix 5.C227.

Let us now start with the following lemma, which gives conditions on $\mathscr{Q}$ to obtain an error bound on linear approximations of the transition matrices of the Markov chains $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, which is uniform with respect to all elements of this set and all time points. We emphasise that this is a technical result that does not have immediate practical (i.e. algorithmic) applicability, since we do not actually have a way to construct the matrix Q in Equation (5.8).

Lemma 5.12. *Let $\mathscr{Q}$ be any non-empty, bounded, and convex set of rate matrices, and choose any $\varepsilon \in \mathbb{R}_{>0}$. Then there is some $\delta \in \mathbb{R}_{>0}$ such that for all $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, all $t \in \mathbb{R}_{\geq 0}$, and all $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta < \delta$, it holds that*

$$(\exists Q \in \mathscr{Q}) \left\| T_t^{t+\Delta} - (I + \Delta Q) \right\| \leq \Delta \varepsilon \,, \tag{5.8}$$

where $T_t^{t+\Delta}$ is the transition matrix corresponding to P.

This result can be generalised to an error bound on linear approximations of the history-dependent transition matrices of the elements $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, which is uniform with respect to all these processes, and all time points and histories leading up to those time points. However, because the set $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ contains non-Markov processes, we need an additional condition on $\mathscr{Q}$ to obtain this result; specifically, we require $\mathscr{Q}$ to have separately specified rows.

Lemma 5.13. *Let $\mathscr{Q}$ be a non-empty, bounded, and convex set of rate matrices that has separately specified rows, and choose any $\varepsilon \in \mathbb{R}_{>0}$. Then there is some $\delta \in \mathbb{R}_{>0}$ such that, for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, all $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$, and all $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta < \delta$, it holds that*

$$(\exists Q \in \mathscr{Q}) \left\| T_{t,x_u}^{t+\Delta} - (I + \Delta Q) \right\| \leq \Delta \varepsilon \,,$$

where $T_{t,x_u}^{t+\Delta}$ is the history-dependent transition matrix corresponding to P.

We will now start by considering the CTIMC $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, which, as we recall from Section 5.2188, is the set of all well-behaved Markov chains that are consistent with both $\mathscr{Q}$ and $\mathscr{M}$. Then, as we know from Proposition 5.1183, each of these Markov chains $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ has a corresponding family of transition matrices $\left({}^{P}T_t^s\right)$ that is a transition matrix system. Using the developments from Section 4.5158, we can consider the restrictions $({}^{P}T_t^s)_{[a,b]}$ of these transition matrix systems to the interval $[a,b]$, with $a,b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Let us now collect all these restricted transition matrix systems into a single set, i.e., let

$$\mathcal{T}^{\mathscr{Q}}_{[a,b]} := \left\{ \left({}^{P}T_t^s\right)_{[a,b]} \,\middle|\, P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}} \right\} . \tag{5.9}$$

Note that this makes no reference to the choice of $\mathscr{M}$. This is because, as the next result makes explicit, the choice of $\mathscr{M}$ does not influence the transition matrix systems of the elements of $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, and hence we do not carry the parameter in the notation for and definition of $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$:

Proposition 5.14. *Consider a non-empty set $\mathscr{Q}$ of rate matrices and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Let $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ be as in Equation* (5.9)$_{\frown}$. *Then it holds that*

$$\mathcal{T}^{\mathscr{Q}}_{[a,b]} = \left\{ \left({}^{P}T_t^s\right)_{[a,b]} \,\middle|\, P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \right\}.$$

Observe that $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ is a subset of the set $\mathcal{T}_{[a,b]}$, as introduced in Section 4.5$_{158}$. Moreover, recall that by Proposition 4.21$_{163}$, and using the metric d defined in Equation (4.15)$_{162}$, the metric space $(\mathcal{T}_{[a,b]}, d)$ is complete. The following result establishes sufficient conditions on $\mathscr{Q}$ for the set $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ to also form a complete metric space under d.

Lemma 5.15. *Consider a non-empty, compact, and convex set of rate matrices $\mathscr{Q}$ and any $a, b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Let d be the metric that is defined in Equation* (4.15)$_{162}$. *The metric space $\left(\mathcal{T}^{\mathscr{Q}}_{[a,b]}, d\right)$ is then complete.*

Moreover, we have the following result.

Lemma 5.16. *Consider a non-empty bounded set of rate matrices $\mathscr{Q}$ and any $a, b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Let d be the metric that is defined in Equation* (4.15)$_{162}$. *The metric space $\left(\mathcal{T}^{\mathscr{Q}}_{[a,b]}, d\right)$ is then totally bounded.*

Together, these two lemmas imply the following result.

Theorem 5.17. *Consider a non-empty, compact, and convex set of rate matrices $\mathscr{Q}$ and any $a, b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Let d be the metric that is defined in Equation* (4.15)$_{162}$. *The metric space $\left(\mathcal{T}^{\mathscr{Q}}_{[a,b]}, d\right)$ is then compact.*

Proof. Since a metric space is compact if and only if it is complete and totally bounded [100, Theorem 5.1.7], this result is an immediate consequence of Lemmas 5.15 and 5.16. □

To see the value of this largely technical result, consider that Lemma 5.15 tells us that a sequence $\left\{\left({}^{P_i}T_t^s\right)_{[a,b]}\right\}_{i \in \mathbb{Z}_{>0}}$ in $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ that is Cauchy with respect to the metric d, converges to a restricted transition matrix system $\left({}^{P_*}T_t^s\right)_{[a,b]}$ that is also in $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$. In particular, due to Proposition 5.14, this implies the existence of a Markov chain $P_* \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ whose corresponding family of transition matrices $\left({}^{P_*}T_t^s\right)$ is an extension of $\left({}^{P_*}T_t^s\right)_{[a,b]}$ to a(n unrestricted) transition matrix system.

Theorem 5.17 strengthens this result, because the compactness implies that every sequence $\left\{\left({}^{P_i}T_t^s\right)_{[a,b]}\right\}_{i \in \mathbb{Z}_{>0}}$ in $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ has a convergent (i.e. Cauchy) subsequence. Hence, this allows us to establish the existence of Markov chains in $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, as limits of (subsequences of) arbitrary sequences of Markov chains in this set. This result is fundamental to proving some of the results further in this section. One example of this is the upcoming Corollary 5.18.

We now move the discussion to sets of transition matrices induced by a given CTIMC which, as mentioned in the introduction to this section, are the objects that we want to study here. Our first result is a straightforward consequence of Theorem 5.17.

Corollary 5.18. *Consider a non-empty, compact, and convex set of rate matrices $\mathscr{Q}$, any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, and any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$. For all $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, let ${}^{P}T_t^s$ denote the transition matrix corresponding to P, and let $\mathscr{T} := \left\{ {}^{P}T_t^s : P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \right\}$ be the induced set of transition matrices. Then $\mathscr{T}$ is a non-empty and compact set of transition matrices.*

Moving on, for any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, let $\left({}^{P}T_{t,x_u}^s\right)$ denote the family of history-dependent transition matrices corresponding to P. Now, for any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, let

$$ {}^{\mathscr{Q}}\mathscr{T}_t^s := \left\{ {}^{P}T_{t,x_u}^s \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}, u \in \mathscr{U}_{<t}, x_u \in \mathscr{X}_u \right\}, \tag{5.10} $$

be the set of all history-dependent transition matrices of the elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, with the transition probabilities from time t to time s, and let

$$ {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s := \left\{ {}^{P}T_{t,x_u}^s \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}, u \in \mathscr{U}_{<t}, x_u \in \mathscr{X}_u \right\}, \tag{5.11} $$

be the set of all history-dependent transition matrices of the elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Perhaps unsurprisingly, under some conditions on $\mathscr{Q}$ the set ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ does not depend on $\mathscr{M}$:

Proposition 5.19. *Consider a non-empty and convex set $\mathscr{Q}$ of rate matrices that has separately specified rows, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then it holds that*

$$ {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = {}^{\mathscr{Q}}\mathscr{T}_t^s . $$

Although the set ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ contains all history-dependent transition matrices ${}^{P}T_{t,x_u}^s$ of the elements $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as the next result shows, under some conditions on $\mathscr{Q}$ we could equivalently only consider the (history-independent) transition matrices of these elements.

Proposition 5.20. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t,s \in \mathbb{R}$ such that $t \leq s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then*

$$ {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \left\{ {}^{P}T_t^s \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \right\}, $$

where, for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, ${}^{P}T_t^s$ is the transition matrix corresponding to P.

Moreover, under these same conditions this set satisfies the following properties.

Theorem 5.21. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t,s \in \mathbb{R}$ such that $t \leq s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is a non-empty, closed, and convex set of transition matrices that has separately specified rows.*

One of the reasons why this result is so interesting, is that it exactly establishes the properties that, by Corollary 3.38120, turn ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ into the unique dominating set of transition matrices corresponding to some lower transition operator. We will revisit this connection in Chapter 6259.

5.4 Lower and Upper Expectations for CTIMCs

Let us conclude this chapter by introducing and discussing the *lower* and *upper expectations* for continuous-time imprecise-Markov chains, which, as discussed in Chapter 129, are fundamentally the inferences we are interested in. Let us start with the general definition.

Definition 5.8. *For any non-empty set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$, its corresponding (conditional)* lower- *and* upper *expectations are defined, respectively, as*

$$\underline{\mathbb{E}}[\cdot\,|\,\cdot] := \inf_{P\in\mathscr{P}} \mathbb{E}_P[\cdot\,|\,\cdot] \quad \textit{and} \quad \overline{\mathbb{E}}[\cdot\,|\,\cdot] := \sup_{P\in\mathscr{P}} \mathbb{E}_P[\cdot\,|\,\cdot]\,, \tag{5.12}$$

whose domain(s) we take to be the intersection of the domains $\mathscr{D}_P$ of the conditional expectations $\mathbb{E}_P$ corresponding to the elements $P \in \mathscr{P}$.

In particular, for any non-empty set of rate matrices $\mathscr{Q}$ and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we let

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] := \inf_{P\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P[\cdot\,|\,\cdot] \quad \textit{and} \quad \overline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] := \sup_{P\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P[\cdot\,|\,\cdot]\,, \tag{5.13}$$

and similarly for $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\overline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, and $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ and $\overline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. If $\mathscr{M}$ is the set of all *probability mass functions on $\mathscr{X}$, then as in Definition 5.5*189*, we will write $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}$ instead of $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, and similarly for the other lower and upper expectations.*

We again recall from Chapter 129 the important conjugacy relation between lower and upper expectations, which in this context states that

$$\overline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] = -\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[-\cdot\,|\,\cdot]\,,$$

and similarly for the other lower and upper expectations. Hence, as before, we will mostly formulate our results in terms of lower expectations, where we keep in mind that these results can always be generalised to upper expectations using the conjugacy relation.

Moreover, as in the discrete-time case, *lower-* and *upper probabilities* can always be expressed using the lower- and upper expectations of indicators of events; for any $(A,C) \in \mathscr{C}^{\mathrm{SP}}$ the lower probability with respect to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies

$$\underline{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}(A\,|\,C) := \inf_{P\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} P(A\,|\,C) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_A\,|\,C]\,,$$

where we used Proposition 2.12$_{56}$ for the equality. Upper probabilities are derived analogously using upper expectations, and can therefore be obtained from lower expectations using the above-mentioned conjugacy relation. Hence, as before, because lower- and upper probabilities can always be derived using lower expectations, we will in the sequel express our results mostly in terms of the latter. Whenever we do explicitly use lower- and upper probabilities, then in addition to the above, we use the intuitive notation $\underline{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}(A\,|\,C)$ and $\underline{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}(A\,|\,C)$ for the lower probabilities with respect to $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, respectively, and similarly for the upper probabilities. As with lower- and upper expectations, we omit the set $\mathscr{M}$ from our notation when it is the set of all probability mass functions on $\mathscr{X}$.

So, let us now consider some first properties of the lower expectations in Definition 5.8. First of all, we know from Section 5.2$_{188}$ that the sets $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ are nested subsets of each other. As an immediate consequence, their corresponding lower expectations provide (lower) bounds for each other.

Proposition 5.22. *Consider any non-empty set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] \leq \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] \leq \underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot]\,.$$

Proof. This is immediate from Proposition 5.9$_{190}$ and Definition 5.8. □

Generally speaking, the inequalities in this proposition can be—and often are—strict; for the first inequality, we will illustrate this further on in this section, in Figure 5.1$_{206}$, whereas for the second inequality, we will give a detailed example in Section 6.6.2$_{297}$. On the other hand, there are also cases where all these quantities coincide; as illustrated in, again, Figure 5.1$_{206}$ further on. Both possibilities are important to keep in mind because, as we will argue further on, we will for practical reasons often want to work with the models $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ or $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. And while these two models can arise naturally, there are also cases where, from an application-specific point of view, $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ might be more natural. In such cases, we would thus be working with *conservative* approximations to the desired model—due to Proposition 5.9190—and Proposition 5.22↶ guarantees that inferences computed from such approximations are bounds on the intended inference of interest. However, the *degree* to which these bounds are conservative, depends strongly on the specific model and inference that one considers: it may be that the lower expectations coincide with that of the intended model, or it may be that the outer approximation is extremely conservative and carries relatively little information.

Next, let us establish that in particular the lower (and hence also upper) expectations of all u-measurable functions are well-defined, provided that care is taken with the conditioning events.

Lemma 5.23. *Let $\mathscr{P} \subseteq \mathbb{P}$ be a non-empty set of stochastic processes, and let $\underline{\mathbb{E}}$ be its corresponding (conditional) lower expectation, as in Definition 5.8*198. *Then for all $u, v \in \mathscr{U}$ such that $u < v$ and $u \cup v \neq \emptyset$, all $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}[f(X_u, X_v) \,|\, X_u = x_u] = \inf_{P \in \mathscr{P}} \sum_{x_v \in \mathscr{X}_v} f(x_u, x_v) P(X_v = x_v \,|\, X_u = x_u)\,.$$

Proof. Because $u < v$ it holds that $u \cup v \subset u \cup \mathbb{R}_{>u}$. Therefore, and because $u \cup v \neq \emptyset$, it follows from Proposition 2.2373 that for any $P \in \mathscr{P}$ it holds that

$$\begin{aligned}\mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u] &= \sum_{x_v \in \mathscr{X}_v} f(x_u, x_v) P(X_u = x_u, X_v = x_v \,|\, X_u = x_u) \\ &= \sum_{x_v \in \mathscr{X}_v} f(x_u, x_v) P(X_v = x_v \,|\, X_u = x_u)\,,\end{aligned}$$

where for the second equality we used Properties F447 and F247. Hence it follows from Definition 5.8198 that

$$\begin{aligned}\underline{\mathbb{E}}[f(X_u, X_v) \,|\, X_u = x_u] &= \inf_{P \in \mathscr{P}} \mathbb{E}_P[f(X_u, X_v) \,|\, X_u = x_u] \\ &= \inf_{P \in \mathscr{P}} \sum_{x_v \in \mathscr{X}_v} f(x_u, x_v) P(X_v = x_v \,|\, X_u = x_u)\,,\end{aligned}$$

which concludes the proof. □

Lemma 5.23 establishes that in particular the lower expectations $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, and $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ are all well-defined for u-measurable functions. Moreover, let us note that such lower expectations are always real-valued:

Proposition 5.24. *Let $\mathscr{P} \subseteq \mathbb{P}$ be a non-empty set of stochastic processes, and let $\underline{\mathbb{E}}$ be its corresponding (conditional) lower expectation, as in Definition 5.8*$_{198}$*. Then for all $u, v \in \mathscr{U}$ such that $u < v$ and $u \cup v \neq \emptyset$, all $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, and all $x_u \in \mathscr{X}_u$, it holds that $\underline{\mathbb{E}}[f(X_{u\cup v}) \,|\, X_u = x_u] \in \mathbb{R}$. In particular,*

$$\min_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) \leq \underline{\mathbb{E}}[f(X_{u\cup v}) \,|\, X_u = x_u] \leq \max_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) .$$

Proof. Because $u < v$ it holds that $u \cup v \subset u \cup \mathbb{R}_{>u}$. Therefore, and because $u \cup v \neq \emptyset$, it follows from Proposition 2.23$_{73}$ that for any $P \in \mathscr{P}$, $\mathbb{E}_P[f(X_{u\cup v}) \,|\, X_u = x_u]$ is well-defined and, due to Property CE1$_{78}$, that

$$\min_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) \leq \mathbb{E}_P[f(X_{u\cup v}) \,|\, X_u = x_u] \leq \max_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) . \tag{5.14}$$

Note that because f is $u \cup v$-measurable, it follows from Proposition 2.21$_{72}$ that it is bounded, i.e. it obtains its extremal values in $\mathbb{R}$, so the minimum and maximum operations in the inequalities above are well-defined. Because Equation (5.24) holds for all $P \in \mathscr{P}$, it follows from Definition 5.8$_{198}$ that

$$\min_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) \leq \underline{\mathbb{E}}[f(X_{u\cup v}) \,|\, X_u = x_u] .$$

Moreover, since $\mathscr{P}$ is non-empty, there is some $P \in \mathscr{P}$ such that, using Definition 5.8$_{198}$, it holds that

$$\underline{\mathbb{E}}[f(X_{u\cup v}) \,|\, X_u = x_u] \leq \mathbb{E}_P[f(X_{u\cup v}) \,|\, X_u = x_u] \leq \max_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} f(y_{u\cup v}) ,$$

and therefore $\underline{\mathbb{E}}[f(X_{u\cup v}) \,|\, X_u = x_u] \in \mathbb{R}$. □

The following result provides an alternative expression for these lower expectations in case the u-measurable function depends only on a single time point, i.e., when u is a singleton. This expression is given in terms of (history-dependent) transition matrices.

Proposition 5.25. *Let $\mathscr{P} \subseteq \mathbb{P}$ be a non-empty set of stochastic processes, and let $\underline{\mathbb{E}}$ be its corresponding (conditional) lower expectation, as in Definition 5.8*$_{198}$*. Then for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$, $x_t \in \mathscr{X}$, and $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\underline{\mathbb{E}}[f(X_s) \,|\, X_t = x_t, X_u = x_u] = \inf_{P\in\mathscr{P}} {}^{P}T_{t,x_u}^{s} f(x_t) ,$$

where, for all $P \in \mathscr{P}$, ${}^{P}T_{t,x_u}^{s}$ is the history-dependent transition matrix corresponding to P.

Proof. It follows from Proposition 4.3$_{149}$ that for all $P \in \mathscr{P}$ it holds that

$$\mathbb{E}_P[f(X_s)\,|\,X_t = x_t, X_u = x_u] = {}^P T_{t,x_u}^s f(x_t)\,.$$

Hence it follows from Definition 5.8$_{198}$ that

$$\begin{aligned}\underline{\mathbb{E}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] &= \inf_{P\in\mathscr{P}} \mathbb{E}_P[f(X_s)\,|\,X_t = x_t, X_u = x_u] \\ &= \inf_{P\in\mathscr{P}} {}^P T_{t,x_u}^s f(x_t)\,,\end{aligned}$$

which concludes the proof. ☐

We now have all the parts to motivate the terminology that (and when) the sets $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ are imprecise-Markov chains; we will show sufficient conditions on $\mathscr{Q}$ for all these models to satisfy an imprecise-Markov property.[3] Let us start with the two obvious models, which are themselves made up solely by (precise) Markov chains.

Proposition 5.26. *Let $\mathscr{Q}$ be a non-empty set of rate matrices, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}_t$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,,$$

and

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,.$$

Proof. It follows from Definition 5.6$_{190}$ that each $P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ is a Markov chain. Hence, by Proposition 5.25$_{\curvearrowleft}$ and Lemma 5.3$_{185}$ we find that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] &= \inf_{P\in\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}} {}^P T_{t,x_u}^s f(x_t) \\ &= \inf_{P\in\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}} {}^P T_t^s f(x_t) \\ &= \underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,,\end{aligned}$$

where, for all $P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, ${}^P T_{t,x_u}^s$ and ${}^P T_t^s$ denote the (history-dependent) transition matrices corresponding to P. This concludes the proof for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. The proof for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ is completely analogous. ☐

[3] In Chapter 6$_{259}$ further on we will derive weaker conditions on $\mathscr{Q}$ for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ to satisfy an imprecise-Markov property, but we feel that the results presented here are also illuminating.

It requires a bit more work to obtain a similar statement for the set $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$. Let us start by noting the following; because the proofs here become a bit more involved we have deferred them to Appendix 5.E$_{253}$.

Lemma 5.27. *Let $\mathcal{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, and all $x_t \in \mathscr{X}_t$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t] = \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t]\,.$$

We use this lemma in the proof of the following result, which states that $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$ also satisfies an imprecise-Markov property when certain conditions on $\mathcal{Q}$ are satisfied.

Proposition 5.28. *Let $\mathcal{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}_t$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t]\,.$$

Of course, this immediately implies that the inferences for $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$ coincide for functions that depend only on a single time point, whenever these conditions on $\mathcal{Q}$ are satisfied:

Corollary 5.29. *Let $\mathcal{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}_t$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u]\,.$$

Proof. Using Propositions 5.26 and 5.28 and Lemma 5.27, we get

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] &= \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t] \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t] \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u]\,,\end{aligned}$$

which concludes the proof. □

This allows us to show that under these conditions, the infimum that characterises the lower expectations is actually a minimum, i.e. the lower expectation is here reached by some process in the CTIMC.

Proposition 5.30. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and all $f \in \mathscr{L}(\mathscr{X})$, there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that for all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}_t$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \mathbb{E}_P[f(X_s)\,|\,X_t = x_t]\,.$$

We note that Proposition 5.30 requires the compactness of $\mathscr{Q}$, but not that of $\mathscr{M}$; for the technical details we refer to the proof in Appendix $5.\mathrm{E}_{253}$, but we can provide some intuition here. Essentially, the point is that the statement is always *conditional* on the state at time t, and once that state is fixed the remainder of the inference only depends on "future" dynamic behaviour—which is completely characterised by $\mathscr{Q}$—and not on "initial" behaviour described by $\mathscr{M}$. This, incidentally, also explains why we did not need to impose many constraints on $\mathscr{M}$, for other results in this section.

Having established these first properties of our lower expectations, it will be of primary interest in the remainder of this dissertation to develop methods to efficiently *compute* them. Due to the different qualitative properties of the sets $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it turns out that the difficulty of computing their lower expectations is also different.

In fact, rather ironically, computing lower expectations for the set $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$—which, intuitively, is the simplest of our three types of CTIMCs—seems to be much harder than for the sets $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ or $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. The problem is, essentially, that while homogeneous Markov chains are easy to work with numerically, the set $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ does not provide enough "degrees of freedom" to easily solve the optimisation problem that is involved in computing $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. Another way to put this, is that the homogeneity condition constrains the canonical parameters Q_P of the processes $P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ to be the same at all points in time. In contrast, the weaker assumptions involved in constructing $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ do not impose such a constraint. This issue is analogous to the difficulty of working with sets of discrete-time homogeneous Markov chains that we mentioned in Section $3.3.1_{102}$. Let us next illustrate this.

Suppose that we want to compute the lower expectation $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_0 = x_0]$ of some function $f \in \mathscr{L}(\mathscr{X})$ at time t, conditional on the information that the state X_0 at time 0 takes the value $x_0 \in \mathscr{X}_0$. It then follows from Definition 5.8_{198}, Proposition 5.25_{201}, Theorem 5.6_{186}, Equation $(5.5)_{191}$, and Corollary 5.5_{186} that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_0=x_0] &= \inf_{P\in\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P[f(X_t)\,|\,X_0=x_0] \\ &= \inf_{P\in\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}} {}^{P}T_0^t f(x_0) \\ &= \inf_{P\in\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}} e^{Q_P t} f(x_0) = \inf_{Q\in\mathscr{Q}} e^{Qt} f(x_0). \end{aligned} \tag{5.15}$$

Therefore, computing $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_0=x_0]$ is at its core a non-linear, constrained optimisation problem over the set $\mathscr{Q}$, where the non-linearity stems from the term e^{Qt}, and the specific form of the constraints depends on the choice of $\mathscr{Q}$. The following example illustrates the non-linearity of this type of optimisation problem in a simple case.

Example 5.3. Consider an ordered ternary state space $\mathscr{X} := \{a,b,c\}$, let $f \in \mathscr{L}(\mathscr{X})$ be defined such that $f(a) := 1/2$, $f(b) := 0$ and $f(c) := 1$, and consider the set of transition rate matrices

$$\mathscr{Q} := \left\{ \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} : \lambda \in [2,5] \right\}. \tag{5.16}$$

Every process P in the imprecise-Markov chain $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q}}$ is then a homogeneous Markov chain of which the unique transition rate matrix Q is completely determined by some λ in $[2,5]$. Furthermore, as we know from Proposition 4.3_{149} and Theorem 5.6_{186}, the conditional expectation $\mathbb{E}_P[f(X_t)|X_0=a]$ that corresponds to this homogeneous Markov chain is equal to $e^{Qt}f(a)$. Due to this equality, it is a matter of applying some basic—yet cumbersome—algebra to find that

$$\mathbb{E}_P[f(X_t)|X_0=a] = 1 + \frac{e^{-\lambda t}}{2} + \frac{e^{-\lambda t} - \lambda e^{-t}}{\lambda - 1}. \tag{5.17}$$

Obtaining the value of $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q}}[f(X_t)\,|\,X_0=a]$ now corresponds to minimising this expression as λ ranges over the interval $[2,5]$. Figure $5.1_{\curvearrowright}$ illustrates that, even in this simple ternary case, this minimisation problem is already non-trivial, because—depending on the value of t—the minimum is not guaranteed to be obtained for one of the end points of $[2,5]$, but may only be achieved by an internal point. Which is to say, for certain values of t the right-hand side of Equation (5.17) is not a monotone function of λ.

In particular, we see from Figure $5.1_{\curvearrowright}$ that the lower expectation with respect to this set is initially reached by choosing $\lambda = 5$. This changes around the time point $t \approx 0.5$, after which the minimising value

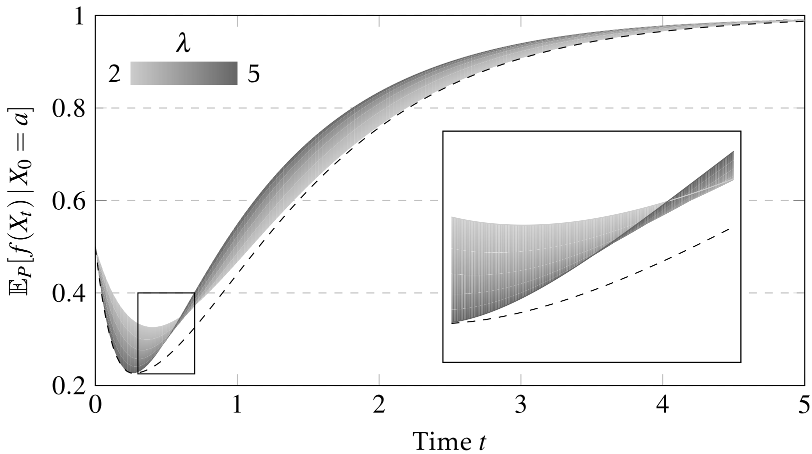

Figure 5.1: Plot of the set of expected values $\mathbb{E}_P[f(X_t)\,|\,X_0 = a]$, for time points $t \in [0,5]$, corresponding to all $P \in \mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q}}$ obtained as we vary $\lambda \in [2,5]$. The inset highlights the region of the function where the minimising value of λ becomes an internal point of the interval $[2,5]$. The dashed line corresponds to the lower expectation with respect to the sets $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q}}$.

of λ becomes a (changing) internal point of the interval $[2,5]$. This region of the function is highlighted in the inset, which also shows the extremal values of λ as emphasised lines, illustrating that the minimum is obtained by an internal point.

The dashed line corresponds to the lower expectation with respect to the sets $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q}}$, which also include non-homogeneous Markov chains and, for $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q}}$, even more general processes. The fact that they coincide is explained by Corollary 5.29$_{203}$; it is straightforward to check that $\mathcal{Q}$ satisfies the required properties of that statement. The method by which we have computed these quantities will be explained in Section 6.4$_{279}$. Note that the lower expectations with respect to $\mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q}}$ are all equal for $t < 0.3$ but, as time evolves, the lower expectation with respect to $\mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q}}$ diverges from the other two. This illustrates that the first inequality in Proposition 5.22$_{199}$ can indeed be strict, but can also be satisfied with equality. ◊

Although the inference problem in the previous example can be solved numerically by, essentially, exhaustively searching for the minimising value of λ, this approach clearly does not generalise well to higher dimensions, for which the corresponding computational complexity would grow exponentially. Moreover, in general a simple expression for the element-wise conditional expectation, such as given in Equation $(5.17)_{205}$, will typically not exist or at least be difficult to find. Hence, in general one does not have access to analytical methods to simplify this minimisation problem. One must then resort to approximative methods to even evaluate the function that is to be minimised; see e.g. Reference [76] for details on how this might be done. It should also be mentioned that this optimisation problem is in general not convex—nor monotone, as already illustrated in Example 5.3_{205}—and that the objective surface over which the minimisation is to take place, can in general be very irregular, having multiple local optima.

All of these considerations suggest that computing lower expectations for CTIMCs that are of the type $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ is typically very difficult, and it will therefore often be necessary to resort to approximation methods. This could for example be accomplished by using the methods to compute (generalised) exponentials of *interval matrices* as explored in References [37, 83]. Although our notion of sets of rate matrices is not really compatible with the notion of an interval matrix[4] as these authors use it, one could always include a (bounded) set of rate matrices in an interval matrix and then apply these methods, in order to compute conservative bounds with respect to exponentials of the elements of the original set of rate matrices.

In the remainder of this work, we will mostly ignore CTIMCs that are of the type $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. Instead, we will focus on the lower expectations that correspond to CTIMCs that are of the type $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ or $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, and we will develop efficient methods for computing them.

At their core, these methods are based on a *law of iterated lower expectations*, which is analogous to the one we discussed for discrete-time imprecise-Markov chains in Chapter 3_{83}. To make sense of this statement, we first need to formalise the following.

Definition 5.9. *For any stochastic process $P \in \mathbb{P}$, any $u, v \in \mathscr{U}_{\supset\emptyset}$ such that $u < v$, and any $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, we consider the function $\mathbb{E}_P[f(X_{u\cup v})\,|\,X_u]$ in $\mathscr{L}(\mathscr{X}_u)$ whose value in $x_u \in \mathscr{X}_u$ is given by $\mathbb{E}_P[f(X_{u\cup v})\,|\,X_u = x_u]$.*

Now consider a function $f \in \mathscr{L}(\mathscr{X}_{u\cup v\cup w})$ defined on the union $u \cup v \cup w$ of three finite sets of time points $u, v, w \in \mathscr{U}$ such that $u < v < w$. The precise version of the law of iterated expectations—which is well-

[4]References [37, 83] call a set of matrices $\mathscr{Q}$ an *interval matrix* if there are matrices $Q_1, Q_2 \in \mathbb{M}$ such that $\mathscr{Q} = \{Q \in \mathbb{M} : Q_1 \leq Q \leq Q_2\}$, where the inequalities are taken to be componentwise. We stress that this is different from the terminology of e.g. Škulj [102], who calls $\mathscr{Q}$ an interval matrix if it is compact, convex, and has separately specified rows. Similarly, Troffaes *et al.* [110] call $\mathscr{Q}$ an *interval rate matrix* if it is a set of rate matrices that satisfies these properties. Note that the set $\mathscr{Q}$ in Equation $(5.16)_{205}$ satisfies these second and third definitions but not the first one; in particular, Property $\mathrm{R1}_{150}$ will usually not be satisfied for all elements in a set constructed according to the first definition.

known, and which was already given in Proposition 2.26$_{77}$—then states that for any stochastic process P, the corresponding expectation of f, conditional on X_u, decomposes over these time points. The following statement verifies that this property also holds here.

Proposition 5.31. *Let $P \in \mathbb{P}$ be a continuous-time stochastic process, and consider any $u, v, w \in \mathscr{U}$ such that $u < v < w$, $v \neq \emptyset$ and $w \neq \emptyset$. Then for all $f \in \mathscr{L}(\mathscr{X}_{u\cup v\cup w})$ and all $x_u \in \mathscr{X}_u$, it holds that*

$$\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_u = x_u] = \mathbb{E}_P\left[\mathbb{E}_P[f(X_{u\cup v\cup w}) \,|\, X_{u\cup v}] \,\middle|\, X_u = x_u\right].$$

Proof. This is an immediate consequence of Proposition 2.26$_{77}$. □

Rather remarkably, if $\mathscr{Q}$ is convex and has separately specified rows, then the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies a similar property. This is the so-called law of iterated *lower* expectations, and our proof of it is based on Theorem 5.11$_{193}$.

Theorem 5.32. *Let $\mathscr{Q}$ be a non-empty and convex set of rate matrices that has separately specified rows, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for all $u, v, w \in \mathscr{U}$ such that $u < v < w$, $v \neq \emptyset$, and $w \neq \emptyset$, all $f \in \mathscr{L}(\mathscr{X}_{u\cup v\cup w})$, and all $x_u \in \mathscr{X}_u$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v\cup w}) \,|\, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v\cup w}) \,|\, X_{u\cup v}] \,\middle|\, X_u = x_u\right]. \tag{5.18}$$

The reason that this result is so useful, is that it essentially allows us to compute lower expectations recursively. This is analogous to our discussion of iterated lower expectations in Chapter 3$_{83}$, although the notation here is a bit more cumbersome. Essentially, for any $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$ with $u < v$ and $v = t_0, \ldots, t_n$, $n \in \mathbb{Z}_{>0}$, we can repeatedly invoke Theorem 5.32 to rewrite $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v}) \,|\, X_u = x_u\right]$ as

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\cdots \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v}) \,|\, X_u, X_{\{t_0,\ldots,t_{n-1}\}}] \cdots \,\middle|\, X_u, X_{\{t_0\}}\right] \,\middle|\, X_u = x_u\right].$$

Hence, instead of computing the lower expectation for all time points simultaneously, we can focus on each of the time points separately, and eliminate them one by one. In particular, if we start this process with the innermost lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v}) \,|\, X_u, X_{\{t_0,\ldots,t_{n-1}\}}]$, we see that for all $x_u \in \mathscr{X}_u$ and $x_{\{t_0,\ldots,t_{n-1}\}} \in \mathscr{X}_{\{t_0,\ldots,t_{n-1}\}}$, the quantity

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v}) \,|\, X_u = x_u, X_{\{t_0,\ldots,t_{n-1}\}} = x_{\{t_0,\ldots,t_{n-1}\}}] \\
&\quad = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P[f(X_{u\cup v}) \,|\, X_u = x_u, X_{\{t_0,\ldots,t_{n-1}\}} = x_{\{t_0,\ldots,t_{n-1}\}}] \\
&\quad = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P\left[f(x_u, x_{\{t_0,\ldots,t_{n-1}\}}, X_{t_n}) \,\middle|\, X_u = x_u, X_{\{t_0,\ldots,t_{n-1}\}} = x_{\{t_0,\ldots,t_{n-1}\}}\right] \\
&\quad = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(x_u, x_{\{t_0,\ldots,t_{n-1}\}}, X_{t_n}) \,\middle|\, X_u = x_u, X_{\{t_0,\ldots,t_{n-1}\}} = x_{\{t_0,\ldots,t_{n-1}\}}\right]
\end{aligned}$$

is a lower expectation of a function that only depends on the single time point t_n, where we used Proposition 2.2575 for the second equality. This puts these lower expectations into a regime where e.g. Proposition 5.25201 is applicable; in Chapter 6259 we will develop algorithms that can efficiently solve inferences of this type. Working backwards, and analogously to the above, the next lower expectation to be computed will then be of a function that only depends on the time point t_{n-1} (although the conditioning event references a more complicated history), and so on.

The above considerations make the computations of these lower expectations somewhat easier but, nevertheless, the number of conditioning events to take into account is of order $|\mathscr{X}_{u\cup v}|$, which is exponential in n. This combinatorial explosion cannot be prevented in general: it is equally complex to even *specify* an arbitrary $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, so we can hardly expect the computation of its lower expectation to be easier than that. However, as we will illustrate in Chapter 7335, for specific classes of functions it is possible to develop efficient algorithms. These algorithms crucially depend on Theorem 5.32, which, in summary, lets us rewrite a “global” lower expectation, where the expectation is taken over multiple time points, as a composition of “local” lower expectations, which each take the expectation over only a single time point. What we require, then, is to find an efficient method to compute such “local” lower expectations; this is the topic of Chapter 6259.

Appendix

5.A Proofs of Results in Section 5.1

This appendix starts with some technical lemmas that are required for the proof of Theorem 5.2184.

Lemma 5.33. *Let $w = \{w_0, w_1, \ldots, w_m\} \subset \mathbb{R}_{\geq 0}$ be a finite set of time points, with $m \in \mathbb{Z}_{\geq 0}$, such that $w_0 < w_1 < \cdots < w_m$. Let (T_t^s) be a transition matrix system and let $\tilde{P}_w$ be any full conditional probability such that for all $j \in \{1, \ldots, m\}$ and $x_{w_\ell} \in \mathscr{X}$, $\ell \in \{0, \ldots, j\}$:*

$$\tilde{P}_w(X_{w_j} = x_{w_j} | X_{w_0} = x_{w_0}, \ldots, X_{w_{j-1}} = x_{w_{j-1}}) = T_{w_{j-1}}^{w_j}(x_{w_{j-1}}, x_{w_j}).$$

Then for any $s \in w$ and $u \subseteq w$ such that $s > u$ and $u \neq \emptyset$, any $y \in \mathscr{X}$ and any $x_u \in \mathscr{X}_u$, we have that

$$\tilde{P}_w(X_s = y | X_u = x_u) = T_{\max u}^s(x_{\max u}, y).$$

Proof. We provide a proof by induction. If $s = w_1$, then because $s > u$ and $w \supseteq u \neq \emptyset$, it follows that $u = \{w_0\}$; and therefore, the result follows trivially from the assumptions in this lemma. Assume now that the result is true for $s = w_j$, with $1 \leq j < m$. We will prove that this implies that it is also true for $s = w_{j+1}$. Since $\emptyset \neq u \subseteq w$, $s \in w$ and $s > u$, it follows that in fact $u \subseteq \{w_0, \ldots, w_j\}$. We consider two cases: $\max u = w_j$ and $\max u < w_j$.

If $\max u = w_j$, then with $v := \{w_0, \ldots, w_j\} \setminus u$ it holds that

$$\begin{aligned}
&\tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u) \\
&\qquad = \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y, X_v = z_v | X_u = x_u) \\
&\qquad = \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u, X_v = z_v) \tilde{P}_w(X_v = z_v | X_u = x_u) \\
&\qquad = \sum_{z_v \in \mathscr{X}_v} T_{\max u}^{w_{j+1}}(x_{\max u}, y) \tilde{P}_w(X_v = z_v | X_u = x_u) \\
&\qquad = T_{\max u}^{w_{j+1}}(x_{\max u}, y) \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_v = z_v | X_u = x_u) \\
&\qquad = T_{\max u}^{w_{j+1}}(x_{\max u}, y),
\end{aligned}$$

where the first equality follows from F347, the second equality follows from F447, the third equality follows from the assumptions in this lemma and the fact that $\max u = w_j$, and the last equality follows from F347 and F847.

If $\max u < w_j$, then with $v \coloneqq \{w_0, \dots, w_{j-1}\} \setminus u$ it holds that

$$
\begin{aligned}
&\tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y, X_{w_j} = z_{w_j}, X_v = z_v | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u, X_{w_j} = z_{w_j}, X_v = z_v) \\
&\qquad\qquad \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \tilde{P}_w(X_{w_j} = z_{w_j} | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} T_{w_j}^{w_{j+1}}(z_{w_j}, y) \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \\
&\qquad\qquad \tilde{P}_w(X_{w_j} = z_{w_j} | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} T_{w_j}^{w_{j+1}}(z_{w_j}, y) \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) \\
&= \sum_{z_{w_j} \in \mathscr{X}} T_{w_j}^{w_{j+1}}(z_{w_j}, y) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \\
&= \sum_{z_{w_j} \in \mathscr{X}} T_{w_j}^{w_{j+1}}(z_{w_j}, y) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) = T_{\max u}^{w_{j+1}}(x_{\max u}, y),
\end{aligned}
$$

where the first equality follows from $F3_{47}$, the second equality follows from $F4_{47}$, the third equality follows from the assumptions in this lemma, the fourth equality follows from the induction hypothesis, the sixth equality follows from $F3_{47}$ and $F8_{47}$, and the last equality follows from Equation $(4.9)_{156}$. □

Lemma 5.34. *Consider two Markov chains $P_1, P_2 \in \mathbb{P}^{\mathrm{M}}$ with corresponding families of transition matrices $({}^1T_t^s)$ and $({}^2T_t^s)$, respectively, such that $({}^1T_t^s) = ({}^2T_t^s)$ and, for all $y \in \mathscr{X}$, $P_1(X_0 = y) = P_2(X_0 = y)$. Then $P_1 = P_2$.*

Proof. Let $(T_t^s) \coloneqq ({}^1T_t^s) = ({}^2T_t^s)$ be the common transition matrix system of P_1 and P_2 and let p be their common initial probability mass function, as defined by $p(y) \coloneqq P_1(X_0 = y) = P_2(X_0 = y)$ for all $y \in \mathscr{X}$. Let

$$
\begin{aligned}
\mathscr{C} \coloneqq \{(X_s = y, X_u = x_u) \in \mathscr{C}^{\mathrm{SP}} \colon u \in \mathscr{U}_{\supset\emptyset},\ s > u,\ x_u \in \mathscr{X}_u,\ y \in \mathscr{X}\} \\
\cup \{(X_0 = y, X_\emptyset = x_\emptyset) \in \mathscr{C}^{\mathrm{SP}} \colon y \in \mathscr{X}\}
\end{aligned}
$$

and consider the function $\tilde{P}$ on $\mathscr{C}$ that is defined, for all $(X_s = y, X_u = x_u) \in \mathscr{C}$, as

$$
\tilde{P}(X_s = y | X_u = x_u) \coloneqq \begin{cases} p(y) & \text{if } u = \emptyset, \text{ and} \\ T_{\max u}^s(x_{\max u}, y) & \text{otherwise.} \end{cases} \tag{5.19}
$$

It then follows from Definition 5.1182 that the restriction of P_1 and P_2 to $\mathscr{C}$ is equal to $\tilde{P}$. Furthermore, for any $s > 0$, $y \in \mathscr{X}$ and $j \in \{1,2\}$, we find that

$$P_j(X_s = y) = \sum_{x \in \mathscr{X}} P_j(X_s = y, X_0 = x) = \sum_{x \in \mathscr{X}} P_j(X_s = y | X_0 = x) P_j(X_0 = x)$$
$$= \sum_{x \in \mathscr{X}} \tilde{P}(X_s = y | X_0 = x) \tilde{P}(X_0 = x).$$

Hence, the restrictions of P_1 and P_2 to

$$\mathscr{C}^* := \mathscr{C} \cup \{(X_s = y, X_\emptyset = x_\emptyset) \colon s \in \mathbb{R}_{>0},\ y \in \mathscr{X}\}$$
$$= \{(X_s = y, X_u = x_u) \colon u \in \mathscr{U},\ s \in \mathbb{R}_{\geq 0},\ s > u,\ x_u \in \mathscr{X}_u,\ y \in \mathscr{X}\}$$

are identical. We denote this common restriction by $\tilde{P}^*$.

Consider now any $(A, X_u = x_u) \in \mathscr{C}^{\mathrm{SP}}$. Then since $A \in \mathscr{A}_u$, due to Proposition 2.1866 there is some finite set $w \subset u \cup \mathbb{R}_{>u}$ and some set $S' \subseteq \mathscr{X}_w$ such that $A = \cup_{z_w \in S'} (X_w = z_w)$. Let $S := \left\{ z_{u\cup w} \in \mathscr{X}_{u \cup w} : z_w \in S' \right\}$. Then, clearly,

$$\bigcup_{z_{u\cup w} \in S} (X_{u\cup w} = z_{u \cup w}) = \bigcup_{z_w \in S'} (X_w = z_w) = A\,.$$

Let $v = \{t \in w : u < t\}$ be the subset of w that contains all time points greater than $\max u$; then $u \cup w = u \cup v$ since $w \subset u \cup \mathbb{R}_{>u}$. Hence, it follows that $A = \cup_{z_{u \cup v} \in S} (X_{u \cup v} = z_{u \cup v})$. Let $S_v := \{z_v \in \mathscr{X}_v \colon (x_u, z_v) \in S\}$. For any $j \in \{1,2\}$, we then find that

$$\begin{aligned}
P_j(A | X_u = x_u) &= \sum_{z_{u \cup v} \in S} P_j(X_{u \cup v} = z_{u \cup v} | X_u = x_u) \\
&= \sum_{z_v \in S_v} P_j(X_{v_1} = z_{v_1}, X_{v_2} = z_{v_2}, \dots, X_{v_n} = z_{v_n} | X_u = x_u) \\
&= \sum_{z_v \in S_v} \prod_{i=1}^{n} P_j(X_{v_i} = z_{v_i} | X_u = x_u, X_{v_1} = z_{v_1}, \dots, X_{v_{i-1}} = z_{v_{i-1}}) \\
&= \sum_{z_v \in S_v} \prod_{i=1}^{n} \tilde{P}^*(X_{v_i} = z_{v_i} | X_u = x_u, X_{v_1} = z_{v_1}, \dots, X_{v_{i-1}} = z_{v_{i-1}}),
\end{aligned}$$

which implies that $P_1(A | X_u = x_u) = P_2(A | X_u = x_u)$. Since this is true for any $(A, X_u = x_u) \in \mathscr{C}^{\mathrm{SP}}$, it follows that $P_1 = P_2$. □

Proof of Theorem 5.2184. Let

$$\mathscr{C} := \{(X_s = y, X_u = x_u) \in \mathscr{C}^{\mathrm{SP}} \colon u \in \mathscr{U}_{\supset \emptyset},\ s > u,\ x_u \in \mathscr{X}_u,\ y \in \mathscr{X}\}$$
$$\cup \{(X_0 = y, X_\emptyset = x_\emptyset) \in \mathscr{C}^{\mathrm{SP}} \colon y \in \mathscr{X}\}$$

and consider the function $\tilde{P}$ on $\mathscr{C}$ that is defined, for all $(X_s = y, X_u = x_u) \in \mathscr{C}$, as

$$\tilde{P}(X_s = y | X_u = x_u) := \begin{cases} p(y) & \text{if } u = \emptyset \text{, and} \\ T^s_{\max u}(x_{\max u}, y) & \text{otherwise.} \end{cases} \tag{5.20}$$

We will first prove that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$. To this end, consider any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \ldots, n\}$, choose any $(A_i, C_i) = (X_{s_i} = y_i, X_{u_i} = x_{u_i}) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega) \big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C_0 \right\} \geq 0, \tag{5.21}$$

with $C_0 := \cup_{i=1}^{n} C_i$. Since every sequence u_i is finite, there is some finite set $w = \{w_0, w_1, \ldots, w_m\} \subset \mathbb{R}_{\geq 0}$ of time points, with $m \in \mathbb{Z}_{>0}$, such that $0 = w_0 < w_1 < \cdots < w_m$ and, for all $i \in \{1, \ldots, n\}$, $u_i \subseteq w$ and $s_i \in w$. Let

$$\mathscr{C}_w := \big\{ (X_{w_j} = y, X_u = x_u) \colon j \in \{0, \ldots, m\},\ u = \{w_0, \ldots, w_{j-1}\}, \\ y \in \mathscr{X},\ x_u \in \mathscr{X}_u \big\},$$

and let P_w be the restriction of $\tilde{P}$ to $\mathscr{C}_w$. Then since P_w clearly satisfies the conditions of Lemma 3.45_{127}, it follows that P_w is a coherent conditional probability. Because of Theorem 2.3_{49}, this implies that P_w can be extended to a coherent conditional probability $\tilde{P}_w$ on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$, which, because of Theorem 2.2_{49}, is also a full conditional probability. Since $\tilde{P}_w$ is a coherent conditional probability, it now follows from Definition 2.2_{48} that

$$\max \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega) \big(\tilde{P}_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C_0 \right\} \geq 0. \tag{5.22}$$

By comparing Equations (5.21) and (5.22), we see that in order to prove that $\tilde{P}$ is coherent, it suffices to show that $\tilde{P}_w(A_i|C_i) = \tilde{P}(A_i|C_i)$ for all $i \in \{1, \ldots, n\}$.

So fix any $i \in \{1, \ldots, n\}$. If $u_i = \emptyset$, then $s_i = 0 = w_0$ and therefore $(A_i, C_i) \in \mathscr{C}_w$, which implies that $\tilde{P}_w(A_i|C_i) = P_w(A_i|C_i) = \tilde{P}(A_i|C_i)$. If $u_i \neq \emptyset$, then since $u_i \subseteq w$, $s_i \in w$ and $s_i > u_i$, it follows from Lemma 5.33_{210} that $\tilde{P}_w(A_i|C_i) = \tilde{P}(A_i|C_i)$. Hence, $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$.

Therefore, due to Theorem 2.3_{49}, and because $\mathscr{C} \subseteq \mathscr{C}^{\mathrm{SP}}$, $\tilde{P}$ can be extended to a coherent conditional probability P on $\mathscr{C}^{\mathrm{SP}}$, which, according to Definition 2.12_{68}, is a stochastic process. Due to Equation (5.20), this implies that P is a Markov chain with corresponding family of transition matrices (T^s_t) and, for all $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$. Lemma 5.34_{211}

implies that this Markov chain is unique and, since (T_t^s) is the family of transition matrices corresponding to P, Proposition 5.1$_{183}$ implies that (T_t^s) is well-behaved if and only if P is well-behaved. □

Proof of Theorem 5.6$_{186}$. Because of Proposition 4.22$_{169}$, we know that $\overline{\partial}_+ T_0^0$ is a non-empty bounded set of rate matrices, which implies that there is some real $B > 0$ such that $\|Q'\| \leq B$ for all $Q' \in \overline{\partial}_+ T_0^0$. Let Q be any element of $\overline{\partial}_+ T_0^0$.

Fix any $c \geq 0$, $\varepsilon > 0$ and $\delta > 0$. It then follows from Proposition 4.23$_{171}$ and N9$_{64}$ that there is some $\delta^* > 0$ such that

$$(\forall 0 < \Delta^* < \delta^*)\ (\exists Q^* \in \overline{\partial}_+ T_0^0)\ \left\| T_0^{\Delta^*} - (I + \Delta^* Q^*) \right\| < \Delta^* \varepsilon. \tag{5.23}$$

Furthermore, because of Equation (4.22)$_{168}$ and N9$_{64}$, there is some $0 < \Delta < \min\{\delta, \delta^*\}$ such that

$$\left\| T_0^{\Delta} - (I + \Delta Q) \right\| < \Delta \varepsilon. \tag{5.24}$$

If we now define $n := \lfloor c/\Delta \rfloor$[5] and $d := c - n\Delta$, then $n\Delta \leq c < (n+1)\Delta$ and therefore also $0 \leq d < \Delta$. Because of Proposition 5.1$_{183}$, Equation (4.9)$_{156}$ and Definition 5.2$_{185}$, we know that

$$T_0^c = \left(\prod_{j=1}^{n} T_{(j-1)\Delta}^{j\Delta} \right) T_{n\Delta}^c = \left(T_0^{\Delta} \right)^n T_0^d$$

and therefore, it follows from Lemma B.5$_{393}$ that

$$\left\| e^{Qc} - T_0^c \right\| = \left\| \left(T_0^{\Delta} \right)^n T_0^d - \left(e^{Q\Delta} \right)^n e^{Qd} \right\| \leq n \left\| T_0^{\Delta} - e^{Q\Delta} \right\| + \left\| T_0^d - e^{Qd} \right\|. \tag{5.25}$$

From Equation (5.24) and Lemma B.8$_{394}$, we infer that

$$\left\| T_0^{\Delta} - e^{Q\Delta} \right\| \leq \left\| T_0^{\Delta} - (I + \Delta Q) \right\| + \left\| (I + \Delta Q) - e^{Q\Delta} \right\| \leq \Delta \varepsilon + \Delta^2 \|Q\|^2. \tag{5.26}$$

Since $d < \Delta < \delta^*$, we infer from Equation (5.23) that there is some $Q^* \in \overline{\partial}_+ T_0^0$ such that $\|T_0^d - (I + dQ^*)\| < d\varepsilon$. Hence, also using Lemma B.8$_{394}$, we find that

$$\begin{aligned} \left\| T_0^d - e^{Qd} \right\| &\leq \left\| T_0^d - (I + dQ^*) \right\| + \|(I + dQ^*) - (I + dQ)\| \\ &\quad + \left\| (I + dQ) - e^{Qd} \right\| \\ &\leq d\varepsilon + d\|Q^* - Q\| + d^2 \|Q\|^2 \\ &\leq d\varepsilon + d\|Q^*\| + d\|Q\| + d^2 \|Q\|^2. \end{aligned} \tag{5.27}$$

[5] We use $\lfloor \cdot \rfloor$ to denote the *floor* function, i.e., it denotes the largest integer that is not greater than its argument.

By combining Equations (5.25), (5.26) and (5.27), it follows that

$$\left\|e^{Qc} - T_0^c\right\| \leq n\Delta\varepsilon + n\Delta^2 \left\|Q\right\|^2 + d\varepsilon + d\left\|Q^*\right\| + d\left\|Q\right\| + d^2\left\|Q\right\|^2.$$

Taking into account that $\|Q\| \leq B$, $\|Q^*\| \leq B$, $n\Delta \leq c$ and $d < \Delta < \delta$, this implies that

$$\left\|e^{Qc} - T_0^c\right\| \leq c\varepsilon + c\delta B^2 + \delta\varepsilon + 2\delta B + \delta^2 B^2.$$

Since this is true for any $\varepsilon > 0$ and $\delta > 0$, it follows that $\left\|e^{Qc} - T_0^c\right\| \leq 0$, which implies that $T_0^c = e^{Qc}$. Since this is true for all $c \geq 0$, it follows from Definition 5.2_{185} that

$$T_t^s = T_0^{s-t} = e^{Q(s-t)} \text{ for all } 0 \leq t \leq s, \tag{5.28}$$

or equivalently, that $(T_t^s) = (e^{Q(s-t)})$.

Finally, we prove that Q is unique. Assume *ex absurdo* that this is not the case, or equivalently, that there are rate matrices Q_1 and Q_2, with $Q_1 \neq Q_2$, such that $(T_t^s) = (e^{Q_1(s-t)})$ and $(T_t^s) = (e^{Q_2(s-t)})$. For all $\Delta > 0$, we then have that $T_0^\Delta = e^{Q_1\Delta} = e^{Q_2\Delta}$, and therefore, it follows from Lemma 4.14_{155} that $\partial_+ T_0^0 = Q_1$ and $\partial_+ T_0^0 = Q_2$, which implies that $Q_1 = Q_2$. From this contradiction, it follows that Q is indeed unique. □

5.B Proofs of Results in Section 5.2

The following lemma is essentially a slightly stronger but more technical statement than Theorem 5.11_{193}. We prove the latter as a special case of this result.

Lemma 5.35. *Consider a non-empty and convex set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ that has separately specified rows, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Fix a finite sequence of time points $u \in \mathscr{U}_{\supset\emptyset}$, and let*

$$\mathscr{C}_\emptyset := \{(A, X_v = x_v) \in \mathscr{C}^{\mathrm{SP}} \colon v \in \mathscr{U}_{<\max u} \text{ and } A \in \langle\{(X_t = x) \colon x \in \mathscr{X}, t \in [0, \max u]\}\rangle\}, \tag{5.29}$$

and, for all $x_u \in \mathscr{X}_u$, let

$$\mathscr{C}_{x_u} := \{(A, X_v = x_v) \in \mathscr{C}^{\mathrm{SP}} \colon u \subseteq v \in \mathscr{U}, x_{v\setminus u} \in \mathscr{X}_{v\setminus u}, A \in \mathscr{A}_{u\cup(v\setminus[0,\max u])}\}. \tag{5.30}$$

Choose any $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and, for all $x_u \in \mathscr{X}_u$, any $P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then there is a stochastic process $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that, for all $(A, C) \in \mathscr{C}_\emptyset$,

$$P(A \,|\, C) = P_\emptyset(A \,|\, C), \tag{5.31}$$

and, for all $x_u \in \mathscr{X}_u$ and all $(A, X_v = x_v) \in \mathscr{C}_{x_u}$,

$$P(A \,|\, X_v = x_v) = P_{x_u}(A \,|\, X_v = x_v). \tag{5.32}$$

Proof. This proof is rather lengthy, and consists of two parts. First, we will show that there is a stochastic process P that satisfies Equations (5.31)$_\curvearrowleft$ and (5.32)$_\curvearrowleft$, by constructing it as the extension of a coherent conditional probability on a set of events $\mathscr{C} \subset \mathscr{C}^{\mathrm{SP}}$. Next, we will finish the proof by showing that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as desired.

Let $\mathscr{C} := \mathscr{C}_\emptyset \cup (\bigcup_{x_u \in \mathscr{X}_u} \mathscr{C}_{x_u})$, with $\mathscr{C}_\emptyset$ as in Equation (5.29)$_\curvearrowleft$ and, for all $x_u \in \mathscr{X}$, $\mathscr{C}_{x_u}$ as in Equation (5.30)$_\curvearrowleft$. Consider a real-valued function $\tilde{P}$ on $\mathscr{C}$ that is defined, for all $(A, X_v = x_v) \in \mathscr{C}$, by

$$\tilde{P}(A|X_v = x_v) := \begin{cases} P_\emptyset(A|X_v = x_v) & \text{if } (A, X_v = x_v) \in \mathscr{C}_\emptyset \\ P_{x_u}(A|X_{u\cup(v\setminus[0,\max u])} = x_{u\cup(v\setminus[0,\max u])}) & \text{if } (A, X_v = x_v) \in \mathscr{C}_{x_u} \end{cases} \tag{5.33}$$

We first prove that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$. So consider any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \dots, n\}$, choose $(A_i, C_i) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max \left\{ \sum_{i=1}^n \lambda_i \mathbb{I}_{C_i}(\omega)\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C_0 \right\} \geq 0, \tag{5.34}$$

with $C_0 := \cup_{i=1}^n C_i$.

Let $S^* := \{i \in \{1, \dots, n\} : (A_i, C_i) \in \mathscr{C}_\emptyset\}$ be the index set for the events that are in $\mathscr{C}_\emptyset$. We consider two cases. First, if $S^* \neq \emptyset$, then since $P_\emptyset$ is a stochastic process, it follows from Equation (5.33) and Definitions 2.12_{68} and 2.2_{48} that

$$\max \left\{ \sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega)\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big) \;\middle|\; \omega \in C^* \right\} \geq 0,$$

with $C^* := \cup_{i \in S^*} C_i$. Therefore, there is some $\omega^* \in C^*$ such that

$$\sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega^*)\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^*)\big) \geq 0. \tag{5.35}$$

For the other case, i.e. if $S^* = \emptyset$, we let ω^* be any element of C_0 (this is always possible, because $C_0 \neq \emptyset$). Clearly, this path ω^* will then also satisfy Equation (5.35)—because the left-hand side is a sum over an empty set and therefore zero.

Now we consider the states of the path ω^* at the time points u, i.e. we let $x_u^* \in \mathscr{X}_u$ be defined by $x_u^* := \omega^*|_u$. Then for all $i \in \{1, \dots, n\}$ such that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$, we know from Equation (5.30)$_\curvearrowleft$ that there are $u \subseteq v_i \in \mathscr{U}$ and $x_{v_i \setminus u} \in \mathscr{X}_{v_i \setminus u}$ such that

$$C_i = (X_u = x_u^*) \cap (X_{v_i \setminus u} = x_{v_i \setminus u}) = C_i^* \cap C_i^{**}, \tag{5.36}$$

with $C_i^* := (X_{(v_i \setminus u)\cap[0,\max u]} = x_{(v_i\setminus u)\cap[0,\max u]})$, and

$$C_i^{**} := (X_u = x_u^*) \cap (X_{v_i\setminus[0,\max u]} = x_{v_i\setminus[0,\max u]})\,. \tag{5.37}$$

To explain this somewhat opaque notation in words, we split the conditioning events C_i into a part C_i^* that captures the time points up to time $\max u$, but excluding the time points u themselves; and a part C_i^{**}, capturing the remaining time points.

Using this notation, we define

$$S^{**} := \{i \in \{1,\dots,n\} \colon (A_i, C_i) \in \mathscr{C}_{x_u^*} \text{ and } \mathbb{I}_{C_i^*}(\omega^*) = 1\}. \tag{5.38}$$

Thus, S^{**} is the index set of conditional events (A_i, C_i) that are in $\mathscr{C}_{x_u^*}$—such that C_i is compatible with ω^* on the time points u—*and* such that $\mathbb{I}_{C_i^*}(\omega^*) = 1$, which means that C_i is also compatible with ω^* on all *other* time points up to time $\max u$. In short, S^{**} simply contains the indices of the conditional events for which C_i is compatible with ω^* for all states up to time $\max u$.

We first consider the case $S^{**} \neq \emptyset$. Since $P_{x_u^*}$ is a stochastic process, it then follows from Definitions 2.12_{68} and 2.2_{48} that

$$\max\left\{\sum_{i\in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega)\big(P_{x_u^*}(A_i|C_i^{**}) - \mathbb{I}_{A_i}(\omega)\big)\;\middle|\;\omega\in C^{**}\right\} \geq 0,$$

with $C^{**} := \cup_{i\in S^{**}} C_i^{**}$. Because of Equation (5.33), this implies that

$$\max\left\{\sum_{i\in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega)\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega)\big)\;\middle|\;\omega\in C^{**}\right\} \geq 0,$$

which allows us to infer that there is some $\omega^{**} \in C^{**}$ such that

$$\sum_{i\in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega^{**})\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{**})\big) \geq 0. \tag{5.39}$$

Furthermore, since $\omega^{**} \in C^{**}$, Equation (5.37) implies that

$$\omega^{**}|_u = x_u^* = \omega^*|_u\,. \tag{5.40}$$

If $S^{**} = \emptyset$, we let $\omega^{**} = \omega^*$. Clearly, also in this case, ω^{**} satisfies Equations (5.39) and (5.40).

For any $i \in \{1,\dots,n\}$, because $(A_i, C_i) \in \mathscr{C}$, there is some finite sequence of time points $w_{C_i} \in \mathscr{U}$ such that C_i only depends on the time points in w_{C_i}. Furthermore, it follows from Equations $(5.29)_{215}$ and $(5.30)_{215}$ that A_i is an element of some algebra $\mathscr{A}$ that is generated by a set of events that only depend on a finite number of time points. Therefore, there is also some finite sequence of time points $w_{A_i} \in \mathscr{U}$, such that A_i only depends on the time points in w_{A_i}. If we now let $w_i \coloneqq w_{A_i} \cup w_{C_i}$, then (A_i, C_i) only depends on the (finite) sequence of time points w_i.

Because this holds for any $i \in \{1, \ldots, n\}$, this implies the existence of some finite sequence $w \in \mathscr{U}$ such that $w_i \subseteq w$ for all $i \in \{1, \ldots, n\}$.

Now let $\omega^{***} \in \Omega$ be any path such that, for all $s \in w$,

$$\omega^{***}(s) \coloneqq \begin{cases} \omega^*(s) & \text{if } s < \max u \\ \omega^{**}(s) & \text{if } s \geq \max u \end{cases}$$

Equation $(2.8)_{65}$ guarantees that this $\omega^{***} \in \Omega$ exists. Furthermore, because of Equation $(5.40)_{\curvearrowleft}$, we know that, for all $s \in w$,

$$\omega^{***}(s) = \omega^*(s) \ \text{ if } s \in [0, \max u] \tag{5.41}$$

and

$$\omega^{***}(s) = \omega^{**}(s) \ \text{ if } s \in u \cup [\max u, +\infty) \tag{5.42}$$

and therefore, it follows from Equation $(5.36)_{216}$ that

$$\omega^{***} \in C_i \Leftrightarrow (\omega^{***} \in C_i^* \text{ and } \omega^{***} \in C_i^{**}) \Leftrightarrow (\omega^* \in C_i^* \text{ and } \omega^{**} \in C_i^{**}) \tag{5.43}$$

for all $i \in \{1, \ldots, n\}$ such that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$.

Next, for any $i \in S^*$, we infer from Equation $(5.29)_{215}$ that the value of $\mathbb{I}_{A_i}(\omega^{***})$ and $\mathbb{I}_{C_i}(\omega^{***})$ is completely determined by $\omega^{***}(t)$, with $t \in (w \cap [0, \max u])$. Therefore, it follows from Equations $(5.35)_{216}$ and (5.41) that

$$\sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega^{***}) \big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{***})\big) \geq 0. \tag{5.44}$$

Similarly, for any $i \in S^{**}$, Equations (5.43) and $(5.38)_{\curvearrowleft}$ imply that $\mathbb{I}_{C_i}(\omega^{***}) = \mathbb{I}_{C_i^{**}}(\omega^{**})$, and Equations $(5.30)_{215}$ and (5.42) imply that $\mathbb{I}_{A_i}(\omega^{***}) = \mathbb{I}_{A_i}(\omega^{**})$. Therefore, it follows from Equation $(5.39)_{\curvearrowleft}$ that

$$\sum_{i \in S^{**}} \lambda_i \mathbb{I}_{C_i}(\omega^{***}) \big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{***})\big) \geq 0. \tag{5.45}$$

In summary, we have found a path ω^{***} that, by Equations (5.44) and (5.45), satisfies the coherence requirement when we only look at the events indexed by S^* and S^{**}. Let us next establish that the remaining events have no contribution to the coherence requirement for the path ω^{***}.

To this end, consider any $i \in \{1, \ldots, n\}$ such that $i \notin S^*$ and $i \notin S^{**}$. Since $i \notin S^*$, there is some $x_u \in \mathscr{X}_u$ such that $(A_i, C_i) \in \mathscr{C}_{x_u}$. If $x_u = x_u^*$, then since $i \notin S^{**}$, it follows from Equation $(5.38)_{\curvearrowleft}$ that $\mathbb{I}_{C_i^*}(\omega^*) = 0$, and therefore, Equation (5.43) implies that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. If $x_u \neq x_u^*$, then $(X_u = x_u) \cap (X_u = x_u^*) = \emptyset$, and therefore, since $(A_i, C_i) \in \mathscr{C}_{x_u}$ implies that $C_i \subseteq (X_u = x_u)$, it follows that $C_i \cap (X_u = x_u^*) = \emptyset$. Since it follows from Equations $(5.40)_{217}$ and (5.42) that $\omega^{***}(t) = x_t^*$ for all $t \in u$, this implies that $\omega^{***} \notin C_i$, and therefore, we find that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. Hence, in all cases, we find that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. Since this is true for any $i \in \{1, \dots, n\}$ such that $i \notin S^*$ and $i \notin S^{**}$, it follows from Equations (5.44) and (5.45) that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega^{***}) \big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{***})\big) \geq 0. \tag{5.46}$$

It now remains to prove that $\omega^{***} \in C_0$. We consider two cases: $S^* \neq \emptyset$ and $S^* = \emptyset$. First assume that $S^* \neq \emptyset$. In this case, we have that $\omega^* \in C^*$, which implies that there is some $i \in S^*$ such that $\omega^* \in C_i$. It then follows from Equations $(5.29)_{215}$ and (5.41) that $\omega^{***} \in C_i \subseteq C_0$. Next, assume that $S^* = \emptyset$. In this case, we have that $\omega^* \in C_0$, which implies that there is some $i \in \{1, \dots, n\}$ such that $\omega^* \in C_i$. Since $(A_i, C_i) \in \mathscr{C}$ and $S^* = \emptyset$, there is some $x_u \in \mathscr{X}_u$ such that $(A_i, C_i) \in \mathscr{C}_{x_u}$ and, since Equation $(5.30)_{215}$ implies that $x_t = \omega^*(t)$ for all $t \in u$, it follows that $x_u = x_u^*$. We conclude from this that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$. Furthermore, since $\omega^* \in C_i \subseteq C_i^*$, we know that $\mathbb{I}_{C_i^*}(\omega^*) = 1$. Therefore, it follows from Equation $(5.38)_{217}$ that $S^{**} \neq \emptyset$, which implies that $\omega^{**} \in C^{**}$. Hence, there is some $j \in S^{**}$ such that $\omega^{**} \in C_j^{**}$ and, since $j \in S^{**}$, we also know that $\mathbb{I}_{C_j^*}(\omega^*) = 1$, or equivalently, that $\omega^* \in C_j^*$. By combining this with Equation (5.43), it follows that $\omega^{***} \in C_j \subseteq C_0$. So, in all cases, we find that $\omega^{***} \in C_0$. By combining this with Equation (5.46), it follows that Equation $(5.34)_{216}$ holds, and therefore, that $\tilde{P}$ is coherent.

Since $\tilde{P}$ is coherent, and because $\mathscr{C} \subseteq \mathscr{C}^{\mathrm{SP}}$, it now follows from Theorem 2.3_{49} and Definition 2.12_{68} that $\tilde{P}$ can be extended to a stochastic process P. Furthermore, since P coincides with $\tilde{P}$ on $\mathscr{C}$, it follows from Equation $(5.33)_{216}$ that P satisfies Equations $(5.31)_{215}$ and $(5.32)_{215}$. This concludes the first part of this proof.

To conclude this proof, we will show that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as desired. First, observe that due to Equation $(5.29)_{215}$, we have for all $x \in \mathscr{X}$ that $(X_0 = y, X_\emptyset = x_\emptyset) \in \mathscr{C}_\emptyset$. Therefore, and because of Equation $(5.33)_{216}$, we find that for all $y \in \mathscr{X}$ it holds that $P(X_0 = y) = P(X_0 = y \,|\, X_\emptyset = x_\emptyset) = P_\emptyset(X_0 = y \,|\, X_\emptyset = x_\emptyset) = P_\emptyset(X_0 = y)$ which together with the fact that $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, implies that $P \sim \mathscr{M}$. Hence, in order to prove that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it remains to show that P is well-behaved as well as consistent with $\mathscr{Q}$.

In order to do this, we start by establishing an important equality. To this end, let P^* be any full conditional probability that coincides with P on $\mathscr{C}^{\mathrm{SP}}$; Corollary 2.20_{68} implies that such a full conditional probability always exists. Consider any $w \in \mathscr{U}$ and $s \in \mathbb{R}_{\geq 0}$ such that

$w < s$ and $u < s$. Then for all $x_w \in \mathscr{X}_w$ and $y \in \mathscr{X}$, we have that

$$\begin{aligned}
P(X_s = y|X_w = x_w) &= P^*(X_s = y|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y, X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y|X_{u\setminus w} = x_{u\setminus w}, X_w = x_w)P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} \tilde{P}(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P_{x_u}(X_s = y|X_u = x_u, X_{w\setminus[0,\max u]} = x_{w\setminus[0,\max u]}) \\
&\qquad\qquad P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w)
\end{aligned} \tag{5.47}$$

Using this equality, we will next show that for all $x \in \mathscr{X}$, and for any small enough $\Delta \in \mathbb{R}_{>0}$, the x-rows of the transition matrices $T_{t,x_v}^{t+\Delta}$ and $T_{t-\Delta,x_v}^{t}$ corresponding to P can each be written as a (different) convex combination of the x-rows of transition matrices corresponding to processes in $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. We note that the elements that make up this convex combination may depend on both x and on Δ.

Formally, we will show that for any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, there is some finite index set $\mathscr{I}$, some $v^* \in \mathscr{U}_{<t}$, and some $\delta > 0$, such that for all $x \in \mathscr{X}$ there is a family of stochastic processes $({}^iP \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, a family of state instantiations $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i\in\mathscr{I}}$, and a family of non-negative coefficients $(\lambda_i)_{i\in\mathscr{I}}$ that sum to one, such that for all $0 < \Delta < \delta$, it holds that

$$T_{t,x_v}^{t+\Delta}(x,\cdot) = \sum_{i\in\mathscr{I}} \lambda_i {}^iT_{t,{}^ix_{v^*}}^{t+\Delta}(x,\cdot)\,. \tag{5.48}$$

This has a lot of moving parts, and it is important to note that the index set $\mathscr{I}$ and the maximum difference δ only depend on t, v and x_v, while the families that are indexed by $\mathscr{I}$ can also depend on x.

Similarly, we will show that for any $t > 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, there is some finite index set $\mathscr{I}$, some $v^* \in \mathscr{U}_{<t}$, a $\delta > 0$, a family of stochastic processes $({}^iP \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, and a family of state instantiations $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i\in\mathscr{I}}$, such that for all $x \in \mathscr{X}$ and all $0 < \Delta < \delta$, there is a family of non-negative coefficients $(\lambda_i)_{i\in\mathscr{I}}$ that sum to one, such that

$$T_{t-\Delta,x_v}^{t}(x,\cdot) = \sum_{i\in\mathscr{I}} \lambda_i {}^iT_{t-\Delta,{}^ix_{v^*}}^{t}(x,\cdot). \tag{5.49}$$

Note that the coefficients $(\lambda_i)_{i\in\mathscr{I}}$ here depend on both x and on Δ, but that the other families do not depend on these quantities.

We start by constructing the convex combination that satisfies Equation (5.48). So consider any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$. We distinguish between two cases: $t < \max u$ and $t \geq \max u$. If $t < \max u$, then for all $\Delta \in (0, \max u - t)$ and $x, y \in \mathscr{X}$, we see that $(X_{t+\Delta} = y, (X_t = x, X_v = x_v)) \in \mathscr{C}_{\emptyset}$, and therefore, since P is an extension of $\tilde{P}$, it follows from Equation $(5.33)_{216}$ that

$$P(X_{t+\Delta} = y | X_t = x, X_v = x_v) = P_{\emptyset}(X_{t+\Delta} = y | X_t = x, X_v = x_v).$$

Hence, if we let $\mathscr{I} := \{1\}$, $v^* := v$, and $\delta := \max u - t$, and if for any $x \in \mathscr{X}$ we let ${}^1P := P_{\emptyset}$, ${}^1x_{v^*} := x_v$, and $\lambda_1 := 1$, then we see that Equation (5.48) is satisfied for any $0 < \Delta < \delta$. If $t \geq \max u$, then for all $\Delta > 0$ and $y \in \mathscr{X}$, it follows from Equation (5.47) (with $s := t + \Delta$ and $w := v \cup t$) that, for all $x_t \in \mathscr{X}_t$,

$$\begin{aligned} &P(X_{t+\Delta} = y | X_t = x_t, X_v = x_v) \\ &\quad = \sum_{x_{u\setminus(v\cup t)} \in \mathscr{X}_{u\setminus(v\cup t)}} P_{x_u}(X_{t+\Delta} = y | X_t = x_t, X_{(u\setminus t)\cup(v\setminus[0,\max u])} = x_{(u\setminus t)\cup(v\setminus[0,\max u])}) \\ &\qquad\qquad P^*(X_{u\setminus(v\cup t)} = x_{u\setminus(v\cup t)} | X_t = x_t, X_v = x_v). \end{aligned}$$

Now let $\mathscr{I} := \mathscr{X}_{u\setminus(v\cup t)}$, $v^* := (u \setminus t) \cup (v \setminus [0, \max u])$, and choose $\delta > 0$ arbitrarily. Fix any $x \in \mathscr{X}$, and for all $x_{u\setminus(v\cup t)} \in \mathscr{I}$, let ${}^{x_{u\setminus(v\cup t)}}P = P_{x_u}$—with $x_t := x$ if $t = \max u$, so P_{x_u} depends on x in this case—and ${}^{x_{u\setminus(v\cup t)}}x_{v^*} := x_{(u\setminus t)\cup(v\setminus[0,\max u])}$, and let

$$\lambda_{x_{u\setminus(v\cup t)}} := P^*(X_{u\setminus(v\cup t)} = x_{u\setminus(v\cup t)} | X_t = x, X_v = x_v).$$

Then Equation (5.48) is satisfied for any $0 < \Delta < \delta$. Hence, Equation (5.48) is satisfied both when $t < \max u$ and when $t \geq \max u$.

We will next construct the convex combination that satisfies Equation (5.49). So, consider any $t > 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$. We again distinguish between two cases: $t \leq \max u$ and $t > \max u$. If $t \leq \max u$, then for all $\Delta \in (0, t - \max v)$ and $x, y \in \mathscr{X}$, we see that $(X_t = y, (X_{t-\Delta} = x, X_v = x_v)) \in \mathscr{C}_{\emptyset}$, and therefore, since P is an extension of $\tilde{P}$, it follows from Equation $(5.33)_{216}$ that

$$P(X_t = y | X_{t-\Delta} = x, X_v = x_v) = P_{\emptyset}(X_t = y | X_{t-\Delta} = x, X_v = x_v).$$

Hence, if we let $\mathscr{I} := \{1\}$, $v^* := v$, $\delta := t - \max v$, ${}^1P := P_{\emptyset}$ and ${}^1x_{v^*} := x_v$, and if for all $x \in \mathscr{X}$ and all $0 < \Delta < \delta$ we let $\lambda_1 := 1$, then Equation (5.49) is satisfied.

If $t > \max u$, then for all $\Delta \in (0, t - \max(v \cup u))$ and $y \in \mathscr{X}$, it follows from Equation (5.47) (with $s := t$ and $w := v \cup \{t - \Delta\}$) that, for all $x_{t-\Delta} \in \mathscr{X}_{t-\Delta}$,

$$P(X_t = y | X_{t-\Delta} = x_{t-\Delta}, X_v = x_v) = \sum_{x_{u\setminus v} \in \mathscr{X}_{u\setminus v}} P_{x_u}(X_t = y | X_{t-\Delta} = x_{t-\Delta}, X_{u\cup(v\setminus[0,\max u])} = x_{u\cup(v\setminus[0,\max u])}) \\ P^*(X_{u\setminus v} = x_{u\setminus v} | X_{t-\Delta} = x_{t-\Delta}, X_v = x_v).$$

Now, let $\mathscr{I} := \mathscr{X}_{u\setminus v}$, $v^* := u \cup (v \setminus [0, \max u])$, $\delta := t - \max(v \cup u)$, and for all $x_{u\setminus v} \in \mathscr{I}$, let ${}^{x_{u\setminus v}}P = P_{x_u}$ and ${}^{x_{u\setminus v}}x_{v^*} := x_{u\cup(v\setminus[0,\max u])}$. If for all $x \in \mathscr{X}$ and $0 < \Delta < \delta$ we let $x_{t-\Delta} := x$ and

$$\lambda_{x_{u\setminus v}} := P^*(X_{u\setminus v} = x_{u\setminus v} | X_{t-\Delta} = x_{t-\Delta}, X_v = x_v),$$

then Equation $(5.49)_{220}$ is again satisfied. Hence, Equation $(5.49)_{220}$ can be satisfied both when $t \leq \max u$ and when $t > \max u$.

Therefore, indeed, as claimed before, the x-rows of both $T_{t,x_v}^{t+\Delta}$ and $T_{t-\Delta,x_v}^{t}$ can be written as a convex combination of the x-rows of transition matrices corresponding to elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$—assuming that Δ is small enough. We will now use this fact to prove that P is well-behaved and consistent with $\mathscr{Q}$.

We start by proving that P is well-behaved. First fix any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, and consider the quantities $\mathscr{I}$, v^*, and δ leading to the convex combination that satisfies Equation $(5.48)_{220}$. Fix any $x \in \mathscr{X}$, and consider the indexed families $({}^iP \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i\in\mathscr{I}}$, and $(\lambda_i)_{i\in\mathscr{I}}$ that together satisfy Equation $(5.48)_{220}$ for any $0 < \Delta < \delta$.

Fix any $y \in \mathscr{X}$. Then for all $i \in \mathscr{I}$, because ${}^iP$ is well-behaved, it follows from Definition 4.1_{145} that there is some $B_i \in \mathbb{R}$ such that

$$\limsup_{\Delta\to 0^+} \frac{1}{\Delta} \left| {}^iP(X_{t+\Delta} = y \,|\, X_t = x, X_{v^*} = {}^ix_{v^*}) - \mathbb{I}_x(y) \right| \leq B_i \,.$$

Because this is true for all $i \in \mathscr{I}$, and because $\mathscr{I}$ is finite, it follows that $B := \max_{i\in\mathscr{I}} B_i$ is in $\mathbb{R}$—in particular, that it is finite. Using that Equation $(5.48)_{220}$ holds for all $0 < \Delta < \delta$, together with Definition 4.2_{148} and the fact that $\mathscr{I}$ is finite, it follows that

$$\begin{aligned} &\limsup_{\Delta\to 0^+} \frac{1}{\Delta} \left| P(X_{t+\Delta} = y \,|\, X_t = x, X_v = x_v) - \mathbb{I}_x(y) \right| \\ &= \limsup_{\Delta\to 0^+} \frac{1}{\Delta} \left| \sum_{i\in\mathscr{I}} \lambda_i \left({}^iP(X_{t+\Delta} = y \,|\, X_t = x, X_{v^*} = {}^ix_{v^*}) - \mathbb{I}_x(y) \right) \right| \\ &\leq \limsup_{\Delta\to 0^+} \sum_{i\in\mathscr{I}} \lambda_i \frac{1}{\Delta} \left| {}^iP(X_{t+\Delta} = y \,|\, X_t = x, X_{v^*} = {}^ix_{v^*}) - \mathbb{I}_x(y) \right| \\ &\leq \sum_{i\in\mathscr{I}} \lambda_i \limsup_{\Delta\to 0^+} \frac{1}{\Delta} \left| {}^iP(X_{t+\Delta} = y \,|\, X_t = x, X_{v^*} = {}^ix_{v^*}) - \mathbb{I}_x(y) \right| \leq \sum_{i\in\mathscr{I}} \lambda_i B = B \,. \end{aligned}$$

In summary, we have found that

$$\limsup_{\Delta\to 0^+} \frac{1}{\Delta}\left|P(X_{t+\Delta}=y\,|\,X_t=x, X_v=x_v) - \mathbb{I}_x(y)\right| \le B < +\infty. \tag{5.50}$$

Similarly, fix any $t>0$, $v\in\mathscr{U}_{<t}$ and $x_v\in\mathscr{X}_v$, and consider the quantities $\mathscr{I}$, v^*, δ, $({}^iP\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, and $({}^ix_{v^*}\in\mathscr{X}_{v^*})_{i\in\mathscr{I}}$ leading to the convex combination that satisfies Equation $(5.49)_{220}$.

Fix any $x,y\in\mathscr{X}$. Then for all $i\in\mathscr{I}$, because ${}^iP$ is well-behaved, it follows from Definition 4.1_{145} that there is some $B_i\in\mathbb{R}$ such that

$$\limsup_{\Delta\to 0^+} \frac{1}{\Delta}\left|{}^iP(X_t=y\,|\,X_{t-\Delta}=x, X_{v^*}={}^ix_{v^*}) - \mathbb{I}_x(y)\right| \le B_i\,.$$

This implies that there is some $\delta_i'>0$ such that for all $0<\Delta<\delta_i'$ it holds that

$$\frac{1}{\Delta}\left|{}^iP(X_t=y\,|\,X_{t-\Delta}=x, X_{v^*}={}^ix_{v^*}) - \mathbb{I}_x(y)\right| \le B_i\,.$$

Now let $\delta' \coloneqq \min_{i\in\mathscr{I}}\delta_i'$ and $B\coloneqq \max_{i\in\mathscr{I}} B_i$. Then because $\mathscr{I}$ is finite it follows that $B\in\mathbb{R}$—in particular, that it is finite—and that $0<\delta'$.

Now fix any $0<\Delta<\min\{\delta,\delta'\}$, and consider the family $(\lambda_i)_{i\in\mathscr{I}}$ that satisfies Equation $(5.49)_{220}$. Then it follows from Definition 4.2_{148} that

$$\begin{aligned}
&\frac{1}{\Delta}\left|P(X_t=y\,|\,X_{t-\Delta}=x, X_v=x_v) - \mathbb{I}_x(y)\right| \\
&= \frac{1}{\Delta}\left|\sum_{i\in\mathscr{I}}\lambda_i\big({}^iP(X_t=y\,|\,X_{t-\Delta}=x, X_{v^*}={}^ix_{v^*}) - \mathbb{I}_x(y)\big)\right| \\
&\le \sum_{i\in\mathscr{I}}\lambda_i\frac{1}{\Delta}\left|{}^iP(X_t=y\,|\,X_{t-\Delta}=x, X_{v^*}={}^ix_{v^*}) - \mathbb{I}_x(y)\right| \le \sum_{i\in\mathscr{I}}\lambda_i B = B\,,
\end{aligned}$$

where for the final inequality we used that $\Delta<\delta'\le\delta_i'$ for all $i\in\mathscr{I}$. Because this is true for all $0<\Delta<\min\{\delta,\delta'\}$, it follows that

$$\limsup_{\Delta\to 0^+} \frac{1}{\Delta}\left|P(X_t=y\,|\,X_{t-\Delta}=x, X_v=x_v) - \mathbb{I}_x(y)\right| \le B < +\infty. \tag{5.51}$$

Because the $t\in\mathbb{R}_{\ge 0}$, $v\in\mathscr{U}_{<t}$ and $x_v\in\mathscr{X}_v$ in Equation (5.50), and the $t\in\mathbb{R}_{>0}$, $v\in\mathscr{U}_{<t}$ and $x_v\in\mathscr{X}_v$ in Equation (5.51) are arbitrary, it follows from Definition 4.1_{145} that P is well-behaved.

We end by proving that P is consistent with $\mathscr{Q}$. To this end, fix any $t\ge 0$, $v\in\mathscr{U}_{<t}$, and $x_v\in\mathscr{X}_v$. We need to show that for all $Q^*\in\overline{\partial}T^t_{t,x_v}$ it holds that $Q^*\in\mathscr{Q}$. We consider two (possibly overlapping) cases: $Q^*\in\overline{\partial}_+T^t_{t,x_v}$ and $Q^*\in\overline{\partial}_-T^t_{t,x_v}$.

If $Q^* \in \overline{\partial}_+ T_{t,x_v}^t$, it follows from Equation (4.22)$_{168}$ that there is a sequence $\{\Delta_j\}_{j\in\mathbb{Z}_{>0}} \to 0^+$ such that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}\left(T_{t,x_v}^{t+\Delta_j} - I\right) = Q^* . \tag{5.52}$$

Consider the $\mathscr{I}$, $v^* \in \mathscr{U}_{<t}$, and δ that lead to the convex combination satisfying Equation (5.48)$_{220}$. Fix any $x \in \mathscr{X}$, and consider the indexed families $({}^iP \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, $({}^ix_{v^*})_{i\in\mathscr{I}}$, and $(\lambda_i)_{i\in\mathscr{I}}$ that together satisfy Equation (5.48)$_{220}$ for any $0 < \Delta < \delta$.

Fix any $i \in \mathscr{I}$. Since ${}^iP$ is well-behaved, the sequence

$$\left\{\frac{1}{\Delta_j}\left({}^iT_{t,{}^ix_{v^*}}^{t+\Delta_j} - I\right)\right\}_{j\in\mathbb{Z}_{>0}}$$

is bounded, and therefore, Corollary A.14$_{378}$ implies that it has a convergent subsequence, of which we denote the limit by $Q_{i,x}$. Hence, without loss of generality—simply remove the indices j that do not correspond to the subsequence—we may assume that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}\left({}^iT_{t,{}^ix_{v^*}}^{t+\Delta_j} - I\right) = Q_{i,x}. \tag{5.53}$$

Because $i \in \mathscr{I}$ is arbitrary, we can now repeat this argument for the next $i' \in \mathscr{I}$, and so on—each time removing indices j from the remaining *sub*-sequence—until eventually, we can assume without loss of generality that Equation (5.53) holds for all $i \in \mathscr{I}$.

Then, for any $i \in \mathscr{I}$, since we know from Proposition 4.10$_{153}$ that $Q_{i,x}$ is a limit of rate matrices, $Q_{i,x}$ is also a rate matrix due to Proposition 4.6$_{151}$, and therefore, it follows from Equation (4.22)$_{168}$ that $Q_{i,x} \in \overline{\partial}_+ {}^iT_{t,{}^ix_{v^*}}^t$, which, since ${}^iP$ is consistent with $\mathscr{Q}$, implies that $Q_{i,x} \in \mathscr{Q}$. In this manner, we obtain a family of rate matrices $(Q_{i,x} \in \mathscr{Q})_{i\in\mathscr{I}}$ such that Equation (5.53) holds for every $i \in \mathscr{I}$. Additionally, since $\lim_{j\to+\infty}\Delta_j = 0$, we may assume without loss of generality that $0 < \Delta_j < \delta$ for all $j \in \mathbb{Z}_{>0}$. Equations (5.48)$_{220}$ and (5.53) now imply that

$$\begin{aligned}\lim_{j\to+\infty} \frac{1}{\Delta_j}\left(T_{t,x_v}^{t+\Delta_j}(x,\cdot) - I(x,\cdot)\right) &= \lim_{j\to+\infty} \frac{1}{\Delta_j}\left(\sum_{i\in\mathscr{I}} \lambda_i {}^iT_{t,{}^ix_{v^*}}^{t+\Delta_j}(x,\cdot) - I(x,\cdot)\right) \\ &= \sum_{i\in\mathscr{I}} \lambda_i Q_{i,x}(x,\cdot),\end{aligned}$$

which, because of Equation (5.52), implies that $Q^*(x,\cdot) = \sum_{i\in\mathscr{I}} \lambda_i Q_{i,x}(x,\cdot) = Q_x(x,\cdot)$, with $Q_x := \sum_{i\in\mathscr{I}} \lambda_i Q_{i,x}$. Since $\mathscr{Q}$ is convex, we also know that $Q_x \in \mathscr{Q}$. Hence, since $\mathscr{Q}$ has separately specified rows, and because $x \in \mathscr{X}$ is arbitrary, it follows from Definition 5.7_{193} that $Q^* \in \mathscr{Q}$. This concludes the argument in the case when $Q^* \in \overline{\partial}_+ T^t_{t,x_v}$.

Next, suppose that $Q^* \in \overline{\partial}_- T^t_{t,x_v}$. Then, due to Equation $(4.23)_{168}$, there is a sequence $\{\Delta_j\}_{j\in\mathbb{Z}_{>0}} \to 0^+$ such that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}\left(T^t_{t-\Delta_j,x_v} - I\right) = Q^* . \tag{5.54}$$

Consider the quantities $\mathscr{I}$, v^*, δ, $({}^iP \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}})_{i\in\mathscr{I}}$, and $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i\in\mathscr{I}}$ leading to the convex combination that satisfies Equation $(5.49)_{220}$. Because $\lim_{j\to+\infty}\Delta_j = 0$ we can assume without loss of generality that $0 < \Delta_j < \delta$ for all $j \in \mathbb{Z}_{>0}$. Fix any $x \in \mathscr{X}$. Then for all $j \in \mathbb{Z}_{>0}$, there is a family $({}^j\lambda_i)_{i\in\mathscr{I}}$ of non-negative coefficients that sum to one, that satisfies Equation $(5.49)_{220}$ for $\Delta = \Delta_j$.

Now consider the sequence $\left\{({}^j\lambda_i)_{i\in\mathscr{I}}\right\}_{j\in\mathbb{Z}_{>0}}$ of these families of coefficients. Let $k := |\mathscr{I}|$ denote the size of the index set $\mathscr{I}$. Then the sequence $\left\{({}^j\lambda_i)_{i\in\mathscr{I}}\right\}_{j\in\mathbb{Z}_{>0}}$ lives in $[0,1]^{\mathscr{I}}$, which is clearly a closed and bounded subset of the k-dimensional vector space $\mathbb{R}^{\mathscr{I}}$ of real-valued functions on $\mathscr{I}$, which we endow with the supremum norm. Hence, by Corollary $\mathrm{A}.12_{378}$, there is a convergent subsequence, whose limit we denote by $({}^*\lambda_i)_{i\in\mathscr{I}}$, that also lives in $[0,1]^{\mathscr{I}}$. Without loss of generality—simply remove the indexes j that do not correspond to the subsequence—we may therefore assume that

$$\lim_{j\to+\infty} \max_{i\in\mathscr{I}} \left|{}^j\lambda_i - {}^*\lambda_i\right| = 0 . \tag{5.55}$$

It clearly holds that ${}^*\lambda_i \geq 0$ for all $i \in \mathscr{I}$. Moreover, for any $j \in \mathbb{Z}_{>0}$ it holds that

$$\begin{aligned}\left|\sum_{i\in\mathscr{I}} {}^*\lambda_i - 1\right| &\leq \left|\sum_{i\in\mathscr{I}} {}^*\lambda_i - {}^j\lambda_i\right| + \left|\sum_{i\in\mathscr{I}} {}^j\lambda_i - 1\right| \\ &= \left|\sum_{i\in\mathscr{I}} {}^*\lambda_i - {}^j\lambda_i\right| \leq \sum_{i\in\mathscr{I}} \left|{}^*\lambda_i - {}^j\lambda_i\right| \leq k \max_{i\in\mathscr{I}} \left|{}^*\lambda_i - {}^j\lambda_i\right| ,\end{aligned}$$

where for the equality we used that $\sum_{i\in\mathscr{I}} {}^j\lambda_i = 1$. Since this is true for all $j \in \mathbb{Z}_{>0}$, and since $k \in \mathbb{Z}_{>0}$, it follows from Equation (5.55) that $|\sum_{i\in\mathscr{I}} {}^*\lambda_i - 1| = 0$, or in other words, that $\sum_{i\in\mathscr{I}} {}^*\lambda_i = 1$. So, in summary, we have found that the (sub-)sequence $\left\{({}^j\lambda_i)_{i\in\mathscr{I}}\right\}_{j\in\mathbb{Z}_{>0}}$ converges to a limit $({}^*\lambda_i)_{i\in\mathscr{I}}$ which is a family of non-negative coefficients that sum to one.

Now fix any $i \in \mathscr{I}$. Since ${}^iP$ is well-behaved, the sequence

$$\left\{\frac{1}{\Delta_j}\left({}^iT^t_{t-\Delta_j,{}^ix_{v^*}} - I\right)\right\}_{j\in\mathbb{Z}_{>0}}$$

is bounded, and therefore, Corollary A.14$_{378}$ implies that it has a convergent subsequence, of which we denote the limit by $Q_{i,x}$. Hence, without loss of generality—again simply remove the indexes j that do not correspond to the subsequence—we may assume that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}\left({}^{i}T^{t}_{t-\Delta_j,{}^{i}x_{v^*}} - I\right) = Q_{i,x}. \tag{5.56}$$

Because $i \in \mathscr{I}$ is arbitrary, we can now repeat this argument for the next $i' \in \mathscr{I}$, and so on—each time removing indices j from the remaining *sub*-sequence—until eventually, we can assume without loss of generality that Equation (5.56) holds for all $i \in \mathscr{I}$.

Then, for any $i \in \mathscr{I}$, since we know from Proposition 4.10$_{153}$ that $Q_{i,x}$ is a limit of rate matrices, $Q_{i,x}$ is also a rate matrix due to Proposition 4.6$_{151}$, and therefore, it follows from Equation (4.23)$_{168}$ that $Q_{i,x} \in \overline{\partial}_{-}{}^{i}T^{t}_{t,{}^{i}x_{v^*}}$, which, since ${}^{i}P$ is consistent with $\mathscr{Q}$, implies that $Q_{i,x} \in \mathscr{Q}$. In this manner, we obtain a family of rate matrices $(Q_{i,x} \in \mathscr{Q})_{i\in\mathscr{I}}$ such that Equation (5.56) holds for every $i \in \mathscr{I}$.

Because we already assumed that $0 < \Delta_j < \delta$ for all $j \in \mathbb{Z}_{>0}$, Equation (5.49)$_{220}$ implies that

$$\begin{aligned}
\lim_{j\to+\infty} \frac{1}{\Delta_j}\left(T^{t}_{t-\Delta_j,x_v}(x,\cdot) - I(x,\cdot)\right) &= \lim_{j\to+\infty} \frac{1}{\Delta_j}\left(\sum_{i\in\mathscr{I}} {}^{j}\lambda_i {}^{i}T^{t}_{t-\Delta_j,x_v}(x,\cdot) - I(x,\cdot)\right)\\
&= \lim_{j\to+\infty} \sum_{i\in\mathscr{I}} {}^{j}\lambda_i \frac{1}{\Delta_j}\left({}^{i}T^{t}_{t-\Delta_j,x_v}(x,\cdot) - I(x,\cdot)\right)\\
&= \sum_{i\in\mathscr{I}} {}^{*}\lambda_i Q_{i,x}(x,\cdot),
\end{aligned}$$

where for the last step we used the individual limits that we established above, i.e. that for all $i \in \mathscr{I}$ it holds that $\lim_{j\to+\infty} {}^{j}\lambda_i = {}^{*}\lambda_i$ and $\lim_{j\to+\infty} {}^{1}/_{\Delta_j}\left({}^{i}T^{t}_{t-\Delta_j,x_v}(x,\cdot) - I(x,\cdot)\right) = Q_{i,x}(x,\cdot)$—this last property follows from Proposition A.33$_{390}$ together with Equation (5.56). Due to Equation (5.54)$_{\frown}$, this implies that $Q^*(x,\cdot) = \sum_{i\in\mathscr{I}} {}^{*}\lambda_i Q_{i,x}(x,\cdot) = Q_x(x,\cdot)$, with $Q_x := \sum_{i\in\mathscr{I}} {}^{*}\lambda_i Q_{i,x}$. Because $\mathscr{Q}$ is convex, and because the coefficients $({}^{*}\lambda_i)_{i\in\mathscr{I}}$ are non-negative and sum to one, we also know that $Q_x \in \mathscr{Q}$. Hence, since $\mathscr{Q}$ has separately specified rows and because $x \in \mathscr{X}$ is arbitrary, it follows from Definition 5.7$_{193}$ that $Q^* \in \mathscr{Q}$.

Since $Q^* \in \mathscr{Q}$ both when $Q^* \in \overline{\partial}_{+}T^{t}_{t,x_v}$ and $Q^* \in \overline{\partial}_{-}T^{t}_{t,x_v}$, and because $Q^* \in \overline{\partial}T^{t}_{t,x_v}$ is arbitrary, we conclude that $\overline{\partial}T^{t}_{t,x_v} \subseteq \mathscr{Q}$. Since this is true for all $t \in \mathbb{R}_{\geq 0}$, $v \in \mathscr{U}_{<t}$, and $x_v \in \mathscr{X}_v$, we find that P is consistent with $\mathscr{Q}$. Since we already know that P is a well-behaved stochastic process such that $P \sim \mathscr{M}$, it follows that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. □

Proof of Theorem 5.11$_{193}$. If $u = \emptyset$ then the statement is satisfied by taking $P = P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. To see this, observe that Equation (5.6)$_{194}$ is vacuously satisfied since there is no $u_2 \subseteq u$ for which $u_2 \neq \emptyset$, while Equation $(5.7)_{194}$ is trivially satisfied.

For the other case, suppose that $u \neq \emptyset$; then this result is a special case of Lemma 5.35_{215}. To see this, first note that, for all $u_1, u_2 \subseteq u$ such that $u_2 \neq \emptyset$ and $u_1 < u_2$, all $x_{u_1} \in \mathscr{X}_{u_1}$, and all $x_{u_2} \in \mathscr{X}_{u_2}$, the conditional event $(X_{u_2} = x_{u_2}, X_{u_1} = x_{u_1})$ is an element of $\mathscr{C}_\emptyset$, as defined in the statement of Lemma 5.35_{215}. Moreover, observe that for all $x_u \in \mathscr{X}_u$ and all $A \in \mathscr{A}_u$, it follows from Definition 2.10_{67} that $(A, X_u = x_u) \in \mathscr{C}_{x_u}$, as defined in the statement of Lemma 5.35_{215}. $\square$

5.C Technical (In)Equalities for CT(I)MCs

This appendix contains a number of identities that we require for some of our proofs, but which we think are too technical to explain in the main text of this dissertation.

First of all, recall from Proposition 4.23_{171} that, for any $\varepsilon \in \mathbb{R}_{>0}$, there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$, the transition matrix $T_{t,x_u}^{t+\Delta}$ of a fixed stochastic process $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ can be approximated by $(I + \Delta Q)$ with an error of at most $\Delta\varepsilon$, using a rate matrix $Q \in \overline{\partial} T_{t,x_u}^{t} \subseteq \mathscr{Q}$. Quite similarly, the following lemma states that, for any given time interval $[t,s]$, there is a finite partition $v \in \mathscr{U}_{[t,s]}$, $v = t_0, \dots, t_n$, such that the transition matrices $T_{t_i,x_u}^{t_{i+1}}$ can all be approximated by $(I + \Delta_{i+1}^{v} Q_{i+1})$, for some $Q_{i+1} \in \mathscr{Q}$ and with $\Delta_{i+1}^{v} = t_{i+1} - t_i$.

The reason that this result does not follow trivially from Proposition 4.23_{171} is because the δ—and hence also Δ—in Proposition 4.23_{171} depends on the particular time point that is considered. For this reason, the intuitive idea of using Proposition 4.23_{171} to first find some Δ_1 and Q_1 such that $T_{t_0,x_u}^{t_1} = T_{t_0,x_u}^{t_0+\Delta_1}$ can be approximated by $I + \Delta_1 Q_1$, and to then continue in this way to find some Δ_2 and Q_2, and then some Δ_3 and Q_3, and so on, is not feasible, because this process may continue indefinitely if $\sum_{i=1}^{\infty} \Delta_i$ is finite. In order to make this work, we need some kind of guarantee that it suffices to consider a finite number of Δ_i, and this is exactly what the following lemma establishes.

Lemma 5.36. *Let $\mathscr{Q}$ be a non-empty set of rate matrices, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Consider any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding family of history-dependent transition matrices (T_{t,x_u}^{s}), any $t, s \in \mathbb{R}_{\geq 0}$ such that $t < s$, any $u \in \mathscr{U}_{<t}$ and any $x_u \in \mathscr{X}_u$. Then for all $\varepsilon > 0$ and $\delta > 0$, there is some $v \in \mathscr{U}_{[t,s]}$, $v = t_0, \dots, t_n$, such that $\sigma(v) < \delta$ and, for all $i \in \{0, \dots, n-1\}$, it holds that*

$$(\exists Q \in \mathscr{Q}) \; \left\| T_{t_i,x_u}^{t_{i+1}} - (I + \Delta_{i+1}^{v} Q) \right\| < \Delta_{i+1}^{v} \varepsilon \,.$$

Proof. Fix any $\varepsilon > 0$ and $\delta > 0$. It then follows from Proposition 4.23171 and Definition 5.3189 that there is some $0 < \delta^* < \min\{\delta, 1/2(s-t)\}$ such that, for all $0 < \Delta < \delta^*$,

$$(\exists Q \in \mathscr{Q}) \left\| \frac{1}{\Delta}(T_{t,x_u}^{t+\Delta} - I) - Q \right\| < \varepsilon \text{ and } (\exists Q \in \mathscr{Q}) \left\| \frac{1}{\Delta}(T_{s-\Delta,x_u}^{s} - I) - Q \right\| < \varepsilon. \tag{5.57}$$

Let $t^* := t + \delta^*$ and $s^* := s - \delta^*$. Then clearly, $t < t^* < s^* < s$. For any $r \in [t^*, s^*]$, it follows from Proposition 4.23171 and Definition 5.3189 that there is some $0 < \delta_r < \delta^*$ such that, for all $0 < \Delta < \delta_r$,

$$(\exists Q \in \mathscr{Q}) \left\| \frac{1}{\Delta}(T_{r,x_u}^{r+\Delta} - I) - Q \right\| < \varepsilon \text{ and } (\exists Q \in \mathscr{Q}) \left\| \frac{1}{\Delta}(T_{r-\Delta,x_u}^{r} - I) - Q \right\| < \varepsilon. \tag{5.58}$$

Let $U_r := (r - \delta_r, r + \delta_r)$. Then the set $C := \{U_r : r \in [t^*, s^*]\}$ is an open cover of $[t^*, s^*]$. By the Heine-Borel theorem, C contains a finite subcover C^* of $[t^*, s^*]$. Without loss of generality, we can take this subcover to be minimal, in the sense that if we remove any of its elements, it is no longer a cover. Let m be the cardinality of C^* and let $r_1 < r_2 < \cdots < r_m$ be the ordered sequence of the midpoints of the intervals in C^*.

We will now prove that

$$r_i - \delta_{r_i} < r_j - \delta_{r_j} \text{ and } r_i + \delta_{r_i} < r_j + \delta_{r_j} \text{ for all } 1 \leq i < j \leq m. \tag{5.59}$$

Assume *ex absurdo* that this statement is not true. Then this implies that there are $1 \leq i < j \leq m$ such that either $r_i - \delta_{r_i} \geq r_j - \delta_{r_j}$ or $r_i + \delta_{r_i} \geq r_j + \delta_{r_j}$. If $r_i - \delta_{r_i} \geq r_j - \delta_{r_j}$, then since $i < j$ implies that $r_i < r_j$, it follows that $\delta_{r_j} \geq \delta_{r_i} + r_j - r_i > \delta_{r_i}$ and therefore, that $r_j + \delta_{r_j} > r_i + \delta_{r_i}$. Hence, we find that $U_{r_i} \subseteq U_{r_j}$. Since C^* was taken to be a minimal cover, this is a contradiction. Similarly, if $r_i + \delta_{r_i} \geq r_j + \delta_{r_j}$, then since $i < j$ implies that $r_i < r_j$, it follows that $\delta_{r_i} \geq \delta_{r_j} + r_j - r_i > \delta_{r_j}$ and therefore, that $r_i - \delta_{r_i} < r_j - \delta_{r_i}$. Hence, we find that $U_{r_j} \subseteq U_{r_i}$. Since C^* was taken to be a minimal cover, this is again a contradiction. From these two contradictions, it follows that Equation (5.59) is indeed true.

Next, we prove that

$$r_{k+1} - \delta_{r_{k+1}} < r_k + \delta_{r_k} \text{ for all } k \in \{1, \ldots, m-1\}. \tag{5.60}$$

Assume *ex absurdo* that this statement is not true or, equivalently, that there is some $k \in \{1, \ldots, m-1\}$ such that $r_k + \delta_{r_k} \leq r_{k+1} - \delta_{r_{k+1}}$. For all $i \in \{k+1, \ldots, m\}$, it then follows from Equation (5.59) that $r_k + \delta_{r_k} \leq r_i - \delta_{r_i}$, which implies that $r_k + \delta_{r_k} \notin U_{r_i}$. Furthermore, for all $i \in \{1, \ldots, k\}$, it follows from Equation (5.59) that $r_i + \delta_{r_i} \leq r_k + \delta_{r_k}$, which again implies that $r_k + \delta_{r_k} \notin U_{r_i}$. Hence, for all $i \in \{1, \ldots, m\}$, we have found that $r_k + \delta_{r_k} \notin U_{r_i}$. Since C^* is a cover of $[t^*, s^*]$, this implies that $r_k + \delta_{r_k} \notin [t^*, s^*]$, which, since $r_k \in [t^*, s^*]$, implies that $r_k + \delta_{r_k} > s^*$. Hence, since we know from Equation (5.59) that $r_k - \delta_{r_k} < r_m - \delta_{r_m}$, it follows that $U_{r_m} \cap [t^*, s^*] \subseteq U_{r_k} \cap [t^*, s^*]$. This contradicts the fact that C^* was taken to be a minimal cover, and therefore, Equation (5.60) must indeed be true.

For all $k \in \{1, \ldots, m-1\}$, we define $q_k := 1/2(r_k + \delta_{r_k} + r_{k+1} - \delta_{r_{k+1}})$. Using Equation (5.59), it then follows that

$$q_k < \frac{r_{k+1} + \delta_{r_{k+1}} + r_{k+1} - \delta_{r_{k+1}}}{2} = r_{k+1} \text{ and } q_k > \frac{r_k + \delta_{r_k} + r_k - \delta_{r_k}}{2} = r_k,$$

and Equation (5.60) trivially implies that $r_{k+1} - \delta_{r_{k+1}} < q_k < r_k + \delta_{r_k}$. Hence,

$$r_k < q_k < r_k + \delta_{r_k} \text{ and } r_{k+1} - \delta_{r_{k+1}} < q_k < r_{k+1}.$$

Due to Equation (5.58), and with $\Delta_k^* := q_k - r_k$ and $\Delta_k^{**} := r_{k+1} - q_k$, this implies that

$$(\exists Q \in \mathscr{Q}) \left\| \frac{(T_{r_k, x_u}^{q_k} - I)}{\Delta_k^*} - Q \right\| < \varepsilon \text{ and } (\exists Q \in \mathscr{Q}) \left\| \frac{(T_{q_k, x_u}^{r_{k+1}} - I)}{\Delta_k^{**}} - Q \right\| < \varepsilon. \tag{5.61}$$

For all $k \in \{1, \ldots, m\}$, we now let $t_{2k} := r_k$ and, for all $k \in \{1, \ldots, m-1\}$, we let $t_{2k+1} := q_k$. For the resulting sequence $t_2 < t_3 < \cdots < t_{2m-1} < t_{2m}$, it then follows from Equation (5.61) and N9$_{64}$ that, for all $i \in \{2, \ldots, 2m-1\}$:

$$(\exists Q \in \mathscr{Q}) \left\| T_{t_i, x_u}^{t_{i+1}} - (I + \Delta_{i+1} Q) \right\| < \Delta_{i+1} \varepsilon, \tag{5.62}$$

with $\Delta_{i+1} := t_{i+1} - t_i < \delta$.

Next, since C^* is a minimal cover, and because of Equation (5.59), we know that $r_1 - \delta_{r_1} < t^* \leq r_1 = t_2$ and, since $\delta_{r_1} < \delta^*$, we also know that $r_1 - \delta_{r_1} > t$. Therefore, it follows that there is some $t_1 \in \mathbb{R}$ such that $t < r_1 - \delta_{r_1} < t_1 < t^* \leq r_1$. If we now let $t_0 := t$, then $\Delta_1 := t_1 - t_0 < \delta^*$ and $\Delta_2 := t_2 - t_1 = r_1 - t_1 < \delta_{r_1}$, and therefore, it follows from Equations (5.57) and (5.58) and N9$_{64}$ that Equation (5.62) is also true for $i = 0$ and $i = 1$.

Finally, again since C^* is a minimal cover and because of Equation (5.59), we know that $t_{2m} = r_m \leq s^* < r_m + \delta_{r_m}$ and, since $\delta_{r_m} < \delta^*$, we also know that $r_m + \delta_{r_m} < s$. Therefore, it follows that there is some $t_{2m+1} \in \mathbb{R}$ such that $t_{2m} = r_m \leq s^* < t_{2m+1} < r_m + \delta_{r_m} < s$. If we now let $t_{2m+2} := s$, then $\Delta_{2m+2} := t_{2m+2} - t_{2m+1} < \delta^*$ and $\Delta_{2m+1} := t_{2m+1} - t_{2m} = t_{2m+1} - r_m < \delta_{r_m}$, and therefore, it follows from Equations (5.57) and (5.58) and N9$_{64}$ that Equation (5.62) is also true for $i = 2m$ and $i = 2m+1$.

Hence, we conclude that Equation (5.62) holds for all $i \in \{0, 1, \ldots, 2m, 2m+1\}$. The result now follows by letting $n := 2m+2$. □

Roughly speaking, the next result establishes that the (history-dependent) transition matrix T_{t,x_u}^s of a stochastic process $P \in \mathbb{P}^{\mathrm{W}}$ can be decomposed into a number of other transition matrices corresponding to this P, such that—crucially—these individual transition matrices do not depend on the exact way that the decomposition was performed—that is, the other transition matrices in the decomposition—despite P not necessarily being a Markov process.

Lemma 5.37. *Consider any $P \in \mathbb{P}^{\mathrm{W}}$ with corresponding family of history-dependent transition matrices (T_{t,x_u}^s), any $t,s \in \mathbb{R}_{\geq 0}$ such that $t < s$, any $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ and any sequence $t = t_0 < t_1 < \cdots < t_n = s$, with $n \in \mathbb{Z}_{>0}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $x_t \in \mathscr{X}$, it holds that*

$$T_{t,x_u}^s f(x_t) = \left(T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right) f(x_t)$$

Proof. We provide a proof by induction. For $n = 1$, the result holds trivially. So consider now any $n > 1$ and assume that the result is true for $n-1$.

For any $g \in \mathscr{L}(\mathscr{X})$, and because $t_0 = t$, it then follows from Definition 4.2$_{148}$ that

$$\begin{aligned}
T_{t_0,x_u}^{t_2} g(x_t) &= \sum_{x_{t_2}\in\mathscr{X}} g(x_{t_2}) P(X_{t_2} = x_{t_2} \,|\, X_t = x_t, X_u = x_u) \\
&= \sum_{x_{t_2}\in\mathscr{X}} g(x_{t_2}) \sum_{x_{t_1}\in\mathscr{X}} P(X_{t_2} = x_{t_2}, X_{t_1} = x_{t_1} \,|\, X_t = x_t, X_u = x_u) \\
&= \sum_{x_{t_2}\in\mathscr{X}} g(x_{t_2}) \sum_{x_{t_1}\in\mathscr{X}} P(X_{t_2} = x_{t_2} \,|\, X_{t_1} = x_{t_1}, X_t = x_t, X_u = x_u) \\
&\qquad\qquad\qquad\qquad P(X_{t_1} = x_{t_1} \,|\, X_t = x_t, X_u = x_u) \\
&= \sum_{x_{t_1}\in\mathscr{X}} T_{t_1,x_{u\cup\{t\}}}^{t_2} g(x_{t_1}) P(X_{t_1} = x_{t_1} \,|\, X_t = x_t, X_u = x_u) \\
&= \big(T_{t_0,x_u}^{t_1} T_{t_1,x_{u\cup\{t\}}}^{t_2} \big) g(x_t)\,,
\end{aligned}$$

where the second equality used F3$_{47}$ and the third equality used F4$_{47}$. Hence, it follows that

$$\begin{aligned}
T_{t,x_u}^s f(x_t) &= T_{t_0,x_u}^{t_2} \left(\prod_{i=3}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} f \right)(x_t) \\
&= T_{t_0,x_u}^{t_1} T_{t_1,x_{u\cup\{t\}}}^{t_2} \left(\prod_{i=3}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} f \right)(x_t) \\
&= \left(T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right) f(x_t)\,,
\end{aligned}$$

using the induction hypothesis for the first equality. □

The previous result can also be phrased as an identity in terms of the rows of these history-dependent transition matrices, as follows.

Corollary 5.38. *Consider any $P \in \mathbb{P}^{\mathrm{W}}$ with corresponding family of history-dependent transition matrices (T_{t,x_u}^{s}), any $t,s \in \mathbb{R}_{\geq 0}$ such that $t < s$, any $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ and any sequence $t = t_0 < t_1 < \cdots < t_n = s$, with $n \in \mathbb{Z}_{>0}$. Then for any $x_t \in \mathscr{X}$ it holds that*

$$T_{t,x_u}^{s}(x_t,\cdot) = \left(T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right)(x_t,\cdot) \tag{5.63}$$

Proof. Fix any $x_t \in \mathscr{X}$. Then for all $x_s \in \mathscr{X}$, it follows from Lemma 5.37 that

$$T_{t,x_u}^{s} \mathbb{I}_{x_s}(x_t) = \left(T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right) \mathbb{I}_{x_s}(x_t)\,,$$

and therefore, in particular, that

$$T_{t,x_u}^{s}(x_t,x_s) = \left(T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right)(x_t,x_s)\,.$$

Because the $x_s \in \mathscr{X}$ is arbitrary, this concludes the proof. □

A word of warning about Corollary 5.38 is in order, because an important subtlety may get lost in the verbose notation: because the matrices $T_{t_{i-1},x_{u\cup\{t\}}}^{t_i}$ on the right-hand side of Equation (5.63) depend on the value of x_t, this does *not* imply that the left-hand matrix is equal to the composite matrix on the right-hand side. Indeed, the row-wise identity stated by Corollary 5.38 is the strongest possible statement because the process $P \in \mathbb{P}^{\mathrm{W}}$ may not be Markovian.

The next result gives a bound on the rate of change of the transition matrices of a well-behaved Markov chain that is consistent with a given bounded set $\mathscr{Q}$. Effectively, this bound can be understood as being uniform with respect to all Markov chains in $\mathbb{P}_{\mathscr{Q}}^{\mathrm{W}}$ and all time points.

Lemma 5.39. *Consider a non-empty bounded set $\mathscr{Q}$ of rate matrices and let $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ be a well-behaved Markov chain that is consistent with $\mathscr{Q}$, with corresponding family of transition matrices (T_t^s). Then, for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, it holds that $\|T_t^s - I\| \leq (s-t)\,\|\mathscr{Q}\|$.*

Proof. If $t = s$, the result is trivial because we know from Proposition 5.1_{183} and Definition 4.6_{156} that $T_t^s = I$. Hence, without loss of generality, we may assume that $t < s$. Fix any $\varepsilon > 0$. Since $t < s$, it follows from Lemma 5.36_{227} that there is some $u \in \mathscr{U}_{[t,s]}$ with $u = t_0, \ldots, t_n$

and $t = t_0 < \ldots < t_n = s$ such that, for all $i \in \{1, \ldots, n\}$, there is some $Q_i \in \mathscr{Q}$ such that $\left\|T_{t_{i-1}}^{t_i} - (I + \Delta_i^u Q_i)\right\| < \Delta_i^u \varepsilon$, and therefore also

$$\left\|T_{t_{i-1}}^{t_i} - I\right\| \leq \left\|T_{t_{i-1}}^{t_i} - (I + \Delta_i^u Q_i)\right\| + \Delta_i^u \left\|Q_i\right\| < \Delta_i^u \varepsilon + \Delta_i^u \left\|\mathscr{Q}\right\| = \Delta_i^u (\varepsilon + \left\|\mathscr{Q}\right\|),$$

with $\Delta_i^u := t_i - t_{i-1}$. Since it follows from Proposition 5.1$_{183}$ and Definition 4.6$_{156}$ that $T_t^s = \prod_{i=1}^n T_{t_{i-1}}^{t_i}$ and that, for all $i \in \{1, \ldots, n\}$, $T_{t_{i-1}}^{t_i}$ and I are transition matrices, we infer from Lemma B.5$_{393}$ that

$$\begin{aligned} \left\|T_t^s - I\right\| = \left\|\prod_{i=1}^n T_{t_{i-1}}^{t_i} - \prod_{i=1}^n I\right\| &\leq \sum_{i=1}^n \left\|T_{t_{i-1}}^{t_i} - I\right\| \\ &< \sum_{i=1}^n \Delta_i^u (\varepsilon + \left\|\mathscr{Q}\right\|) = (s-t)(\varepsilon + \left\|\mathscr{Q}\right\|). \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\left\|T_t^s - I\right\| \leq (s-t)\left\|\mathscr{Q}\right\|$. □

5.D Proofs of Results in Section 5.3

This appendix contains the proofs for the results in Section 5.3$_{194}$, as well as some auxiliary lemmas. Before we proceed, let us remark that many of the results in this appendix use (norms of) the rows of matrices, the technical details of which are introduced and discussed in Appendices A.2$_{380}$ and A.3$_{383}$.

Proof of Lemma 5.12$_{195}$. Choose $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\|\mathscr{Q}\right\| < 1$ and $\delta \left\|\mathscr{Q}\right\|^2 < \varepsilon/2$; because $\mathscr{Q}$ is bounded this is always possible. Now choose any $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$, any $t \in \mathbb{R}_{\geq 0}$, and any $\Delta < \delta$. Let (T_t^s) be the family of transition matrices corresponding to P which, as we know from Proposition 5.1$_{183}$, is a transition matrix system. Then if $\Delta = 0$ the result is trivial because we know from Proposition 5.1$_{183}$ and Definition 4.6$_{156}$ that $T_t^t = I$. So, without loss of generality, we may assume that $\Delta > 0$.

It now follows from Lemma 5.36$_{227}$ that there is some $v \in \mathscr{U}_{[t,t+\Delta]}$ with $v = t_0, \ldots, t_n$ and $t = t_0 < t_1 < \cdots < t_n = t + \Delta$ and, for all $i \in \{1, \ldots, n\}$, some $Q_i \in \mathscr{Q}$ such that

$$\left\|T_{t_{i-1}}^{t_i} - (I + \Delta_i Q_i)\right\| < \Delta_i \frac{\varepsilon}{2} \text{ and } \Delta_i \left\|Q_i\right\| \leq \Delta \left\|\mathscr{Q}\right\| \leq \delta \left\|\mathscr{Q}\right\| < 1.$$

Therefore, it follows from Propositions 4.9$_{153}$ and 5.1$_{183}$, Definition 4.6$_{156}$ and Lemma B.5$_{393}$ that

$$\left\|T_t^{t+\Delta} - \prod_{i=1}^n (I + \Delta_i Q_i)\right\| = \left\|\prod_{i=1}^n T_{t_{i-1}}^{t_i} - \prod_{i=1}^n (I + \Delta_i Q_i)\right\| < \sum_{i=1}^n \Delta_i \frac{\varepsilon}{2} = \Delta \frac{\varepsilon}{2}.$$

and, with $Q \coloneqq {}^1\!/_\Delta \sum_{i=1}^n \Delta_i Q_i$, it follows from Lemma B.12$_{395}$ that

$$\left\| \prod_{i=1}^n (I + \Delta_i Q_i) - (I + \Delta Q) \right\| = \left\| \prod_{i=1}^n (I + \Delta_i Q_i) - (I + \sum_{i=1}^n \Delta_i Q_i) \right\| \leq \Delta^2 \left\| \mathscr{Q} \right\|^2 .$$

By combining these two inequalities, we find that

$$\left\| T_t^{t+\Delta} - (I + \Delta Q) \right\| < \Delta \frac{\varepsilon}{2} + \Delta^2 \left\| \mathscr{Q} \right\|^2 \leq \Delta \frac{\varepsilon}{2} + \Delta \delta \left\| \mathscr{Q} \right\|^2 < \Delta \frac{\varepsilon}{2} + \Delta \frac{\varepsilon}{2} = \Delta \varepsilon.$$

The result follows because the convexity of $\mathscr{Q}$ implies that $Q \in \mathscr{Q}$. □

Proof of Lemma 5.13$_{195}$. Fix any $\varepsilon > 0$ and let $\delta \in \mathbb{R}_{>0}$ be such that $\delta \left\| \mathscr{Q} \right\| \leq 1$ and $\delta \left\| \mathscr{Q} \right\|^2 \leq \varepsilon/2$; because $\mathscr{Q}$ is bounded this is always possible. Now choose any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, any $t \in \mathbb{R}_{\geq 0}$, any $u \in \mathscr{U}_{<t}$, any $x_u \in \mathscr{X}$, and any $\Delta < \delta$. Let (T^s_{t,x_u}) be the family of history-dependent transition matrices corresponding to P. Then if $\Delta = 0$ the result is trivial because we know from Proposition 4.2$_{149}$ and Definition 4.2$_{148}$ that $T^t_{t,x_u} = I$. So, without loss of generality, we may assume that $\Delta > 0$.

Because $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, by Proposition 4.23$_{171}$ and Definition 5.3$_{189}$, there is some $\delta' > 0$ such that with $0 < \Delta_1 < \min\{\delta, \delta', \Delta\}$, there is some $Q_1 \in \mathscr{Q}$ such that

$$\left\| T^{t+\Delta_1}_{t,x_u} - (I + \Delta_1 Q_1) \right\| < \Delta_1 \frac{\varepsilon}{2} . \tag{5.64}$$

Now fix any $x_t \in \mathscr{X}$. Then, by Lemma 5.36$_{227}$, there is some $v \in \mathscr{U}_{[t+\Delta_1, t+\Delta]}$, $v = t_1, \ldots, t_n$, with $\sigma(v) < \delta$, such that, for all $i \in \{2, \ldots, n\}$, with $\Delta_i \coloneqq \Delta^v_i = t_i - t_{i-1}$, it holds that

$$(\exists Q_i \in \mathscr{Q}) \left\| T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} - (I + \Delta_i Q_i) \right\| < \Delta_i \frac{\varepsilon}{2} . \tag{5.65}$$

Noting that $t_1 = t + \Delta_1$ and $\Delta = \sum_{i=1}^n \Delta_i$, and using Corollary 5.38$_{231}$, it holds that

$$T^{t+\Delta}_{t,x_u}(x_t, \cdot) = \left(T^{t+\Delta_1}_{t,x_u} \prod_{i=2}^n T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} \right) (x_t, \cdot) .$$

Now let $T_\varepsilon \coloneqq \prod_{i=1}^n (I + \Delta_i Q_i)$. Then, for all $i \in \{1, \ldots, n\}$, it holds that $\Delta_i \leq \delta$ and hence, because $\delta \left\| \mathscr{Q} \right\| \leq 1$, it follows from Proposition 4.9$_{153}$

that $(I+\Delta_i Q_i)$ is a transition matrix. We now find that

$$\begin{aligned}
&\left\| T_{t,x_u}^{t+\Delta}(x_t,\cdot) - T_\varepsilon(x_t,\cdot) \right\|_* \\
&\quad = \left\| \left(T_{t,x_u}^{t+\Delta_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right)(x_t,\cdot) - T_\varepsilon(x_t,\cdot) \right\|_* \\
&\quad \leq \left\| T_{t,x_u}^{t+\Delta_1} \prod_{i=2}^{n} T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} - \prod_{i=1}^{n}(I+\Delta_i Q_i) \right\| \\
&\quad \leq \left\| T_{t,x_u}^{t+\Delta_1} - (I+\Delta_1 Q_1) \right\| + \sum_{i=2}^{n} \left\| T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} - (I+\Delta_i Q_i) \right\| \\
&\quad < \sum_{i=1}^{n} \Delta_i \frac{\varepsilon}{2} = \Delta \frac{\varepsilon}{2},
\end{aligned}$$

where the first inequality follows from Proposition A.33390, the second inequality follows from Lemma B.5393 and the fact that all matrices involved are transition matrices, and the third inequality follows from Equations (5.64)↶ and (5.65)↶.

Moreover, because δ is such that $\delta \|\mathscr{Q}\| \leq 1$ and $\delta \|\mathscr{Q}\|^2 \leq \varepsilon/2$, and because $\mathscr{Q}$ is non-empty, bounded, and convex, it follows from Lemma B.12395 that

$$\left\| \prod_{i=1}^{n}(I+\Delta_i Q_i) - \left(I + \sum_{i=1}^{n} \Delta_i Q_i \right) \right\| \leq \Delta^2 \|\mathscr{Q}\|^2 \leq \Delta\delta \|\mathscr{Q}\|^2 \leq \Delta \frac{\varepsilon}{2}.$$

Hence, with $Q_{x_t} := 1/\Delta \sum_{i=1}^{n} \Delta_i Q_i$, we have $Q_{x_t} \in \mathscr{Q}$ because $\mathscr{Q}$ is convex, and

$$\begin{aligned}
&\left\| T_{t,x_u}^{t+\Delta}(x_t,\cdot) - (I+\Delta Q_{x_t})(x_t,\cdot) \right\|_* \\
&\quad \leq \left\| T_{t,x_u}^{t+\Delta}(x_t,\cdot) - T_\varepsilon(x_t,\cdot) \right\|_* + \left\| T_\varepsilon(x_t,\cdot) - (I+\Delta Q_{x_t})(x_t,\cdot) \right\|_* \\
&\quad < \Delta \frac{\varepsilon}{2} + \left\| T_\varepsilon - (I+\Delta Q_{x_t}) \right\| \leq \Delta\varepsilon,
\end{aligned}$$

using Proposition A.33390 for the second inequality.

If we now repeat the above construction of such a Q_{x_t} for all $x_t \in \mathscr{X}$, we can construct a new matrix Q such that $Q(x_t,\cdot) := Q_{x_t}(x_t,\cdot)$ for all $x_t \in \mathscr{X}$. Then $Q \in \mathscr{Q}$ because $\mathscr{Q}$ has separately specified rows, and

$$\left\| T_{t,x_u}^{t+\Delta} - (I+\Delta Q) \right\| = \max_{x_t \in \mathscr{X}} \left\| T_{t,x_u}^{t+\Delta}(x_t,\cdot) - (I+\Delta Q_{x_t})(x_t,\cdot) \right\|_* < \Delta\varepsilon,$$

using Equation (A.4)390 for the equality. □

Proof of Proposition 5.14$_{196}$. Let first

$$\mathcal{T}^{\mathscr{Q},\mathscr{M}}_{[a,b]} := \left\{ \left({}^{P}T_t^s\right)_{[a,b]} \,\middle|\, P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \right\}.$$

It is clear from the definition of $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ that

$$\mathcal{T}^{\mathscr{Q}}_{[a,b]} = \left\{ \left({}^{P}T_t^s\right)_{[a,b]} \,\middle|\, P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}} \right\} \supseteq \mathcal{T}^{\mathscr{Q},\mathscr{M}}_{[a,b]},$$

so let us prove the inclusion in the other direction. So, fix any $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. By Proposition 5.1$_{183}$, this P has a corresponding transition matrix system $({}^{P}T_t^s)$. Choose an arbitrary $p \in \mathscr{M}$; this is possible because $\mathscr{M}$ is non-empty. Due to Theorem 5.2$_{184}$, there is a Markov chain $P' \in \mathbb{P}^{\mathrm{M}}$ that has $({}^{P}T_t^s)$ as its corresponding transition matrix system, and that satisfies $P'(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$. Hence it holds that $P' \sim \mathscr{M}$ due to Definition 5.4$_{189}$. For ease of notation, let $(T_t^s) := ({}^{P}T_t^s)$ denote the transition matrix system of P'. Because $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ it follows from Proposition 5.1$_{183}$ that $(T_t^s) = ({}^{P}T_t^s)$ is a well-behaved transition matrix system and therefore, again by Proposition 5.1$_{183}$, that P' is a well-behaved Markov chain.

Let us now show that $P' \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. Because P' is a Markov chain, and due to Equation (5.4)$_{191}$, we need to show that, for all $r \in \mathbb{R}_{\geq 0}$, it holds that $\overline{\partial} T_r^r \subseteq \mathscr{Q}$. So fix any $r \in \mathbb{R}_{\geq 0}$. Because $(T_t^s) = ({}^{P}T_t^s)$ it follows that $\overline{\partial} T_r^r = \overline{\partial}\,{}^{P}T_r^r$ due to Definition 4.11$_{168}$. Since $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ it holds that $P \sim \mathscr{Q}$, which by Definition 5.3$_{189}$ implies that $\overline{\partial}\,{}^{P}T_r^r \subseteq \mathscr{Q}$, and therefore, that $\overline{\partial} T_r^r = \overline{\partial}\,{}^{P}T_r^r \subseteq \mathscr{Q}$. Because this is true for all $r \in \mathbb{R}_{\geq 0}$, and since we already know that $P' \sim \mathscr{M}$, it follows from Equation (5.4)$_{191}$ that $P' \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$.

Now let $(T_t^s)_{[a,b]}$ and $({}^{P}T_t^s)_{[a,b]}$ be the restrictions of (T_t^s) and $({}^{P}T_t^s)$, respectively, to the interval $[a,b]$. Then, clearly, $({}^{P}T_t^s)_{[a,b]} = (T_t^s)_{[a,b]}$. Moreover, since $P' \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, it follows that $(T_t^s)_{[a,b]} \in \mathcal{T}^{\mathscr{Q},\mathscr{M}}_{[a,b]}$, which, since $({}^{P}T_t^s)_{[a,b]} = (T_t^s)_{[a,b]}$ and because $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ is arbitrary, implies that $\mathcal{T}^{\mathscr{Q}}_{[a,b]} \subseteq \mathcal{T}^{\mathscr{Q},\mathscr{M}}_{[a,b]}$. □

Proof of Lemma 5.15$_{196}$. Consider any sequence $\{({}^{i}T_t^s)_{[a,b]}\}_{i \in \mathbb{Z}_{>0}}$ in $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$ that is Cauchy. We need to prove that this sequence converges to a limit $({}^{*}T_t^s)_{[a,b]}$ that belongs to $\mathcal{T}^{\mathscr{Q}}_{[a,b]}$. Since we already know from Proposition 4.21$_{163}$ that this limit exists and belongs to $\mathcal{T}_{[a,b]}$, the only thing that remains to be shown is that $({}^{*}T_t^s)_{[a,b]} \in \mathcal{T}^{\mathscr{Q}}_{[a,b]}$.

Fix any $t \in [a,b]$ and consider any $\Delta \in \mathbb{R}_{>0}$. Then, if $t + \Delta \leq b$, because $\mathscr{Q}$ is compact and therefore bounded, due to Lemma 5.39$_{231}$ it holds for all $i \in \mathbb{Z}_{>0}$ that

$$\left\| {}^{*}T_t^{t+\Delta} - I \right\| \leq \left\| {}^{*}T_t^{t+\Delta} - {}^{i}T_t^{t+\Delta} \right\| + \left\| {}^{i}T_t^{t+\Delta} - I \right\| \leq \left\| {}^{*}T_t^{t+\Delta} - {}^{i}T_t^{t+\Delta} \right\| + \Delta \left\| \mathscr{Q} \right\|.$$

Since $\lim_{i\to+\infty} {}^{i}T_t^{t+\Delta} = {}^{*}T_t^{t+\Delta}$, this implies that $\left\|{}^{*}T_t^{t+\Delta} - I\right\| \leq \Delta \|\mathscr{Q}\|$. Similarly, if $t-\Delta \geq a$, we also find that $\left\|{}^{*}T_{t-\Delta}^{t} - I\right\| \leq \Delta \|\mathscr{Q}\|$. Hence, it follows that, if $t \in [a,b)$,

$$\limsup_{\Delta\to 0^+} \left\| \frac{1}{\Delta} \left({}^{*}T_t^{t+\Delta} - I \right) \right\| \leq \|\mathscr{Q}\|$$

and, if $t \in (a,b]$,

$$\limsup_{\Delta\to 0^+} \left\| \frac{1}{\Delta} \left({}^{*}T_{t-\Delta}^{t} - I \right) \right\| \leq \|\mathscr{Q}\|,$$

which, since $t \in [a,b]$ was arbitrary, implies that $({}^{*}T_t^{s})_{[a,b]}$ is well-behaved.

Now consider any $Q \in \mathscr{Q}$, and let $\left(e^{Q(s-t)}\right)$ be the corresponding exponential transition matrix system, as given by Definition 4.8_{158}. Let

$$({}^{*}T_t^{s}) := \left(e^{Q(s-t)}\right)_{[0,a]} \otimes ({}^{*}T_t^{s})_{[a,b]} \otimes \left(e^{Q(s-t)}\right)_{[b,\infty)}.$$

Then clearly, $({}^{*}T_t^{s})$ is a transition matrix system and, since $({}^{*}T_t^{s})_{[a,b]}$ is well-behaved, it follows from Propositions 4.16_{158} and 4.19_{161} that $({}^{*}T_t^{s})$ is also well-behaved. Therefore, and because of Theorem 5.2_{184}, there is some $P_* \in \mathbb{P}^{\mathrm{WM}}$ whose corresponding family of transition matrices is given by $({}^{*}T_t^{s})$. In the remainder of this proof, we will show that $P_* \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, which then implies that $({}^{*}T_t^{s})_{[a,b]} \in \mathcal{T}^{\mathscr{Q}}_{[a,b]}$ due to Equation $(5.9)_{195}$.

To this end, let $(\overline{\partial}\,{}^{*}T_t^{t})$ be the family of outer partial derivatives corresponding to P_*, and fix any $t \in \mathbb{R}_{\geq 0}$. We need to show that $\overline{\partial}\,{}^{*}T_t^{t} \subseteq \mathscr{Q}$, or equivalently, that $\overline{\partial}_{+}{}^{*}T_t^{t} \subseteq \mathscr{Q}$ and, if $t > 0$, that $\overline{\partial}_{-}{}^{*}T_t^{t} \subseteq \mathscr{Q}$. We will only prove that $\overline{\partial}_{+}{}^{*}T_t^{t} \subseteq \mathscr{Q}$. The proof for $\overline{\partial}_{-}{}^{*}T_t^{t} \subseteq \mathscr{Q}$ in the case that $t > 0$ is completely analogous. So, fix any $Q_* \in \overline{\partial}_{+}{}^{*}T_t^{t}$. We will show that $Q_* \in \mathscr{Q}$. If $t < a$ or $b \leq t$, this holds trivially because in that case, we know from Lemma 4.14_{155} that $\partial_{+}{}^{*}T_t^{t} = Q$, which, because of Corollary 4.24_{171}, implies that $\overline{\partial}_{+}{}^{*}T_t^{t} = \{Q\} \subseteq \mathscr{Q}$. Therefore, without loss of generality, we may assume that $t \in [a,b)$.

Fix any $m \in \mathbb{Z}_{>0}$. Because $\mathscr{Q}$ is non-empty, convex and, since it is compact, bounded (by Corollary $\mathrm{A}.12_{378}$), Lemma 5.12_{195} then implies the existence of some $\delta \in \mathbb{R}_{>0}$ such that

$$(\forall 0 < \Delta < \delta)\,(\forall i \in \mathbb{Z}_{>0})\,(\exists Q \in \mathscr{Q})\, \left\| \frac{{}^{i}T_t^{t+\Delta} - I}{\Delta} - Q \right\| \leq \frac{1}{m}. \tag{5.66}$$

Since $Q_* \in \overline{\partial}_{+}{}^{*}T_t^{t}$, Equation $(4.22)_{168}$ implies the existence of some $0 < \Delta_m < \delta$ such that

$$\left\| \frac{{}^{*}T_t^{t+\Delta_m} - I}{\Delta_m} - Q_* \right\| < \frac{1}{m}$$

and, since $\lim_{i \to +\infty} {}^{i}T_t^{t+\Delta_m} = {}^{*}T_t^{t+\Delta_m}$, there is some $i_m \in \mathbb{Z}_{>0}$ such that

$$\left\| \frac{{}^{*}T_t^{t+\Delta_m} - {}^{i_m}T_t^{t+\Delta_m}}{\Delta_m} \right\| \leq \frac{1}{m}.$$

Therefore, we can apply Equation (5.66)—with $\Delta := \Delta_m$ and $i := i_m$—to infer that there is some $Q_m \in \mathcal{Q}$ such that

$$\|Q_* - Q_m\| \leq \left\| Q_* - \frac{{}^{*}T_t^{t+\Delta_m} - I}{\Delta_m} \right\| + \left\| \frac{{}^{*}T_t^{t+\Delta_m} - {}^{i_m}T_t^{t+\Delta_m}}{\Delta_m} \right\| + \left\| \frac{{}^{i_m}T_t^{t+\Delta_m} - I}{\Delta_m} - Q_m \right\| \leq \frac{3}{m}.$$

By repeating this procedure for every $m \in \mathbb{Z}_{>0}$, we obtain a sequence $\{Q_m\}_{m \in \mathbb{Z}_{>0}}$ in $\mathcal{Q}$ such that $\|Q_* - Q_m\| \leq 3/m$ for all $m \in \mathbb{Z}_{>0}$, which clearly implies that $\lim_{m \to +\infty} Q_m = Q_*$. Since the sequence $\{Q_m\}_{m \in \mathbb{Z}_{>0}}$ belongs to the set $\mathcal{Q}$, and because $\mathcal{Q}$ is compact and therefore closed, it follows that $Q_* \in \mathcal{Q}$. $\square$

Lemma 5.40. *Consider a non-empty, compact, and convex set of rate matrices $\mathcal{Q}$ and any $a, b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Let d be the metric that is defined in Equation* (4.15)$_{162}$. *The metric space $\left(\mathcal{T}^{\mathcal{Q}}_{[a,b]}, d\right)$ is then totally bounded.*

Proof. In order to prove that $\left(\mathcal{T}^{\mathcal{Q}}_{[a,b]}, d\right)$ is totally bounded, it suffices to prove that, for all $\varepsilon \in \mathbb{R}_{>0}$, there is a finite collection of open ε-balls centered on elements of $\mathcal{T}^{\mathcal{Q}}_{[a,b]}$, such that this collection covers $\mathcal{T}^{\mathcal{Q}}_{[a,b]}$, or equivalently, that there is a finite subset $\mathbb{C}_\varepsilon$ of $\mathcal{T}^{\mathcal{Q}}_{[a,b]}$ such that

$$(\forall (T_t^s)_{[a,b]} \in \mathcal{T}^{\mathcal{Q}}_{[a,b]})\, (\exists (S_t^s)_{[a,b]} \in \mathbb{C}_\varepsilon)\; d\big((T_t^s)_{[a,b]}, (S_t^s)_{[a,b]}\big) < \varepsilon . \tag{5.67}$$

So fix any $\varepsilon > 0$. We will now construct such a set $\mathbb{C}_\varepsilon$ and prove that it satisfies Equation (5.67).

First suppose that $a = b$, fix any $Q \in \mathcal{Q}$—this is always possible since $\mathcal{Q}$ is non-empty—and let

$$\mathbb{C}_\varepsilon := \left\{ (e^{Q(s-t)})_{[a,b]} \right\} .$$

Then $\mathbb{C}_\varepsilon$ contains a single element (and hence is clearly finite) and, due to Proposition 5.10$_{192}$, $\mathbb{C}_\varepsilon$ is a subset of $\mathcal{T}^{\mathcal{Q}}_{[a,b]}$. Fix any $(T_t^s)_{[a,b]} \in \mathcal{T}^{\mathcal{Q}}_{[a,b]}$. Then for any $t, s \in [a,b]$ with $t \leq s$ it holds that $s = t$ since $a = b$, and hence it follows from Proposition 4.17$_{159}$ that $e^{Q(s-t)} = I$ and that $T_t^s = I$, whence $\left\| T_t^s - e^{Q(s-t)} \right\| = 0$, which implies that

$$d\big((T_t^s)_{[a,b]}, (e^{Q(s-t)})_{[a,b]}\big) = 0 < \varepsilon .$$

Because this is true for all $(T_t^s)_{[a,b]} \in \mathcal{T}_{[a,b]}^{\mathcal{Q}}$ it follows that $\mathbb{C}_\varepsilon$ indeed satisfies Equation (5.67)$_{\curvearrowleft}$. This concludes the proof for the case where $a = b$.

Now suppose that $a < b$. Fix $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \in \mathbb{R}_{>0}$ such that $(b-a)(\varepsilon_1+\varepsilon_2+\varepsilon_3)+2\varepsilon_4 < \varepsilon/2$ and choose $\delta_3, \delta_4 \in \mathbb{R}_{>0}$ such that $\delta_3 \left\|\mathcal{Q}\right\|^2 < \varepsilon_3$ and $2\delta_4 \left\|\mathcal{Q}\right\| < \varepsilon_4$; this is clearly always possible since $\mathcal{Q}$ is compact and therefore bounded by Corollary A.12$_{378}$.

Because $\mathcal{Q}$ is compact, it admits a finite cover of open balls with radius ε_2. In other words, there is some finite set $\mathcal{Q}_{\varepsilon_2} \subseteq \mathcal{Q}$ such that

$$(\forall Q \in \mathcal{Q})(\exists \tilde{Q} \in \mathcal{Q}_{\varepsilon_2}) : \left\|Q - \tilde{Q}\right\| < \varepsilon_2. \tag{5.68}$$

Furthermore, because $\mathcal{Q}$ is non-empty, bounded and convex, we know from Lemma 5.12$_{195}$ that there is some $\delta_1 \in \mathbb{R}_{>0}$ such that for all $P \in \mathbb{P}_{\mathcal{Q}}^{\mathrm{WM}}$, all $t \in \mathbb{R}_{\geq 0}$, and all $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta < \delta_1$:

$$(\exists Q \in \mathcal{Q}) \left\|T_t^{t+\Delta} - (I + \Delta Q)\right\| \leq \Delta\varepsilon_1. \tag{5.69}$$

Now let $\delta := \min\{\delta_1, \delta_3, \delta_4\}$ and consider any $u \in \mathcal{U}_{[a,b]}$, $u = t_0, \ldots, t_n$ with $n \in \mathbb{Z}_{>0}$, such that $\sigma(u) < \delta$. Let

$$\mathbb{C}_\varepsilon := \left\{ \left(e^{\tilde{Q}_1(s-t)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{\tilde{Q}_n(s-t)}\right)_{[t_{n-1},t_n]} \,\middle|\, (\forall i \in \{1,\ldots,n\})\ \tilde{Q}_i \in \mathcal{Q}_{\varepsilon_2} \right\}.$$

Since n and $|\mathcal{Q}_{\varepsilon_2}|$ are finite, $\mathbb{C}_\varepsilon$ is clearly finite and, due to Proposition 5.10$_{192}$, $\mathbb{C}_\varepsilon$ is a subset of $\mathcal{T}_{[a,b]}^{\mathcal{Q}}$. The only thing that we still need to prove is that $\mathbb{C}_\varepsilon$ satisfies Equation (5.67)$_{\curvearrowleft}$.

So fix any $(T_t^s)_{[a,b]} \in \mathcal{T}_{[a,b]}^{\mathcal{Q}}$. For all $i \in \{1,\ldots,n\}$, since $\Delta_i^u \leq \sigma(u) < \delta \leq \delta_1$, it follows from Equations (5.69) and (5.68)—in that order—and Lemma B.8$_{394}$ that there are $Q_i \in \mathcal{Q}$ and $\tilde{Q}_i^* \in \mathcal{Q}_{\varepsilon_2}$ such that

$$\begin{aligned} \left\|T_{t_{i-1}}^{t_i} - e^{\tilde{Q}_i^*\Delta_i^u}\right\| &\leq \left\|T_{t_{i-1}}^{t_i} - (I + \Delta_i^u Q_i)\right\| + \left\|\Delta_i^u(Q_i - \tilde{Q}_i^*)\right\| + \left\|I + \Delta_i^u\tilde{Q}_i^* - e^{\tilde{Q}_i^*\Delta_i^u}\right\| \\ &\leq \Delta_i^u\varepsilon_1 + \Delta_i^u\varepsilon_2 + (\Delta_i^u)^2\left\|\mathcal{Q}\right\|^2 \leq \Delta_i^u(\varepsilon_1+\varepsilon_2+\varepsilon_3), \end{aligned} \tag{5.70}$$

where the final inequality holds since $\Delta_i^u \left\|\mathcal{Q}\right\|^2 \leq \delta \left\|\mathcal{Q}\right\|^2 \leq \delta_3 \left\|\mathcal{Q}\right\|^2 < \varepsilon_3$. We now use these $\tilde{Q}_i^* \in \mathcal{Q}_{\varepsilon_2}$, $i \in \{1,\ldots,n\}$, to define

$$(S_t^s)_{[a,b]} := \left(e^{\tilde{Q}_1^*(s-t)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{\tilde{Q}_n^*(s-t)}\right)_{[t_{n-1},t_n]} \in \mathbb{C}_\varepsilon.$$

For any $i \in \{1,\ldots,n\}$, it follows from Definition 4.8$_{158}$ and Equation (5.70) that

$$\left\|T_{t_{i-1}}^{t_i} - S_{t_{i-1}}^{t_i}\right\| = \left\|T_{t_{i-1}}^{t_i} - e^{\tilde{Q}_i^*\Delta_i^u}\right\| \leq \Delta_i^u(\varepsilon_1+\varepsilon_2+\varepsilon_3) \tag{5.71}$$

and from Lemma 5.39[231] that, for all $t,s \in [t_{i-1},t_i]$ such that $t \leq s$,

$$\|T_t^s - S_t^s\| \leq \|T_t^s - I\| + \|S_t^s - I\| \leq 2(s-t)\,\|\mathscr{Q}\| < \varepsilon_4, \tag{5.72}$$

where the final inequality holds since $s-t \leq t_i - t_{i-1} = \Delta_i^u \leq \sigma(u) \leq \delta \leq \delta_4$ and because $2\delta_4\,\|\mathscr{Q}\| < \varepsilon_4$.

Consider now any $t,s \in [a,b]$ such that $t \leq s$. We will prove that $\|T_t^s - S_t^s\| < \varepsilon/2$. If $t,s \in [t_{i-1},t_i]$ for some $i \in \{1,\ldots,n\}$, this follows trivially from Equation (5.72). In any other case, there must be some $k,\ell \in \{1,\ldots,n-1\}$ such that $k \leq \ell$, $t \in [t_{k-1},t_k]$ and $s \in [t_\ell,t_{\ell+1}]$, and we then find that, again,

$$\begin{aligned}\|T_t^s - S_t^s\| &= \left\| T_t^{t_k}\left(\prod_{i=k+1}^{\ell} T_{t_{i-1}}^{t_i}\right) T_{t_\ell}^s - S_t^{t_k}\left(\prod_{i=k+1}^{\ell} S_{t_{i-1}}^{t_i}\right) S_{t_\ell}^s \right\| \\ &\leq \left\|T_t^{t_k} - S_t^{t_k}\right\| + \sum_{i=k+1}^{\ell} \left\|T_{t_{i-1}}^{t_i} - S_{t_{i-1}}^{t_i}\right\| + \left\|T_{t_\ell}^s - S_{t_\ell}^s\right\| \\ &\leq 2\varepsilon_4 + \sum_{i=k+1}^{\ell} \Delta_i^u(\varepsilon_1 + \varepsilon_2 + \varepsilon_3) < \frac{\varepsilon}{2},\end{aligned}$$

where the equality follows from Definition 4.6[156], the first inequality follows from Definition 4.6[156] and Lemma B.5[393], the second inequality follows from Equations (5.71) and (5.72), and the final inequality holds because $\sum_{i=k+1}^{\ell} \Delta_i^u = t_\ell - t_k \leq b-a$. Hence, in all cases, it holds that $\|T_t^s - S_t^s\| < \varepsilon/2$.

Since this is true for all $t,s \in [a,b]$ such that $t \leq s$, we immediately find that

$$d\big((T_t^s)_{[a,b]},(S_t^s)_{[a,b]}\big) = \sup\{\|T_t^s - S_t^s\| \,:\, t,s \in [a,b], t \leq s\} \leq \frac{\varepsilon}{2} < \varepsilon.$$

Since $(T_t^s)_{[a,b]} \in \mathcal{T}_{[a,b]}^{\mathscr{Q}}$ is arbitrary, it follows that $\mathbb{C}_\varepsilon$ satisfies Equation (5.67)[237]. □

Proof of Lemma 5.16[196]. Let $\mathscr{Q}'$ denote the closed convex hull of $\mathscr{Q}$. Then, clearly, $\mathscr{Q}'$ is a non-empty, bounded, closed, and convex set of rate matrices. Since $\mathscr{Q}'$ is bounded and closed, it is compact by Corollary A.12[378]. Furthermore, since $\mathscr{Q} \subseteq \mathscr{Q}'$, we know that $\mathcal{T}_{[a,b]}^{\mathscr{Q}} \subseteq \mathcal{T}_{[a,b]}^{\mathscr{Q}'}$. Because any subspace of a totally bounded metric space is itself totally bounded [100, top of p.175], the result now follows trivially from Lemma 5.40[237]. □

Proof of Corollary 5.18[197]. That $\mathscr{T}$ is non-empty follows from the fact that $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ is non-empty. That its elements are transition matrices follows from the fact that, for all $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$, ${}^{P}T_t^s$ is a transition matrix due to Corollary 4.4$_{150}$. It therefore also follows from Lemma 3.9$_{92}$ that $\mathscr{T}$ is bounded. To establish the compactness, it follows from Corollary A.12$_{378}$ that we need to show that $\mathscr{T}$ is also closed. So, consider any convergent sequence $\{T_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{T}$ such that $\lim_{i\to+\infty} T_i = T_*$; we will now show that $T_* \in \mathscr{T}$.

As in Equation (5.9)$_{195}$, let $\boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}$ denote the set of restricted transition matrix systems corresponding to the elements of $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. Because $\mathscr{Q}$ is non-empty, compact, and convex, Theorem 5.17$_{196}$ implies that $\boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}$ is compact under the metric d defined in Equation (4.15)$_{162}$. Now, for any $i \in \mathbb{Z}_{>0}$, since $T_i \in \mathscr{T}$, there is some $P_i \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ such that ${}^{P_i}T_t^s = T_i$. In turn, this implies that $\left({}^{P_i}T_a^b\right)_{[t,s]} \in \boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}$. Hence, we obtain a co-sequence $\left\{\left({}^{P_i}T_a^b\right)_{[t,s]}\right\}_{i\in\mathbb{Z}_{>0}}$ in $\boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}$, such that $\lim_{i\to+\infty} {}^{P_i}T_t^s = \lim_{i\to+\infty} T_i = T_*$.

Because $\boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}$ is a compact metric space, it is sequentially compact by Proposition A.10$_{377}$. Hence, we get the existence of a convergent subsequence $\left\{\left({}^{P_{i_j}}T_a^b\right)_{[t,s]}\right\}_{j\in\mathbb{Z}_{>0}}$ such that

$$\lim_{j\to+\infty} \left({}^{P_{i_j}}T_a^b\right)_{[t,s]} =: \left({}^{P_{i_*}}T_a^b\right)_{[t,s]} \in \boldsymbol{\mathcal{T}}^{\mathscr{Q}}_{[t,s]}.$$

Using Proposition 5.14$_{196}$, this implies the existence of some $P_{i_*} \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P_{i_*}}T_t^s$ such that

$$\lim_{j\to+\infty} \left\| {}^{P_{i_j}}T_t^s - {}^{P_{i_*}}T_t^s \right\| \leq \lim_{j\to+\infty} d\left(\left({}^{P_{i_j}}T_a^b\right)_{[t,s]}, \left({}^{P_{i_*}}T_a^b\right)_{[t,s]} \right) = 0\,,$$

using the definition of the metric d in Equation (4.15)$_{162}$. Since $\lim_{i\to+\infty} T_i = T_*$, it holds that also $\lim_{j\to+\infty} T_{i_j} = T_*$ which, since ${}^{P_{i_j}}T_t^s = T_{i_j}$ for all $j \in \mathbb{Z}_{>0}$, implies that $T_* = {}^{P_{i_*}}T_t^s$. Since $P_{i_*} \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ it holds that ${}^{P_{i_*}}T_t^s \in \mathscr{T}$, from which we conclude that $T_* \in \mathscr{T}$, which is what we wanted to prove. Since the convergent sequence $\{T_i\}_{i\in\mathbb{Z}_{>0}}$ was arbitrary, Proposition A.8$_{376}$ now implies that $\mathscr{T}$ is closed, and hence we conclude that $\mathscr{T}$ is compact. □

Proof of Proposition 5.19$_{197}$. It is immediate from the definitions in Equation (5.10)$_{197}$ and (5.11)$_{197}$ that

$${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s \subseteq {}^{\mathscr{Q}}\mathscr{T}_t^s\,,$$

so let us prove the inclusion in the other direction. To this end, fix any $T \in {}^{\mathscr{Q}}\mathscr{T}_t^s$; we will show that $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. By Equation (5.10)$_{197}$, there is some $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ such that, for some $v \in \mathscr{U}_{<t}$ and $y_v \in \mathscr{X}$, it holds that $T = {}^{P_*}T_{t,y_v}^s$.

Next, choose any $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. By Definitions 5.6$_{190}$ and 5.4$_{189}$, this implies that there is some $p \in \mathscr{M}$ such that $P_\emptyset(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$.

Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, it follows from Lemma 5.35$_{215}$—with $u = v \cup \{t\}$, $P_\emptyset$ as chosen above and, for all $x_u \in \mathscr{X}_u$, $P_{x_u} = P_*$—that there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ such that, for all $x \in \mathscr{X}$,

$$P(X_0 = x) = P_\emptyset(X_0 = x) = p(x)\,, \tag{5.73}$$

and, for all $x_u \in \mathscr{X}_u$ and all $y \in \mathscr{X}$,

$$P(X_s = y \,|\, X_u = x_u) = P_{x_u}(X_s = y \,|\, X_u = x_u) = P_*(X_s = y \,|\, X_u = x_u)\,. \tag{5.74}$$

Because $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ it holds that $P \sim \mathscr{Q}$. Moreover, it follows from Equation (5.73) that $P \sim \mathscr{M}$, which implies that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Because $u = v \cup \{t\}$, it follows from Definition 4.2$_{148}$ and Equation (5.74) that the history-dependent transition matrix ${}^{P}T^{s}_{t,y_v}$ corresponding to P satisfies, for all $x, y \in \mathscr{X}$,

$$\begin{aligned} {}^{P}T^{s}_{t,y_v}(x,y) &= P(X_s = y \,|\, X_t = x, X_v = y_v) \\ &= P_*(X_s = y \,|\, X_t = x, X_v = y_v) = T(x,y)\,, \end{aligned}$$

and hence $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}$. □

Lemma 5.41. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t, s \in \mathbb{R}_{\geq 0}$ such that $t < s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as in Equation* (5.11)$_{197}$.

Then for all $\varepsilon > 0$ and $\delta > 0$, there is some $v \in \mathscr{U}_{[t,s]}$ with $v = t_0, \ldots, t_n$, $n \in \mathbb{Z}_{>0}$, and $\sigma(v) < \delta$, such that for all $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}$ and all $x \in \mathscr{X}$, there are $Q_1, \ldots, Q_n \in \mathscr{Q}$, such that

$$\left\| T(x,\cdot) - \left(\prod_{i=1}^{n} (I + \Delta^{v}_{i} Q_i) \right)(x,\cdot) \right\|_* < \varepsilon\,.$$

Proof. Fix any $\varepsilon > 0$ and $\delta > 0$, and let $\varepsilon_* := \varepsilon/(s-t)$. Because $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, and because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, we can use Lemma 5.13$_{195}$ to find $\delta' \in \mathbb{R}_{>0}$ such that, for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding family of history-dependent transition matrices $\left({}^{P}T^{s}_{t,x_u}\right)$, all $r \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<r}$, $x_u \in \mathscr{X}_u$, and $\Delta \in \mathbb{R}_{>0}$ such that $\Delta < \delta'$, it holds that

$$(\exists Q \in \mathscr{Q}) \left\| {}^{P}T^{r+\Delta}_{r,x_u} - (I + \Delta Q) \right\| < \Delta \varepsilon_*\,. \tag{5.75}$$

Let $\delta^* \in \mathbb{R}_{>0}$ be such that $\delta^* \leq \min\{\delta, \delta'\}$ and $\delta^* \|\mathscr{Q}\| \leq 1$, let $n := 1 + \lceil (s-t)/\delta^* \rceil$, let $\Delta := (s-t)/n$ and, for all $i \in \{0, \ldots, n\}$, let

$t_i := t + i\Delta$. Then it holds that $v := t_0, \dots, t_n \in \mathscr{U}_{[t,s]}$, and $\Delta_i^v = \Delta < \delta^*$ for all $i \in \{1, \dots, n\}$, and hence $\sigma(v) < \delta$, $\Delta < \delta'$, and $\Delta \|\mathscr{Q}\| \leq 1$.

Now choose any $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. There must then be some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding family of history-dependent transition matrices $({}^P T_{t,x_u}^s)$, such that, for some $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$, it holds that ${}^P T_{t,x_u}^s = T$.

Fix any $x \in \mathscr{X}$ and, for notational convenience, let $x_t := x$. Then, due to Corollary 5.38$_{231}$, it holds that

$$ {}^P T_{t,x_u}^s(x_t, \cdot) = \left({}^P T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} {}^P T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right)(x_t, \cdot) . $$

Because $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, and because $t_1 - t_0 = \Delta_1^v = \Delta < \delta'$, it follows from Equation (5.75)$_\frown$ that there is some $Q_1 \in \mathscr{Q}$ such that

$$ \left\| {}^P T_{t_0,x_u}^{t_1} - (I + \Delta Q_1) \right\| < \Delta \epsilon_* . $$

Similarly, for all $i \in \{2, \dots, n\}$, because $t_i - t_{i-1} = \Delta_i^v = \Delta < \delta'$, there is some $Q_i \in \mathscr{Q}$ such that

$$ \left\| {}^P T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} - (I + \Delta Q_i) \right\| < \Delta \epsilon_* . $$

For all $i \in \{1, \dots, n\}$, because $\Delta \|Q_i\| \leq \Delta \|\mathscr{Q}\| \leq 1$, it follows from Proposition 4.9$_{153}$ that $(I + \Delta Q_i)$ is a transition matrix. Therefore, it holds that

$$ \begin{aligned} &\left\| T(x, \cdot) - \left(\prod_{i=1}^{n} (I + \Delta Q_i) \right)(x, \cdot) \right\|_* \\ &= \left\| {}^P T_{t,x_u}^s(x_t, \cdot) - \left(\prod_{i=1}^{n} (I + \Delta Q_i) \right)(x_t, \cdot) \right\|_* \\ &= \left\| \left({}^P T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} {}^P T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} \right)(x_t, \cdot) - \left(\prod_{i=1}^{n} (I + \Delta Q_i) \right)(x_t, \cdot) \right\|_* \\ &\leq \left\| {}^P T_{t_0,x_u}^{t_1} \prod_{i=2}^{n} {}^P T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} - \prod_{i=1}^{n} (I + \Delta Q_i) \right\| \\ &\leq \left\| {}^P T_{t_0,x_u}^{t_1} - (I + \Delta Q_1) \right\| + \sum_{i=2}^{n} \left\| {}^P T_{t_{i-1},x_{u\cup\{t\}}}^{t_i} - (I + \Delta Q_i) \right\| \\ &< \sum_{i=1}^{n} \Delta \epsilon_* = (s - t)\epsilon_* = \epsilon , \end{aligned} $$

where we used Proposition A.33$_{390}$ for the first inequality and Lemma B.5$_{393}$ for the second inequality. □

Lemma 5.42. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t,s \in \mathbb{R}$ such that $t < s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then for all $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ and $x \in \mathscr{X}$, there is a Markov chain $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P}T_t^s$ such that ${}^{P}T_t^s(x,\cdot) = T(x,\cdot)$.*

Proof. Fix any $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. This implies the existence of some $P' \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that, for some $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$, its history-dependent transition matrix satisfies ${}^{P'}T_{t,x_u}^s = T$. Now fix any $x \in \mathscr{X}$.

Choose any $\varepsilon > 0$, let $\varepsilon_* := \varepsilon/2(s-t)$, and let $\delta > 0$ be such that $\delta \|\mathscr{Q}\| \leq 1$ and $\delta \|\mathscr{Q}\|^2 < \varepsilon_*$; because $\mathscr{Q}$ is compact and therefore bounded by Corollary A.12$_{378}$, this is always possible. Moreover, because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, using Lemma 5.41$_{241}$, we find that there is some $v \in \mathscr{U}_{[t,s]}$ with $v = t_0,\dots,t_n$, $n \in \mathbb{Z}_{>0}$, such that $\sigma(v) < \delta$, and some $Q_1,\dots,Q_n \in \mathscr{Q}$, such that

$$\left\| {}^{P'}T_{t,x_u}^s(x,\cdot) - \left(\prod_{i=1}^n (I + \Delta_i^v Q_i)\right)(x,\cdot) \right\|_* < \frac{\varepsilon}{2}\,. \tag{5.76}$$

Because $\delta \|\mathscr{Q}\| \leq 1$ it holds for all $i \in \{1,\dots,n\}$, since $\Delta_i^v \leq \sigma(v) < \delta$ and $\|Q_i\| \leq \|\mathscr{Q}\|$, that, by Proposition 4.9$_{153}$, $(I + \Delta_i^v Q_i)$ is a transition matrix.

Consider the restricted transition matrix system

$$({}^{\varepsilon}T_r^q)_{[t,s]} := \left(e^{Q_1(q-r)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{Q_n(q-r)}\right)_{[t_{n-1},t_n]}\,. \tag{5.77}$$

Now choose any $Q_0 \in \mathscr{Q}$, and consider the transition matrix system

$$({}^{\varepsilon}T_r^q) := \left(e^{Q_0(q-r)}\right)_{[0,t_0]} \otimes ({}^{\varepsilon}T_r^q)_{[t,s]} \otimes \left(e^{Q_0(q-r)}\right)_{[t_n,+\infty)}\,.$$

Then, clearly, $({}^{\varepsilon}T_r^q)$ is the extension of $({}^{\varepsilon}T_r^q)_{[t,s]}$ to an (unrestricted) transition matrix system, or in other words, $({}^{\varepsilon}T_r^q)_{[t,s]}$ is the restriction of $({}^{\varepsilon}T_r^q)$ to the interval $[t,s]$.

Now fix any $p \in \mathscr{M}$. Since $Q_0, Q_1, \dots, Q_n \in \mathscr{Q}$, and using Proposition 5.10$_{192}$, we find a Markov chain $P_\varepsilon \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ with corresponding family of transition matrices $({}^{\varepsilon}T_r^q)$. Due to Equation (5.77), its corresponding transition matrix ${}^{\varepsilon}T_t^s$ satisfies

$${}^{\varepsilon}T_t^s = \prod_{i=1}^n e^{Q_i \Delta_i^v}\,,$$

and hence

$$\begin{aligned}
\left\| {}^{\varepsilon}T_t^s - \prod_{i=1}^{n}(I + \Delta_i^v Q_i) \right\| &= \left\| \prod_{i=1}^{n} e^{Q_i \Delta_i^v} - \prod_{i=1}^{n}(I + \Delta_i^v Q_i) \right\| \\
&\leq \sum_{i=1}^{n} \left\| e^{Q_i \Delta_i^v} - (I + \Delta_i^v Q_i) \right\| \\
&\leq \sum_{i=1}^{n} (\Delta_i^v)^2 \left\| Q_i \right\|^2 \\
&\leq \sum_{i=1}^{n} \Delta_i^v \delta \left\| \mathscr{Q} \right\|^2 \\
&< \sum_{i=1}^{n} \Delta_i^v \varepsilon_* \\
&= (s-t)\varepsilon_* = \frac{\varepsilon}{2}\,,
\end{aligned} \tag{5.78}$$

where the first inequality follows from Lemma B.5~393~, which we can use because all matrices involved are transition matrices; the second inequality follows from Lemma B.8~394~; the third inequality follows because $\Delta_i^v \leq \sigma(v) < \delta$ and $\|Q_i\| \leq \|\mathscr{Q}\|$ for all $i \in \{1, \ldots, n\}$; the last inequality follows because $\delta \|\mathscr{Q}\|^2 < \varepsilon_*$; and the last equality follows because $\varepsilon_* = \varepsilon/2(s-t)$.

Moreover, it now holds that

$$\begin{aligned}
\left\| T(x,\cdot) - {}^{\varepsilon}T_t^s(x,\cdot) \right\|_* &= \left\| {}^{P'}T_{t,x_u}^s(x,\cdot) - {}^{\varepsilon}T_t^s(x,\cdot) \right\|_* \\
&\leq \left\| {}^{P'}T_{t,x_u}^s(x,\cdot) - \left(\prod_{i=1}^{n}(I + \Delta_i^v Q_i) \right)(x,\cdot) \right\|_* \\
&\qquad + \left\| \left(\prod_{i=1}^{n}(I + \Delta_i^v Q_i) \right)(x,\cdot) - {}^{\varepsilon}T_t^s(x,\cdot) \right\|_* \\
&< \frac{\varepsilon}{2} + \left\| \left(\prod_{i=1}^{n}(I + \Delta_i^v Q_i) \right)(x,\cdot) - {}^{\varepsilon}T_t^s(x,\cdot) \right\|_* \\
&\leq \frac{\varepsilon}{2} + \left\| \prod_{i=1}^{n}(I + \Delta_i^v Q_i) - {}^{\varepsilon}T_t^s \right\| \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,,
\end{aligned}$$

where the second inequality follows from Equation (5.76)↶, the third inequality follows from Equation (2.7)~63~, and the final inequality follows from Equation (5.78).

Now, consider a sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+$ and, for all $i \in \mathbb{Z}_{>0}$, repeat the above construction to find a restricted transition matrix system $({}^{\varepsilon_i}T_r^q)_{[t,s]}$ whose corresponding transition matrix ${}^{\varepsilon_i}T_t^s$ satisfies

$$\|T(x,\cdot) - {}^{\varepsilon_i}T_t^s(x,\cdot)\|_* < \varepsilon_i. \tag{5.79}$$

Then because, as we have seen above, each of these $({}^{\varepsilon_i}T_r^q)_{[t,s]}$ is the restriction of a family $({}^{\varepsilon_i}T_r^q)$ of transition matrices corresponding to a Markov chain $P_{\varepsilon_i} \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, it follows from Equation $(5.9)_{195}$ that $({}^{\varepsilon_i}T_r^q)_{[t,s]} \in \mathcal{T}^{\mathscr{Q}}_{[t,s]}$ for all $i \in \mathbb{Z}_{>0}$.

Because $\mathscr{Q}$ is non-empty, compact, and convex, it follows from Theorem 5.17_{196} that $\mathcal{T}^{\mathscr{Q}}_{[t,s]}$ is compact under the metric d defined in Equation $(4.15)_{162}$. Therefore, the sequence $\{({}^{\varepsilon_i}T_r^q)_{[t,s]}\}_{i\in\mathbb{Z}_{>0}}$ has a convergent subsequence $\{({}^{\varepsilon_{i_k}}T_r^q)_{[t,s]}\}_{k\in\mathbb{Z}_{>0}}$, whose limit $({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]} := \lim_{k\to+\infty}({}^{\varepsilon_{i_k}}T_r^q)_{[t,s]}$ also belongs to $\mathcal{T}^{\mathscr{Q}}_{[t,s]}$. Because $({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]} \in \mathcal{T}^{\mathscr{Q}}_{[t,s]}$, it follows from Equation $(5.9)_{195}$ and Proposition 5.14_{196} that there is some Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding family of transition matrices $({}^PT_r^q)$ such that $({}^PT_r^q)_{[t,s]} = ({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]}$, and hence in particular its corresponding transition matrix satisfies ${}^PT_t^s = {}^{\varepsilon_{i_*}}T_t^s$.

It remains to show that ${}^PT_t^s(x,\cdot) = T(x,\cdot)$. To this end, fix any $\varepsilon > 0$. Because $\{\varepsilon_i\}_{i\in\mathbb{Z}_{>0}} \to 0^+$, and because $({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]} := \lim_{k\to+\infty}({}^{\varepsilon_{i_k}}T_r^q)_{[t,s]}$, there is some $n \in \mathbb{Z}_{>0}$ such that, for all $k > n$, it holds that $\varepsilon_{i_k} < \varepsilon/2$ and

$$d\Big(({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]}, ({}^{\varepsilon_{i_k}}T_r^q)_{[t,s]}\Big) < \frac{\varepsilon}{2}.$$

For any $k > n$ it then holds that

$$\begin{aligned}
\big\|T(x,\cdot) - {}^PT_t^s(x,\cdot)\big\|_* &= \|T(x,\cdot) - {}^{\varepsilon_{i_*}}T_t^s(x,\cdot)\|_* \\
&\leq \big\|T(x,\cdot) - {}^{\varepsilon_{i_k}}T_t^s(x,\cdot)\big\|_* + \big\|{}^{\varepsilon_{i_k}}T_t^s(x,\cdot) - {}^{\varepsilon_{i_*}}T_t^s(x,\cdot)\big\|_* \\
&\leq \varepsilon_{i_k} + \big\|{}^{\varepsilon_{i_k}}T_t^s - {}^{\varepsilon_{i_*}}T_t^s\big\| \\
&\leq \varepsilon_{i_k} + d\Big(({}^{\varepsilon_{i_*}}T_r^q)_{[t,s]}, ({}^{\varepsilon_{i_k}}T_r^q)_{[t,s]}\Big) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

where for the second inequality we used Equation (5.79) and Proposition $A.33_{390}$. Since the $\varepsilon > 0$ is arbitrary, this means that $\big\|T(x,\cdot) - {}^PT_t^s(x,\cdot)\big\|_* = 0$, or in other words, that $T(x,\cdot) = {}^PT_t^s(x,\cdot)$. □

Proof of Proposition 5.20_{197}. It trivially follows from Equation $(5.11)_{197}$ that $\{{}^PT_t^s \,|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\} \subseteq {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$, so it suffices to prove the inclusion in the other direction. If $t = s$ then for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it follows from Proposition 4.2_{149} that ${}^PT_t^s = I = {}^PT_{t,x_u}^s$ for all $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$, whence in that case the result is trivial. So, for the remainder of this proof, let us suppose that $t < s$.

Now, fix any $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$; we will show that $T \in \left\{ {}^{P}T_t^s \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \right\}$. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 5.42243 and the fact that $t < s$ that for all $x \in \mathscr{X}$ there is a Markov chain $P_x \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P_x}T_t^s$ such that ${}^{P_x}T_t^s(x,\cdot) = T(x,\cdot)$. For all $y \in \mathscr{X}$ it therefore holds that

$$P_x(X_s = y \,|\, X_t = x) = {}^{P_x}T_t^s(x,y) = T(x,y)\,.$$

Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, and due to Theorem 5.11$_{193}$—with $u = \{t\}$ and $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ chosen arbitrarily—there is some $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that, for all $x,y \in \mathscr{X}$,

$$P_*(X_s = y \,|\, X_t = x) = P_x(X_s = y \,|\, X_t = x)\,.$$

Hence the transition matrix ${}^{P_*}T_t^s$ of P_* satisfies ${}^{P_*}T_t^s(x,\cdot) = {}^{P_x}T_t^s(x,\cdot) = T(x,\cdot)$ for all $x \in \mathscr{X}$, which implies that ${}^{P_*}T_t^s = T$. It follows that $T \in \left\{ {}^{P}T_t^s \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \right\}$ because $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. □

Lemma 5.43. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t,s \in \mathbb{R}$ such that $t < s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ has separately specified rows.*

Proof. For any $x \in \mathscr{X}$, let $\mathscr{T}_x$ denote the set of x-rows of elements of ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. Let T be any matrix such that, for all $x \in \mathscr{X}$, $T(x,\cdot) \in \mathscr{T}_x$. We need to show that $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$.

Now first fix any $x \in \mathscr{X}$. Because $T(x,\cdot) \in \mathscr{T}_x$, there is some $S \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ such that $S(x,\cdot) = T(x,\cdot)$. Because $\mathscr{Q}$ is non-empty, compact, convex and has separately specified rows, due to Lemma 5.42$_{243}$, this implies the existence of a Markov chain $P_x \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ whose corresponding transition matrix satisfies ${}^{P_x}T_t^s(x,\cdot) = S(x,\cdot)$. For all $y \in \mathscr{X}$ it therefore holds that

$$P_x(X_s = y \,|\, X_t = x) = {}^{P_x}T_t^s(x,y) = S(x,y) = T(x,y)\,.$$

Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, and due to Theorem 5.11$_{193}$—with $u = \{t\}$ and $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ chosen arbitrarily—there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that, for all $x,y \in \mathscr{X}$,

$$P(X_s = y \,|\, X_t = x) = P_x(X_s = y \,|\, X_t = x)\,.$$

Hence the transition matrix ${}^{P}T_t^s$ of P satisfies ${}^{P}T_t^s(x,\cdot) = {}^{P_x}T_t^s(x,\cdot) = T(x,\cdot)$ for all $x \in \mathscr{X}$, which implies that ${}^{P}T_t^s = T$. Hence it follows that $T = {}^{P}T_t^s \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ because $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. □

Lemma 5.44. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t,s \in \mathbb{R}$ such that $t < s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$. For any $x \in \mathscr{X}$, let*

$$\mathscr{T}_x := \left\{ T(x,\cdot) \,\middle|\, T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s \right\}$$

denote the set of x-rows of elements of ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. Then $\mathscr{T}_x$ is closed.

Proof. Fix any $x \in \mathscr{X}$, and consider any convergent sequence $\left\{T_i(x,\cdot)\right\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{T}_x$ such that $\lim_{i\to+\infty} T_i(x,\cdot) = T(x,\cdot)$. We need to show that $T(x,\cdot) \in \mathscr{T}_x$.

For all $i \in \mathbb{Z}_{>0}$, because $T_i(x,\cdot) \in \mathscr{T}_x$, there is some $T_i \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ whose x-row is $T_i(x,\cdot)$. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, due to Lemma 5.42$_{243}$, there is some Markov chain $P_i \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P_i}T_t^s$ such that ${}^{P_i}T_t^s(x,\cdot) = T_i(x,\cdot)$.

Let $\mathscr{T} := \left\{{}^{P}T_t^s : P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}\right\}$ be the set of transition matrices induced by the set of Markov chains $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. Then the sequence $\{{}^{P_i}T_t^s\}_{i\in\mathbb{Z}_{>0}}$ is in $\mathscr{T}$. Moreover, because $\mathscr{Q}$ is non-empty, compact, and convex, it follows from Corollary 5.18$_{197}$ that $\mathscr{T}$ is compact. Hence by Corollary A.12$_{378}$ $\mathscr{T}$ is sequentially compact, which implies that there is a convergent subsequence $\{{}^{P_{i_j}}T_t^s\}_{j\in\mathbb{Z}_{>0}}$ with limit $T_* := \lim_{j\to+\infty} {}^{P_{i_j}}T_t^s$ that satisfies $T_* \in \mathscr{T}$. This implies the existence of a Markov chain $P_* \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ such that ${}^{P_*}T_t^s = T_*$.

Because $\lim_{i\to+\infty} T_i(x,\cdot) = T(x,\cdot)$ it holds that also $\lim_{j\to+\infty} T_{i_j}(x,\cdot) = T(x,\cdot)$. Fix any $\varepsilon > 0$. Because also $\lim_{j\to+\infty} {}^{P_{i_j}}T_t^s = {}^{P_*}T_t^s$, there is some $n \in \mathbb{Z}_{>0}$ such that, for all $j > n$, it holds that

$$\left\|T_{i_j}(x,\cdot) - T(x,\cdot)\right\|_* < \frac{\varepsilon}{2} \quad \text{and} \quad \left\|{}^{P_{i_j}}T_t^s - {}^{P_*}T_t^s\right\| < \frac{\varepsilon}{2}. \tag{5.80}$$

Now fix any $j > n$. Then it holds that

$$\begin{aligned}
\left\|T(x,\cdot) - {}^{P_*}T_t^s(x,\cdot)\right\|_* &\leq \left\|T(x,\cdot) - T_{i_j}(x,\cdot)\right\|_* + \left\|T_{i_j}(x,\cdot) - {}^{P_*}T_t^s(x,\cdot)\right\|_* \\
&< \frac{\varepsilon}{2} + \left\|{}^{P_{i_j}}T_t^s(x,\cdot) - {}^{P_*}T_t^s(x,\cdot)\right\|_* \\
&\leq \frac{\varepsilon}{2} + \left\|{}^{P_{i_j}}T_t^s - {}^{P_*}T_t^s\right\| < \varepsilon,
\end{aligned}$$

where for the second inequality we used Equation (5.80) and that $T_{i_j}(x,\cdot) = {}^{P_{i_j}}T_t^s(x,\cdot)$, for the third inequality we used Proposition A.33$_{390}$, and for the final inequality we used Equation (5.80). Because $\varepsilon > 0$ is arbitrary, this implies that $\left\|T(x,\cdot) - {}^{P_*}T_t^s(x,\cdot)\right\|_* = 0$, or equivalently, that $T(x,\cdot) = {}^{P_*}T_t^s(x,\cdot)$. Because $P_* \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it holds that ${}^{P_*}T_t^s \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$, and hence it follows that $T(x,\cdot) = {}^{P_*}T_t^s(x,\cdot) \in \mathscr{T}_x$. □

Lemma 5.45 ([101, Lemma 1]). *Let $\mathscr{M}$ be a non-empty convex set of probability mass functions on $\mathscr{X}$ and let $\mathscr{T}$ be a non-empty and convex set of transition matrices that has separately specified rows. Then the elementwise product*

$$\mathscr{M}\mathscr{T} := \left\{ \sum_{x\in\mathscr{X}} p(x)T(x,\cdot) \,\middle|\, p \in \mathscr{M}, T \in \mathscr{T} \right\}$$

is a non-empty and convex set of probability mass functions on $\mathscr{X}$.

Lemma 5.46. *For any $n \in \mathbb{Z}_{>0}$, let $\mathscr{T}_1,\dots,\mathscr{T}_n$ be non-empty and closed sets of transition matrices. Then their elementwise composition,*

$$\mathscr{T} := \left\{ \prod_{i=1}^{n} T_i \,\middle|\, \forall i \in \{1,\dots,n\} : T_i \in \mathscr{T}_i \right\},$$

is a non-empty and closed set of transition matrices.

Proof. That $\mathscr{T}$ is non-empty follows trivially from the fact that $\mathscr{T}_i$ is non-empty for all $i \in \{1,\dots,n\}$. That its elements are transition matrices follows from Proposition 3.8₉₁. To establish the closure, let $\{T_k\}_{k\in\mathbb{Z}_{>0}}$ be a sequence in $\mathscr{T}$ such that $\lim_{k\to+\infty} T_k = T$. We need to show that $T \in \mathscr{T}$.

To this end, first consider any $k \in \mathbb{Z}_{\geq 0}$. Then there are $T_1^{(k)},\dots,T_n^{(k)}$ such that $T_i^{(k)} \in \mathscr{T}_i$ for all $i \in \{1,\dots,n\}$, and such that $T_k = \prod_{i=1}^{n} T_i^{(k)}$.

Now, fix any $i \in \{1,\dots,n\}$. Due to Lemma 3.9₉₂ it holds that $\|\mathscr{T}_i\| = 1$, whence $\mathscr{T}_i$ is bounded. Because $\mathscr{T}_i$ is also closed, $\mathscr{T}_i$ is sequentially compact by Corollary A.12₃₇₈. Hence, the sequence $\{T_i^{(k)}\}_{k\in\mathbb{Z}_{\geq 0}}$ contains a convergent subsequence whose limit T_i^* is also in $\mathscr{T}_i$. Let us assume without loss of generality that in fact $T_i^* = \lim_{k\to+\infty} T_i^{(k)}$; simply remove the indices that do not correspond to the subsequence.

By repeating the above for all $i \in \{1,\dots,n\}$, each time taking subsequences of the previous (sub)sequence, we find the matrices $T_1^*,\dots,T_n^*$, and $T^* := \prod_{i=1}^{n} T_i^*$ is an element of $\mathscr{T}$. As we are about to show, it holds that $T^* = T$. To this end, fix any $\varepsilon > 0$.

Then, because $\lim_{k\to+\infty} T_k = T$, there is some $m \in \mathbb{Z}_{>0}$ such that $\|T_k - T\| < \varepsilon/2$ for all $k > m$. Moreover, for all $i \in \{1,\dots,n\}$, there is some $m_i \in \mathbb{Z}_{\geq 0}$ such that $\|T_i^{(k)} - T_i^*\| < \varepsilon/2n$ for all $k > m_i$. Now let $k \in \mathbb{Z}_{>0}$ be such that $k > m$ and $k > \max_{i=1,\dots,n} m_i$. Then it holds that

$$\begin{aligned}
\|T - T^*\| &\leq \|T - T_k\| + \|T_k - T_*\| \\
&= \|T - T_k\| + \left\| \prod_{i=1}^{n} T_i^{(k)} - \prod_{i=1}^{n} T_i^* \right\| \\
&\leq \|T - T_k\| + \sum_{i=1}^{n} \left\| T_i^{(k)} - T_i^* \right\| \\
&< \frac{\varepsilon}{2} + \sum_{i=1}^{n} \frac{\varepsilon}{2n} = \varepsilon\,,
\end{aligned}$$

where we used Lemma B.5$_{393}$ for the second inequality. Because $\varepsilon > 0$ is arbitrary, we conclude that $\|T - T^*\| = 0$, or in other words, that $T = T^*$. Hence $T \in \mathscr{T}$, whence $\mathscr{T}$ is closed. □

Lemma 5.47. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. For any $t, s \in \mathbb{R}$ such that $t < s$, let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ denote the set of (history-dependent) transition matrices of elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. For any $x \in \mathscr{X}$, let*

$$\mathscr{T}_x := \left\{ T(x,\cdot) \,\middle|\, T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s \right\}$$

denote the set of x-rows of elements of ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$. Then $\mathscr{T}_x$ is convex.

Proof. The proof works by constructing a set $\mathscr{M}_\varepsilon$ such that $\lim_{\varepsilon \to 0^+} \mathscr{M}_\varepsilon = \mathscr{T}_x$, and such that $\mathscr{M}_\varepsilon$ is convex for all $\varepsilon > 0$. Since, as we will show, convexity is preserved in the limit, the result follows.

So, first fix any $\varepsilon > 0$, let $\varepsilon_* := \varepsilon/2(s-t)$, and let $\delta > 0$ be such that $\delta \|\mathscr{Q}\| \leq 1$ and $\delta \|\mathscr{Q}\|^2 < \varepsilon_*$; because $\mathscr{Q}$ is compact and therefore bounded by Corollary A.12$_{378}$, this is always possible. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, Lemma 5.41$_{241}$ then implies that there is some $v \in \mathscr{U}_{[t,s]}$ with $v = t_0, \dots, t_n$, $n \in \mathbb{Z}_{>0}$ and $\sigma(v) < \delta$, such that for all $T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ there are $Q_1, \dots, Q_n \in \mathscr{Q}$ such that

$$\left\| T(x,\cdot) - \left(\prod_{i=1}^{n} (I + \Delta_i^v Q_i) \right)(x,\cdot) \right\|_* < \frac{\varepsilon}{2}\,. \tag{5.81}$$

Now, with $v \in \mathscr{U}_{[t,s]}$ as above, for all $i \in \{1, \dots, n\}$, let

$$\mathscr{T}_i := \left\{ (I + \Delta_i^v Q) \,\middle|\, Q \in \mathscr{Q} \right\}.$$

Then, because $\Delta_i^v \leq \sigma(v) < \delta$ and because $\delta \|\mathscr{Q}\| \leq 1$, it holds that $\Delta_i^v Q \leq \delta \|\mathscr{Q}\| \leq 1$, and so it follows from Proposition 4.9$_{153}$ that $\mathscr{T}_i$ is a set of transition matrices. Also note that, clearly, $\mathscr{T}_i$ is non-empty because $\mathscr{Q}$ is non-empty. Moreover, because $\mathscr{Q}$ is convex, it is easy to see that $\mathscr{T}_i$ is convex. Similarly, because $\mathscr{Q}$ is closed, also $\mathscr{T}_i$ is closed; and because $\mathscr{Q}$ has separately specified rows, also $\mathscr{T}_i$ has separately specified rows.

Now, let $\mathscr{M}_1 := \{T(x,\cdot) \,|\, T \in \mathscr{T}_1\}$ denote the set of x-rows of the elements of $\mathscr{T}_1$. Then, because $\mathscr{T}_1$ is convex, clearly also $\mathscr{M}_1$ is convex. Moreover, because the elements of $\mathscr{T}_1$ are transition matrices, it follows that the elements of $\mathscr{M}_1$ are probability mass functions on $\mathscr{X}$. Now, for all $i \in \{2,\dots,n\}$, let

$$\mathscr{M}_i := \left\{ \sum_{y\in\mathscr{X}} p(y)T(y,\cdot) \,\middle|\, p \in \mathscr{M}_{i-1}, T \in \mathscr{T}_i \right\}.$$

Then, because each $\mathscr{T}_i$ is a non-empty and convex set of transition matrices that has separately specified rows, and because $\mathscr{M}_1$ is a non-empty and convex set of probability mass functions on $\mathscr{X}$, it follows from Lemma 5.45$_{248}$ and induction on i that, for all $i \in \{1,\dots,n\}$, $\mathscr{M}_i$ is a non-empty and convex set of probability mass functions on $\mathscr{X}$.

Let us also note that

$$\begin{aligned}\mathscr{M}_n &= \left\{ \left(\prod_{i=1}^n T_i\right)(x,\cdot) \,\middle|\, T_i \in \mathscr{T}_i \text{ for all } i \in \{1,\dots,n\} \right\}\\ &= \left\{ \left(\prod_{i=1}^n (I + \Delta_i^\nu Q_i)\right)(x,\cdot) \,\middle|\, Q_1,\dots,Q_n \in \mathscr{Q} \right\}, \end{aligned} \tag{5.82}$$

and therefore, using this first equality, and since $\mathscr{T}_i$ is closed for all $i \in \{1,\dots,n\}$, we conclude from Lemmas 3.46$_{137}$ and 5.46$_{248}$ that $\mathscr{M}_n$ is also closed.

To complete the construction, let $\mathscr{M}_\varepsilon := \mathscr{M}_n$. We will now set out to prove that $\mathscr{M}_\varepsilon$ converges to $\mathscr{T}_x$ as we take ε to zero. The notion of convergence that we use is with respect to the (Pompeiu-)Hausdorff distance [89, Section 4.C] d_H between two sets,

$$d_H(\mathscr{M}_\varepsilon, \mathscr{T}_x) := \sup_{p\in\mathscr{M}_\varepsilon\cup\mathscr{T}_x} \left| d_{\mathscr{M}_\varepsilon}(p) - d_{\mathscr{T}_x}(p) \right|, \tag{5.83}$$

where, for all $p \in \mathscr{M}_\varepsilon \cup \mathscr{T}_x$,

$$d_{\mathscr{M}_\varepsilon}(p) := \inf_{q\in\mathscr{M}_\varepsilon} \|p-q\|_* \quad\text{and}\quad d_{\mathscr{T}_x}(p) := \inf_{q\in\mathscr{T}_x} \|p-q\|_* .$$

It is well-known (and easy to see) that we can also write

$$d_H(\mathscr{M}_\varepsilon, \mathscr{T}_x) = \max\left\{ \sup_{p\in\mathscr{M}_\varepsilon} d_{\mathscr{T}_x}(p),\ \sup_{p\in\mathscr{T}_x} d_{\mathscr{M}_\varepsilon}(p) \right\},$$

which will be a more convenient expression for our purposes. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, $\mathscr{T}_x$ is closed due to Lemma 5.44247. Moreover, as we have seen above, $\mathscr{M}_\varepsilon$ is also closed, and clearly both $\mathscr{T}_x$ and $\mathscr{M}_\varepsilon$ are non-empty. Hence by [89, Section 4.C], d_H is a metric on these sets, and we can consider the convergence of $\lim_{\varepsilon\to 0^+} \mathscr{M}_\varepsilon$ with respect to d_H. To this end, we will show that

$$d_H\left(\mathscr{M}_\varepsilon, \mathscr{T}_x\right) < \varepsilon\,,$$

from which it will follow that $\lim_{\varepsilon\to 0^+} \mathscr{M}_\varepsilon = \mathscr{T}_x$ with respect to d_H. In order to establish this inequality, let us first show that $\sup_{T(x,\cdot)\in\mathscr{T}_x} d_{\mathscr{M}_\varepsilon}\big(T(x,\cdot)\big) < \varepsilon$.

To this end, fix any $T(x,\cdot) \in \mathscr{T}_x$. There are then $Q_1,\dots,Q_n \in \mathscr{Q}$ satisfying Equation (5.81)249. Due to Equation (5.82), there is some corresponding

$$p \coloneqq \left(\prod_{i=1}^{n}(I+\Delta_i^{\nu} Q_i)\right)(x,\cdot) \in \mathscr{M}_\varepsilon\,,$$

and hence

$$\begin{aligned} d_{\mathscr{M}_\varepsilon}\big(T(x,\cdot)\big) &= \inf_{q\in\mathscr{M}_\varepsilon} \left\lVert T(x,\cdot) - q\right\rVert_* \\ &\leq \left\lVert T(x,\cdot) - p\right\rVert_* \\ &= \left\lVert T(x,\cdot) - \left(\prod_{i=1}^{n}(I+\Delta_i^{\nu} Q_i)\right)(x,\cdot)\right\rVert_* < \frac{\varepsilon}{2}\,. \end{aligned}$$

Because the $T(x,\cdot) \in \mathscr{T}_x$ is arbitrary, this implies that

$$\sup_{T(x,\cdot)\in\mathscr{T}_x} d_{\mathscr{M}_\varepsilon}\big(T(x,\cdot)\big) \leq \frac{\varepsilon}{2} < \varepsilon\,.$$

Now for the other direction, choose any $p \in \mathscr{M}_\varepsilon$. Then there are $Q_1,\dots,Q_n \in \mathscr{Q}$ such that

$$p = \left(\prod_{i=1}^{n}(I+\Delta_i^{\nu} Q_i)\right)(x,\cdot)\,.$$

Consider the restricted transition matrix system

$$(T_r^q)_{[t,s]} \coloneqq \left(e^{Q_1(q-r)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{Q_n(q-r)}\right)_{[t_{n-1},t_n]}.$$

Choose an arbitrary $Q_0 \in \mathscr{Q}$, and consider the transition matrix system

$$(T_r^q) \coloneqq \left(e^{Q_0(q-r)}\right)_{[0,t]} \otimes (T_r^q)_{[t,s]} \otimes \left(e^{Q_0(q-r)}\right)_{[s,+\infty)}\,.$$

Then, because $Q_0, Q_1, \dots, Q_n \in \mathcal{Q}$, it follows from Proposition 5.10$_{192}$—with the initial distribution chosen arbitrarily in $\mathcal{M}$—that there is some Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ with corresponding family of transition matrices (T_r^q). Hence in particular, its corresponding transition matrix T_t^s satisfies

$$T_t^s = \prod_{i=1}^{n} e^{Q_i \Delta_i^v}\,.$$

Moreover, because $P \in \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$, it follows from Equation (5.11)$_{197}$ that $T_t^s \in {}^{\mathcal{Q}}_{\mathcal{M}}\mathcal{T}_t^s$, and hence in particular, that $T_t^s(x,\cdot) \in \mathcal{T}_x$. We now find that

$$\begin{aligned}
d_{\mathcal{T}_x}(p) = \inf_{q \in \mathcal{T}_x} \|p - q\|_* \leq \|p - T_t^s(x,\cdot)\|_* &\\
&= \left\| \left(\prod_{i=1}^{n}(I + \Delta_i^v Q_i)\right)(x,\cdot) - \left(\prod_{i=1}^{n} e^{Q_i \Delta_i^v}\right)(x,\cdot) \right\|_* \\
&\leq \left\| \prod_{i=1}^{n}(I + \Delta_i^v Q_i) - \prod_{i=1}^{n} e^{Q_i \Delta_i^v} \right\| \\
&\leq \sum_{i=1}^{n} \left\| (I + \Delta_i^v Q_i) - e^{Q_i \Delta_i^v} \right\| \\
&\leq \sum_{i=1}^{n} (\Delta_i^v)^2 \|Q_i\|^2 \\
&< \sum_{i=1}^{n} \Delta_i^v \delta \|\mathcal{Q}\|^2 \\
&= (s-t)\delta \|\mathcal{Q}\|^2 < (s-t)\epsilon_* = (s-t)\frac{\epsilon}{2(s-t)} = \frac{\epsilon}{2}\,,
\end{aligned}$$

where the second inequality follows from Proposition A.33$_{390}$; the third inequality follows from Lemma B.5$_{393}$, which we can use because as established towards the beginning of this proof, all matrices involved are transition matrices; the fourth inequality follows from Lemma B.8$_{394}$; the fifth inequality follows because $\Delta_i^v \leq \sigma(v) < \delta$ and $\|Q_i\| \leq \|\mathcal{Q}\|$ for all $i \in \{1, \dots, n\}$; and the final inequality follows because $\delta \|\mathcal{Q}\|^2 < \epsilon_*$. Because $p \in \mathcal{M}_\epsilon$ is arbitrary, this implies that

$$\sup_{p \in \mathcal{M}_\epsilon} d_{\mathcal{T}_x}(p) \leq \frac{\epsilon}{2} < \epsilon\,,$$

and hence we conclude that, indeed,

$$d_H\big(\mathcal{M}_\epsilon, \mathcal{T}_x\big) < \epsilon\,,$$

which, since $\epsilon > 0$ is arbitrary, implies the convergence $\lim_{\epsilon \to 0^+} \mathcal{M}_\epsilon = \mathcal{T}_x$ with respect to d_H.

As both $\mathscr{T}_x$ and $\mathscr{M}_\varepsilon$ are non-empty sets of probability mass functions, these sets are bounded, whence the convergence with respect to d_H is equivalent to set-convergence in the Painlevé-Kuratowski sense [89, Section 4.B-4.C]. Because the limit of a sequence of convex sets that converges in the Painlevé-Kuratowski sense is itself a convex set [89, Proposition 4.15], and because, as we have seen above, $\mathscr{M}_\varepsilon$ is convex for any $\varepsilon > 0$, it follows that $\lim_{\varepsilon\to 0^+} \mathscr{M}_\varepsilon = \mathscr{T}_x$ is also convex. □

Proof of Theorem 5.21$_{198}$. That ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is non-empty is an immediate consequence of the fact that $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is non-empty.

If $t = s$ then for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it follows from Proposition 4.2$_{149}$ that ${}^{P}T^s_{t,x_u} = I$ for all $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$, whence in that case ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \{I\}$, and the result is then trivial because the singleton set $\{I\}$ is clearly closed, convex, and has separately specified rows.

Conversely, suppose that $t < s$. That ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is closed, convex, and has separately specified rows then follows from Lemmas 3.46$_{137}$, 5.43$_{246}$, 5.44$_{247}$, and 5.47$_{249}$. □

5.E Proofs of Results in Section 5.4

Proof of Lemma 5.27$_{203}$. Fix any $\varepsilon > 0$. Then due to Propositions 5.25$_{201}$ and 5.24$_{201}$, and the properties of the infimum, there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P}T_t^s$ such that

$$ {}^{P}T_t^s f(x_t) < \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s) \,|\, X_t = x_t] + \varepsilon\,. \tag{5.84}$$

This transition matrix ${}^{P}T_t^s$ is an element of the set ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ defined in Equation (5.11)$_{197}$. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 5.42$_{243}$ that there is a Markov chain $P' \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P'}T_t^s$ such that ${}^{P'}T_t^s(x_t,\cdot) = {}^{P}T_t^s(x_t,\cdot)$. Hence it follows that also

$$ {}^{P'}T_t^s f(x_t) = {}^{P}T_t^s f(x_t)\,. \tag{5.85}$$

Because $P' \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is a Markov chain, it follows that $P' \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ due to Proposition 5.9$_{190}$. Hence, by using Proposition 5.25$_{201}$ we obtain

$$\begin{aligned}
\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s) \,|\, X_t = x_t] &= \inf_{P_* \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}} {}^{P_*}T_t^s f(x_t) \\
&\leq {}^{P'}T_t^s f(x_t) \\
&= {}^{P}T_t^s f(x_t) < \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s) \,|\, X_t = x_t] + \varepsilon\,,
\end{aligned}$$

where we used Equation (5.85)↶ for the second equality and Equation (5.84)↶ for the last inequality. Because $\varepsilon > 0$ is arbitrary, we conclude that

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t] \leq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,.$$

However, we already know from Proposition 5.22₁₉₉ that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t] \leq \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]$, whence the equality follows. □

*Proof of Proposition 5.28*₂₀₃. We will start by proving that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] \geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]$. To this end, first note that

$$\left\{{}^{P}T^{s}_{t,x_u}\,\middle|\,P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\right\} \subseteq \left\{{}^{P}T^{s}_{t,y_v}\,\middle|\,P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}},\ v \in \mathscr{U}_{<t},\ y_v \in \mathscr{X}_v\right\} = {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}\,, \tag{5.86}$$

with ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}$ as defined in Equation (5.11)₁₉₇. Hence it follows from Proposition 5.25₂₀₁ that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u]\\
&\qquad= \inf_{P\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} {}^{P}T^{s}_{t,x_u} f(x_t)\\
&\qquad\geq \inf\left\{{}^{P}T^{s}_{t,y_v} f(x_t)\,\middle|\,P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}},\ v \in \mathscr{U}_{<t},\ y_v \in \mathscr{X}_v\right\}\\
&\qquad= \inf\left\{{}^{P}T^{s}_{t} f(x_t)\,\middle|\,P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\right\}\\
&\qquad= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,,
\end{aligned}$$

where the inequality follows from Equation (5.86), the second equality follows from Proposition 5.20₁₉₇ and the fact that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, and the final equality again follows from Proposition 5.25₂₀₁. Using this inequality, the remainder of this proof is now straightforward.

Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, we can use Lemma 5.27₂₀₃ to find

$$\begin{aligned}
\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t] &= \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\\
&= \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u]\\
&\geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u]\\
&\geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]\,,
\end{aligned}$$

where the second equality follows from Proposition 5.26₂₀₂, the first inequality follows from Proposition 5.22₁₉₉, and the second inequality was derived in the first part of this proof. □

Proof of Proposition 5.30$_{204}$. Fix any $x \in \mathscr{X}$. Then it holds by Proposition 5.25$_{201}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x] = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} {}^{P}T_t^s f(x) = \inf_{T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s} Tf(x)\,,$$

where the second equality follows from Proposition 5.20$_{197}$ and the fact that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows. Due to these same properties, it follows from Theorem 5.21$_{198}$ that ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is non-empty, closed, convex, and has separately specified rows. It therefore follows from Lemma 3.46$_{137}$ that the set of x-rows $\mathscr{T}_x := \left\{T(x,\cdot)\,|\,T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s\right\}$ of ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is also closed. This implies that for all $x \in \mathscr{X}$ there is some $T_x(x,\cdot) \in \mathscr{T}_x$ such that

$$\inf_{T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s} Tf(x) = \inf_{T(x,\cdot) \in \mathscr{T}_x} \sum_{y \in \mathscr{X}} T(x,y)f(y) = \sum_{y \in \mathscr{X}} T_x(x,y)f(y)\,.$$

Because this is true for all $x \in \mathscr{X}$, and because ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ has separately specified rows, it follows that there is some $T_* \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ such that $T_*(x,\cdot) = T_x(x,\cdot)$ for all $x \in \mathscr{X}$. For all $x \in \mathscr{X}$ this matrix T_* then satisfies

$$T_*f(x) = \sum_{y \in \mathscr{X}} T_*(x,y)f(y) = \sum_{y \in \mathscr{X}} T_x(x,y)f(y) = \inf_{T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s} Tf(x)\,,$$

and hence

$$T_*f(x) = \inf_{T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s} Tf(x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x]\,.$$

Moreover, because $T_* \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ and ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \left\{{}^{P}T_t^s\,\middle|\,P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\right\}$ due to Proposition 5.20$_{197}$, it follows that there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P}T_t^s = T_*$. Hence it follows from Corollary 4.4$_{150}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x] = T_*f(x) = {}^{P}T_t^s f(x) = \mathbb{E}_P[f(X_s)\,|\,X_t = x]\,.$$

Therefore, and because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Proposition 5.28$_{203}$ that for all $u \in \mathscr{U}_{<t}$, all $x \in \mathscr{X}_t$, and all $x_u \in \mathscr{X}_u$, it holds that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x] = \mathbb{E}_P[f(X_s)\,|\,X_t = x]\,,$$

which concludes the proof. □

Proof of Theorem 5.32$_{208}$. Let $g \in \mathscr{L}(\mathscr{X}_{u\cup v})$ be defined, for all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$, as

$$g(y_{u\cup v}) := \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})\,|\,X_{u\cup v} = y_{u\cup v}\right]\,.$$

Then it follows from Definition 5.8$_{198}$ that for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$, it holds that

$$g(y_{u\cup v}) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_{u\cup v} = y_{u\cup v}\right] \leq \mathbb{E}_P\left[f(X_{u\cup v\cup w})|X_{u\cup v} = y_{u\cup v}\right]. \tag{5.87}$$

For any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it now follows from Proposition 5.31$_{208}$ that

$$\begin{aligned}
&\mathbb{E}_P\left[f(X_{u\cup v\cup w})|X_u = x_u\right] \\
&\quad= \mathbb{E}_P\Big[\mathbb{E}_P[f(X_{u\cup v\cup w})\,|X_{u\cup v}]\,\Big|X_u = x_u\Big] \\
&\quad= \sum_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} \mathbb{E}_P[f(X_{u\cup v\cup w})\,|X_{u\cup v} = y_{u\cup v}]P(X_{u\cup v} = y_{u\cup v}\,|X_u = x_u) \\
&\quad\geq \sum_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} g(y_{u\cup v})P(X_{u\cup v} = y_{u\cup v}\,|X_u = x_u) \\
&\quad= \mathbb{E}_P\big[g(X_{u\cup v})|X_u = x_u\big] \\
&\quad\geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[g(X_{u\cup v})|X_u = x_u\big],
\end{aligned}$$

where the second equality used Proposition 2.23$_{73}$, the first inequality used Equation (5.87), the third equality used Proposition 2.23$_{73}$, and the final inequality used Definition 5.8$_{198}$.

Since this holds for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, this implies that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_u = x_u\right] \geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[g(X_{u\cup v})|X_u = x_u\big]. \tag{5.88}$$

Now fix any $\varepsilon \in \mathbb{R}_{>0}$. Then due to Definition 5.8$_{198}$ and Proposition 5.24$_{201}$, there is some $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that

$$\mathbb{E}_{P_\emptyset}\big[g(X_{u\cup v})|X_u = x_u\big] < \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[g(X_{u\cup v})\big|X_u = x_u\big] + \frac{\varepsilon}{2}. \tag{5.89}$$

Moreover, it follows from Proposition 2.25$_{75}$ and the fact that $u < v$, that

$$\mathbb{E}_{P_\emptyset}\big[g(x_u, X_v)|X_u = x_u\big] = \mathbb{E}_{P_\emptyset}\big[g(X_{u\cup v})|X_u = x_u\big],$$

and hence it follows from Equation (5.89) that

$$\mathbb{E}_{P_\emptyset}\big[g(X_{u\cup v})|X_u = x_u\big] < \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[g(X_{u\cup v})\big|X_u = x_u\big] + \frac{\varepsilon}{2}. \tag{5.90}$$

Similarly, for all $x_v \in \mathscr{X}_v$, and due to Proposition 5.24$_{201}$, there is some $P_{x_v} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that

$$\begin{aligned}
\mathbb{E}_{P_{x_v}}\left[f(X_{u\cup v\cup w})|X_{u\cup v} = x_{u\cup v}\right] &< \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_{u\cup v} = x_{u\cup v}\right] + \frac{\varepsilon}{2} \\
&= g(x_{u\cup v}) + \frac{\varepsilon}{2}.
\end{aligned} \tag{5.91}$$

Now, for all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$, let $P_{y_{u\cup v}} := P_{y_v}$. Since $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, Theorem 5.11_{193} now implies the existence of a process $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that for all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$ and $x_w \in \mathscr{X}_w$:

$$P(X_v = y_v \,|\, X_u = y_u) = P_\emptyset(X_v = y_v \,|\, X_u = y_u)$$

and

$$\begin{aligned} P(X_w = x_w \,|\, X_{u\cup v} = y_{u\cup v}) &= P_{y_{u\cup v}}(X_w = x_w \,|\, X_{u\cup v} = y_{u\cup v}) \\ &= P_{y_v}(X_w = x_w \,|\, X_{u\cup v} = y_{u\cup v})\,. \end{aligned}$$

Hence in particular, due to Equation (5.90) and Proposition 2.23_{73}, this P satisfies

$$\mathbb{E}_P[g(x_u, X_v)|X_u = x_u] = \mathbb{E}_{P_\emptyset}[g(x_u, X_v)|X_u = x_u] < \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{u\cup v})\middle|X_u = x_u\right] + \frac{\varepsilon}{2}$$

and, for all $x_v \in \mathscr{X}_v$, due to Propositions 2.25_{75} and 2.23_{73},

$$\begin{aligned} \mathbb{E}_P[f(X_{u\cup v\cup w})|X_{u\cup v} = x_{u\cup v}] &= \mathbb{E}_P[f(x_{u\cup v}, X_w)|X_{u\cup v} = x_{u\cup v}] \\ &= \mathbb{E}_{P_{x_v}}[f(x_{u\cup v}, X_w)|X_{u\cup v} = x_{u\cup v}] \\ &= \mathbb{E}_{P_{x_v}}[f(X_{u\cup v\cup w})|X_{u\cup v} = x_{u\cup v}] < g(x_{u\cup v}) + \frac{\varepsilon}{2}\,, \end{aligned}$$

where we used Equation (5.91) for the final inequality.

Hence it follows from Properties $\mathrm{CE4}_{78}$ and $\mathrm{CE6}_{79}$ that

$$\begin{aligned} \mathbb{E}_P\left[\mathbb{E}_P\left[f(X_{u\cup v\cup w})|x_u, X_v\right]|X_u = x_u\right] &\leq \mathbb{E}_P\left[g(x_u, X_v)\,\middle|\,X_u = x_u\right] + \frac{\varepsilon}{2} \\ &< \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{u\cup v})\middle|X_u = x_u\right] + \varepsilon. \end{aligned}$$

Since ε is arbitrary, and because

$$\begin{aligned} \mathbb{E}_P\left[\mathbb{E}_P\left[f(X_{u\cup v\cup w})|x_u, X_v\right]|X_u = x_u\right] &= \mathbb{E}_P\left[\mathbb{E}_P\left[f(X_{u\cup v\cup w})|X_{u\cup v}\right]|X_u = x_u\right] \\ &= \mathbb{E}_P\left[f(X_{u\cup v\cup w})|X_u = x_u\right] \\ &\geq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_u = x_u\right], \end{aligned}$$

where the first equality used Proposition 2.25_{75}, the second equality used Proposition 5.31_{208}, and the inequality used Definition 5.8_{198}, this implies that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_u = x_u\right] \leq \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{u\cup v})\middle|X_u = x_u\right]$. Hence, and because of Equation (5.88), we find that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_u = x_u\right] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{u\cup v})\middle|X_u = x_u\right].$$

The result now follows because $g(X_{u\cup v}) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v\cup w})|X_{u\cup v}\right]$. □

6

Lower Transition Operators for Continuous-Time Imprecise-Markov Chains

"It's dangerous to go alone! Take this."

The Legend of Zelda

When we discussed discrete-time imprecise-Markov chains in Chapter 3$_{83}$, we also considered the corresponding *lower transition operators*. We saw how these operators generalise the transition matrices corresponding to (precise) stochastic processes, and how they can be used to represent the (conditional) lower expectations for these imprecise models. It is the aim of this chapter to also introduce lower transition operators for *continuous*-time imprecise-Markov chains.

To this end, we will start in Section 6.1$_{\curvearrowright}$ by defining them directly using the lower expectations for these models. The remainder of this chapter is dedicated to deriving an alternative characterisation of such lower transition operators, which we can eventually use as a computational tool. We do this by introducing lower transition *rate* operators in Section 6.2$_{265}$, and subsequently consider exponentials of such operators in Section 6.3$_{269}$; we also discuss there an algorithm that can be used to evaluate these exponentials numerically. In Section 6.4$_{279}$, we show that, under some conditions, these exponentials coincide with the lower transition operators of imprecise-Markov chains, as we define them in Section 6.1. We bring all these parts together in Section 6.5$_{284}$, where we consider a general computational (i.e. algorithmic) framework for numerically evaluating lower expectations of arbitrary u-measurable functions for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, under some structural assumptions on its characterising set $\mathscr{Q}$ of rate matrices. We conclude this chapter with Section 6.6$_{290}$, where we demonstrate the methods developed earlier using a simple numerical example.

6.1 Induced Lower Transition Operators

In Section 3.4$_{116}$, and with Definition 3.15$_{116}$ in particular, we gave the general definition of lower transition operators. For our present purposes, it remains to define them as objects corresponding to a set of (continuous-time) stochastic processes. The following result will allow us to do this.

Proposition 6.1. *Let $\mathscr{P} \subseteq \mathbb{P}$ be a non-empty set of stochastic processes, fix any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, and for all $f \in \mathscr{L}$, let $\underline{T}_t^s f \in \mathscr{L}(\mathscr{X})$ be defined for all $x \in \mathscr{X}$ as $\underline{T}_t^s f(x) := \underline{\mathbb{E}}[f(X_s)\,|\,X_t = x]$, where $\underline{\mathbb{E}}$ is the lower expectation corresponding to $\mathscr{P}$, as in Definition 5.8$_{198}$. Then the map $\underline{T}_t^s : f \mapsto \underline{T}_t^s f$ on $\mathscr{L}(\mathscr{X})$ is a lower transition operator.*

Proof. Fix any $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, any $f,g \in \mathscr{L}(\mathscr{X})$, any $\lambda \in \mathbb{R}_{\geq 0}$, and any $x \in \mathscr{X}$. It follows from Proposition 5.24$_{201}$ that $\min_{y\in\mathscr{X}} f(y) \leq \underline{T}_t^s f(x) \leq \max_{y\in\mathscr{X}} f(y)$, which immediately implies that $\underline{T}_t^s$ satisfies Property LT1$_{116}$. Moreover, because $f \in \mathscr{L}(\mathscr{X})$ and since $\mathscr{X}$ is finite, it follows that $\underline{T}_t^s f(x)$ is real-valued, and hence $\underline{T}_t^s$ maps $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$.

Next, it follows from Definition 5.8$_{198}$ together with Property CE2$_{78}$ and the properties of the infimum, that

$$\begin{aligned}
\underline{T}_t^s(f+g)(x) &= \underline{\mathbb{E}}[(f+g)(X_s)\,|\,X_t = x] \\
&= \inf_{P\in\mathscr{P}} \mathbb{E}_P[(f+g)(X_s)\,|\,X_t = x] \\
&= \inf_{P\in\mathscr{P}} \mathbb{E}_P[f(X_s)\,|\,X_t = x] + \mathbb{E}_P[g(X_s)\,|\,X_t = x] \\
&\geq \inf_{P\in\mathscr{P}} \mathbb{E}_P[f(X_s)\,|\,X_t = x] + \inf_{P\in\mathscr{P}} \mathbb{E}_P[g(X_s)\,|\,X_t = x] \\
&= \underline{\mathbb{E}}[f(X_s)\,|\,X_t = x] + \underline{\mathbb{E}}[g(X_s)\,|\,X_t = x] = \underline{T}_t^s f(x) + \underline{T}_t^s g(x)\,,
\end{aligned}$$

and therefore $\underline{T}_t^s$ satisfies Property LT2$_{116}$. Finally, from Definition 5.8$_{198}$, Property CE3$_{78}$ and the properties of the infimum, and using that $\lambda \geq 0$, we find that

$$\begin{aligned}\underline{T}_t^s(\lambda f)(x) &= \underline{\mathbb{E}}[\lambda f(X_s) \,|\, X_t = x] \\ &= \inf_{P\in\mathscr{P}} \mathbb{E}_P[\lambda f(X_s) \,|\, X_t = x] \\ &= \inf_{P\in\mathscr{P}} \lambda \mathbb{E}_P[\lambda f(X_s) \,|\, X_t = x] \\ &= \lambda \inf_{P\in\mathscr{P}} \mathbb{E}_P[\lambda f(X_s) \,|\, X_t = x] = \lambda \underline{\mathbb{E}}[f(X_s) \,|\, X_t = x] = \lambda \underline{T}_t^s f(x)\,,\end{aligned}$$

so $\underline{T}_t^s$ satisfies Property LT3$_{116}$.

Hence, we have established that $\underline{T}_t^s$ satisfies Properties LT1$_{116}$–LT3$_{116}$ and that it is a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$. Therefore, it is a lower transition operator by Definition 3.15$_{116}$. □

This allows us to introduce the following generic definition, which provides the basis for the developments in this chapter.

Definition 6.1. *For any non-empty set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$, we define the corresponding* family of lower transition operators $(\underline{T}_t^s)$, *which is a two-parameter family of lower transition operators $\underline{T}_t^s$ that are defined for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, as*

$$\underline{T}_t^s f(x) := \underline{\mathbb{E}}[f(X_s) \,|\, X_t = x] \quad \textit{for all } f \in \mathscr{L}(\mathscr{X}) \textit{ and } x \in \mathscr{X}\,,$$

where $\underline{\mathbb{E}}$ is the lower expectation for $\mathscr{P}$, as in Definition 5.8$_{198}$.

Proposition 6.1 ensures that these corresponding lower transition operators are all, indeed, lower transition operators. Moreover, in Definition 6.1, we have defined lower transition operators using the lower expectations with respect to sets of stochastic processes. This is analogous to how, e.g., De Cooman *et al.* [22] defined lower transition operators corresponding to discrete-time imprecise-Markov chains. However, it is notably different from our own developments in Chapter 3$_{83}$, where we took sets of transition matrices as our starting point, and derived both discrete-time imprecise-Markov chains and their corresponding lower transition operators from these sets. Moreover, in Section 5.3$_{194}$ we also discussed the sets of transition matrices that are induced by a given continuous-time imprecise-Markov chain, so it seems natural to investigate how these different objects are related. The following result establishes this connection between these sets of transition matrices ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$, and the lower transition operators $(\underline{T}_t^s)$ corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, under some structural assumptions on $\mathscr{Q}$.

Theorem 6.2. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $(\underline{T}_t^s)$ denote the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, $\underline{T}_t^s$*

is the lower transition operator corresponding to the set ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ *of (history-dependent) transition matrices induced by* $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$*, as in Equation* (5.11)$_{197}$*. Moreover, it holds that* ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \mathscr{T}_{\underline{T}_t^s}$*, where* $\mathscr{T}_{\underline{T}_t^s}$ *is the set of transition matrices that dominate* $\underline{T}_t^s$*, as in Definition 3.17*$_{120}$.

Proof. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Definition 6.1$_{\curvearrowleft}$ and Propositions 5.25$_{201}$ and 5.20$_{197}$—in that order—that, for any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$,

$$\underline{T}_t^s f(x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x] = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} {}^{P}T_t^s f(x) = \inf_{T \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s} Tf(x)\,,$$

so $\underline{T}_t^s$ is the lower transition operator corresponding to ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$.

Moreover, because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Theorem 5.21$_{198}$ that ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ is a non-empty, closed, and convex set of transition matrices that has separately specified rows. Because we have already established that ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ has $\underline{T}_t^s$ as its corresponding lower transition operator, it follows from Corollary 3.38$_{120}$ that ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \mathscr{T}_{\underline{T}_t^s}$. □

Moving on, using these corresponding lower transition operators we can derive expressions for the lower expectations of continuous-time imprecise-Markov chains, that are analogous to the expressions obtained in Section 3.5$_{121}$ for discrete-time imprecise-Markov chains. As with our developments in Section 3.5$_{121}$, we again abuse our notation to write, for any lower transition operator $\underline{T}$, any $t \in \mathbb{R}_{>0}$, any $u \in \mathscr{U}_{<t}$ with $u \neq \emptyset$, any $f \in \mathscr{L}(\mathscr{X}_{u\cup\{t\}})$, and any $x_u \in \mathscr{X}_u$,

$$\underline{T}\,f(x_u) := \big[\underline{T}\,f(x_u,\cdot)\big](x_{\max u})\,,$$

where, on the right-hand side, we have applied the (original) operator $\underline{T}$ to $f(x_u,\cdot)$, which is the corresponding "projection" of f onto $\mathscr{L}(\mathscr{X}_t)$, i.e. the element of $\mathscr{L}(\mathscr{X}_t)$ corresponding to the t-measurable function $f(x_u,X_t)$. This notation allows us to formulate lower expectations of the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, using its corresponding lower transition operators, as follows.

Lemma 6.3. *Let* $\mathscr{Q}$ *be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let* $\mathscr{M}$ *be a non-empty set of probability mass functions on* $\mathscr{X}$*. Let* $(\underline{T}_t^s)$ *be the family of lower transition operators corresponding to* $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$*, as in Definition 6.1*$_{\curvearrowleft}$*. Then for all* $t,s \in \mathbb{R}_{\geq 0}$ *with* $t < s$*, all* $u \in \mathscr{U}_{<t}$*, all* $f \in \mathscr{L}(\mathscr{X}_{u\cup\{t,s\}})$*, and all* $x_u \in \mathscr{X}_u$ *and* $x_t \in \mathscr{X}_t$*, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Big[f\big(X_{u\cup\{t\}},X_s\big)\,\Big|\,X_t = x_t, X_u = x_u\Big] = \underline{T}_t^s f(x_{u\cup\{t\}})\,.$$

Proof. Fix any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Because $u < t < s$ it follows from Proposition 2.25₇₅ that

$$\mathbb{E}_P\left[f\left(X_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right] = \mathbb{E}_P\left[f\left(x_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right].$$

Because this is true for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it follows from Definition 5.8_{198} that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}&\left[f\left(X_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right]\\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f\left(x_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right].\end{aligned}$$

Since $f(x_{u\cup\{t\}},X_s)$ is an $\{s\}$-measurable function (that depends on the single time point s), and because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Proposition 5.28_{203} and Definition 6.1_{261} that, for all $x_u \in \mathscr{X}_u$ and all $x_t \in \mathscr{X}_t$,

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f\left(x_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right] &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f\left(x_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t\right]\\ &= \left[\underline{T}_t^s f(x_{u\cup\{t\}},\cdot)\right](x_t).\end{aligned}$$

Using the notation established above, we have that for all $x_u \in \mathscr{X}_u$ and all $x_t \in \mathscr{X}_t$, since $u < t$ and so $t = \max u \cup \{t\}$, that

$$\left[\underline{T}_t^s f(x_{u\cup\{t\}},\cdot)\right](x_t) = \underline{T}_t^s f(x_{u\cup\{t\}}),$$

and hence

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f\left(X_{u\cup\{t\}},X_s\right)\,\middle|\,X_t = x_t, X_u = x_u\right] = \underline{T}_t^s f(x_{u\cup\{t\}}),$$

which concludes the proof. □

Theorem 6.4. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Let $(\underline{T}_t^s)$ be the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then for all $u,v \in \mathscr{U}_{\supset\emptyset}$ such that $u < v$, with $v = t_0,\ldots,t_n$, $n \in \mathbb{Z}_{\geq 0}$, all $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$, and all $x_u \in \mathscr{X}_u$, it holds that*[1]

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right] = \underline{T}^{t_0}_{\max u}\underline{T}^{t_1}_{t_0}\cdots\underline{T}^{t_n}_{t_{n-1}}f(x_u). \tag{6.1}$$

[1] We have over-expanded the composition on the right-hand side of Equation (6.1) for clarity of exposition, but this means it is only correct for $n \geq 2$. If $n = 0$ the right-hand side should simply read $\underline{T}^{t_0}_{\max u}f(x_u)$, and if $n = 1$ it should read $\underline{T}^{t_0}_{\max u}\underline{T}^{t_1}_{t_0}f(x_u)$.

Proof. If $n = 0$ then $v = \{t_0\}$ and then the result follows trivially from Lemma 6.3₂₆₂ and the fact that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows. So let us suppose for the remainder of this proof that $n \geq 1$.

Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, it follows from Theorem 5.32₂₀₈ that

$$\begin{aligned}
&\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup v})\,\big|\,X_u = x_u\bigr] \\
&\quad= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup\{t_0,\ldots,t_n\}})\,\big|\,X_u = x_u\bigr] && (6.2)\\
&\quad= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Bigl[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup\{t_0,\ldots,t_n\}})\,\big|\,X_u, X_{t_0,\ldots,t_{n-1}}\bigr]\,\Big|\,X_u = x_u\Bigr]\,. && (6.3)
\end{aligned}$$

Consider the inner lower expectation in Equation (6.3). This is a $u \cup \{t_0,\ldots,t_{n-1}\}$-measurable function, and it follows from Lemma 6.3₂₆₂—using that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows—that, for all $x_u \in \mathscr{X}_u$ and all $x_{\{t_0,\ldots,t_{n-1}\}} \in \mathscr{X}_{\{t_0,\ldots,t_{n-1}\}}$, it holds that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup\{t_0,\ldots,t_n\}})\,\big|\,X_{u\cup\{t_0,\ldots,t_{n-1}\}} = x_{u\cup\{t_0,\ldots,t_{n-1}\}}\bigr] = \underline{T}_{t_{n-1}}^{t_n} f(x_{u\cup\{t_0,\ldots,t_{n-1}\}})\,.$$

Hence we can replace the inner conditional lower expectation in Equation (6.3) with the $u \cup \{t_0,\ldots,t_{n-1}\}$-measurable function $\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0,\ldots,t_{n-1}\}})$, to obtain

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup\{t_0,\ldots,t_n\}})\,\big|\,X_u = x_u\bigr] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Bigl[\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0,\ldots,t_{n-1}\}})\,\Big|\,X_u = x_u\Bigr]\,.$$

Comparing this to Equation (6.2), we see that we have reduced a lower expectation of the $u \cup \{t_0,\ldots,t_n\}$-measurable function f, to a lower expectation of the $u \cup \{t_0,\ldots,t_{n-1}\}$-measurable function $\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0,\ldots,t_{n-1}\}})$. Hence, if $n > 1$, we now iteratively repeat the above process, first substituting the $u \cup \{t_0,\ldots,t_{n-2}\}$-measurable function $\underline{T}_{t_{n-2}}^{t_{n-1}}\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0,\ldots,t_{n-2}\}})$, and so forth, until we are eventually left with

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\bigl[f(X_{u\cup\{t_0,\ldots,t_n\}})\,\big|\,X_u = x_u\bigr] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Bigl[\underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0\}})\,\Big|\,X_u = x_u\Bigr]\,.$$

The remaining lower expectation on the right-hand side of this expression is taken over a $u \cup \{t_0\}$-measurable function, and one last application of Lemma 6.3₂₆₂—again using that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows—now yields that, for all $x_u \in \mathscr{X}_u$,

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Bigl[\underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(X_{u\cup\{t_0\}})\,\Big|\,X_u = x_u\Bigr] = \underline{T}_{\max u}^{t_0}\underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_u)\,,$$

which completes the proof. □

This result suggests that, if we want to compute the lower expectation of any u-measurable function for the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it suffices to be able to evaluate its corresponding lower transition operators. Of course, since this family $(\underline{T}_t^s)$ is defined in terms of lower expectations, this does not immediately get us anywhere. However, in the remainder of this chapter we will develop an alternative characterisation of these lower transition operators that, as it turns out, allows them to be efficiently evaluated.

6.2 Lower Transition Rate Operators

We recall from Section 3.4_{116} that there is a strong connection between sets $\mathscr{T}$ of transition matrices, and lower transition operators $\underline{T}$. Specifically, these notions are dual, in the sense that for any non-empty set $\mathscr{T}$, we can obtain a lower transition operator $\underline{T}$ as the lower envelope $\underline{T}f := \inf_{T\in\mathscr{T}} Tf$. Conversely, any lower transition operator $\underline{T}$ has a dominating set $\mathscr{T}_{\underline{T}}$ of transition matrices. We established with Proposition 3.37_{120} and Corollary 3.38_{120} that $\mathscr{T} = \mathscr{T}_{\underline{T}}$ if and only if $\mathscr{T}$ is non-empty, closed, convex, has separately specified rows, and has $\underline{T}$ as its lower envelope.

It is the purpose of this section to introduce the notion of *lower transition rate operators*, which are to transition rate matrices as lower transition operators are to transition matrices. That is, they are a super-additive and non-negatively homogeneous generalisation of the linear maps represented by rate matrices. In full analogy with the developments in Section 3.4_{116}, we will show that they can be obtained as lower envelopes of sets $\mathscr{Q}$ of rate matrices. Moreover, we will establish a duality between lower transition rate operators and sets of rate matrices that is analogous to that for lower transition operators and sets of transition matrices.

Lower transition rate operators will form the core of our developments of efficient computational methods in later sections. For now, let us start with the general definition.

Definition 6.2. *A map $\underline{Q} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{Q}f$ is called a* lower transition rate operator *if, for all $f, g \in \mathscr{L}(\mathscr{X})$, all $\lambda \in \mathbb{R}_{\geq 0}$, all constant functions $\mu \in \mathscr{L}(\mathscr{X})$, and all $x \in \mathscr{X}$:*

LR1: $\underline{Q}\mu(x) = 0$;

LR2: $\underline{Q}\mathbb{I}_y(x) \geq 0$ *for all $y \in \mathscr{X}$ such that $x \neq y$;*

LR3: $\underline{Q}(f+g)(x) \geq \underline{Q}f(x) + \underline{Q}g(x)$;

LR4: $\underline{Q}(\lambda f)(x) = \lambda \underline{Q}f(x)$.

Such lower transition rate operators furthermore satisfy the following properties—see Reference [17] for a proof.

Proposition 6.5. *For any lower transition rate operator $\underline{Q}$ and any two non-negatively homogeneous operators A, B from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$:*

LR5: $\|\underline{Q}\| \leq 2\max_{x\in\mathscr{X}} |\underline{Q}\mathbb{I}_x(x)| < +\infty$;

LR6: $\|\underline{Q}A - \underline{Q}B\| \leq 2\|\underline{Q}\| \|A - B\|$.

Note that Properties LR1$_{\curvearrowleft}$ and LR2$_{\curvearrowleft}$ essentially preserve Properties R1$_{150}$ and R2$_{150}$ from Definition 4.4$_{150}$. The main difference lies in the fact that a rate matrix Q is a linear map, whereas Properties LR3$_{\curvearrowleft}$ and LR4$_{\curvearrowleft}$ merely require that a lower transition rate operator $\underline{Q}$ should be super-additive and non-negatively homogeneous. Therefore, every rate matrix is clearly a lower transition rate operator, with the latter concept providing a generalisation of the former.

A first reason why this specific generalisation is of interest, is that it extends the relation between transition matrices and rate matrices that was established in Propositions 4.9$_{153}$ and 4.10$_{153}$, to a relation between lower transition operators and lower transition rate operators. The following two results formalise this.

Proposition 6.6 ([17, Proposition 5])**.** *Consider any lower transition rate operator $\underline{Q}$, and any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta\|\underline{Q}\| \leq 1$. Then $(I + \Delta\underline{Q})$ is a lower transition operator.*

Proposition 6.7 ([17, Proposition 6])**.** *Consider any lower transition operator $\underline{T}$, and any $\Delta \in \mathbb{R}_{>0}$. Then $1/\Delta(\underline{T} - I)$ is a lower transition rate operator.*

Let us now introduce the lower transition rate operator corresponding to a given set $\mathscr{Q} \subset \mathscr{R}$ of rate matrices, which is the lower envelope of $\mathscr{Q}$. We need the following result.

Lemma 6.8. *For any non-empty and bounded set $\mathscr{Q}$ of rate matrices, any $f \in \mathscr{L}(\mathscr{X})$, and any $x \in \mathscr{X}$, it holds that $\inf_{Q\in\mathscr{Q}} Qf(x) \in \mathbb{R}$.*

Proof. Because $\mathscr{Q}$ is bounded (see Definition A.12$_{376}$), there is some $B \in \mathbb{R}$ such that $\sup_{Q\in\mathscr{Q}} \|Q\| = \|\mathscr{Q}\| < B$. Therefore, and using Property N11$_{64}$ it holds for any $Q \in \mathscr{Q}$ that $|Qf(x)| \leq \|Qf\| \leq \|Q\| \|f\| \leq \|\mathscr{Q}\| \|f\| \leq B\|f\|$, which implies that $-B\|f\| \leq Qf(x) \leq B\|f\|$. Because this is true for all $Q \in \mathscr{Q}$, and because $\mathscr{Q}$ is non-empty, it follows that $-B\|f\| \leq \inf_{Q\in\mathscr{Q}} Qf(x) \leq B\|f\|$. Because $B\|f\| \in \mathbb{R}$, this implies that also $\inf_{Q\in\mathscr{Q}} Qf(x)$ is real-valued. □

This allows us to define the lower envelope of $\mathscr{Q}$ as follows; Lemma 6.8 guarantees that the codomain of this map is indeed $\mathscr{L}(\mathscr{X})$.

Definition 6.3. *For any non-empty and bounded set $\mathscr{Q}$ of rate matrices, we define its* lower envelope $\underline{Q} : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto \underline{Q}f$ *where, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$, we let $\underline{Q}f(x) := \inf_{Q\in\mathscr{Q}} Qf(x)$.*

It is a matter of straightforward verification to see that the lower envelope $\underline{Q}$ of $\mathscr{Q}$ is a lower transition rate operator:

Proposition 6.9. *For any non-empty and bounded set $\mathscr{Q} \subset \mathscr{R}$ of rate matrices, its lower envelope $\underline{Q}$ is a lower transition rate operator.*

Proof. Consider any $Q \in \mathscr{Q}$. Because, as we have seen above, Q is a lower transition rate operator, it satisfies Properties LR1$_{265}$ and LR2$_{265}$. Now fix any constant function $\mu \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$. Then because any $Q \in \mathscr{Q}$ satisfies Property LR1$_{265}$, it holds that

$$\underline{Q}\mu(x) = \inf_{Q\in\mathscr{Q}} Q\mu(x) = 0\,,$$

and therefore $\underline{Q}$ also satisfies Property LR1$_{265}$.

Next, fix any $x, y \in \mathscr{X}$ such that $x \neq y$. Then because any $Q \in \mathscr{Q}$ satisfies Property LR2$_{265}$, it holds that

$$\underline{Q}\mathbb{I}_y(x) = \inf_{Q\in\mathscr{Q}} Q\mathbb{I}_y(x) \geq 0\,,$$

and therefore $\underline{Q}$ also satisfies Property LR2$_{265}$.

Properties LR3$_{265}$ and LR4$_{265}$ follow directly from the properties of the infimum and the fact that the elements $Q \in \mathscr{Q}$ are linear maps. That is, for any $f, g \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$,

$$\begin{aligned}\underline{Q}(f+g)(x) = \inf_{Q\in\mathscr{Q}} Q(f+g)(x) &= \inf_{Q\in\mathscr{Q}} \big(Qf(x) + Qg(x)\big) \\ &\geq \inf_{Q\in\mathscr{Q}} Qf(x) + \inf_{Q\in\mathscr{Q}} Qg(x) \\ &= \underline{Q}f(x) + \underline{Q}g(x)\,,\end{aligned}$$

so $\underline{Q}$ satisfies Property LR3$_{265}$. Similarly, for any $f \in \mathscr{L}(\mathscr{X})$, $\lambda \in \mathbb{R}_{\geq 0}$, and $x \in \mathscr{X}$ it holds that

$$\begin{aligned}\underline{Q}(\lambda f)(x) = \inf_{Q\in\mathscr{Q}} Q(\lambda f)(x) &= \inf_{Q\in\mathscr{Q}} \lambda Qf(x) \\ &= \lambda \inf_{Q\in\mathscr{Q}} Qf(x) = \lambda \underline{Q}f(x)\,,\end{aligned}$$

and therefore $\underline{Q}$ satisfies Property LR4$_{265}$.

Because $\underline{Q}$ satisfies Properties LR1$_{265}$–LR4$_{265}$, it is a lower transition rate operator by Definition 6.2$_{265}$. □

Inspired by this result, we will also refer to the lower envelope of $\mathscr{Q}$ as the *lower transition rate operator corresponding to* $\mathscr{Q}$.

The following result provides sufficient conditions on the set $\mathscr{Q}$ for the value of $\underline{Q}f$ to be reached by Qf for some $Q \in \mathscr{Q}$, where $\underline{Q}$ is the lower transition rate operator corresponding to $\mathscr{Q}$. In other words, under those conditions the lower envelope is actually a minimum, rather than an infimum; and in particular, this minimum is achieved uniformly over all elements of $\mathscr{X}$. We recall from Definition 5.7$_{193}$ that a set $\mathscr{Q}$ of rate matrices has separately specified rows if, essentially, it is closed under the row-wise recombination of its elements. Moreover, using Corollary A.12$_{378}$, $\mathscr{Q}$ is compact if and only if it is closed and bounded.

Proposition 6.10. *Let $\mathscr{Q}$ be a non-empty and compact set of rate matrices that has separately specified rows, and let $\underline{Q}$ be its corresponding lower transition rate operator. Then for all $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}$ such that $Qf(x) = \underline{Q}f(x)$ for all $x \in \mathscr{X}$.*

Proof. Fix any $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$, and consider a sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ such that $\lim_{i\to+\infty}\varepsilon_i = 0$. Then, for all $i \in \mathbb{Z}_{>0}$, due to Definition 6.3$_{\frown}$, there is some $Q_x^{(i)} \in \mathscr{Q}$ such that $\underline{Q}f(x) \leq Q_x^{(i)}f(x) < \underline{Q}f(x) + \varepsilon_i$. The sequence $\{Q_x^{(i)}\}_{i\in\mathbb{Z}_{>0}}$ lives in $\mathscr{Q}$, which is a compact set by assumption. By Corollary A.12$_{378}$, this implies that $\mathscr{Q}$ is sequentially compact, whence there is some convergent subsequence $\{Q_x^{(i_k)}\}_{k\in\mathbb{Z}_{>0}}$ with $\lim_{k\to+\infty} Q_x^{(i_k)} =: Q_x^* \in \mathscr{Q}$. Moreover, for all $k \in \mathbb{Z}_{>0}$ it holds that

$$\underline{Q}f(x) \leq Q_x^{(i_k)}f(x) < \underline{Q}f(x) + \varepsilon_{i_k},$$

and therefore $Q_x^*f(x) = \underline{Q}f(x)$ because $\lim_{k\to+\infty}\varepsilon_{i_k} = 0$.

Because $\mathscr{Q}$ has separately specified rows, there is some $Q \in \mathscr{Q}$ such that $Q(x,\cdot) = Q_x^*(x,\cdot)$ for all $x \in \mathscr{X}$. Because of the above, it therefore holds that $Qf(x) = Q_x^*f(x) = \underline{Q}f(x)$ for all $x \in \mathscr{X}$. □

We have seen above that any set $\mathscr{Q}$ of rate matrices has a corresponding lower transition rate operator. We will now reason in the opposite direction; given an arbitrary lower transition rate operator $\underline{Q}$, is there a set $\mathscr{Q}$ of rate matrices that corresponds to it? To this end, we consider the set of rate matrices that *dominate* this lower transition rate operator, as follows.

Definition 6.4. *For any lower transition rate operator $\underline{Q}$, we define its* dominating set of rate matrices $\mathscr{Q}_{\underline{Q}}$ *as*

$$\mathscr{Q}_{\underline{Q}} := \left\{ Q \in \mathscr{R} \,\middle|\, Qf \geq \underline{Q}f \text{ for all } f \in \mathscr{L}(\mathscr{X}) \right\}. \tag{6.4}$$

It turns out that this set of dominating rate matrices satisfies a number of convenient properties, in analogy with the properties satisfied by the set $\mathscr{T}_{\underline{T}}$ of transition matrices that dominate a given lower transition operator $\underline{T}$; see Proposition 3.37_{120}. Because the proofs of the following statements are somewhat long, we have deferred them to Appendix $6.\mathrm{A}_{311}$.

Proposition 6.11. *For any lower transition rate operator $\underline{Q}$, its dominating set of rate matrices $\mathscr{Q}_{\underline{Q}}$ is a non-empty, compact, and convex set of rate matrices that has separately specified rows, and that has $\underline{Q}$ as its corresponding lower transition rate operator.*

These properties characterise $\mathscr{Q}_{\underline{Q}}$ completely, in the sense that no other set satisfies them.

Proposition 6.12. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and that has $\underline{Q}$ as its corresponding lower transition rate operator. Then $\mathscr{Q} = \mathscr{Q}_{\underline{Q}}$.*

Using this duality between lower transition rate operators and compact convex sets of rate matrices that have separately specified rows, we also obtain a correspondence between lower transition rate operators and continuous-time imprecise-Markov chains which, as we know from Chapter 5_{181}, are parameterised by sets of rate matrices. We also note that the characterising properties of the dominating set $\mathscr{Q}_{\underline{Q}}$ of a given lower transition rate operator $\underline{Q}$, are exactly the properties that we already encountered as preconditions in, e.g., Proposition 5.28_{203} and Theorems 6.2_{261} and 6.4_{263}.

Therefore, and using the connection between lower transition rate operators and lower transition operators—see Propositions 6.6_{266} and 6.7_{266}—we are led to the idea that we can use lower transition rate operators to characterise the lower transition operators corresponding to continuous-time imprecise-Markov chains. This will be the subject of the remainder of this chapter.

6.3 Exponentials of Lower Transition Rate Operators

Proposition 6.6_{266} has already established that we can easily construct a lower transition operator from a given lower transition rate operator $\underline{Q}$: if $\Delta \geq 0$ is sufficiently small, then $I + \Delta\underline{Q}$ will be a lower transition operator. In this section, we construct a somewhat more complicated lower transition operator from a given lower transition rate operator, and it is this specific lower transition operator we will focus on for the remainder of this chapter. In particular, we will introduce the *exponential* $e^{\underline{Q}t}$ of $\underline{Q}t$, where $\underline{Q}$ is a lower transition rate operator and $t \in \mathbb{R}_{\geq 0}$. The fact that we call this object an "exponential" perhaps deserves some motivation. This is because, as we will show throughout this section, it (i) is characterised by straightforward generalisations of the characterisations ME2$_{154}$ and ME3$_{154}$ of the matrix exponential, (ii) forms a (non-linear) semigroup of operators when viewed as a function of t, and (iii) exactly coincides with the matrix exponential e^{Qt} when $Q = \underline{Q}$ is a linear map.

The remainder of this section is devoted to constructing and studying this exponential. In Section 6.4$_{279}$ we show that, under some mild conditions, it coincides with the lower transition operators—and hence, as explained in Section 6.1$_{260}$, with the lower expectations—of the imprecise-Markov chains $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ characterised by a set $\mathscr{Q}$ of which $\underline{Q}$ is the corresponding lower transition rate operator.

6.3.1 *Construction of the Exponential*

Our starting point for constructing this exponential is the characterisation ME3$_{154}$ of the matrix exponential e^{Qt} using what we described as the Euler solution to the ordinary differential equation ME2$_{154}$, i.e.,

$$e^{Qt} = \lim_{k\to+\infty} (I + {}^{t}/{}_{k}Q)^k \quad \text{for all } t \in \mathbb{R}_{\geq 0} \text{ and } Q \in \mathscr{R}\,. \tag{6.5}$$

At its core, the construction that we are about to present consists in replacing the rate matrix Q in the above equation, with a lower transition rate operator $\underline{Q}$. We will show that this converges to a limit, which we denote as $e^{\underline{Q}t}$, and we will prove in Section 6.3.2$_{273}$ that this limit satisfies many properties that are desirable in our setting: $e^{\underline{Q}t}$ is a lower transition operator (c.f. Proposition 4.11$_{154}$); satisfies $e^{\underline{Q}t} = I$ if $t = 0$; satisfies the *semigroup property* $e^{\underline{Q}(t+s)} = e^{\underline{Q}t}e^{\underline{Q}s}$ for all $t, s \in \mathbb{R}_{\geq 0}$; is differentiable in t; and is the solution $e^{\underline{Q}t} = \underline{T}_t$ of the (in general non-linear) operator-valued ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{T}_t = \underline{Q}\,\underline{T}_t \quad \text{for all } t \in \mathbb{R}_{\geq 0}, \text{ with boundary condition } \underline{T}_0 = I\,.$$

So let us now proceed with the construction. We first note that, in Equation (6.5), for any fixed $k \in \mathbb{Z}_{>0}$, the right-hand side essentially involves a *uniform* partition of the interval $[0, t]$, with a length of ${}^{t}/{}_{k}$ for each element of this partition. In contrast, we will—in addition to generalising the rate matrix Q in this expression to a lower transition rate operator $\underline{Q}$—also generalise this partitioning scheme to *non-uniform* partitions. This will yield the slightly stronger result that the limit is independent of the exact partition used.

To this end, for any lower transition rate operator $\underline{Q}$, any $t \in \mathbb{R}_{\geq 0}$, and any finite partition $u \in \mathscr{U}_{[0,t]}$ of the interval $[0, t]$, we will consider an auxiliary operator Φ_u; for notational convenience, we do not carry the $\underline{Q}$ in our notation for this operator, but this should not cause any confusion. For any $u \in \mathscr{U}_{[0,t]}$ such that $u = t_0, \ldots, t_n$, $n \in \mathbb{Z}_{\geq 0}$, we define this operator as

$$\Phi_u := \prod_{i=1}^{n} \left(I + \Delta_i^u \underline{Q}\right), \tag{6.6}$$

where, as in Section 2.2.1$_{58}$, for every $i \in \{1, \ldots, n\}$, $\Delta_i^u = t_i - t_{i-1}$ denotes the difference between two consecutive time points in u, and $\sigma(u) := \max\{\Delta_i^u : i \in \{1, \ldots, n\}\}$ is the maximum such difference, i.e. the *mesh* of the partition. We note that if $n = 0$ then the product on the right-hand side of Equation (6.6) is empty, and in that case $\Phi_u = I$, which is trivially a lower transition operator. Conversely, if $n > 1$ then clearly, if $\sigma(u)$ is small enough, Proposition 6.6$_{266}$ guarantees that each of the terms $I + \Delta_i^u \underline{Q}$ is a lower transition operator, and it then follows from Proposition 3.33$_{117}$ that Φ_u—since it is a composition of lower transition operators—is also a lower transition operator.

The exponential $e^{\underline{Q}t}$ will be defined below as the limit of these lower transition operators Φ_u, obtained as we take u to be an increasingly finer partition of the interval $[0,t]$. Formally, $\mathscr{U}_{[0,t]}$ can be seen as a directed set—we induce a preorder $\preceq$ on $\mathscr{U}_{[0,t]}$ using the mesh of the partitions, i.e. for all $u, v \in \mathscr{U}_{[0,t]}$ we have $u \preceq v$ if $\sigma(u) \geq \sigma(v)$, and $u \cup v \in \mathscr{U}_{[0,t]}$ satisfies $u \preceq u \cup v$ and $v \preceq u \cup v$—and Φ_u defines a *net* on $\mathscr{U}_{[0,t]}$ that is (eventually) in $\underline{\mathbb{T}}$. We will show that this net converges to $e^{\underline{Q}t}$; see e.g. [77, Chapter 3] for some discussion about convergence of nets in topological spaces.

To this end, we start by providing a bound on the distance between two operators Φ_u and Φ_{u*}. Here and in what follows, we have moved some of our proofs to Appendix 6.B$_{314}$.

Proposition 6.13. *Let $\underline{Q}$ be a lower transition rate operator and choose any $t \in \mathbb{R}_{\geq 0}$, any $\delta \in \mathbb{R}_{>0}$ with $\delta \left\|\underline{Q}\right\| \leq 1$, and any $u, u^* \in \mathscr{U}_{[0,t]}$ with $\sigma(u) \leq \delta$ and $\sigma(u^*) \leq \delta$. Then $\|\Phi_u - \Phi_{u^*}\| \leq 2t\delta \left\|\underline{Q}\right\|^2$.*

Note that the distance $\|\Phi_u - \Phi_{u^*}\|$ vanishes as we make $\sigma(u)$ and $\sigma(u^*)$ smaller and smaller. This allows us to state the following result.

Corollary 6.14. *Let $\underline{Q}$ be a lower transition rate operator and choose any $t \in \mathbb{R}_{\geq 0}$. Then for every sequence $\{u_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathscr{U}_{[0,t]}$ with $\lim_{i \to +\infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i \in \mathbb{Z}_{>0}}$ is Cauchy.*

Proof. By definition of a Cauchy sequence, we need to show that

$$(\forall \varepsilon > 0)(\exists n \in \mathbb{Z}_{>0})(\forall i, j \geq n) \left\|\Phi_{u_i} - \Phi_{u_j}\right\| < \varepsilon.$$

So fix any $\varepsilon > 0$ and choose $\delta \in \mathbb{R}_{>0}$ such that $2t\delta \left\|\underline{Q}\right\|^2 < \varepsilon$ and $\delta \left\|\underline{Q}\right\| \leq 1$; this is clearly always possible. Because $\lim_{i\to+\infty} \sigma(u_i) = 0$, there is some $n \in \mathbb{Z}_{>0}$ such that, for all $k \in \mathbb{Z}_{>0}$ with $k > n$, it holds that $\sigma(u_k) \leq \delta$. Consider any $k, \ell \in \mathbb{Z}_{>0}$ such that $k, \ell > n$. It then follows from Proposition 6.13$_{\curvearrowleft}$ that

$$\left\|\Phi_{u_i} - \Phi_{u_j}\right\| \leq 2t\delta \left\|\underline{Q}\right\|^2 < \varepsilon\,,$$

which concludes the proof. □

Since we already know that, for partitions u that are sufficiently fine, Φ_u is a lower transition operator, Proposition 3.41$_{121}$ implies that this Cauchy sequence converges to a lower transition operator.

Corollary 6.15. *Let $\underline{Q}$ be a lower transition rate operator, and choose any $t \in \mathbb{R}_{\geq 0}$. For every sequence $\{u_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{U}_{[0,t]}$ with $\lim_{i\to+\infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i\in\mathbb{Z}_{>0}}$ converges to a lower transition operator.*

Proof. Since $\lim_{i\to+\infty} \sigma(u_i) = 0$, and due to Propositions 6.6$_{266}$ and 3.33$_{117}$, and Property LR5$_{266}$, there is some $n \in \mathbb{Z}_{>0}$ such that the sequence $\Phi_{u_n}, \Phi_{u_{n+1}}, \ldots$ consists of lower transition operators. Due to Corollary 6.14$_{\curvearrowleft}$, this sequence is Cauchy and therefore, because of Proposition 3.41$_{121}$, this sequence has a limit that is also a lower transition operator. Because the first $n-1$ elements of the sequence do not influence the convergence, we find that the sequence $\{\Phi_{u_i}\}_{i\in\mathbb{Z}_{>0}}$ has a limit, and that this limit is a lower transition operator. □

Finally, as our next result establishes, this limit is unique, in the sense that it is independent of the choice of $\{u_i\}_{i\in\mathbb{Z}_{>0}}$.

Theorem 6.16. *Let $\underline{Q}$ be a lower transition rate operator, and choose any $t \in \mathbb{R}_{\geq 0}$. Then there is a unique lower transition operator $e^{\underline{Q}t} \in \underline{\mathbb{T}}$ such that*

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall u \in \mathscr{U}_{[0,t]} \colon \sigma(u) < \delta)\; \left\|e^{\underline{Q}t} - \Phi_u\right\| < \varepsilon. \tag{6.7}$$

Proof. Let $\{u_i\}_{i\in\mathbb{Z}_{>0}}$ be any sequence in $\mathscr{U}_{[0,t]}$ such that $\lim_{i\to+\infty} \sigma(u_i) = 0$. Because of Corollary 6.15, the sequence $\{\Phi_{u_i}\}_{i\in\mathbb{Z}_{>0}}$ then converges to a lower transition operator, which we denote by $e^{\underline{Q}t}$.

Fix any $\varepsilon > 0$, and choose any $\delta > 0$ such that $4t\delta \left\|\underline{Q}\right\|^2 < \varepsilon$ and $\delta \left\|\underline{Q}\right\| \leq 1$; this is clearly always possible. Since $\lim_{i\to+\infty} \Phi_{u_i} = e^{\underline{Q}t}$ and $\lim_{i\to+\infty} \sigma(u_i) = 0$, there is some $n \in \mathbb{Z}_{>0}$ such that

$$\sigma(u_n) < \delta \text{ and } \left\|e^{\underline{Q}t} - \Phi_{u_n}\right\| < \frac{\varepsilon}{2}. \tag{6.8}$$

Consider now any $u \in \mathscr{U}_{[0,t]}$ such that $\sigma(u) < \delta$. Then

$$\left\|e^{\underline{Q}t} - \Phi_u\right\| \leq \left\|e^{\underline{Q}t} - \Phi_{u_n}\right\| + \left\|\Phi_{u_n} - \Phi_u\right\| < \frac{\varepsilon}{2} + 2t\delta \left\|\underline{Q}\right\|^2 < \varepsilon,$$

where the second inequality follows from Equation (6.8) and Proposition 6.13271. This proves that $e^{\underline{Q}t}$ satisfies Equation (6.7).

It remains to show that $e^{\underline{Q}t}$ is unique. Therefore, let $\underline{T}$ be any lower transition operator that satisfies Equation (6.7). Then clearly, for any $\varepsilon > 0$, there is some $u \in \mathscr{U}_{[0,t]}$ such that $\left\|e^{\underline{Q}t} - \Phi_u\right\| < \varepsilon/2$ and $\|\underline{T} - \Phi_u\| < \varepsilon/2$, and therefore $\left\|e^{\underline{Q}t} - \underline{T}\right\| < \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this implies that $e^{\underline{Q}t} = \underline{T}$. Because this holds for every $\underline{T}$ that satisfies Equation (6.7), it follows that $e^{\underline{Q}t}$ is the unique operator satisfying this equation. □

Note that the $\varepsilon - \delta$ expression in Theorem 6.16 is essentially a limit statement; it establishes the convergence of the net Φ_u on $\mathscr{U}_{[0,t]}$ to the operator $e^{\underline{Q}t}$ in $\underline{\mathbb{T}}$ (again, see e.g. [77, Chapter 3] for details on convergence of nets). In the sequel, for any lower transition rate operator $\underline{Q}$ and any $t \in \mathbb{R}_{\geq 0}$, we will always use the notation $e^{\underline{Q}t}$ to denote the corresponding unique lower transition operator identified by Theorem 6.16.

6.3.2 *Semigroups of Lower Transition Operators*

We now move the discussion to the *family* $\left(e^{\underline{Q}t}\right)$ of (generalised) exponentials corresponding to $\underline{Q}$, with $t \in \mathbb{R}_{\geq 0}$. Let us first establish that this family satisfies the same semigroup property that holds for the family of matrix exponentials (e^{Qt}) of a rate matrix Q; that is, that this family is a semigroup.

Proposition 6.17 ([17, Section 6]). *Let $\underline{Q}$ be a lower transition rate operator, and let $(e^{\underline{Q}t})$ be the corresponding family of lower transition operators, with $t \in \mathbb{R}_{\geq 0}$. Then, for all $t, s \in \mathbb{R}_{\geq 0}$, it holds that*

$$e^{\underline{Q}(t+s)} = e^{\underline{Q}t} e^{\underline{Q}s}.$$

Furthermore, if $t = 0$ then it holds that $e^{\underline{Q}t} = I$.

Inspired by this result, we also refer to the family $\left(e^{\underline{Q}t}\right)$ of these operators as the *semigroup of lower transition operators generated by $\underline{Q}$* (c.f. Proposition 4.15157 and the discussion around it).

Next, it turns out that the derivatives of these lower transition operators always exist, and that they satisfy the following equality.

Proposition 6.18 ([17, Proposition 9]). *Let $\underline{Q}$ be a lower transition rate operator, and let $(e^{\underline{Q}t})$ be its generated semigroup of lower transition operators. Then for all $t \in \mathbb{R}_{\geq 0}$ it holds that*[2]

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{\underline{Q}t} = \underline{Q} e^{\underline{Q}t}.$$

[2] If $t = 0$ we take this to only be a right-sided derivative.

Propositions 6.17$_{\curvearrowleft}$ and 6.18$_{\curvearrowleft}$ together show that $e^{\underline{Q}t}$ is a solution $e^{\underline{Q}t} = \underline{T}_t$ of the (in general non-linear) operator-valued initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}s}\underline{T}_s = \underline{Q}\,\underline{T}_s \quad \text{for all } s \in \mathbb{R}_{\geq 0}, \text{ with } \underline{T}_0 = I. \tag{6.9}$$

In fact, due to [103, Corollary 2], it can be shown that $e^{\underline{Q}t}$ is the *unique* solution to this differential equation; see also [17, Section 6] for further discussion. Therefore, and since we already know from Section 6.2$_{265}$ that any rate matrix $Q \in \mathscr{R}$ is also a lower transition rate operator, it follows from Definition 4.5$_{154}$, and the characterisation ME2$_{154}$ in particular, that $e^{\underline{Q}t}$ coincides with the matrix exponential e^{Qt} whenever $Q = \underline{Q}$.

We would like to point out here that the derivatives in Proposition 6.18$_{\curvearrowleft}$ are not taken pointwise, but are taken with respect to the operator norm. For example, with $t > 0$, Proposition 6.18$_{\curvearrowleft}$ does not state that

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{\underline{Q}t} f = \underline{Q} e^{\underline{Q}t} f \ \text{ for all } f \in \mathscr{L}(\mathscr{X}), \tag{6.10}$$

but rather that

$$\lim_{\Delta \to 0} \left\| \frac{e^{\underline{Q}(t+\Delta)} - e^{\underline{Q}t}}{\Delta} - \underline{Q} e^{\underline{Q}t} \right\| = 0. \tag{6.11}$$

Of these two statements, the latter is the strongest one, in the sense that it trivially implies the former. Hence, although from an intuitive point of view, the reader may wish to interpret the results in Proposition 6.18$_{\curvearrowleft}$ as in Equation (6.10)—which would be correct—one should keep in mind that from a technical point of view, the result is in fact stronger, and is intended to be read as in Equation (6.11).[3]

At this point, and as is perhaps already clear from the discussion in the first part of this section, we would like to note that we are not the first to study this semigroup of lower transition operators. In the explicit context of imprecise probabilities, Škulj [103] first studied pointwise solutions of the form (6.10) as bounds on the expectations of continuous-time imprecise-Markov chains. In particular, he studied these solutions without explicitly constructing a set of continuous-time stochastic processes like we did in Chapter 5$_{181}$. De Bock [17] showed the existence and studied the long-term and ergodic behaviour of these operators in the limit where t goes to infinity, and established the uniform (i.e. operator-valued) properties given in Propositions 6.17$_{\curvearrowleft}$ and 6.18$_{\curvearrowleft}$. In our own work in Reference [61], we also described the

[3]Note that Equation (6.11) does not follow trivially from Equation (6.10) and the finite-dimensionality of $\mathscr{L}(\mathscr{X})$, since the operators involved are, in general, not linear.

uniform solution, i.e. the operator-valued semigroup whose elements are characterised by Theorem 6.16$_{272}$, and we showed its relation to continuous-time imprecise-Markov chains as specific sets of stochastic processes; incidentally, this will also be the subject of Section 6.4$_{279}$. We have since learned that this semigroup is essentially a special case of a semigroup of operators that was already described by Nisio [81] in 1976, in the context of stochastic optimal control. Nendel [78] has used these Nisio semigroups to describe Markov chains with non-linear expectations, which are similar to imprecise-Markov chains but are based on a different underlying theory; we refer to Chapter 1$_{29}$ for details.

6.3.3 *Evaluating the Exponential*

Having established the existence and some initial properties of the exponential $e^{\underline{Q}t}$, and with the aim of using it to represent and compute lower expectations for imprecise-Markov chains, let us now turn to how to *evaluate* this operator numerically. That is, we present in this section a simple algorithm to evaluate the quantity $e^{\underline{Q}t}f$ numerically for any $t \in \mathbb{R}_{\geq 0}$ and any $f \in \mathscr{L}(\mathscr{X})$. This is the algorithm that we initially described in Reference [61], and it is easy to understand and implement, but it should by no means be taken to be the most efficient algorithm possible. Indeed, it is essentially simply an instance of the explicit Euler method with an identification of the required stepsize to guarantee a maximum numerical error. Although we argued in Reference [61] that it will generally outperform a method described by Škulj [103], we note that Erreygers and De Bock [32] have since described an algorithm that will often be more efficient than the one we present here. Moreover, more recent work by Škulj [104] also suggests possible improvements to this computational problem.

We want to emphasize that the problem of computing $e^{\underline{Q}t}f$ reduces to solving the non-linear ordinary differential equation (6.10), so if one is not—from a practical point of view—worried about mathematical bounds on the maximum numerical error, then this could also be solved using popular numerical integrators like, e.g., Runge-Kutta methods. Regardless of the particular algorithm that one uses to evaluate the operators $e^{\underline{Q}t}$ numerically, we will show in Sections 6.4$_{279}$ and 6.5$_{284}$ that such methods can be used to compute the lower expectations for imprecise-Markov chains.

So let us now turn to deriving the above-mentioned algorithm for evaluating $e^{\underline{Q}t}f$. As suggested above, our aim here is to provide an explicit Euler method that will compute the desired quantity up to some pre-specified maximum numerical error that can be made arbitrarily small at the expense of more computational effort. The construction of the operator $e^{\underline{Q}t}$ in Section 6.3.1$_{270}$ already suggests a method for accomplishing this; namely, by using a finite-precision approximation of $e^{\underline{Q}t}$ using the auxiliary operator $\Phi_u := \prod_{i=1}^n (I + \Delta_i^u \underline{Q})$. Recall from Section 6.3.1$_{270}$ that the approximation of $e^{\underline{Q}t}$ by Φ_u becomes better as we take $u \in \mathscr{U}_{[0,t]}$ to be an increasingly finer partition of the interval $[0,t]$. The following result gives a bound on how fine this partition needs to be for a specific function $f \in \mathscr{L}(\mathscr{X})$, in order to guarantee an ε-error bound on $e^{\underline{Q}t} f$; the proof can be found in Appendix 6.B$_{314}$.

Proposition 6.19. *Let $\underline{Q}$ be a lower transition rate operator, and let $(e^{\underline{Q}t})$ be its generated semigroup of lower transition operators. Then for all $t \in \mathbb{R}_{\geq 0}$, all $f \in \mathscr{L}(\mathscr{X})$, and all $\varepsilon \in \mathbb{R}_{>0}$, if we choose any $n \in \mathbb{Z}_{>0}$ such that*

$$n \geq \max\left\{ t \left\|\underline{Q}\right\|, \frac{t^2}{2\varepsilon} \left\|\underline{Q}\right\|^2 \|f\|_{\mathrm{v}} \right\},$$

with $\|f\|_{\mathrm{v}} := \max f - \min f$, then it holds that

$$\left\| e^{\underline{Q}t} f - \prod_{i=1}^{n} (I + {}^{t}\!/\!_{n}\underline{Q}) f \right\| \leq \varepsilon .$$

Simply put, this result reduces the problem of evaluating $e^{\underline{Q}t} f$, to n evaluations of $(I + {}^{t}\!/\!_{n}\underline{Q}) g_i$, with $g_i \in \mathscr{L}(\mathscr{X})$ such that $g_i := (I + {}^{t}\!/\!_{n}\underline{Q}) g_{i-1}$ for all $i \in \{1, \ldots, n\}$, and with $g_0 := f$. In particular, for any $i \in \{1, \ldots, n\}$, these g_i simply correspond to the partial compositions

$$g_i = (I + {}^{t}\!/\!_{n}\underline{Q}) g_{i-1} = \prod_{j=1}^{i} (I + {}^{t}\!/\!_{n}\underline{Q}) f .$$

Thus $g_n = \prod_{j=1}^n (I + {}^{t}\!/\!_{n}\underline{Q}) f$, and Proposition 6.19 then guarantees that $\left\| e^{\underline{Q}t} f - g_n \right\| \leq \varepsilon$ provided that n is chosen appropriately.

This yields an iterative procedure for computing $e^{\underline{Q}t} f$ that is outlined in Algorithm 1. The algorithm first finds the number n of steps required to reach the given precision ε (Line 2). Starting with the function $g_0 := f$ (Line 3), the algorithm iteratively computes the function $g_i := (I + {}^{t}\!/\!_{n}\underline{Q}) g_{i-1}$ (Line 5). After n iterations (Line 4), the returned function g_n (Line 7) corresponds to $e^{\underline{Q}t} f \pm \varepsilon$, due to Proposition 6.19.

This algorithm takes for granted that $\left\|\underline{Q}\right\|$ is known and/or can be derived from $\underline{Q}$; Erreygers and De Bock [32, Proposition 4] have shown that the bound in LR5$_{266}$ is actually an equality, i.e. that $\left\|\underline{Q}\right\| = 2 \max_{x \in \mathscr{X}} \left| \underline{Q}\mathbb{I}_x(x) \right|$, so it suffices to be able to evaluate $\underline{Q}\mathbb{I}_x$ for every $x \in \mathscr{X}$ in order to compute $\left\|\underline{Q}\right\|$. Since at every step of the algorithm we also need to evaluate $\underline{Q} g_{i-1}$ in order to obtain g_i, in summary, this tells us that if we can compute $\underline{Q} g$ for all $g \in \mathscr{L}(\mathscr{X})$, then

Algorithm 1 Numerically compute $e^{\underline{Q}t}f$ for any $f \in \mathscr{L}(\mathscr{X})$.

Input: A lower transition rate operator $\underline{Q}$, a scalar $t \in \mathbb{R}_{\geq 0}$, a function $f \in \mathscr{L}(\mathscr{X})$, and a maximum numerical error $\varepsilon \in \mathbb{R}_{>0}$.

Output: A function $e^{\underline{Q}t}f \pm \varepsilon$ in $\mathscr{L}(\mathscr{X})$.

function ComputeExponential($\underline{Q}, t, f, \varepsilon$)
  $n \leftarrow \left\lceil \max\left\{ \left(t^2 \|\underline{Q}\|^2 \|f\|_{\mathrm{v}}\right)/2\varepsilon,\ t\left\|\underline{Q}\right\| \right\} \right\rceil$
  $g_0 \leftarrow f$
  for $i \in \{1, \ldots, n\}$ **do**
    $g_i \leftarrow g_{i-1} + {}^{t}/_{n}\underline{Q}g_{i-1}$
  end for
  return g_n
end function

we can also approximate the quantity $e^{\underline{Q}t}f$ to arbitrary precision, for any $f \in \mathscr{L}(\mathscr{X})$.

So let us conclude this section by considering how $\underline{Q}g$ may be computed for a given $g \in \mathscr{L}(\mathscr{X})$. It is difficult to make any general statements about this, because it strongly depends on the way that $\underline{Q}$ is encoded. If we suppose that we are given a non-empty and bounded set $\mathscr{Q}$ of rate matrices, and that $\underline{Q}$ is the corresponding lower transition rate operator, then using Definition 6.3$_{267}$, evaluating $\underline{Q}g(x)$ reduces to computing $\inf_{Q \in \mathscr{Q}} Qg(x)$. The difficulty of solving *this* problem clearly depends on $\mathscr{Q}$. However, we note that if $\mathscr{Q}$ is also closed then this infimum turns into a minimum, which means that we then need to solve

$$\underline{Q}g(x) = \min_{Q \in \mathscr{Q}} Qg(x) = \min_{Q(x,\cdot) \in \mathscr{Q}_x} \sum_{y \in \mathscr{X}} Q(x,y)g(y),$$

where $\mathscr{Q}_x := \{Q(x,\cdot) : Q \in \mathscr{Q}\}$ is the set of x-rows of $\mathscr{Q}$. A first observation is that this is a *linear* minimisation problem of the function $\sum_{y \in \mathscr{X}} Q(x,y)g(y)$ over the set $\mathscr{Q}_x$. If, therefore, $\mathscr{Q}_x$ is described by a finite number of linear (in)equality constraints, the problem can be written as a linear program and can then be solved by any of the methods that are available for this in the literature.

Although this is, at least theoretically, a fairly straightforward and well-behaved problem, we note that this would be a linear program in $n := |\mathscr{X}| + c$ variables, where $c \in \mathbb{Z}_{>0}$ is the number of constraints describing $\mathscr{Q}_x$. It is well-known that this is solvable in polynomial time, with a very recent result [8] providing a deterministic algorithm that solves it in $\tilde{O}(n^{\omega})$[4] time, where $O(n^{\omega})$ is the complexity of multiplying

[4] Following [8], $\tilde{O}$ hides polylog(n) factors.

two $n \times n$ matrices; the current bound for this exponent is $\omega \approx 2.38$ [8]. Because the solution of this linear program only yields the quantity $\underline{Q}g(x)$, we need to execute it $|\mathscr{X}|$ times—once for each $x \in \mathscr{X}$—in order to obtain the entire vector $\underline{Q}g$. This shows that $\underline{Q}g$ can be computed with a runtime complexity of $\tilde{O}\big(|\mathscr{X}|n^{\omega}\big)$, assuming that the number of constraints specifying each $\mathscr{Q}_x$ is the same for each x. Although polynomial in the size of the problem, this may become unfeasible when $|\mathscr{X}|$ becomes moderately large.

However, it should be noted that the analysis above considers a fairly general setting; it only assumes that $\mathscr{Q}$ can be described using a finite number of linear constraints on its sets $\mathscr{Q}_x$ of rows. When more is known about $\mathscr{Q}$, it may be possible to derive more efficient algorithms. For example, it is possible that $\mathscr{Q}_x$ lives in a subspace that has a dimension (much) less than $|\mathscr{X}|$, which reduces the difficulty of the corresponding minimisation problem proportionally. Moreover, we might simply assume additional structure. References [33–35, 62] illustrate examples where domain-specific knowledge allows one to simplify this problem considerably.

Finally, one specific instance that might be of general practical importance is if $\mathscr{Q}$ is a closed ball in $\mathscr{R}$ with radius r around some given rate matrix $Q \in \mathscr{R}$. For example, this could be relevant from a practitioner's point of view, where one may have a precise estimate for the rate matrix Q of a continuous-time homogeneous Markov chain, and one wants to perform a sensitivity analysis using a continuous-time imprecise-Markov chain that is described by the ball around this estimate. Let us present some results for this setting; the proofs of the following results can be found in Appendix 6.B$_{314}$. The first observation is that such a ball satisfies all the desirable properties that we have encountered for sets of rate matrices:

Proposition 6.20. *Let $Q_* \in \mathscr{R}$ be a rate matrix, fix any $r \in \mathbb{R}_{\geq 0}$, and let $B_r(Q_*) := \big\{Q \in \mathscr{R} : \|Q - Q_*\| \leq r\big\}$ be the closed ball in $\mathscr{R}$ of radius r around Q_*. Then $B_r(Q_*)$ is a non-empty, compact, and convex set of rate matrices that has separately specified rows.*

As the next result shows, one can evaluate the lower transition rate operator $\underline{Q}$ corresponding to the ball around Q in a time complexity order of $O\big(|\mathscr{X}|^2\big)$, although there is some computational overhead of $O(|\mathscr{X}|\log|\mathscr{X}|)$. We note that this result provides a method that is more than an order of magnitude faster than the general linear programming formulation described above, and in particular, in this setting, for any $g \in \mathscr{L}(\mathscr{X})$, computing $\underline{Q}g$ has the same complexity as computing Qg.

Proposition 6.21. *Let $Q_* \in \mathscr{R}$ be a rate matrix, fix any $r \in \mathbb{R}_{\geq 0}$, let $B_r(Q_*) := \big\{Q \in \mathscr{R} : \|Q - Q_*\| \leq r\big\}$ be the closed ball in $\mathscr{R}$ of radius r*

around Q_*, and let $\underline{Q}$ be the lower transition rate operator corresponding to $B_r(Q_*)$. Let $n := |\mathscr{X}|$, fix any $g \in \mathscr{L}(\mathscr{X})$, and let $z_1,\ldots,z_n$ be an ordering of $\mathscr{X}$ such that $g(z_i) \geq g(z_{i+1})$ for all $i \in \{1,\ldots,n-1\}$. Fix any $x \in \mathscr{X}$, let $r_0 := r/2$ and, for all $i \in \{1,\ldots,n\}$, let $\Delta_i = r_{i-1}$ if $z_i = x$ and $\Delta_i := \min\{r_{i-1}, Q_*(x,z_i)\}$ otherwise, and let $r_i := r_{i-1} - \Delta_i$. Then

$$\underline{Q}g(x) = Q_*g(x) - \sum_{i=1}^{n} \Delta_i\big(g(z_i) - g(z_n)\big). \tag{6.12}$$

In this result, obtaining the ordering $z_1,\ldots,z_n$ requires sorting $n = |\mathscr{X}|$ elements, which clearly can be done in $O(n\log n)$ time. Once we have that ordering, computing the $\Delta_1,\ldots,\Delta_n$ and $r_1,\ldots,r_n$ can be done in $O(n)$, and it is clear that the right-hand side of Equation (6.12) can also be evaluated in $O(n)$ time. Only these last steps depend on x, and hence they need to be repeated n times, which yields the runtime complexity of $O(n^2)$.

6.4 Continuous-Time Imprecise-Markov Chains and Semigroups of Lower Transition Operators

Having introduced lower transition rate operators $\underline{Q}$ corresponding to sets $\mathscr{Q}$ of rate matrices, as well as the corresponding generated semigroup $(e^{\underline{Q}t})$ of (generalised) exponentials $e^{\underline{Q}t}$, let us now consider the connection of these objects to the family $(\underline{T}_t^s)$ of lower transition operators corresponding to an imprecise-Markov chain parameterised by $\mathscr{Q}$. Throughout this section, whenever the proof of a result is not given here, it can be found in Appendix 6.C$_{324}$.

Our first result is that these exponentials provide a lower bound on the expectations for the stochastic processes that are elements of such an imprecise-Markov chain, in the following sense:

Proposition 6.22. *Let $\mathscr{Q}$ be a non-empty bounded set of rate matrices with corresponding lower transition rate operator $\underline{Q}$, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition operators. Then, for any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any $u \in \mathscr{U}_{<t}$, any $x_t \in \mathscr{X}$ and $x_u \in \mathscr{X}_u$, and any $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\mathbb{E}_P[f(X_s)\,|\,X_t = x_t, X_u = x_u] \geq e^{\underline{Q}(s-t)}f(x_t).$$

Notice that this result is stated for stochastic processes P in $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, whose initial distributions $P(X_0)$ are not required to belong to some given set of initial distributions $\mathscr{M}$. However, since $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is a clearly a subset of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, the same result also holds for any choice of such $\mathscr{M}$.

Our next result establishes that the bound in Proposition 6.22 is tight if $\mathscr{Q}$ has separately specified rows. Specifically, we show that

$e^{\underline{Q}(s-t)}f$ can then be approximated to arbitrary precision by carefully choosing a Markov chain P from the set $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$.

Proposition 6.23. *Let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathcal{X}$, let $\mathcal{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition operators. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathcal{L}(\mathcal{X})$, and all $\varepsilon \in \mathbb{R}_{>0}$, there is a well-behaved Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ such that*

$$\left| \mathbb{E}_P[f(X_s) \,|\, X_t = x_t] - e^{\underline{Q}(s-t)} f(x_t) \right| < \varepsilon \ \text{ for all } x_t \in \mathcal{X}.$$

Moreover, under some additional assumptions on $\mathcal{Q}$ this lower bound can actually be *reached* by a Markov chain in $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$:

Corollary 6.24. *Let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathcal{X}$, let $\mathcal{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition operators. Then for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and all $f \in \mathcal{L}(\mathcal{X})$, there is a well-behaved Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ such that*

$$\mathbb{E}_P[f(X_s) \,|\, X_t = x_t] = e^{\underline{Q}(s-t)} f(x_t) \ \text{ for all } x_t \in \mathcal{X}.$$

Together, Propositions 6.22 and 6.23 establish a strong connection between the lower transition operator $e^{\underline{Q}(s-t)}$ and the lower expectations for $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$. In particular, for $\mathcal{Q}$ with separately specified rows, and for functions $f(X_s)$ that depend on the state X_s at a single time point s not before the latest time point t in the conditioning event $(X_t = x_t, X_u = x_u)$, these three objects end up being identical.

Corollary 6.25. *Let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathcal{X}$, let $\mathcal{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition operators. Then, for all $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathcal{U}_{<t}$, $x_u \in \mathcal{X}_u$ and $x_t \in \mathcal{X}$, and all $f \in \mathcal{L}(\mathcal{X})$, it holds that*

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t = x_t, X_u = x_u] &= e^{\underline{Q}(s-t)} f(x_t) \\ &= \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t = x_t, X_u = x_u].\end{aligned}$$

Proof. Fix any $\varepsilon > 0$. By Proposition 6.23 there is some $P \in \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ such that $\mathbb{E}_P[f(X_s)|X_t = x_t] < e^{\underline{Q}(s-t)} f(x_t) + \varepsilon$, and hence it follows that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}[f(X_s) \,|\, X_t = x_t, X_u = x_u] &\leq \mathbb{E}_P[f(X_s)|X_t = x_t, X_u = x_u] \\ &= \mathbb{E}_P[f(X_s)|X_t = x_t] < e^{\underline{Q}(s-t)} f(x_t) + \varepsilon,\end{aligned}$$

using Equation (5.13)$_{198}$ for the first inequality and the Markov property of P for the equality. Since $\varepsilon > 0$ is arbitrary, it follows that $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] \leq e^{\underline{Q}(s-t)} f(x_t)$. Now observe that also

$$\begin{aligned} e^{\underline{Q}(s-t)} f(x_t) &\leq \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \mathbb{E}_P[f(X_s)\,|\,X_t = x_t, X_u = x_u] \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] \\ &\leq \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u], \end{aligned}$$

where the first inequality follows from Proposition 6.22$_{279}$ and the last inequality follows from Proposition 5.22$_{199}$. □

Hence, we find that there is indeed a correspondence between the semigroup of lower transition operators $(e^{\underline{Q}t})$ and the lower expectations that correspond to continuous-time imprecise-Markov chains.

This result also helps to further clarify why we choose to call $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ an imprecise-Markov chain, despite the fact that it contains processes that do not satisfy the Markov property. In order to see that, observe that Corollary 6.25 holds for *all* histories $x_u \in \mathscr{X}_u$ and *any* sequence of time points $u \in \mathscr{U}_{<t}$. Therefore, and because the definition of $e^{\underline{Q}(s-t)}$ does not depend on this choice of u and x_u, it follows that for non-empty and bounded $\mathscr{Q}$ with separately specified rows, it holds that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t = x_t]. \qquad (6.13)$$

In other words, the conditional lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies an *imprecise-Markov property*: conditional on the state at time t, the lower expectation of a function $f(X_s)$ at a future time point s is functionally independent of the states at time points u that precede t. Recall that we already established that this was the case using Proposition 5.28$_{203}$. However, that result requires that $\mathscr{Q}$ is non-empty, compact, and convex, and that it has separately specified rows. The observation in Equation (6.13), which is due to Corollary 6.25, is thus much stronger: it turns out that $\mathscr{Q}$ need only be a non-empty and bounded set that has separately specified rows, and we do not need the closure, nor the convexity, of this set for $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ to satisfy this imprecise-Markov property.

This observation also allows us to identify the semigroup $(e^{\underline{Q}t})$ with the family $(\underline{T}_t^s)$ of lower transition operators corresponding to an imprecise-Markov chain, as discussed in Section 6.1$_{260}$.

Proposition 6.26. *Let $\mathscr{Q}$ be a non-empty and bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition* *operators. Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $(\underline{T}_t^s)$ be the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then $\underline{T}_t^s = e^{\underline{Q}(s-t)}$ for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$.*

Proof. Fix any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, and any $f \in \mathscr{L}(\mathscr{X})$. Since $\mathscr{Q}$ is non-empty and bounded with separately specified rows, it follows from Corollary 6.25280 (with $u = \emptyset$) and Definition 6.1261 that, for all $x \in \mathscr{X}$,

$$e^{\underline{Q}(s-t)} f(x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_s) \,|\, X_t = x] = \underline{T}_t^s f(x)\,.$$

Because this is true for all $x \in \mathscr{X}$, it follows that $e^{\underline{Q}(s-t)} f = \underline{T}_t^s f$. Because this is true for all $f \in \mathscr{L}(\mathscr{X})$, it follows that $e^{\underline{Q}(s-t)} = \underline{T}_t^s$. □

Moreover, it easily follows from the above results that $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies an (imprecise) time-homogeneity property:

Corollary 6.27. *Let $\mathscr{Q}$ be a non-empty and bounded set of rate matrices that has separately specified rows, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $(\underline{T}_t^s)$ be the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then $\underline{T}_t^s = \underline{T}_0^{(s-t)}$ for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$.*

Proof. Fix any $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, and let $r := s - t$. Then, because $\mathscr{Q}$ is non-empty and bounded with separately specified rows, it follows from Proposition 6.26↶ that

$$\underline{T}_t^s = e^{\underline{Q}(s-t)} = e^{\underline{Q}r} = \underline{T}_0^r = \underline{T}_0^{(s-t)}\,,$$

which concludes the proof. □

Analogous results hold for the imprecise-Markov chain $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. That is, $e^{\underline{Q}(s-t)} = \underline{T}_t^s$, and $\underline{T}_t^s = \underline{T}_0^{(s-t)}$, where $(\underline{T}_t^s)$ is the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, provided that $\mathscr{Q}$ is non-empty and bounded and has separately specified rows. The proofs of these claims are completely analogous to the proofs of Proposition 6.26↶ and Corollary 6.27, so we omit them here.

Moving on, we note that Corollary 6.25280 also establishes that the correspondence between the semigroup of lower transition operators $(e^{\underline{Q}t})$ and sets of continuous-time stochastic processes is not one-to-one. For starters, $e^{\underline{Q}(s-t)}$ represents the lower expectation for the different sets of processes $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, for *any* choice of $\mathscr{M}$. Moreover, because $(e^{\underline{Q}t})$ only depends on $\underline{Q}$, whenever two sets $\mathscr{Q}_1 \neq \mathscr{Q}_2$ have the same lower transition rate operator $\underline{Q}$ then—assuming the conditions in Corollary 6.25280 are met by both $\mathscr{Q}_1$ and $\mathscr{Q}_2$—$(e^{\underline{Q}t})$ represents the lower expectation with respect to the sets of stochastic processes $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}_1,\mathscr{M}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}_2,\mathscr{M}}$, $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_1,\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_2,\mathscr{M}}$—again, for any choice of $\mathscr{M}$.

A particularly interesting special case is when $\mathscr{Q}_1$ is a non-empty and bounded set of rate matrices $\mathscr{Q}$ that has separately specified rows, and $\mathscr{Q}_2$ is its closed convex hull, which, because of Proposition 6.12$_{269}$, is equal to $\mathscr{Q}_{\underline{Q}}$, where $\underline{Q}$ is the lower transition rate operator that corresponds to $\mathscr{Q}$. The two sets of rate matrices $\mathscr{Q}$ and $\mathscr{Q}_{\underline{Q}}$ then clearly (i) have the same corresponding lower transition rate operator $\underline{Q}$ and (ii) satisfy the conditions in Corollary 6.25$_{280}$. Therefore, it follows from the preceding argument that the resulting lower expectations are identical and, in particular, that, for any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, it holds that

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}[f(X_s)\,|\,X_t=x_t,X_u=x_u]=e^{\underline{Q}(s-t)}f(x_t)=\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}_{\underline{Q}}}[f(X_s)\,|\,X_t=x_t,X_u=x_u]$$

which in turn immediately implies that for any set of stochastic processes $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}\subseteq\mathscr{P}\subseteq\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_{\underline{Q}}}$, it holds that

$$\underline{\mathbb{E}}[f(X_s)\,|\,X_t=x_t,X_u=x_u]=e^{\underline{Q}(s-t)}f(x_t), \tag{6.14}$$

where $\underline{\mathbb{E}}$ is the lower expectation for $\mathscr{P}$, as in Equation (5.12)$_{198}$.

A common feature of these sets of stochastic processes $\mathscr{P}$, is that each of their elements P is well-behaved and consistent with $\mathscr{Q}_{\underline{Q}}$. An obvious question, then, is whether this feature is necessary in order for Equation (6.14) to hold. The following result establishes that this is indeed the case.

Theorem 6.28. *Let $\underline{Q}$ be a lower transition rate operator, with $\mathscr{Q}_{\underline{Q}}$ its set of dominating rate matrices, and let $(e^{\underline{Q}t})$ be the corresponding semigroup of lower transition operators. Then $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_{\underline{Q}}}$ is the largest set of stochastic processes $\mathscr{P}\subseteq\mathbb{P}$ for which the corresponding conditional lower expectation $\underline{\mathbb{E}}[\cdot\,|\,\cdot]$—as defined in Equation* (5.12)$_{198}$*—satisfies*

$$\underline{\mathbb{E}}[f(X_s)\,|\,X_t=x_t,X_u=x_u]=e^{\underline{Q}(s-t)}f(x_t)\qquad\text{for all } f\in\mathscr{L}(\mathscr{X}),$$

for all $t,s\in\mathbb{R}_{\geq 0}$ with $t\leq s$, all $u\in\mathscr{U}_{<t}$, and all $x_t\in\mathscr{X}$ and $x_u\in\mathscr{X}_u$.

We regard this result as a vindication for our choice to focus on *well-behaved* stochastic processes—instead of more restricted ones, such as, say, differentiable stochastic processes. Since one of our aims here is to use the semigroup of lower transition operators $(e^{\underline{Q}t})$ as a representational and computational tool for lower expectations, it follows from this result that in order to be able to do this, it is indeed necessary to impose this minimal property of well-behavedness.

In summary, we have established that—under some conditions on $\mathscr{Q}$—continuous-time imprecise-Markov chains, with our definition, satisfy a number of convenient qualitative properties; in particular, an imprecise-Markov property and an (imprecise) time-homogeneity property. Moreover, by establishing a correspondence between their lower expectations and the semigroup $\left(e^{\underline{Q}t}\right)$ generated by the lower transition rate operator $\underline{Q}$ corresponding to the set of rate matrices $\mathscr{Q}$ that parameterises the imprecise-Markov chain, it follows that we can use the algorithmic result(s) from Section 6.3.3₂₇₅ to efficiently evaluate their lower expectations—at least for functions that only depend on the state at a single point in time. Finally, our choice to focus on well-behaved processes has allowed us to characterise imprecise-Markov chains exactly as the largest—most conservative—set(s) of stochastic processes that satisfy all these properties.

6.5 A General Computational Method

In this section we provide a general algorithm for computing the lower expectation of any u-measurable function, for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ that is parameterised by a set $\mathscr{Q}$ of rate matrices that is non-empty, compact, convex, and has separately specified rows, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$.

We begin by noting that, if $\mathscr{Q}$ is non-empty and convex, it follows from Theorem 5.32₂₀₈ that the lower expectation of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies a law of iterated lower expectations which, using Theorem 6.4₂₆₃, can be expressed using the corresponding lower transition operators whenever $\mathscr{Q}$ is additionally compact and has separately specified rows. Moreover, under these same conditions, these lower transition operators can be identified with the semigroup $(e^{\underline{Q}t})$ due to Proposition 6.26₂₈₁. With the aim of providing a computational method, let us consider the numerical errors that are introduced when approximating the terms $e^{\underline{Q}t}f$ numerically—e.g., using the method discussed in Section 6.3.3₂₇₅—in this application of the law of iterated lower expectations.

We first give the following result, which essentially tells us that if we have a composition of lower transition operators that can each be computed up to a numerical error of ε, then the total error of the combined computation is simply the sum of these individual errors. Note that we explicitly use the form of the composition used in the statement of Theorem 6.4₂₆₃. The proof can be found in Appendix 6.D₃₃₁.

Lemma 6.29. *Let $(\underline{T}_t^s)$ be a family of lower transition operators $\underline{T}_t^s$, defined for all $t,s \in \mathbb{R}_{\geq 0}$ with $t \leq s$. Fix any $u,v \in \mathscr{U}_{\supset\emptyset}$ such that $u < v$ and $v = t_0,\dots,t_n$, $n \in \mathbb{Z}_{\geq 0}$, and any $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$. Fix any $\varepsilon \in \mathbb{R}_{>0}$, let $g_n := f$ and, for all $i \in \{1,\dots,n\}$, let $w_i := u \cup \{t_0,\dots,t_{i-1}\}$, and let $g_{i-1} \in \mathscr{L}(\mathscr{X}_{w_i})$ be such that*

$$\left|g_{i-1}(x_{w_i}) - \underline{T}_{t_{i-1}}^{t_i} g_i(x_{w_i})\right| \leq \varepsilon \quad \textit{for all } x_{w_i} \in \mathscr{X}_{w_i}, \tag{6.15}$$

and let $\tilde{f} \in \mathscr{L}(\mathscr{X}_u)$ *be such that*

$$\left|\tilde{f}(x_u) - \underline{T}_{\max u}^{t_0} g_0(x_u)\right| \leq \varepsilon \quad \text{for all } x_u \in \mathscr{X}_u. \tag{6.16}$$

Then it holds that[5]

$$\left|\tilde{f}(x_u) - \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u)\right| \leq (n+1)\varepsilon \quad \text{for all } x_u \in \mathscr{X}_u.$$

By combining this result with the law of iterated lower expectations expressed using lower transition operators (Theorem 6.4_{263}), the correspondence between lower transition operators and the imprecise semigroup $(e^{\underline{Q}t})$ (Proposition 6.26_{281}), and the fact that we have a numerical method to evaluate the imprecise exponentials $e^{\underline{Q}t}$ (Algorithm 1_{277}), we obtain a numerical method for computing the lower expectation of arbitrary u-measurable functions for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, provided that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows. This method is given by Algorithm $2_{\curvearrowright}$.

This algorithm works by using the composition of lower transition operators provided by Theorem 6.4_{263}, i.e.,

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right] = \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u)\,,$$

where $(\underline{T}_t^s)$ is the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. In particular, the method iteratively resolves the operators $\underline{T}_{t_{i-1}}^{t_i}$ by working backwards from $i = n$ to $i = 1$; this occurs on Lines 2–9 of Algorithm $2_{\curvearrowright}$. Starting from $g_n = f$ (Line 2), for each $i \in \{n, \ldots, 1\}$ (Line 3), the algorithm computes $\underline{T}_{t_{i-1}}^{t_i} g_i(x_{w_i})$ for each $x_{w_i} \in \mathscr{X}_{w_i}$ (Line 5), where $w_i = u \cup \{t_0, \ldots, t_{i-1}\}$ (Line 4). Using the notation introduced in Section 6.1_{260}, this quantity satisfies

$$\underline{T}_{t_{i-1}}^{t_i} g_i(x_{w_i}) = \left[\underline{T}_{t_{i-1}}^{t_i} g_i(x_{w_i}, \cdot)\right](x_{\max w_i})\,,$$

where $g_i(x_{w_i}, \cdot)$ is the element of $\mathscr{L}(\mathscr{X}_{t_i})$ corresponding to the t_i-measurable function $g_i(x_{w_i}, X_{t_i})$. On Line 6, the algorithm computes $h_{x_{w_i}} := e^{\underline{Q}(t_i - t_{i-1})} g_i(x_{w_i}, \cdot)$ by invoking Algorithm 1_{277}, which computes this quantity up to a numerical error of $\varepsilon/n+1$. Due to Proposition 6.26_{281} it holds that $\underline{T}_{t_{i-1}}^{t_i} = e^{\underline{Q}(t_i - t_{i-1})}$, so the quantity g_{i-1} constructed on Line 7 is an element of $\mathscr{L}(\mathscr{X}_{w_i})$ that satisfies

$$g_{i-1}(x_{w_i}) = h_{x_{w_i}}(x_{\max w_i}) = e^{\underline{Q}(t_i - t_{i-1})} g_i(x_{w_i}) \pm \frac{\varepsilon}{n+1} = \underline{T}_{t_{i-1}}^{t_i} g_i(x_{w_i}) \pm \frac{\varepsilon}{n+1}\,.$$

[5] As in the statement of Theorem 6.4_{263}, we have over-expanded the composition of these lower transition operators for clarity of exposition, but the expression as written is only valid for $n \geq 2$. If $n = 0$ the second term should be $\underline{T}_{\max u}^{t_0} f(x_u)$, and if $n = 1$ it should be $\underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} f(x_u)$.

Algorithm 2 Compute $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v})\,|\,X_u]$ for any $f\in\mathscr{L}(\mathscr{X}_{u\cup v})$.

Input: A set $\mathscr{Q}$ of rate matrices that is non-empty, compact, convex, that has separately specified rows and lower transition rate operator $\underline{Q}$, two sequences of time points $u,v\in\mathscr{U}_{\supset\emptyset}$ such that $u<v$, with $v=t_0,\dots,t_n$, $n\in\mathbb{Z}_{\geq 0}$, a function $f\in\mathscr{L}(\mathscr{X}_{u\cup v})$, and a maximum numerical error $\varepsilon\in\mathbb{R}_{>0}$.

Output: A function $\tilde{f}\in\mathscr{L}(\mathscr{X}_u)$ such that, for all $x_u\in\mathscr{X}_u$, it holds that $\tilde{f}(x_u)=\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{u\cup v})\,|\,X_u=x_u]\pm\varepsilon$, for any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$.

1: **function** ComputeLowerExpectation($\mathscr{Q},u,v,f,\varepsilon$)
2:     $g_n \leftarrow f$
3:     **for** $i\in\{n,\dots,1\}$ **do**
4:         $w_i \leftarrow u\cup\{t_0,\dots,t_{i-1}\}$
5:         **for** $x_{w_i}\in\mathscr{X}_{w_i}$ **do**
6:             $h_{x_{w_i}} \leftarrow$ ComputeExponential($\underline{Q},t_i-t_{i-1},g_i(x_{w_i},\cdot),\varepsilon/n+1$)
7:             $g_{i-1}(x_{w_i}) \leftarrow h_{x_{w_i}}(x_{\max w_i})$
8:         **end for**
9:     **end for**
10:     **for** $x_u\in\mathscr{X}_u$ **do**
11:         $h_{x_u} \leftarrow$ ComputeExponential($\underline{Q},t_0-\max u,g_0(x_u,\cdot),\varepsilon/n+1$)
12:         $\tilde{f}(x_u) \leftarrow h_{x_u}(x_{\max u})$
13:     **end for**
14:     **return** $\tilde{f}$
15: **end function**

The final loop of Algorithm 2, on Lines 10–13, repeats this same process to evaluate the operator $\underline{T}^{t_0}_{\max u}$, and the resulting function $\tilde{f}$, which is returned by the algorithm on Line 14, then satisfies

$$\tilde{f}(x_u)=\underline{T}^{t_0}_{\max u}g_0(x_u)\pm\frac{\varepsilon}{n+1}=\underline{T}^{t_0}_{\max u}\underline{T}^{t_1}_{t_0}\cdots\underline{T}^{t_n}_{t_{n-1}}f(x_u)\pm\varepsilon\,,$$

using the additivity of the errors established by Lemma 6.29_{284}.

Let us remark on several important points about this algorithm. First, although we here use Algorithm 1_{277} to numerically compute the terms $e^{\underline{Q}(t_i-t_{i-1})}g_i$, this can of course be replaced by a more efficient algorithm, provided that the error bounds are guaranteed for each step (assuming that we want to guarantee the overall global error of the algorithm). Secondly, the runtime of the algorithm is clearly exponential in $|u\cup v|$; the loops on Lines 5 and 10 iterate over the joint state spaces $\mathscr{X}_{w_i}$ and $\mathscr{X}_u$, respectively, which clearly grow exponentially as functions of w_i and u, respectively. Of course, this is hardly surprising: even *specifying* the function $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$ whose lower expectation is computed by Algorithm 2 is exponentially complex in $|u \cup v|$, so we cannot really expect the computation of its lower expectation to be simpler than that; also see the discussion following Theorem 5.32$_{208}$. Nevertheless, as we will demonstrate in Chapter 7$_{335}$, there are classes of functions for which more efficient algorithms can be developed.

Finally, we note that Algorithm 2 can only be used to compute *conditional* lower expectations, since it requires that the collection of time points u in the conditioning event should be non-empty. In order to complete the discussion on general computational methods, let us therefore also consider how to compute *un*conditional lower expectations. As we will see, however, this is relatively straightforward once we have a method to compute conditional lower expectations, e.g. the one that we provided above. To this end, let us consider the lower expectation corresponding only to the initial model, i.e. the uncertainty model for the state of the system at time zero. Due to our developments in Section 5.2$_{188}$, and Definition 5.4$_{189}$ in particular, this is described using the set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. We have already encountered this exact same object in Definition 3.14$_{114}$ in the context of discrete-time processes, so let us simply extend the definition here with some notation that makes it usable in the continuous-time setting. For convenience, we explicitly recall how it was constructed:

Definition 6.5. *For any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, consider the map $\underline{\mathbb{E}}_{\mathscr{M}} : \mathscr{L}(\mathscr{X}) \to \mathbb{R} : f \mapsto \underline{\mathbb{E}}_{\mathscr{M}}[f]$ from Definition 3.14$_{114}$ which was defined, for all $f \in \mathscr{L}(\mathscr{X})$, by $\underline{\mathbb{E}}_{\mathscr{M}}[f] := \inf_{p\in\mathscr{M}} \sum_{x\in\mathscr{X}} p(x)f(x)$.*

For any $\{0\}$-measurable function $f(X_0) : \Omega \to \mathbb{R}$, we let $\underline{\mathbb{E}}_{\mathscr{M}}[f(X_0)] := \underline{\mathbb{E}}_{\mathscr{M}}[f]$, where f is the element of $\mathscr{L}(\mathscr{X})$ corresponding to $f(X_0)$, as described in Section 2.4$_{71}$.

Proposition 6.30. *Let $\mathscr{Q}$ be a non-empty set of rate matrices, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then, for all $p \in \mathscr{M}$, there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that $P(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$.*

Proof. Fix any $Q \in \mathscr{Q}$ and any $p \in \mathscr{M}$; this is always possible because $\mathscr{Q}$ and $\mathscr{M}$ are non-empty. Because Q is a rate matrix and p is a probability mass function on $\mathscr{X}$, it follows from Corollary 5.5$_{186}$ that there is some homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$ with corresponding rate matrix $Q_P = Q$, that satisfies $P(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$. Because $Q \in \mathscr{Q}$ this implies that $P \sim \mathscr{Q}$ due to Proposition 5.7$_{187}$ and Definition 5.3$_{189}$. Moreover, because $p \in \mathscr{M}$, it follows that $P \sim \mathscr{M}$ due to Definition 5.4$_{189}$. Because $P \in \mathbb{P}^{\mathrm{WHM}}$, this implies that $P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ due to Definition 5.6$_{190}$. Since $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ due to Proposition 5.9$_{190}$, it follows that $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. □

Then we have the following straightforward result.

Proposition 6.31. *Let $\mathscr{Q}$ be a non-empty set of rate matrices, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then, for all $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[f(X_0)]\,.$$

Proof. It follows from Lemma 5.23_{200} (with $u = \emptyset$ and $v = \{0\}$) that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_0)] = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \sum_{x \in \mathscr{X}} f(x) P(X_0 = x)\,.$$

Now, for any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it holds that $P \sim \mathscr{M}$ due to Definition 5.6_{190}, which, by Definition 5.4_{189}, implies that there is some $p \in \mathscr{M}$ such that $p(x) = P(X_0 = x)$ for all $x \in \mathscr{X}$. Conversely, it follows from Proposition $6.30_{\curvearrowleft}$ that for all $p \in \mathscr{M}$, there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that $P(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$. This implies that

$$\inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \sum_{x \in \mathscr{X}} f(x) P(X_0 = x) = \inf_{p \in \mathscr{M}} \sum_{x \in \mathscr{X}} f(x) p(x)\,.$$

Hence it follows from Definition $6.5_{\curvearrowleft}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_0)] = \inf_{P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}} \sum_{x \in \mathscr{X}} f(x) P(X_0 = x) = \inf_{p \in \mathscr{M}} \sum_{x \in \mathscr{X}} f(x) p(x) = \underline{\mathbb{E}}_{\mathscr{M}}[f]\,.$$

Finally, again due to Definition $6.5_{\curvearrowleft}$, it holds that $\underline{\mathbb{E}}_{\mathscr{M}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[f]$ because $f(X_0)$ is clearly $\{0\}$-measurable. □

So, by combining Proposition 6.31 with Definition $6.5_{\curvearrowleft}$, we see that the problem of computing $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_0)]$ essentially reduces to a linear minimisation problem of the function $\sum_{x \in \mathscr{X}} p(x) f(x)$ over $\mathscr{M}$. As with our discussion in Section $6.3.3_{275}$ about computing $\underline{Q}g(x)$, the practical difficulty of this problem will depend strongly on the way that $\mathscr{M}$ is encoded. Again, if $\mathscr{M}$ is specified using a finite number of linear (in)equality constraints, then the problem can be written as a linear program and solved using any of the available methods in the literature. We will not analyse this problem in more detail; in the remainder of this dissertation, we will assume that this minimisation problem is solvable for practical purposes.

Moving on, it remains to connect the computation of conditional expectations, as discussed previously, and the computation of the lower expectation on the initial time point, in order to obtain a general computational method for unconditional lower expectations. We start with the following technical property.

Lemma 6.32. *Fix any $u, v \in \mathscr{U}_{\supset\emptyset}$ such that $u \cap v = \emptyset$ and choose any $f \in \mathscr{L}(\mathscr{X}_v)$. Let $\tilde{f} \in \mathscr{L}(\mathscr{X}_{u\cup v})$ be defined as $\tilde{f}(x_{u\cup v}) := f(x_v)$ for all $x_{u\cup v} \in \mathscr{X}_{u\cup v}$. Then it holds that $f(X_v)(\omega) = \tilde{f}(X_{u\cup v})(\omega)$ for all $\omega \in \Omega$, and hence $f(X_v) = \tilde{f}(X_{u\cup v})$.*

Proof. Fix any $\omega \in \Omega$. Then it follows from Definition 2.15_{72} that $\tilde{f}(X_{u\cup v})(\omega) = \tilde{f}(\omega|_{u\cup v}) = f(\omega|_v) = f(X_v)(\omega)$, where the second equality used the definition of $\tilde{f}$. Because this is true for all $\omega \in \Omega$, the result follows. □

The following proposition now provides the required connection between unconditional lower expectations, conditional lower expectations, and the lower expectation for the initial time point. Since we have already discussed above when and how we can compute the two lower expectations on the right-hand side of Equation (6.17), this result immediately gives us a way to also compute the unconditional lower expectation on the left-hand side.

Theorem 6.33. *Let $\mathscr{Q}$ be a non-empty and convex set of rate matrices that has separately specified rows, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for all $u \in \mathscr{U}_{\supset\emptyset}$ and all $f \in \mathscr{L}(\mathscr{X}_u)$, it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)\,|\,X_0]\right]. \tag{6.17}$$

Proof. We start by considering three cases. First, if $0 \notin u$ let $u' := \{0\} \cup u$ and let $f' \in \mathscr{L}(\mathscr{X}_{u'})$ be defined, for all $x_{\{0\}\cup u} \in \mathscr{X}_{\{0\}\cup u}$, as $f'(x_{\{0\}\cup u}) := f(x_u)$. Because $0 \notin u$ it then follows from Lemma 6.32 that $f'(X_{u'}) = f(X_u)$ and, because $u \neq \emptyset$, it follows that $u' \supset \{0\}$. Conversely, if $0 \in u$, we consider two more cases. If $u = \{0\}$, then fix any $t \in \mathbb{R}_{>0}$, let $u' := u \cup \{t\}$, and let $f' \in \mathscr{L}(\mathscr{X}_{u'})$ be defined, for all $x_{u\cup\{t\}} \in \mathscr{X}_{u\cup\{t\}}$, as $f'(x_{u\cup\{t\}}) := f(x_u)$. Because $u = \{0\}$ and $t > 0$ it then follows that $u' \supset \{0\}$ and, from Lemma 6.32 that $f'(X_{u'}) = f(X_u)$. The final case that we consider is when $0 \in u$ and $u \neq \{0\}$; then simply let $u' := u$ and let $f' := f$; then we trivially have that $f'(X_{u'}) = f(X_u)$ and $u \supset \{0\}$. Hence, in all cases, it holds that $u' \supset \{0\}$, and

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})],$$

and

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)\,|\,X_0] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0].$$

Therefore, it suffices to prove the statement for f' and u', i.e. we will show that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0]\right].$$

To prove this statement we will use Theorem 5.32_{208}, but unfortunately the notation of the time points that we are currently using might make this a bit ambiguous. To be explicit in what follows, let $w := u' \setminus \{0\}$, $v := \{0\}$, and let the *u in the statement of Theorem 5.32*$_{208}$ be empty. Due to our construction of u', it follows that $w \neq \emptyset$ and, clearly, $u' = w \cup v \cup \emptyset$. Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, it follows from Theorem 5.32$_{208}$, using the notation established above, that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})] &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{w\cup v\cup\emptyset})\,|\,X_{\emptyset}] \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{w\cup v\cup\emptyset})\,|\,X_v]\,|\,X_{\emptyset}\right] \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0]\right].\end{aligned}$$

The inner conditional lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0]$ is a $\{0\}$-measurable function, whence it follows from Proposition 6.31$_{288}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0]\right] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f'(X_{u'})\,|\,X_0]\right],$$

which concludes the proof. □

6.6 A Numerical Example

Our aim in this final section of this chapter is twofold. First, in Section 6.6.1, we will numerically illustrate the computational methods developed in Sections 6.3.3$_{275}$ and 6.5$_{284}$, to show how they can be used to compute lower expectations for the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Secondly, in Section 6.6.2$_{297}$, we will illustrate by means of a continuation of this example that the lower expectation for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ does not coincide with the lower expectation for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, thereby finally showing that the first inequality in Proposition 5.22$_{199}$ can indeed be strict.

6.6.1 *Sets of General Processes*

Let us start by setting up the model that will serve as the basis for the running example in this section.

Example 6.1. Fix an ordered binary state space $\mathscr{X} := \{a,b\}$, consider the set of transition rate matrices

$$\mathscr{Q} := \left\{ \begin{bmatrix} -\lambda_a & \lambda_a \\ \lambda_b & -\lambda_b \end{bmatrix} : \lambda_a \in [\underline{\lambda_a}, \overline{\lambda_a}], \lambda_b \in [\underline{\lambda_b}, \overline{\lambda_b}] \right\}, \tag{6.18}$$

where $\underline{\lambda_a}, \overline{\lambda_a}, \underline{\lambda_b}, \overline{\lambda_b} \in \mathbb{R}_{\geq 0}$ are such that $\underline{\lambda_a} < \overline{\lambda_a}$ and $\underline{\lambda_b} < \overline{\lambda_b}$, and let $\underline{Q}$ denote the lower transition rate operator corresponding to $\mathscr{Q}$.

We will begin by analysing some properties of this set $\mathscr{Q}$. First, clearly, every element $Q \in \mathscr{Q}$ is a rate matrix (c.f. Definition 4.4$_{150}$) that is completely determined by the choice of two scalars λ_a, λ_b, which take their values in $[\underline{\lambda_a}, \overline{\lambda_a}]$ and $[\underline{\lambda_b}, \overline{\lambda_b}]$, respectively. Because these intervals are non-empty and bounded, it follows that $\mathscr{Q}$ is also non-empty and, using Proposition 4.8$_{152}$, that $\mathscr{Q}$ is bounded. Moreover, because these intervals are closed and convex, one can relatively easily show that $\mathscr{Q}$ is also closed (with respect to the operator norm; see Proposition A.8$_{376}$), as well as convex; in the interest of brevity, however, we omit the proof here. Because $\mathscr{Q}$ is both closed and bounded, it follows from Corollary A.12$_{378}$ that $\mathscr{Q}$ is compact. Finally, because the values of λ_a and λ_b can be chosen independently from one another, it follows that $\mathscr{Q}$ has separately specified rows (c.f. Definition 5.7$_{193}$).

In summary, we have found that $\mathscr{Q}$ is a non-empty, compact, and convex set of rate matrices that has separately specified rows, and that has $\underline{Q}$ as its corresponding lower transition rate operator. Therefore, it follows from Proposition 6.12$_{269}$ that $\mathscr{Q} = \mathscr{Q}_{\underline{Q}}$.

Next, let us turn our attention to some properties of this $\underline{Q}$. By Definition 6.3$_{267}$, for any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$, we have that $\underline{Q}f(x) := \inf_{Q \in \mathscr{Q}} Qf(x)$. In the exceedingly simple case of a binary state space $\mathscr{X}$ that we are considering here, it turns out that we can derive a very simple explicit expression for the value of $\underline{Q}$ in any $f \in \mathscr{L}(\mathscr{X})$. In particular, some straightforward algebra reveals that, for any $Q \in \mathscr{Q}$ and any $f \in \mathscr{L}(\mathscr{X})$, writing the elements of $\mathscr{L}(\mathscr{X})$ as column vectors,

$$Qf = \begin{bmatrix} Qf(a) \\ Qf(b) \end{bmatrix} = \begin{bmatrix} \lambda_a\big(f(b) - f(a)\big) \\ \lambda_b\big(f(a) - f(b)\big) \end{bmatrix}, \tag{6.19}$$

where λ_a and λ_b are the two parameters determining Q. Hence we see that, minimising this over $\mathscr{Q}$, we obtain

$$\underline{Q}f(a) = \begin{cases} \overline{\lambda_a}\big(f(b) - f(a)\big) & \text{if } f(a) \geq f(b) \text{, and} \\ \underline{\lambda_a}\big(f(b) - f(a)\big) & \text{otherwise} \end{cases} \tag{6.20}$$

and

$$\underline{Q}f(b) = \begin{cases} \overline{\lambda_b}\big(f(a) - f(b)\big) & \text{if } f(a) \leq f(b) \text{, and} \\ \underline{\lambda_b}\big(f(a) - f(b)\big) & \text{otherwise.} \end{cases} \tag{6.21}$$

Using these expressions, we can easily evaluate $\underline{Q}$ in any function $f \in \mathscr{L}(\mathscr{X})$. For instance, we may consider the indicators $\mathbb{I}_a$ and $\mathbb{I}_b$ of $\{a\}$ and $\{b\}$, respectively, and compute that

$$\underline{Q}\mathbb{I}_a(a) = -\overline{\lambda_a} \quad \text{and} \quad \underline{Q}\mathbb{I}_b(b) = -\overline{\lambda_b}\,.$$

Hence, using Property LR5$_{266}$—or rather,[6] using [32, Proposition 4]—

[6] As we already noted in Section 6.3.3$_{275}$, Erreygers and De Bock [32, Proposition 4] have shown that the inequality LR5$_{266}$ is actually an equality.

we can compute the norm $\|\underline{Q}\|$ of $\underline{Q}$ as

$$\|\underline{Q}\| = 2\max\{\overline{\lambda_a}, \overline{\lambda_b}\}. \tag{6.22}$$

In order to complete setting up this example, it only remains to choose numerical values for the domains of λ_a and λ_b. In the remainder of this running example, we will set

$$\underline{\lambda}_a := \frac{1}{2}, \quad \overline{\lambda}_a := 2, \quad \underline{\lambda}_b := \frac{1}{2}, \quad \overline{\lambda}_b := 1\,.$$

Hence, due to Equation (6.22), we get $\|\underline{Q}\| = 4$. Moreover, having determined a way to evaluate $\underline{Q}$ and having computed its norm, we now have all the parts to be able to use Algorithm 1_{277} to evaluate $e^{\underline{Q}t} f$ for any $t \in \mathbb{R}_{\geq 0}$ and any $f \in \mathscr{L}(\mathscr{X})$, up to any precision $\varepsilon \in \mathbb{R}_{>0}$. ◊

The following example illustrates the use of Algorithm 1_{277} to compute the lower expectations of some specific functions, for the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$.

Example 6.2. Let $\mathscr{X}$, $\mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1_{290}. Let $t := 1$ and let $s := 2$. It will be the aim of this example to compute the quantities

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_y(X_s)\,|\,X_t = x\right],$$

for any $x, y \in \mathscr{X}$, and up to a numerical error of $\varepsilon := 10^{-3}$.

First, for any $x, y \in \mathscr{X}$, it follows from Corollary 6.25_{280} (with $u = \emptyset$ and $f = \mathbb{I}_y$) that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_y(X_s)\,|\,X_t = x\right] = \left[e^{\underline{Q}(s-t)}\mathbb{I}_y\right](x)\,,$$

so what we need to do is evaluate $e^{\underline{Q}(s-t)}$ in the function $\mathbb{I}_y$, which we know can be done using Algorithm 1_{277}. It is easily seen that $\|\mathbb{I}_y\|_v = 1$ for any $y \in \mathscr{X}$,[7] and we already computed that $\|\underline{Q}\| = 4$ in Example 6.1_{290}.

So now first let $y := a$. The minimum number of steps n that we need to execute Algorithm 1_{277} is then

$$\begin{aligned} n &:= \left\lceil \max\left\{ \left((s-t)^2 \|\underline{Q}\|^2 \|\mathbb{I}_a\|_v\right)/2\varepsilon,\, (s-t)\|\underline{Q}\| \right\} \right\rceil \\ &= \lceil \max\{8000, 4\} \rceil = 8000\,. \end{aligned}$$

Following Algorithm 1_{277}, we first set $g_0 := \mathbb{I}_a$, and then compute

$$g_1 := g_0 + \frac{s-t}{n}\underline{Q}g_0\,.$$

[7] Recall from e.g. Proposition 6.19_{276} that $\|f\|_v := \max f - \min f$.

Since $g_0(a) = \mathbb{I}_a(a) = 1$ and $g_0(b) = \mathbb{I}_a(b) = 0$, it holds that $g_0(a) > g_0(b)$, and hence it follows from Equations $(6.20)_{291}$ and $(6.21)_{291}$ in Example 6.1_{290} that, using vector notation,

$$\underline{Q}g_0 = \begin{bmatrix} \underline{Q}g_0(a) \\ \underline{Q}g_0(b) \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_a\big(g_0(b) - g_0(a)\big) \\ \underline{\lambda}_b\big(g_0(a) - g_0(b)\big) \end{bmatrix} = \begin{bmatrix} -2 \\ 1/2 \end{bmatrix}.$$

Since $(s-t) = 1$ and $n = 8000$, it now follows that

$$g_1 = \begin{bmatrix} g_1(a) \\ g_1(b) \end{bmatrix} = \begin{bmatrix} g_0(a) + 1/8000\underline{Q}g_0(a) \\ g_0(b) + 1/8000\underline{Q}g_0(b) \end{bmatrix} = \begin{bmatrix} 1 - 2/8000 \\ 0 + 1/16000 \end{bmatrix} = \begin{bmatrix} 7998/8000 \\ 1/16000 \end{bmatrix}.$$

We now proceed iteratively with the execution of Algorithm 1. On the next step, we want to compute $g_2 = g_1 + {}^{(s-t)}/_n\underline{Q}g_1$, and since clearly $g_1(a) > g_1(b)$, we now find from Equations $(6.20)_{291}$ and $(6.21)_{291}$ that

$$\underline{Q}g_1 = \begin{bmatrix} \underline{Q}g_1(a) \\ \underline{Q}g_1(b) \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_a\big(g_1(b) - g_1(a)\big) \\ \underline{\lambda}_b\big(g_1(a) - g_1(b)\big) \end{bmatrix} = \begin{bmatrix} 2\big(1/16000 - 7998/8000\big) \\ 1/2\big(7998/8000 - 1/16000\big) \end{bmatrix},$$

and then compute

$$g_2 = \begin{bmatrix} g_2(a) \\ g_2(b) \end{bmatrix} = \begin{bmatrix} g_1(a) + 1/8000\underline{Q}g_1(a) \\ g_1(b) + 1/8000\underline{Q}g_1(b) \end{bmatrix} = \begin{bmatrix} 12793601/12800000 \\ 6399/51200000 \end{bmatrix}.$$

At this point the usefulness of the numerical part of this example is probably starting to break down, so let us skip to the end of the execution of Algorithm 1_{277}. After 7997 more iterations of the above process, we have just computed $g_{7999} = g_{n-1}$ to be, approximately,

$$g_{n-1} = \begin{bmatrix} g_{n-1}(a) \\ g_{n-1}(b) \end{bmatrix} \approx \begin{bmatrix} 0.2656629 \\ 0.1835843 \end{bmatrix}$$

The condition $g_{n-1}(a) > g_{n-1}(b)$ again allows us to evaluate $\underline{Q}g_{n-1}$ using Equations $(6.20)_{291}$ and $(6.21)_{291}$, so that we may finally compute

$$g_n = g_{n-1} + \frac{1}{8000}\underline{Q}g_{n-1} \approx \begin{bmatrix} 0.2656423 \\ 0.1835894 \end{bmatrix}.$$

It now follows from Proposition 6.19_{276} that

$$\left\| g_n - e^{\underline{Q}(s-t)}\mathbb{I}_a \right\| \leq \epsilon\,,$$

and hence, due to Corollary 6.25_{280}, that, up to an error of at most $\epsilon = 10^{-3}$, we have

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_a(X_s)\,|\,X_t = a\big] \approx 0.266 \quad \text{and} \quad \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_a(X_s)\,|\,X_t = b\big] \approx 0.184\,.$$

Analogously repeating the above process using $y := b$ yields that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_b(X_s)\,|\,X_t = a\big] \approx 0.259 \quad \text{and} \quad \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_b(X_s)\,|\,X_t = b\big] \approx 0.482\,,$$

up to an error no more than $\epsilon = 10^{-3}$. ◊

The previous example demonstrated how to compute lower expectations of functions that depend on the state at only a single time point. In the following example we demonstrate the use of Algorithm 2_{286} to compute the lower expectation of a function that depends on more than one time point, for the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$.

Example 6.3. Let $\mathscr{X}$, $\mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1_{290}, and let $t, s \in \mathbb{R}_{>0}$ be as in Example 6.2_{292}. It will be the aim of this example to compute the lower probability—for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$—of the event $(X_t = X_s)$ conditional on the state of the system at time zero. That is, we will be interested in the lower probability that the underlying system, at time s, is in the same state that it is in at time t, given some starting state of the system. As explained in Section 5.4_{198}, we can write this quantity, for some $x \in \mathscr{X}$, as

$$\underline{P}^{\mathrm{W}}_{\mathscr{Q}}(X_t = X_s \,|\, X_0 = x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x\right],$$

where $\mathbb{I}_{X_t = X_s}$ is the indicator of the event $(X_t = X_s)$, i.e. the function that is defined, for all $\omega \in \Omega$, as

$$\mathbb{I}_{X_t = X_s}(\omega) := \begin{cases} 1 & \text{if } \omega(t) = \omega(s), \\ 0 & \text{otherwise.} \end{cases}$$

Since this function only depends on the time points t and s, it is clearly $\{t,s\}$-measurable (c.f. Definition 2.14_{72}). Hence, for ease of notation, we can replace it with the function $f' \in \mathscr{L}(\mathscr{X}_{\{t,s\}})$ that is defined, for all $x_t, x_s \in \mathscr{X}$, as

$$f'(x_t, x_s) := \begin{cases} 1 & \text{if } x_t = x_s, \\ 0 & \text{otherwise.} \end{cases}$$

Then $f'(X_t, X_s) = \mathbb{I}_{X_t = X_s}$, and the quantity that we want to compute is

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[f'(X_t, X_s) \,\middle|\, X_0 = x\right],$$

for some $x \in \mathscr{X}$. The method by which we will compute this quantity is given by Algorithm 2_{286}. However, that algorithm requires that the function of interest also depends on the state of the system in the conditioning event (so state X_0 in our case), which f' clearly does not. The solution is offered by Lemma 6.32_{289}—with $u = \{0\}$ and $v = \{t,s\}$—which tells us that we can instead use the trivial extension of f' to $f \in \mathscr{L}(\mathscr{X}_{\{0,t,s\}})$, defined for all $x_0, x_t, x_s \in \mathscr{X}$ as $f(x_0, x_t, x_s) := f'(x_t, x_s)$. Since $0 < t < s$ it holds that $\{0\} \cap \{t,s\} = \emptyset$, whence Lemma 6.32_{289} implies that $f(X_0, X_t, X_s) = f'(X_t, X_s)$, so that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[f'(X_t, X_s) \,\middle|\, X_0 = x\right] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[f(X_0, X_t, X_s) \,\middle|\, X_0 = x\right],$$

for any $x \in \mathscr{X}$, and we can now proceed by computing the quantity on the right-hand side of this equation.

Let $(\underline{T}_q^r)$ denote the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$. We know from Example 6.1$_{290}$ that $\mathscr{Q}$ is non-empty, compact, convex, has separately specified rows, and has $\underline{Q}$ as its corresponding lower transition rate operator. Since $\mathscr{Q}$ is compact, it is bounded by Corollary A.12$_{378}$. Hence, it follows from Proposition 6.26$_{281}$ that $\underline{T}_q^r = e^{\underline{Q}(r-q)}$ for all $q,r \in \mathbb{R}_{\geq 0}$ with $q \leq r$. Moreover, it follows from Theorem 6.4$_{263}$ that, for any $x \in \mathscr{X}$,

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[f(X_0,X_t,X_s)\,|\,X_0 = x\right] = \left[\underline{T}_0^t \underline{T}_t^s f\right](x)\,, \tag{6.23}$$

using the usual convention for applying $\underline{T}_t^s$ to $f \in \mathscr{L}(\mathscr{X}_{\{0,t,s\}})$ and $\underline{T}_0^t$ to $\underline{T}_t^s f \in \mathscr{L}(\mathscr{X}_{\{0,t\}})$; see Section 6.1$_{260}$ for details.

In the notation of Algorithm 2$_{286}$, we now have $u = \{0\}$ and $v = \{t_0,t_n\} = \{t_0,t_1\} = \{t,s\}$, so that $n = 1$. Numerically, we will aim to compute the quantity of interest up to a maximum numerical error of $\varepsilon := 2\cdot 10^{-3}$. Executing Algorithm 2$_{286}$, we start by setting $g_n = g_1 = f$, and identify the time points $w_n = w_1 = u \cup \{t_0,\ldots,t_{n-1}\} = \{0,t\}$. For all $x_{\{0,t\}} \in \mathscr{X}_{\{0,t\}}$ we now numerically compute (see Line 5–8 of Algorithm 2$_{286}$) the quantity

$$h_{x_{\{0,t\}}} \approx e^{\underline{Q}(s-t)} g_n(x_{\{0,t\}},\cdot) \tag{6.24}$$

guaranteeing a numerical error of at most $\varepsilon/2 = 10^{-3}$, and then set the value of $g_{n-1} \in \mathscr{L}(\mathscr{X}_{\{0,t\}})$ in $x_{\{0,t\}}$ as

$$g_{n-1}(x_{\{0,t\}}) := h_{x_{\{0,t\}}}(x_t)\,. \tag{6.25}$$

Let us consider these steps in more detail. In Equation (6.24), we apply $e^{\underline{Q}(s-t)}$ to the function $g_n(x_{\{0,t\}},\cdot)$, that is, the function $g_n = f$ projected onto $\mathscr{L}(\mathscr{X}_s)$ by fixing its first two arguments to be $x_{\{0,t\}}$. From the definition of f, we therefore have that, for any $x_s \in \mathscr{X}_s$,

$$g_n(x_{\{0,t\}},x_s) = f(x_0,x_t,x_s) = f'(x_t,x_s) = \begin{cases} 1 & \text{if } x_t = x_s \\ 0 & \text{otherwise.} \end{cases}$$

We should note several things here. First, the value of g_n in $x_{\{0,t,s\}}$ does not depend on x_0; this should be obvious since $g_n = f$ and f was defined by extending f' in such a way that its value was independent of x_0. Secondly, for any $x_t \in \mathscr{X}_t$ (and any $x_0 \in \mathscr{X}_0$), we see that

$$\mathbb{I}_{x_t}(x_s) = \begin{cases} 1 & \text{if } x_t = x_s \\ 0 & \text{otherwise} \end{cases} = g_n(x_{\{0,t\}},x_s)\,. \tag{6.26}$$

Therefore, by combining Equations (6.25), (6.24), and (6.26)—in that order—we find, for any $x_{\{0,t\}} \in \mathscr{X}_{\{0,t\}}$, that

$$g_{n-1}(x_{\{0,t\}}) \approx \left[e^{\underline{Q}(s-t)} g_n(x_{\{0,t\}},\cdot)\right](x_t) = e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)\,.$$

Now observe that the quantities on the right-hand side of this equation, are exactly the quantities that we computed in Example 6.2292 up to the desired numerical error of 10^{-3}. Hence, referring back to this previous example, for any $x_0 \in \mathscr{X}_0$ we find that

$$g_{n-1}(x_0,a) \approx 0.266 \quad \text{and} \quad g_{n-1}(x_0,b) \approx 0.482\,. \tag{6.27}$$

At this point we have finished the computations outlined on Lines 3–9 of Algorithm 2286—having computed g_0—and it remains to perform the iterations specified on Lines 10–13.

Since $u = \{0\}$, we first fix any $x_0 \in \mathscr{X}_0$, and now compute, up to a maximum numerical error of $\varepsilon/2 = 10^{-3}$, the quantity

$$h_{x_0} \approx e^{\underline{Q}(t-0)} g_0(x_0,\cdot) \tag{6.28}$$

and then set

$$\tilde{f}(x_0) \coloneqq h_{x_0}(x_0)\,. \tag{6.29}$$

Since we know that the value of g_0 does not actually depend on x_0, this is relatively straightforward. We simply have

$$g_{x_0}(x_0,\cdot) = \begin{bmatrix} 0.266 \\ 0.482 \end{bmatrix},$$

so in order to compute h_{x_0} we simply execute Algorithm 1277 once more. As before, we have $\|\underline{Q}\| = 4$, but now have $\|g_{x_0}(x_0,\cdot)\|_v = 0.482 - 0.266 = 0.216$. Hence, we can execute Algorithm 1277 with

$$\lceil \max\{(16 \cdot 0.216)/2\varepsilon, 4\} \rceil = 6912$$

iterations. This proceeds completely analogously to our computations in Example 6.2292, so that we eventually find that

$$e^{\underline{Q}(t-0)} g_0(x_0,\cdot) \approx \begin{bmatrix} 0.322 \\ 0.370 \end{bmatrix},$$

up to a numerical error of 10^{-3}. Since we already argued that these values are independent of the choice of $x_0 \in \mathscr{X}_0$, we find that

$$\tilde{f} = \begin{bmatrix} 0.322 \\ 0.370 \end{bmatrix}.$$

The additivity of the errors, as guaranteed by Lemma 6.29284, ensures that $\tilde{f}(x_0)$ will not differ from $\underline{T}_0^t \underline{T}_t^s f(x_0)$ by more than the desired maximum numerical error of $2 \cdot 10^{-3} = \varepsilon$. Hence, substituting through the earlier simplifications of the quantity of interest, we conclude that

$$\underline{P}^{\mathrm{W}}_{\mathscr{Q}}(X_t = X_s \,|\, X_0 = a) \approx 0.322 \quad \text{and} \quad \underline{P}^{\mathrm{W}}_{\mathscr{Q}}(X_t = X_s \,|\, X_0 = b) \approx 0.370\,,$$

up to a numerical error of at most $2 \cdot 10^{-3}$. ◊

6.6.2 *A Counterexample for Sets of Markov Chains*

We concluded the previous section with Example 6.3294, where we illustrated how we can use the machinery developed in this chapter to compute inferences for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, when these inferences depend on more than one time point.

The goal of this current section is to show that we *cannot* in general employ the same methods to compute such inferences for the imprecise-Markov chain $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. To put it differently: the methods that we have developed are valid for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, but up until this point we have not yet shown that the lower expectations for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ can actually be different. Had they been the same—which, again, unfortunately they are not—we could have also used these same methods to compute lower expectations for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. So, our current aim is to show that the first inequality in Proposition 5.22199 can indeed be strict, implying that these lower expectations do not in general agree.

Establishing this property will, unfortunately, be fairly involved; since the crucial results that we have derived so far only apply to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, we cannot directly use them to obtain properties for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. Therefore, we will start by deriving some *ad hoc* results for the special case that we are considering here, that is, where $\mathscr{X}$ is binary and $\mathscr{Q}$ is of the form (6.18)290.

Before we begin, however, let us provide some intuition for what we are about to prove. The function for which we will show that the two models have a different lower expectation, is the function $\mathbb{I}_{X_t=X_s}$ for which we were able to compute the lower expectation with respect to the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ in Example 6.3294. The intuitive reason that we cannot do this efficiently when working with sets of Markov chains is the following. In order to compute the lower expectation $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}[\mathbb{I}_{X_t=X_s} \,|\, X_0=x_0]$, we must effectively minimise the probability that the system is in the same state at time t and s; put differently, we should here choose a precise model P so as to "steer away" from the state X_t that is occupied at time t. This "steering away" is done by choosing the dynamic behaviour of the system *between* the time points t and s. However, in the computation of $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}[\mathbb{I}_{X_t=X_s} \,|\, X_0=x]$, we do not know the state X_t at time $t>0$, whence we must do this "steering away" for every possible state at time t jointly. The crux is now that Markov chains, by their Markovian nature, will "forget" at time $r>t$ the state that they had occupied at time t, and so the dynamic behaviour at the time points $r\in(t,s)$ cannot be chosen to depend on the state X_t. Hence, the minimisation with respect to the set of Markov chains cannot attain the minimum that is achieved over a set of processes that can depend intricately on their historic behaviour—and that can, in particular, account for the historic state X_t. The remainder of this section is devoted to formalising this intuition and proving the above-mentioned inequality, in the form of a continued running example.

The first step to this end is an alternate formulation of some results in Example 6.1$_{290}$; we there already derived explicit expressions for $\underline{Q}f$, but we will now show that $\underline{Q}f$ is obtained by some Qf, where Q is chosen from a subset of two rate matrices in $\mathscr{Q}$, depending on f.

Example 6.4. Let $\mathscr{X}$, $\mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1$_{290}$. Consider the two rate matrices

$$Q_{a\geq b} := \begin{bmatrix} -\overline{\lambda}_a & \overline{\lambda}_a \\ \underline{\lambda}_b & -\underline{\lambda}_b \end{bmatrix} \quad \text{and} \quad Q_{a\leq b} := \begin{bmatrix} -\underline{\lambda}_a & \underline{\lambda}_a \\ \overline{\lambda}_b & -\overline{\lambda}_b \end{bmatrix}. \tag{6.30}$$

Inspection of Equation (6.18)$_{290}$ shows that $Q_{a\geq b}, Q_{a\leq b} \in \mathscr{Q}$. Next, fix any $f \in \mathscr{L}(\mathscr{X})$. Then, using the properties of matrix-vector multiplication, we have

$$Q_{a\geq b}f = \begin{bmatrix} Q_{a\geq b}f(a) \\ Q_{a\geq b}f(b) \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_a\big(f(b)-f(a)\big) \\ \underline{\lambda}_b\big(f(a)-f(b)\big) \end{bmatrix}$$

and

$$Q_{a\leq b}f = \begin{bmatrix} Q_{a\leq b}f(a) \\ Q_{a\leq b}f(b) \end{bmatrix} = \begin{bmatrix} \underline{\lambda}_a\big(f(b)-f(a)\big) \\ \overline{\lambda}_b\big(f(a)-f(b)\big) \end{bmatrix}.$$

Comparing this to Equations (6.20)$_{291}$ and (6.21)$_{291}$, we see that

$$\underline{Q}f = \begin{cases} Q_{a\geq b}f & \text{if } f(a) \geq f(b), \\ Q_{a\leq b}f & \text{otherwise.} \end{cases} \tag{6.31}$$

We conclude that $\underline{Q}f$ is obtained by Qf, with $Q \in \{Q_{a\geq b}, Q_{a\leq b}\} \subseteq \mathscr{Q}$, depending on whether $f(a) \geq f(b)$ or $f(a) < f(b)$. ◊

The next property that we need is that, for any $t \in \mathbb{R}_{\geq 0}$ and any $f \in \mathscr{L}(\mathscr{X})$, it holds that $e^{\underline{Q}t}f = e^{Qt}f$, where $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ is such that $\underline{Q}f = Qf$, as in Example 6.4. Put differently, this tells us that—in the very specific case that we are considering here—the generalised exponential $e^{\underline{Q}t}f$ evaluated in f, coincides with a (normal) matrix exponential $e^{Qt}f$ evaluated in f, with Q depending only on f.

Since we will be relying heavily on an analysis of matrix exponentials, let us start by stating a closed-form expression that holds in the binary-state case that we are considering here.

Lemma 6.34 ([108, Theorem 2]). *Let $\mathscr{X} = \{a,b\}$ be an ordered binary state space, fix any $\lambda_a, \lambda_b \in \mathbb{R}_{>0}$, and define the 2×2 rate matrix Q by*

$$Q := \begin{bmatrix} -\lambda_a & \lambda_a \\ \lambda_b & -\lambda_b \end{bmatrix}.$$

Then for all $t \in \mathbb{R}_{>0}$ it holds that

$$e^{Qt} = I + \frac{1-e^{-t(\lambda_a+\lambda_b)}}{\lambda_a+\lambda_b} Q\,.$$

From this result, it immediately follows that for any $f \in \mathscr{L}(\mathscr{X})$, it holds that

$$e^{Qt} f = f + \frac{1-e^{-t(\lambda_a+\lambda_b)}}{\lambda_a+\lambda_b} Qf\,, \tag{6.32}$$

which is an equality that we will use in Example 6.5 below.

Let us now show that, as already claimed above, the generalised exponential coincides with such precise exponentials under the conditions of the running example.

Example 6.5. Let $\mathscr{X}$, $\mathcal{Q}$, and $\underline{Q}$ be as in Example 6.1$_{290}$. For any $f \in \mathscr{L}(\mathscr{X})$, let $\lambda_f \in \mathbb{R}_{>0}$ be defined by

$$\lambda_f := \begin{cases} \overline{\lambda}_a + \underline{\lambda}_b & \text{if } f(a) \geq f(b), \\ \underline{\lambda}_a + \overline{\lambda}_b & \text{otherwise.} \end{cases}$$

As shown in e.g. [32, 108], it then holds for any $t \in \mathbb{R}_{\geq 0}$ that

$$e^{\underline{Q}t} f = f + \frac{1-e^{-t\lambda_f}}{\lambda_f} \underline{Q}f = f + \frac{1-e^{-t\lambda_f}}{\lambda_f} Qf\,. \tag{6.33}$$

where $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ is such that $\underline{Q}f = Qf$, as in Example 6.4. Moreover, for this same matrix Q, it follows from Lemma 6.34, and from Equation (6.32) in particular, that

$$e^{Qt} f = f + \frac{1-e^{-t\lambda_f}}{\lambda_f} Qf\,,$$

which implies that $e^{\underline{Q}t} f = e^{Qt} f$. ◊

In summary, we have shown that $e^{\underline{Q}t} f$ is obtained by $e^{Qt} f$, with $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ such that $\underline{Q}f = Qf$. Up to this point these results have been phrased in terms of (generalised) exponentials, i.e. we have only shown an identity between operators. However, this result also has an interesting interpretation in the context of imprecise-Markov chains: since we know from Corollary 6.25$_{280}$ that $e^{\underline{Q}t} f(x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}\left[f(X_t)\,|\,X_0 = x\right]$, and since we know from Equation (5.15)$_{205}$ that

$$\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}}\left[f(X_t)\,|\,X_0 = x\right] = \inf_{Q\in\mathcal{Q}} e^{Qt} f(x)\,,$$

Example 6.5$_{\curvearrowleft}$ and Proposition 5.22$_{199}$ together imply that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}\big[f(X_t)\,|\,X_0=x\big]=\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}}\big[f(X_t)\,|\,X_0=x\big]\,, \tag{6.34}$$

under—and we really want to emphasize this point—*the assumptions made in Examples 6.1*$_{290}$ *and 6.5*$_{\curvearrowleft}$, and with $\mathcal{M}$ an arbitrary non-empty set of probability mass functions on $\mathcal{X}$. Specifically, the fact that this f only depends on the state at a single time point t, together with the fact that we are dealing with a binary state space, are the crucial assumptions that lead to this property; as we have already established in Example 5.3$_{205}$, where we used a ternary state space, the identity (6.34) does not hold in general, even for functions that only depend on the state at a single time point.

Before moving on, it will be useful to establish that the quantities $e^{\underline{Q}t}\mathbb{I}_x$ are strictly positive for all $t\in\mathbb{R}_{>0}$ and all $x\in\mathcal{X}$, as shown in the following example.

Example 6.6. Let $\mathcal{X}$, $\mathcal{Q}$, and $\underline{Q}$ be as in Example 6.1$_{290}$, and let $Q_{a\geq b}, Q_{a\leq b}$ be as in Example 6.4$_{298}$. It holds that $\mathbb{I}_a(a)>\mathbb{I}_a(b)$ and $\mathbb{I}_b(a)<\mathbb{I}_b(b)$, and therefore we know from Example 6.4$_{298}$ that $\underline{Q}\mathbb{I}_a=Q_{a\geq b}\mathbb{I}_a$ and $\underline{Q}\mathbb{I}_b=Q_{a\leq b}\mathbb{I}_b$. Now fix any $t>0$. Then we know from Example 6.5$_{\curvearrowleft}$ that $e^{\underline{Q}t}\mathbb{I}_a=e^{Q_{a\geq b}t}\mathbb{I}_a$ and $e^{\underline{Q}t}\mathbb{I}_b=e^{Q_{a\leq b}t}\mathbb{I}_b$.

Using the closed-form expression from Lemma 6.34$_{298}$, and writing $\gamma_a:=\overline{\lambda}_a+\underline{\lambda}_b$, it follows after some straightforward algebra that

$$e^{\underline{Q}t}\mathbb{I}_a=e^{Q_{a\geq b}t}\mathbb{I}_a=\begin{bmatrix}1-\frac{1-e^{-t\gamma_a}}{\gamma_a}\overline{\lambda}_a\\ \frac{1-e^{-t\gamma_a}}{\gamma_a}\underline{\lambda}_b\end{bmatrix}.$$

Since $\gamma_a>\overline{\lambda}_a>0$ and $t>0$, we find that $e^{\underline{Q}t}\mathbb{I}_a(a)>0$ and $e^{\underline{Q}t}\mathbb{I}_a(b)>0$.

Similarly, writing $\gamma_b:=\underline{\lambda}_a+\overline{\lambda}_b$, Lemma 6.34$_{298}$ implies that

$$e^{\underline{Q}t}\mathbb{I}_b=e^{Q_{a\leq b}t}\mathbb{I}_b=\begin{bmatrix}\frac{1-e^{-t\gamma_b}}{\gamma_b}\underline{\lambda}_a\\ 1-\frac{1-e^{-t\gamma_b}}{\gamma_b}\overline{\lambda}_b\end{bmatrix}.$$

Since $\gamma_b>\overline{\lambda}_b>0$ and $t>0$, we find that $e^{\underline{Q}t}\mathbb{I}_b(a)>0$ and $e^{\underline{Q}t}\mathbb{I}_b(b)>0$. $\Diamond$

Let us now consider the imprecise-Markov chain $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q}}$. We know from Corollary 6.25$_{280}$ that $e^{\underline{Q}t}f(x)=\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathcal{Q}}\big[f(X_t)\,|\,X_0=x\big]$, or in other words, that the lower expectation of f for $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q}}$ is given by $e^{\underline{Q}t}f(x)$; and therefore, due to Example 6.5$_{\curvearrowleft}$, by $e^{Qt}f(x)$, with Q depending only on f in the sense that $\underline{Q}f=Qf$. Since, by Equation (5.5)$_{191}$, $Q\in\mathcal{Q}$ induces a homogeneous Markov chain $P\in\mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q}}$ with characterising rate matrix $Q_P=Q$, (with the initial distribution chosen arbitrarily), and since

$\mathbb{P}_{\mathscr{Q}}^{\mathrm{WHM}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ by Proposition 5.9$_{190}$, this implies that the lower expectation $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}\left[f(X_t) \,|\, X_0 = x\right]$ is reached by some $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$. That is, we know that there is some $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ with conditional expectation $\mathbb{E}_P$, such that

$$\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}\left[f(X_t) \,|\, X_0 = x\right] = \mathbb{E}_P\left[f(X_t) \,|\, X_0 = x\right].$$

In particular, due to the above, we know that this lower expectation is always reached by one of two homogeneous Markov chains, *viz.* the ones characterised by the rate matrices $Q_{a \geq b}$ and $Q_{a \leq b}$. However, this does not immediately imply that these are the *only* two Markov chains that reach the lower expectation. Hence, our next goal will be to strengthen this result, in the following sense: we will show that, provided that $f(a) \neq f(b)$, any Markov chain $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ that satisfies $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}\left[f(X_s) \,|\, X_t = x\right] = \mathbb{E}_P\left[f(X_s) \,|\, X_t = x\right]$ for some $x \in \mathscr{X}$, has a corresponding transition matrix that satisfies ${}^{P}T_t^s = e^{Q(s-t)}$, with $Q \in \{Q_{a \geq b}, Q_{a \leq b}\}$ such that $\underline{Q}f = Qf$. Note that we here also generalise the time interval of the inference from $[0,t]$ to $[t,s]$, which will be helpful down the line.

To this end, we first need a relatively straightforward—if a bit abstract—property of transition matrices in the context of binary state spaces. Essentially, this result tells us that the images of a non-constant function $f \in \mathscr{L}(\mathscr{X})$ under two transition matrices are different, whenever these transition matrices are different.

Example 6.7. Let $\mathscr{X} = \{a,b\}$ be a binary state space, fix any $f \in \mathscr{L}(\mathscr{X})$ such that $f(a) \neq f(b)$, and let T,S be two (2×2) transition matrices such that $T \neq S$. We will establish that $Tf \neq Sf$.

Since $T \neq S$ there is some $x \in \mathscr{X}$ such that $T(x,\cdot) \neq S(x,\cdot)$. In turn, this means that there is some $y \in \mathscr{X}$ such that $T(x,y) \neq S(x,y)$. However, because T,S are transition matrices, it follows from Definition 3.5$_{91}$ that $T(x,a) = 1 - T(x,b)$ and $S(x,a) = 1 - S(x,b)$, so together this implies that $T(x,a) \neq S(x,a)$ and $T(x,b) \neq S(x,b)$.

Now let $z \in \mathscr{X}$ be such that $f(z) = \max f$, and let $y \in \mathscr{X}$ be such that $y \neq z$; this is clearly always possible since $\mathscr{X}$ is binary. Using the variation norm $\|f\|_v = \max f - \min f$, it follows that

$$\begin{aligned} Tf(x) &= T(x,y)f(y) + T(x,z)f(z) \\ &= T(x,y)\min f + T(x,z)\max f \\ &= T(x,y)\min f + T(x,z)\min f - T(x,z)\min f + T(x,z)\max f \\ &= \min f + \|f\|_v T(x,z), \end{aligned}$$

where in the last equality we used that T satisfies Property T1$_{91}$. Completely analogously, we find that $Sf(x) = \min f + \|f\|_v S(x,z)$. Since $f(a) \neq f(b)$ it holds that $\|f\|_v \neq 0$, and since we already know that $T(x,z) \neq S(x,z)$, it follows that $Tf \neq Sf$. ◊

Let us now prove that, as claimed above, any Markov chain in $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ whose conditional expectation $\mathbb{E}_P[f(X_s)\,|\,X_t = x]$ agrees with the lower expectation $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}[f(X_s)\,|\,X_t = x]$ for some $x \in \mathscr{X}$, has a corresponding transition matrix T_t^s that satisfies $T_t^s = e^{Q(s-t)}$, where $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ is such that $\underline{Q}f = Qf$; provided, again, that the specific assumptions of the running example are satisfied.

Example 6.8. Let $\mathscr{X}, \mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1₂₉₀, and let $Q_{a\geq b}, Q_{a\leq b}$ be as in Example 6.4₂₉₈. Let $f \in \mathscr{L}(\mathscr{X})$ be such that $f(a) \neq f(b)$. Let $t, s \in \mathbb{R}_{>0}$ be such that $t < s$, and let $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ be a Markov chain with corresponding transition matrix T_t^s such that, for some $x \in \mathscr{X}$, it holds that $T_t^s f(x) = e^{\underline{Q}(s-t)} f(x)$. Let $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ be such that $\underline{Q}f = Qf$; this is possible by Example 6.4₂₉₈. Moreover, we have seen in Example 6.5₂₉₉ that then $e^{\underline{Q}(s-t)} f = e^{Q(s-t)} f$, and hence it also holds that $T_t^s f(x) = e^{Q(s-t)} f(x)$. We will now show that $T_t^s = e^{Q(s-t)}$.

We start by showing that $T_r^s = e^{Q(s-r)}$ for all $r \in (t,s)$, where T_r^s is the transition matrix corresponding to P. Assume *ex absurdo* that there is some $r \in (t,s)$ such that $T_r^s \neq e^{Q(s-r)}$. Because $f(a) \neq f(b)$, as we have seen in Example 6.7_↶, this implies that also $T_r^s f \neq e^{Q(s-r)} f$.

Because T_r^s corresponds to P, and since $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, it follows from Proposition 5.25₂₀₁ that for all $y \in \mathscr{X}$, it holds that

$$T_r^s f(y) \geq \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}[f(X_s)\,|\,X_r = y] = e^{\underline{Q}(s-r)} f(y)\,,$$

where we used Corollary 6.25₂₈₀ for the equality. Because $Q \in \{Q_{a\geq b}, Q_{a\leq b}\}$ is such that $\underline{Q}f = Qf$, it follows from Example 6.5₂₉₉ that $e^{\underline{Q}(s-r)} f = e^{Q(s-r)} f$. By combining the above properties, it follows that there must be some $y \in \mathscr{X}$ such that $T_r^s f(y) > e^{Q(s-r)} f(y) = e^{\underline{Q}(s-r)} f(y)$. Let $y' \in \mathscr{X}$ be such that $y' \neq y$.

Next let T_t^r be the transition matrix corresponding to P that contains the transition probabilities over the time interval $[t, r]$. Then, as we know from Proposition 5.1₁₈₃ and Definition 4.6₁₅₆, it holds that $T_t^s = T_t^r T_r^s$. Furthermore, since $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, it follows from Proposition 5.25₂₀₁ and Corollary 6.25₂₈₀ that, for any $g \in \mathscr{L}(\mathscr{X})$ and any $z \in \mathscr{X}$,

$$T_t^r g(z) \geq e^{\underline{Q}(r-t)} g(z)\,. \tag{6.35}$$

Hence in particular, we find that

$$T_t^r(x,y) = T_t^r \mathbb{I}_y(x) \geq e^{\underline{Q}(r-t)} \mathbb{I}_y(x) > 0\,,$$

where for the strict inequality we used that $t < r$, whence $(r - t) > 0$, together with the results derived in Example 6.6₃₀₀.

By combining the above properties, we now find that

$$\begin{aligned} T_t^s f(x) &= T_t^r T_r^s f(x) \\ &= T_t^r(x,y)T_r^s f(y) + T_t^r(x,y')T_r^s f(y') \\ &> T_t^r(x,y)e^{\underline{Q}(s-r)} f(y) + T_t^r(x,y')T_r^s f(y') \\ &\geq T_t^r(x,y)e^{\underline{Q}(s-r)} f(y) + T_t^r(x,y')e^{\underline{Q}(s-r)} f(y') \\ &= T_t^r e^{\underline{Q}(s-r)} f(x) \geq e^{\underline{Q}(r-t)} e^{\underline{Q}(s-r)} f(x) = e^{\underline{Q}(s-t)} f(x), \end{aligned}$$

where we used that $T_t^r(x,y) > 0$ and $T_r^s f(y) > e^{\underline{Q}(s-r)} f(y)$ for the strict inequality, that $T_r^s f(y') \geq e^{\underline{Q}(s-r)} f(y')$ for the first non-strict inequality, Equation (6.35) for the final inequality, and Proposition 6.17_{273} for the final equality. Hence we have found that $T_t^s f(x) > e^{\underline{Q}(s-t)} f(x)$, which is a contradiction since $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ was selected so that $T_t^s f(x) = e^{\underline{Q}(s-t)} f(x)$. Hence our assumption must be wrong, and it must hold that $T_r^s = e^{Q(s-r)}$ for all $r \in (t,s)$. It remains to establish the desired equality in the limit where r goes to t.

So fix any ϵ. Because $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ it holds that $P \sim \mathscr{Q}$, which means that $\overline{\partial} T_t^t \subseteq \mathscr{Q}$. Because P is well-behaved, it follows from Proposition 4.23_{171} that there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$, there is some $Q' \in \overline{\partial} T_t^t \subseteq \mathscr{Q}$ such that

$$\left\| T_t^{t+\Delta} - (I + \Delta Q') \right\| < \Delta\epsilon. \tag{6.36}$$

So fix any such $0 < \Delta < \delta$ with $t + \Delta < s$ and a corresponding $Q' \in \mathscr{Q}$, and let $r := t + \Delta$; then $r \in (t,s)$ and $\Delta = r - t$. It now follows that

$$\begin{aligned} \left\| T_t^s - e^{Q(s-t)} \right\| &= \left\| T_t^r T_r^s - e^{Q(r-t)} e^{Q(s-r)} \right\| \\ &\leq \left\| T_t^r - e^{Q(r-t)} \right\| + \left\| T_r^s - e^{Q(s-r)} \right\| \\ &= \left\| T_t^{t+\Delta} - e^{Q\Delta} \right\| \\ &= \left\| T_t^{t+\Delta} - (I + \Delta Q') + (I + \Delta Q') - e^{Q\Delta} \right\| \\ &\leq \left\| T_t^{t+\Delta} - (I + \Delta Q') \right\| + \Delta \left\| Q' \right\| + \left\| I - e^{Q\Delta} \right\| \\ &\leq \Delta\epsilon + \Delta \left\| Q' \right\| + \Delta \left\| Q \right\| \leq \Delta\epsilon + 2\Delta \left\| \mathscr{Q} \right\|, \end{aligned}$$

where for the first inequality we used Lemma $\mathrm{B}.5_{393}$, which we can do because all matrices involved are transition matrices; for the second equality we used that $\Delta = r - t$ and that $\|T_r^s - e^{Q(s-r)}\| = 0$ because, as established above, $T_r^s = e^{Q(s-r)}$ since $r \in (t,s)$; for the third inequality we used Equation (6.36) and Lemma $\mathrm{B}.10_{394}$; and for the final inequality we used that $Q, Q' \in \mathscr{Q}$ and hence $\|Q\| \leq \|\mathscr{Q}\|$ and $\|Q'\| \leq \mathscr{Q}$.

Because ε and $\|\mathscr{Q}\|$ are bounded and independent of Δ, and because this inequality holds for all $0 < \Delta < \delta$ with $t + \Delta < s$, this implies that $\|T_t^s - e^{Q(s-t)}\| = 0$, or equivalently, that $T_t^s = e^{Q(s-t)}$. $\Diamond$

The crucial point of the previous example deserves some emphasis. Suppose that some Markov chain $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ has a transition matrix T_t^s such that, for some $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$, $T_t^s f(x) = e^{\underline{Q}(s-t)} f(x)$. From the properties of matrix-vector multiplication, the value $T_t^s f(x)$ is completely determined by the x-row of T_t^s, that is, it holds that $T_t^s f(x) = \sum_{y \in \mathscr{X}} T_t^s(x,y) f(y)$. Hence, that the equality $T_t^s f(x) = e^{\underline{Q}(s-t)} f(x)$ should imply something about, at least, the x-row of T_t^s, seems obvious. However, the previous example has shown that this condition implies that there is some $Q \in \mathscr{Q}$ such that $T_t^s = e^{Q(s-t)}$, which means that this *additionally* enforces the value of the y-row of T_t^s (with $y \neq x$).

This has an interesting and important consequence that cuts to the heart of why the lower expectations of $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ and $\mathbb{P}_{\mathscr{Q}}^{\mathrm{W}}$ do not always coincide: it implies that the set $\mathscr{T} := \left\{ {}^P T_t^s \,:\, P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}} \right\}$ of transition matrices induced by $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ *does not have separately specified rows*. Contrast this with the sets of transition matrices induced by $\mathbb{P}_{\mathscr{Q}}^{\mathrm{W}}$; we established with Theorem 5.21$_{198}$ that those do have separately specified rows.

To see that this claim is true, consider that, since $\mathbb{I}_a(a) > \mathbb{I}_a(b)$ and $\mathbb{I}_b(a) < \mathbb{I}_b(b)$, it follows from Example 6.8$_{302}$ that any Markov chain with transition matrix T_t^s that satisfies $T_t^s \mathbb{I}_a(a) = e^{\underline{Q}(s-t)} \mathbb{I}_a(a)$ must satisfy $T_t^s = e^{Q_{a \geq b}(s-t)}$, while if it would satisfy $T_t^s \mathbb{I}_b(b) = e^{\underline{Q}(s-t)} \mathbb{I}_b(b)$ it would hold that $T_t^s = e^{Q_{a \leq b}(s-t)}$. Since $e^{Q_{a \geq b}(s-t)} \neq e^{Q_{a \leq b}(s-t)}$[8] this implies that there are no Markov chains in $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ whose transition matrix T_t^s satisfies both $T_t^s \mathbb{I}_a(a) = e^{\underline{Q}(s-t)} \mathbb{I}_a(a)$ and $T_t^s \mathbb{I}_b(b) = e^{\underline{Q}(s-t)} \mathbb{I}_b(b)$. This means that the transition matrix T that is defined as $T(a,\cdot) := e^{Q_{a \geq b}(s-t)}(a,\cdot)$ and $T(b,\cdot) := e^{Q_{a \leq b}(s-t)}(b,\cdot)$, is not in the set $\mathscr{T}$. On the other hand, it holds that $e^{Q_{a \geq b}(s-t)}, e^{Q_{a \leq b}(s-t)} \in \mathscr{T}$, since they are the transition matrices of the homogeneous Markov chains characterised by $Q_{a \geq b}$ and $Q_{a \leq b}$, respectively, which are clearly in $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ since $Q_{a \geq b}, Q_{a \leq b} \in \mathscr{Q}$. Hence we conclude that $\mathscr{T}$ does not have separately specified rows.

It is the aim of the following example to formalise this property in a way that we can use for the final example of this section.

Example 6.9. Let $\mathscr{X}, \mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1$_{290}$, and let $Q_{a \geq b}$ and $Q_{a \leq b}$ be as in Example 6.4$_{298}$. Let $t := 1$ and let $s := 2$, as in Example 6.2$_{292}$. Consider the imprecise-Markov chain $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ and, for any $P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$, let ${}^P T_t^s$ be the corresponding transition matrix containing the transition probabilities over the interval $[t,s]$. Consider the induced set

[8] For details, see Example 6.9 further on.

$\mathscr{T}$ of transition matrices, defined as

$$\mathscr{T} := \left\{ {}^{P}T_t^s \,:\, P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}} \right\}.$$

Then, because we know from Example 6.1$_{290}$ that $\mathscr{Q}$ is non-empty, compact, and convex, it follows from Corollary 5.18$_{197}$ that $\mathscr{T}$ is compact.

Now, for any $T \in \mathscr{T}$, we define the function $\varepsilon_T \in \mathscr{L}(\mathscr{X})$, for all $x \in \mathscr{X}$, as

$$\varepsilon_T(x) := T\mathbb{I}_x(x) - e^{\underline{Q}(s-t)}\mathbb{I}_x(x).$$

We first note that, for every $T \in \mathscr{T}$, it follows from Proposition 5.25$_{201}$ and Corollary 6.25$_{280}$ that $T\mathbb{I}_x(x) \geq e^{\underline{Q}(s-t)}\mathbb{I}_x(x)$, and hence that $\varepsilon_T \geq 0$. Using the compactness of $\mathscr{T}$, we will now show that

$$\inf_{T \in \mathscr{T}} \big(\varepsilon_T(a) + \varepsilon_T(b)\big) > 0. \tag{6.37}$$

Suppose *ex absurdo* that this is false. Since we know that $\varepsilon_T \geq 0$ for all $T \in \mathscr{T}$, it then holds that

$$\inf_{T \in \mathscr{T}} \big(\varepsilon_T(a) + \varepsilon_T(b)\big) = 0.$$

This implies the existence of a sequence $\{T_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathscr{T}$ such that $\lim_{i \to +\infty}\big(\varepsilon_{T_i}(a) + \varepsilon_{T_i}(b)\big) = 0$. Since $\mathscr{T}$ is compact, it is sequentially compact by Corollary A.12$_{378}$, which implies the existence of a convergent subsequence $\{T_{i_j}\}_{j \in \mathbb{Z}_{>0}}$ such that $\lim_{j \to +\infty} T_{i_j} =: T_{i_*} \in \mathscr{T}$. This implies that also $\varepsilon_{T_{i_*}}(a) + \varepsilon_{T_{i_*}}(b) = 0$. Since $\varepsilon_{T_{i_*}} \geq 0$, this means that $\varepsilon_{T_{i_*}}(a) = 0$ and $\varepsilon_{T_{i_*}}(b) = 0$, or, using the definition of $\varepsilon_{T_{i_*}}$, that

$$T_{i_*}\mathbb{I}_a(a) = e^{\underline{Q}(s-t)}\mathbb{I}_a(a) \quad \text{and} \quad T_{i_*}\mathbb{I}_b(b) = e^{\underline{Q}(s-t)}\mathbb{I}_b(b). \tag{6.38}$$

Since $\mathbb{I}_a(a) > \mathbb{I}_a(b)$ and $\mathbb{I}_b(a) < \mathbb{I}_b(b)$, it holds that $\underline{Q}\mathbb{I}_a = Q_{a \geq b}\mathbb{I}_a$ and $\underline{Q}\mathbb{I}_b = Q_{a \leq b}\mathbb{I}_b$. Equation (6.38) together with the results from Example 6.8$_{302}$ therefore imply that

$$T_{i_*} = e^{Q_{a \geq b}(s-t)} \quad \text{and} \quad T_{i_*} = e^{Q_{b \geq a}(s-t)}. \tag{6.39}$$

Now, using that $s - t = 1$, together with the definition of $Q_{a \geq b}$ and $Q_{a \leq b}$ in Example 6.4$_{298}$ and the values $\underline{\lambda}_a = 1/2$, $\overline{\lambda}_a = 2$, $\underline{\lambda}_b = 1/2$, $\overline{\lambda}_b = 1$ from Example 6.1$_{290}$, the closed form expression in Lemma 6.34$_{298}$ for $e^{Q(s-t)}$, with $Q \in \{Q_{a \geq b}, Q_{a \leq b}\}$, implies that

$$e^{Q_{a \geq b}(s-t)}(a,b) = \frac{1 - e^{-2.5}}{2.5} 2 \approx 0.73$$

and

$$e^{Q_{a \leq b}(s-t)}(a,b) = \frac{1 - e^{-1.5}}{1.5} \frac{1}{2} \approx 0.26,$$

which clearly implies that $e^{Q_{a\geq b}(s-t)} \neq e^{Q_{a\leq b}(s-t)}$. In turn, this implies that Equation (6.39)$_{\curvearrowleft}$ must be false. From this contradiction, we conclude that our assumption must be wrong, or in other words, that Equation (6.37)$_{\curvearrowleft}$ indeed holds. In conclusion, this implies the existence of some $\varepsilon > 0$ such that

$$\inf_{T\in\mathscr{T}} \big(\varepsilon_T(a) + \varepsilon_T(b)\big) = \varepsilon > 0\,, \tag{6.40}$$

a property on which we will rely in Example 6.10 below. ◊

We now finally have all the pieces to show that, under the specific conditions of the running example in this section, there are functions for which the lower expectations for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ disagree. The final example of this section details the full argument.

Example 6.10. Let $\mathscr{X}, \mathscr{Q}$, and $\underline{Q}$ be as in Example 6.1$_{290}$, and let $t := 1$ and $s := 2$ as in Example 6.2$_{292}$. Fix any $x_0 \in \mathscr{X}_0$. As in Example 6.3$_{294}$, we consider the lower probability of the event $(X_t = X_s)$, conditional on the state x_0 of the system at time zero, but now for the imprecise-Markov chain $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. Our goal with this example is to show that

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\big[\mathbb{I}_{X_t=X_s} \,\big|\, X_0 = x_0\big] > \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_{X_t=X_s} \,\big|\, X_0 = x_0\big]\,,$$

thereby establishing that the first inequality in Proposition 5.22$_{199}$ can indeed be strict. To this end, let $f \in \mathscr{L}(\mathscr{X}_{\{0,t,s\}})$ be as in Example 6.3$_{294}$, i.e. such that, for all $x_t, x_s \in \mathscr{X}$,

$$f(x_0,x_t,x_s) := \begin{cases} 1 & \text{if } x_t = x_s, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Then, as in Example 6.3$_{294}$, it holds that $f(X_0,X_t,X_s) = \mathbb{I}_{X_t=X_s}$. Let us first simplify the quantity $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_{X_t=X_s} \,\big|\, X_0 = x_0\big]$. To this end, let $(\underline{T}^r_q)$ denote the family of lower transition operators corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$. We know from Example 6.1$_{290}$ that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows. Hence, it follows from Theorem 6.4$_{263}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[\mathbb{I}_{X_t=X_s} \,\big|\, X_0 = x_0\big] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\big[f(X_0,X_t,X_s) \,\big|\, X_0 = x_0\big] = \underline{T}^t_0\underline{T}^s_t f(x_0)\,, \tag{6.41}$$

using the usual convention for applying lower transition operators to functions that depend on the state at more than one time point; see e.g. Section 6.1$_{260}$ for details. Due to this convention, it holds for any $x_t \in \mathscr{X}_t$ that

$$\underline{T}^s_t f(x_0,x_t) = \big[\underline{T}^s_t f(x_0,x_t,\cdot)\big](x_0,x_t) = \underline{T}^s_t \mathbb{I}_{x_t}(x_0,x_t) = e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)\,, \tag{6.42}$$

where for the second equality we used that $f(x_0,x_t,\cdot) = \mathbb{I}_{x_t}$ due to the definition of f, and where for the final equality we used Proposition 6.26$_{281}$, which we can do because $\mathscr{Q}$ is non-empty and bounded (by Corollary A.12$_{378}$, since it is compact) and has separately specified rows, and has $\underline{Q}$ as its corresponding lower transition rate operator.

Next, we will move on to focus on the quantity

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t=X_s}\,\middle|\,X_0=x_0\right]=\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]$$
$$=\inf_{P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}}\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]. \qquad (6.43)$$

First fix any $P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. Then it follows from Proposition 5.31$_{208}$ (with $u=\{0\}$, $v=\{t\}$, and $w=\{s\}$) that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]=\mathbb{E}_P\left[\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0,X_t\right]\,\middle|\,X_0=x_0\right].$$

Moreover, using Proposition 2.25$_{75}$ and the fact that $0<t$ it follows that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]=\mathbb{E}_P\left[\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t,X_0\right]\,\middle|\,X_0=x_0\right]$$
$$=\mathbb{E}_P\left[\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t,X_0=x_0\right]\,\middle|\,X_0=x_0\right]. \qquad (6.44)$$

Now fix any $x_t\in\mathscr{X}_t$. Then, again due to Proposition 2.25$_{75}$, and because $0<t<s$, it holds that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t=x_t,X_0=x_0\right]$$
$$=\mathbb{E}_P\left[f(x_0,x_t,X_s)\,\middle|\,X_t=x_t,X_0=x_0\right]$$
$$=\sum_{x_s\in\mathscr{X}_s}f(x_0,x_t,x_s)P(X_s=x_s\,|\,X_t=x_t,X_0=x_0),$$

where we used Proposition 2.23$_{73}$ for the second equality. Because we know that P is a Markov chain since $P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, it follows from Definition 5.1$_{182}$ that $P(X_s=x_s\,|\,X_t=x_t,X_0=x_0)=P(X_s=x_s\,|\,X_t=x_t)$, and hence that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t=x_t,X_0=x_0\right]=\sum_{x_s\in\mathscr{X}_s}f(x_0,x_t,x_s)P(X_s=x_s\,|\,X_t=x_t).$$

Using the definition of f, we conclude that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t=x_t,X_0=x_0\right]=P(X_s=x_t\,|\,X_t=x_t)$$
$$={}^PT_t^s(x_t,x_t)={}^PT_t^s\mathbb{I}_{x_t}(x_t),$$

where ${}^PT_t^s$ is the transition matrix corresponding to P that contains the transition probabilities over the time interval $[t,s]$. Now for any transition matrix T, define the function $f_T\in\mathscr{L}(\mathscr{X})$, for all $x\in\mathscr{X}$, as

$$f_T(x):=T\mathbb{I}_x(x).$$

Then it holds that

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_t=x_t,X_0=x_0\right]={}^PT_t^s\mathbb{I}_{x_t}(x_t)=f_{{}^PT_t^s}(x_t)\,,$$

and since this is true for all $x_t\in\mathscr{X}_t$, we can substitute back into Equation (6.44)↶ to obtain

$$\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]=\mathbb{E}_P\left[f_{{}^PT_t^s}(X_t)\,\middle|\,X_0=x_0\right].$$

As in Example 6.9$_{304}$, let $\mathscr{T}$ denote the set of transition matrices over the interval $[t,s]$ corresponding to all elements of $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$. Because $P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ it follows that ${}^PT_t^s\in\mathscr{T}$, and hence it follows that

$$\begin{aligned}\mathbb{E}_P\left[f(X_0,X_t,X_s)\,\middle|\,X_0=x_0\right]&=\mathbb{E}_P\left[f_{{}^PT_t^s}(X_t)\,\middle|\,X_0=x_0\right]\\&\geq\inf_{T\in\mathscr{T}}\mathbb{E}_P\left[f_T(X_t)\,\middle|\,X_0=x_0\right].\end{aligned}\tag{6.45}$$

Now, for all $T\in\mathscr{T}$ let $\varepsilon_T\in\mathscr{L}(\mathscr{X})$ be as in Example 6.9$_{304}$. Then, as we know from that example, there is some $\varepsilon>0$ such that

$$\inf_{T\in\mathscr{T}}\big(\varepsilon_T(a)+\varepsilon_T(b)\big)=\varepsilon>0\,.$$

Fix any $T\in\mathscr{T}$. Then for all $x_t\in\mathscr{X}$ it holds that

$$\varepsilon_T(x_t)=T\mathbb{I}_{x_t}(x_t)-e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)=f_T(x_t)-e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)\,,$$

whence it follows that $f_T(x_t)=\varepsilon_T(x_t)+e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)$. Moreover, we know from Equation (6.42)$_{306}$ that $e^{\underline{Q}(s-t)}\mathbb{I}_{x_t}(x_t)=\underline{T}_t^sf(x_0,x_t)$, and hence we find that

$$f_T(x_t)=\varepsilon_T(x_t)+\underline{T}_t^sf(x_0,x_t)\,.$$

Because this is true for all $x_t\in\mathscr{X}$, it holds that $f_T(X_t)=\varepsilon_T(X_t)+\underline{T}_t^sf(x_0,X_t)$, from which it follows that

$$\begin{aligned}\mathbb{E}_P\left[f_T(X_t)\,\middle|\,X_0=x_0\right]&=\mathbb{E}_P\left[\varepsilon_T(X_t)+\underline{T}_t^sf(x_0,X_t)\,\middle|\,X_0=x_0\right]\\&=\mathbb{E}_P\left[\varepsilon_T(X_t)\,\middle|\,X_0=x_0\right]+\mathbb{E}_P\left[\underline{T}_t^sf(x_0,X_t)\,\middle|\,X_0=x_0\right],\end{aligned}\tag{6.46}$$

where we used Property CE2$_{78}$ for the second equality. We proceed by first bounding the first summand on the right-hand side of this expression. Using Proposition 2.23$_{73}$, it holds that

$$\begin{aligned}\mathbb{E}_P\left[\varepsilon_T(X_t)\,\middle|\,X_0=x_0\right] &= \sum_{x_t\in\mathscr{X}} \varepsilon_T(x_t)P(X_t=x_t\,|\,X_0=x_0)\\ &= \sum_{x_t\in\mathscr{X}} \varepsilon_T(x_t)\mathbb{E}_P[\mathbb{I}_{x_t}(X_t)\,|\,X_0=x_0]\\ &\geq \sum_{x_t\in\mathscr{X}} \varepsilon_T(x_t)\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}[\mathbb{I}_{x_t}(X_t)\,|\,X_0=x_0]\\ &= \sum_{x_t\in\mathscr{X}} \varepsilon_T(x_t)e^{\underline{Q}t}\mathbb{I}_{x_t}(x_0)\\ &= \varepsilon_T(a)e^{\underline{Q}t}\mathbb{I}_a(x_0)+\varepsilon_T(b)e^{\underline{Q}t}\mathbb{I}_b(x_0)\\ &\geq \big(\varepsilon_T(a)+\varepsilon_T(b)\big)\min\Big\{e^{\underline{Q}t}\mathbb{I}_a(x_0),e^{\underline{Q}t}\mathbb{I}_b(x_0)\Big\}\,,\end{aligned}$$

where for the first inequality we used that $P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, for the third equality we used Corollary 6.25$_{280}$ together with the properties of $\mathscr{Q}$ established above, and where for the final inequality we used that $\varepsilon_T\geq 0$ as in Example 6.9$_{304}$. Now, as we have seen in Example 6.6$_{300}$, since $t>0$ it holds that $e^{\underline{Q}t}\mathbb{I}_a(x_0)>0$ and $e^{\underline{Q}t}\mathbb{I}_b(x_0)>0$, whence also $C:=\min\{e^{\underline{Q}t}\mathbb{I}_a(x_0),e^{\underline{Q}t}\mathbb{I}_b(x_0)\}>0$.

Substituting this back into Equation (6.46) we find that

$$\mathbb{E}_P\left[f_T(X_t)\,\middle|\,X_0=x_0\right]\geq C\big(\varepsilon_T(a)+\varepsilon_T(b)\big)+\mathbb{E}_P\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right].$$

Because this is true for all $T\in\mathscr{T}$, this implies that

$$\begin{aligned}&\inf_{T\in\mathscr{T}}\mathbb{E}_P\left[f_T(X_t)\,\middle|\,X_0=x_0\right]\\ &\qquad\geq \inf_{T\in\mathscr{T}} C\big(\varepsilon_T(a)+\varepsilon_T(b)\big)+\mathbb{E}_P\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right]\\ &\qquad= C\varepsilon+\mathbb{E}_P\left[\underline{T}_t^s f(x,X_t)\,\middle|\,X_0=x_0\right],\end{aligned}$$

where we used that $\varepsilon=\inf_{T\in\mathscr{T}}\big(\varepsilon_T(a)+\varepsilon_T(b)\big)$. Substituting back into Equation (6.45) we find that

$$\mathbb{E}_P[f(X_0,X_t,X_s)\,|\,X_0=x_0]\geq C\varepsilon+\mathbb{E}_P\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right].$$

Because this is true for all $P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$, substituting this inequality into Equation (6.43)$_{307}$ we obtain

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t=X_s}\,\middle|\,X_0=x_0\right] &= \inf_{P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}}\mathbb{E}_P[f(X_0,X_t,X_s)\,|\,X_0=x_0]\\ &\geq \inf_{P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}}\left(C\varepsilon+\mathbb{E}_P\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right]\right)\\ &= C\varepsilon+\inf_{P\in\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}}\mathbb{E}_P\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right]\\ &= C\varepsilon+\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\underline{T}_t^s f(x_0,X_t)\,\middle|\,X_0=x_0\right].\end{aligned}\tag{6.47}$$

It now remains to simplify the second summand on the right-hand side of this expression. Using the previously established properties of $\mathscr{Q}$, it follows from Corollary 6.25280 that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\underline{T}_t^s f(x_0, X_t) \,\middle|\, X_0 = x_0\right] &= \left[e^{\underline{Q}t}\underline{T}_t^s f(x_0, \cdot)\right](x_0) \\ &= \left[\underline{T}_0^t \underline{T}_t^s f(x_0, \cdot)\right](x_0) \\ &= \underline{T}_0^t \underline{T}_t^s f(x_0) \\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x_0\right],\end{aligned}$$

where $\underline{T}_0^t$ is the lower transition operator corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ and where we used Proposition 6.26281 for the second equality, the usual convention of applying $\underline{T}_0^t$ to the function $\underline{T}_t^s f \in \mathscr{L}(\mathscr{X}_{\{0,t\}})$ in the third equality, and Equation (6.41)306 in the final equality.

Substituting this back into Equation (6.47)↶ we find that

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x_0\right] \geq C\varepsilon + \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x_0\right],$$

which, since $C > 0$ and $\varepsilon > 0$, implies that

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x_0\right] > \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}}\left[\mathbb{I}_{X_t = X_s} \,\middle|\, X_0 = x_0\right],$$

which finally shows that the first inequality in Proposition 5.22199 can indeed be strict, which is what we wanted to demonstrate. ◊

Appendix

6.A Proofs of Results in Section 6.2

Proof of Proposition 6.11$_{269}$. It is immediate from Definition 6.4$_{268}$ that $\mathscr{Q}_{\underline{Q}}$ is a set of rate matrices. Let us show that, for every $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$. To this end, fix any $f \in \mathscr{L}(\mathscr{X})$. Now choose $\Delta > 0$ small enough such that $0 \leq \Delta \|\underline{Q}\| \leq 1$—this always possible because of Property LR5$_{266}$—and define $\underline{T} := I + \Delta\underline{Q}$. Since $\underline{Q}$ is a lower transition rate operator, it then follows from Proposition 6.6$_{266}$ that $\underline{T}$ is a lower transition operator.

Using Definition 3.17$_{120}$, $\underline{T}$ has a dominating set $\mathscr{T}_{\underline{T}}$ of transition matrices. By Proposition 3.37$_{120}$, $\mathscr{T}_{\underline{T}}$ is non-empty, closed, has separately specified rows, and has $\underline{T}$ as its corresponding lower transition operator; therefore, and due to Proposition 3.36$_{119}$, there is some $T \in \mathscr{T}_{\underline{T}}$ such that $Tf = \underline{T}f$. Moreover, for any $g \in \mathscr{L}(\mathscr{X})$ it holds that $Tg \geq \underline{T}g$ because $T \in \mathscr{T}_{\underline{T}}$ and $\underline{T}$ is the lower envelope of this set.

Now let $Q := {}^1\!/\!_\Delta(T - I)$. Then Q is a rate matrix by Proposition 4.10$_{153}$. Because $Tf = \underline{T}f$, it follows that

$$Qf = \frac{1}{\Delta}(Tf - f) = \frac{1}{\Delta}(\underline{T}f - f) = \underline{Q}f.$$

Similarly, because $Tg \geq \underline{T}g$ for all $g \in \mathscr{L}(\mathscr{X})$, it follows that $Qg \geq \underline{Q}g$, or in other words, since Q is a rate matrix, that $Q \in \mathscr{Q}_{\underline{Q}}$. Because f was arbitrary, this proves that, for all $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $Qf = \underline{Q}f$. Since $\mathscr{L}(\mathscr{X})$ is non-empty, this clearly implies that $\mathscr{Q}_{\underline{Q}}$ is non-empty.

We will next show that $\mathscr{Q}_{\underline{Q}}$ is bounded. Consider any $x \in \mathscr{X}$. Then for all $Q \in \mathscr{Q}_{\underline{Q}}$, we have that $Q(x,x) = Q\mathbb{I}_x(x) \geq \underline{Q}\mathbb{I}_x(x)$, which implies that

$$\inf_{Q \in \mathscr{Q}_{\underline{Q}}} Q(x,x) \geq \underline{Q}\mathbb{I}_x(x) > -\infty,$$

where the second inequality used Property LR5$_{266}$. Since $x \in \mathscr{X}$ is arbitrary, Proposition 4.8$_{152}$ now guarantees that $\mathscr{Q}_{\underline{Q}}$ is bounded.

To establish that $\mathscr{Q}_{\underline{Q}}$ has $\underline{Q}$ as its corresponding lower transition rate operator, we need to show that $\underline{Q}$ is the lower envelope of this set. Because we already established that $\mathscr{Q}_{\underline{Q}}$ is non-empty and bounded, this lower envelope is well-defined by Definition 6.3$_{267}$. Now fix any $f \in \mathscr{L}(\mathscr{X})$. By Definition 6.4$_{268}$, it holds that $\underline{Q}f(x) \leq Qf(x)$ for all $Q \in \mathscr{Q}_{\underline{Q}}$, so $\underline{Q}f(x)$ is a lower bound on $\{Qf(x) : Q \in \mathscr{Q}_{\underline{Q}}\}$. Moreover, we already established that there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$, so this lower bound is tight. In other words, it holds that $\underline{Q}f(x) = \inf_{Q \in \mathscr{Q}_{\underline{Q}}} Qf(x)$, whence $\underline{Q}$ is the lower transition rate operator corresponding to $\mathscr{Q}_{\underline{Q}}$.

Let us next show that $\mathscr{Q}_{\underline{Q}}$ is closed, or in other words, using Proposition A.8$_{376}$, that for any convergent sequence $\{Q_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{Q}_{\underline{Q}}$ its limit belongs to $\mathscr{Q}_{\underline{Q}}$. So fix any convergent sequence $\{Q_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{Q}_{\underline{Q}}$, and let $Q_* := \lim_{i\to+\infty} Q_i$. By Proposition 4.6$_{151}$ the metric space $\mathscr{R}$ of all rate matrices is complete, so it follows that, because $\mathscr{Q}_{\underline{Q}} \subseteq \mathscr{R}$ and $\{Q_i\}_{i\in\mathbb{Z}_{>0}}$ lies in $\mathscr{Q}_{\underline{Q}}$, also Q_* is a rate matrix. Now suppose *ex absurdo* that $Q_* \notin \mathscr{Q}_{\underline{Q}}$. Then, using Definition 6.4$_{268}$, there is some $f \in \mathscr{L}(\mathscr{X})$ and some $x \in \mathscr{X}$, such that $Q_*f(x) < \underline{Q}f(x)$. Let $\varepsilon := \underline{Q}f(x) - Q_*f(x)$; then $\varepsilon > 0$ because $Q_*f(x) < \underline{Q}f(x)$, and $\underline{Q}f(x) = Q_*f(x) + \varepsilon$.

Since $\lim_{i\to+\infty} Q_i = Q_*$, it follows from Lemma A.34$_{390}$ that there is some $n \in \mathbb{Z}_{>0}$ such that $\|Q_nf - Q_*f\| < \varepsilon$. Because $Q_n \in \mathscr{Q}_{\underline{Q}}$ it holds that $Q_*f(x) < \underline{Q}f(x) \leq Q_nf(x)$ by Definition 6.4$_{268}$. This implies that

$$Q_nf(x) - Q_*f(x) = |Q_nf(x) - Q_*f(x)| \leq \|Q_nf - Q_*f\| < \varepsilon,$$

which means that $Q_nf(x) < Q_*f(x) + \varepsilon = \underline{Q}f(x)$. This contradicts the fact that $Q_n \in \mathscr{Q}_{\underline{Q}}$, and hence we must have that $Q_* \in \mathscr{Q}_{\underline{Q}}$.

Because, as established above, $\mathscr{Q}_{\underline{Q}}$ is both closed and bounded, it follows that it is compact by Corollary A.12$_{378}$.

Next, we show that $\mathscr{Q}_{\underline{Q}}$ is convex, or equivalently, that for any two rate matrices $Q_1, Q_2 \in \mathscr{Q}_{\underline{Q}}$, and any $\lambda \in [0,1]$, the matrix $Q_\lambda := \lambda Q_1 + (1-\lambda)Q_2$ is again an element of $\mathscr{Q}_{\underline{Q}}$. Because Q_1 and Q_2 are both rate matrices, and because $\lambda \geq 0$ and $(1-\lambda) \geq 0$, it follows from Proposition 4.5$_{151}$ that Q_λ is a rate matrix. Furthermore, for any $f \in \mathscr{L}(\mathscr{X})$, we find that

$$Q_\lambda f = \lambda Q_1 f + (1-\lambda)Q_2 f \geq \lambda \underline{Q}f + (1-\lambda)\underline{Q}f = \underline{Q}f,$$

where the inequality holds because Q_1 and Q_2 belong to $\mathscr{Q}_{\underline{Q}}$. Hence, it follows from Definition 6.4$_{268}$ that $Q_\lambda \in \mathscr{Q}_{\underline{Q}}$.

We finally show that $\mathscr{Q}_{\underline{Q}}$ has separately specified rows. For all $x \in \mathscr{X}$, let $\mathscr{Q}_x := \{Q(x,\cdot) \,|\, Q \in \mathscr{Q}_{\underline{Q}}\}$. Consider now any rate matrix Q such that $Q(x,\cdot) \in \mathscr{Q}_x$ for all $x \in \mathscr{X}$, and assume *ex absurdo* that $Q \notin \mathscr{Q}_{\underline{Q}}$. Definition 6.4$_{268}$ then implies the existence of some $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$ such that $Qf(x) < \underline{Q}f(x)$. Since $Q(x,\cdot) \in \mathscr{Q}_x$, this in turn implies that there is some $Q' \in \mathscr{Q}_{\underline{Q}}$ such that $Q'f(x) < \underline{Q}f(x)$, which is a contradiction. Hence we find that $Q \in \mathscr{Q}_{\underline{Q}}$. ☐

The following lemma uses the interpretation of the rows of a matrix as being elements of the dual space $\mathscr{L}(\mathscr{X})^\top$ of $\mathscr{L}(\mathscr{X})$; see Appendix A$_{369}$ for details.

Lemma 6.35. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices. Then, for all $x \in \mathscr{X}$, $\mathscr{Q}_x := \{Q(x,\cdot) : Q \in \mathscr{Q}\}$ is a non-empty, compact, and convex subset of $\mathscr{L}(\mathscr{X})^\top$.*

Proof. Fix any $x \in \mathscr{X}$. The non-emptiness of $\mathscr{Q}_x$ then follows trivially from the fact that $\mathscr{Q}$ is non-empty.

To see that $\mathscr{Q}_x$ is convex, for any $Q_1(x,\cdot), Q_2(x,\cdot) \in \mathscr{Q}_x$, let Q_1 and Q_2 be the corresponding elements of $\mathscr{Q}$ whose x-row is given by $Q_1(x,\cdot)$ and $Q_2(x,\cdot)$ respectively. Fix any $\lambda \in [0,1]$, and let $Q_\lambda := \lambda Q_1 + (1-\lambda)Q_2$. Then $Q_\lambda \in \mathscr{Q}$ because $\mathscr{Q}$ is convex. This implies that $Q_\lambda(x,\cdot) \in \mathscr{Q}_x$, and $Q_\lambda(x,\cdot) = \lambda Q_1(x,\cdot) + (1-\lambda)Q_2(x,\cdot)$, so $\lambda Q_1(x,\cdot) + (1-\lambda)Q_2(x,\cdot) \in \mathscr{Q}_x$, whence $\mathscr{Q}_x$ is convex.

Because $\mathscr{Q}$ is bounded, by Definition A.12$_{376}$, there is some $B \in \mathbb{R}$ such that $\sup_{Q\in\mathscr{Q}} \|Q\| = \|\mathscr{Q}\| < B$. Using Proposition A.33$_{390}$, it holds that

$$\|Q(x,\cdot)\|_* \leq \|Q\| \leq \|\mathscr{Q}\| < B,$$

from which it follows that $\|\mathscr{Q}_x\|_* = \sup_{Q(x,\cdot)\in\mathscr{Q}_x} \|Q(x,\cdot)\| \leq B$, whence $\mathscr{Q}_x$ is bounded.

Let us next show that $\mathscr{Q}_x$ is closed, or in other words, using Proposition A.8$_{376}$, that for any convergent sequence $\{Q_i(x,\cdot)\}_{i\in\mathbb{Z}_{>0}}$ in $\mathscr{Q}_x$, the limit $Q_*(x,\cdot) := \lim_{i\to+\infty} Q_i(x,\cdot)$ is also an element of $\mathscr{Q}_x$. Let $\{Q_i\}_{i\in\mathbb{Z}_{>0}}$ be the corresponding sequence in $\mathscr{Q}$ such that, for all $i \in \mathbb{Z}_{>0}$, the x-row of Q_i is given by $Q_i(x,\cdot)$. Because $\mathscr{Q}$ is compact, it is sequentially compact by Corollary A.12$_{378}$, which means that there is a convergent subsequence $\{Q_{i_k}\}_{k\in\mathbb{Z}_{>0}}$ such that $\lim_{k\to+\infty} Q_{i_k} =: Q'_* \in \mathscr{Q}$. This means that also $Q'_*(x,\cdot) \in \mathscr{Q}_x$. Let us now show that $Q_*(x,\cdot) = Q'_*(x,\cdot)$.

So fix any $\varepsilon > 0$. Then because $\lim_{k\to+\infty} Q_{i_k} = Q'_*$, there is some $n \in \mathbb{Z}_{>0}$ such that, for all $\ell \in \mathbb{Z}_{>0}$ with $\ell > n$, it holds that $\left\|Q_{i_\ell} - Q'_*\right\| < \varepsilon$. Using Proposition A.33$_{390}$, this implies that also $\left\|Q_{i_k}(x,\cdot) - Q'_*(x,\cdot)\right\|_* < \varepsilon$. Because this is true for any $\varepsilon > 0$, we find that $\lim_{k\to+\infty} Q_{i_k}(x,\cdot) = Q'_*(x,\cdot)$. Because the original sequence $\{Q_i(x,\cdot\}_{i\in\mathbb{Z}_{>0}}$ was convergent with limit $Q_*(x,\cdot)$, it follows that $Q'_*(x,\cdot) = Q_*(x,\cdot)$ and therefore, since $Q'_*(x,\cdot) \in \mathscr{Q}_x$, that $Q_*(x,\cdot) \in \mathscr{Q}_x$.

Hence we conclude that $\mathscr{Q}_x$ is closed and, because we already established that it is bounded, $\mathscr{Q}_x$ is compact by Corollary A.12$_{378}$. □

Proof of Proposition 6.12$_{269}$. Since $\mathscr{Q}$ has $\underline{Q}$ as its lower envelope, it follows from Definition (6.4)$_{268}$ that $\mathscr{Q} \subseteq \mathscr{Q}_{\underline{Q}}$. Assume *ex absurdo* that $\mathscr{Q} \subset \mathscr{Q}_{\underline{Q}}$; then there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $Q \notin \mathscr{Q}$.

Because $\mathscr{Q}$ has separately specified rows, it follows from $Q \notin \mathscr{Q}$ that there is some $x \in \mathscr{X}$ such that $Q(x,\cdot) \notin \mathscr{Q}_x := \{Q'(x,\cdot) \,|\, Q' \in \mathscr{Q}\}$. We note that, as in Appendix A.3$_{383}$, this x-row $Q(x,\cdot)$ of Q is interpreted as an element of the dual space $\mathscr{L}(\mathscr{X})^\top$ of $\mathscr{L}(\mathscr{X})$, whose value in any $f \in \mathscr{L}(\mathscr{X})$ is given by $Q(x,\cdot)f = Qf(x)$.

Now, since $\mathscr{Q}$ is non-empty, compact, and convex, it follows from Lemma 6.35 that $\mathscr{Q}_x$ is a non-empty, compact (and hence closed by Corollary A.12$_{378}$), and convex subset of $\mathscr{L}(\mathscr{X})^\top$, and therefore, since

$\mathscr{L}(\mathscr{X})^\top$ is a normed vector space (see Appendix A.2$_{380}$) and because $Q(x,\cdot) \notin \mathscr{Q}_x$, it follows from a separating hyperplane version of the Hahn-Banach theorem [93, Chapter 14, Corollary 25] that there is some $f \in \mathscr{L}(\mathscr{X})$[9] such that

$$Q(x,\cdot)f < \inf_{Q'(x,\cdot)\in\mathscr{Q}_x} Q'(x,\cdot)f. \tag{6.48}$$

Because $\mathscr{Q}$ is non-empty and compact, has separately specified rows, and has $\underline{Q}$ as its corresponding lower transition rate operator, it follows from Proposition 6.10$_{268}$ that there is some $Q^* \in \mathscr{Q}$ such that $Q^* f = \underline{Q}f$.

Because $Q^* \in \mathscr{Q}$ it holds that $Q^*(x,\cdot) \in \mathscr{Q}_x$ and therefore, we find that

$$\inf_{Q'(x,\cdot)\in\mathscr{Q}_x} Q'(x,\cdot)f \leq Q^*(x,\cdot)f = Q^* f(x) = \underline{Q}f(x)\,.$$

Combining this with Equation (6.48), we get

$$Qf(x) = Q(x,\cdot)f < \inf_{Q'(x,\cdot)\in\mathscr{Q}_x} Q'(x,\cdot)f \leq \underline{Q}f(x)\,,$$

whence $Qf \not\geq \underline{Q}f$, and therefore $Q \notin \mathscr{Q}_{\underline{Q}}$, a contradiction. □

6.B Proofs of Results in Section 6.3

Lemma 6.36. *Fix $n \in \mathbb{Z}_{>0}$ and, for all $k \in \{1,\ldots,n\}$, consider a sequence $\Delta_{k,i} \geq 0$, with $i \in \{1,\ldots,n_k\}$ such that $n_k \in \mathbb{Z}_{>0}$, and let $\Delta_k := \sum_{i=1}^{n_k} \Delta_{k,i}$. Let $C := \sum_{k=1}^{n} \Delta_k$ and let $\delta := \max_{k\in\{1,\ldots,n\}} \Delta_k$. Then for any lower transition rate operator $\underline{Q}$ such that $\delta \left\|\underline{Q}\right\| \leq 1$, it holds that*

$$\left\| \prod_{k=1}^{n} \left(\prod_{i=1}^{n_k} (I + \Delta_{k,i}\underline{Q}) \right) - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) \right\| \leq \delta C \left\|\underline{Q}\right\|^2 .$$

Proof. For any $k \in \{1,\ldots,n\}$, we know that $\Delta_{k,i} \left\|\underline{Q}\right\| \leq \Delta_k \left\|\underline{Q}\right\| \leq \delta \left\|\underline{Q}\right\| \leq 1$ for all $i \in \{1,\ldots,n_k\}$, and therefore, it follows from Propositions 6.6$_{266}$ and 3.33$_{117}$ that $\prod_{i=1}^{n_k}(I + \Delta_{k,i}\underline{Q})$ and $(I + \Delta_k\underline{Q})$ are lower transition oper-

[9] This reference states that f lives in the dual space $\mathscr{L}(\mathscr{X})^{\top\top}$ of $\mathscr{L}(\mathscr{X})^\top$ but, since $\mathscr{L}(\mathscr{X})^\top$ is the dual space of $\mathscr{L}(\mathscr{X})$, and because $\mathscr{L}(\mathscr{X})$ is reflexive by Lemma A.21$_{382}$, it follows that $\mathscr{L}(\mathscr{X}) = \mathscr{L}(\mathscr{X})^{\top\top}$, and hence we get $f \in \mathscr{L}(\mathscr{X})$.

ators. Hence it follows that

$$\begin{aligned}\left\|\prod_{k=1}^{n}\left(\prod_{i=1}^{n_k}(I+\Delta_{k,i}\underline{Q})\right)-\prod_{k=1}^{n}(I+\Delta_k\underline{Q})\right\| \\ &\leq \sum_{k=1}^{n}\left\|\left(\prod_{i=1}^{n_k}(I+\Delta_{k,i}\underline{Q})\right)-(I+\Delta_k\underline{Q})\right\| \\ &\leq \sum_{k=1}^{n}\Delta_k^2\left\|\underline{Q}\right\|^2 \\ &\leq \sum_{k=1}^{n}\delta\Delta_k\left\|\underline{Q}\right\|^2=\delta C\left\|\underline{Q}\right\|^2,\end{aligned}$$

where the first inequality follows from Lemma B.4392 and the second inequality follows from Lemma B.6393. □

Lemma 6.37. *Let $\underline{Q}$ be a lower transition rate operator. Consider any $t \in \mathbb{R}_{\geq 0}$ and any $u \in \mathscr{U}_{[0,t]}$ such that $\sigma(u)\left\|\underline{Q}\right\| \leq 1$. Then for all $u^* \in \mathscr{U}_{[0,t]}$ such that $u \subseteq u^*$, it holds that*

$$\|\Phi_u-\Phi_{u^*}\| \leq t\sigma(u)\left\|\underline{Q}\right\|^2.$$

Proof. This result is trivial if $t=0$, because as discussed in Section 2.258, it then holds that $u=u^*=\{0\}$, which implies that the product in Equation (6.6)271 is empty, whence $\Phi_u=I=\Phi_{u^*}$ and in that case

$$\|\Phi_u-\Phi_{u^*}\|=0\leq t\sigma(u)\left\|\underline{Q}\right\|^2=0.$$

Hence, without loss of generality, we assume that $t>0$, which implies that $u=t_0,\dots,t_n$, with $n\in\mathbb{Z}_{>0}$, $t_0=0$ and $t_n=t$. Since $u\subseteq u^*$, we know that, for all $k\in\{1,\dots,n\}$, there is some sequence $\Delta_{k,i}>0$, $i\in\{1,\dots,n_k\}$, with $n_k\in\mathbb{Z}_{>0}$ and $\Delta_k^u=\sum_{i=1}^{n_k}\Delta_{k,i}\leq\sigma(u)$, such that $\sum_{k=1}^{n}\Delta_k=t$,

$$\Phi_{u^*}:=\prod_{k=1}^{n}\left(\prod_{i=1}^{n_k}(I+\Delta_{k,i}\underline{Q})\right) \text{ and } \Phi_u:=\prod_{k=1}^{n}(I+\Delta_k^u\underline{Q}).$$

Because of Lemma 6.36, this implies that $\|\Phi_{u^*}-\Phi_u\|\leq t\sigma(u)\left\|\underline{Q}\right\|^2$. □

*Proof of Proposition 6.13*271. Let $u'\in\mathscr{U}_{[0,t]}$ be the ordered union of u and u^*. Then, clearly, $u\subseteq u'$ and $u^*\subseteq u'$. Moreover, clearly $\sigma(u')\leq\sigma(u)\leq\delta$ and $\sigma(u')\leq\sigma(u^*)\leq\delta$. Therefore, and because $\delta\left\|\underline{Q}\right\|\leq 1$, Lemma 6.37 implies that $\|\Phi_u-\Phi_{u'}\|\leq t\delta\left\|\underline{Q}\right\|^2$ and $\|\Phi_{u^*}-\Phi_{u'}\|\leq t\delta\left\|\underline{Q}\right\|^2$, and therefore, it follows that

$$\|\Phi_u-\Phi_{u^*}\|\leq\|\Phi_u-\Phi_{u'}\|+\|\Phi_{u'}-\Phi_{u^*}\|\leq 2t\delta\left\|\underline{Q}\right\|^2,$$

which concludes the proof. □

Lemma 6.38. *Let $\underline{Q}$ be a lower transition rate operator, fix any $t \in \mathbb{R}_{\geq 0}$, and consider any $u \in \mathscr{U}_{[0,t]}$ such that $\sigma(u)\left\|\underline{Q}\right\| \leq 1$. Then, with Φ_u as in Equation* (6.6)$_{271}$*, it holds that*

$$\left\|e^{\underline{Q}t} - \Phi_u\right\| \leq t\sigma(u)\left\|\underline{Q}\right\|^2.$$

Proof. Fix any $\varepsilon > 0$. Because of Theorem 6.16$_{272}$, there is some $u^* \in \mathscr{U}_{[0,t]}$ such that $u \subseteq u^*$ and $\left\|e^{\underline{Q}t} - \Phi_{u^*}\right\| < \varepsilon$. By combining this with Lemma 6.37$_{\curvearrowleft}$, it follows that

$$\left\|e^{\underline{Q}t} - \Phi_u\right\| \leq \left\|e^{\underline{Q}t} - \Phi_{u^*}\right\| + \left\|\Phi_{u^*} - \Phi_u\right\| < \varepsilon + t\sigma(u)\left\|\underline{Q}\right\|^2.$$

The result now follows because $\varepsilon > 0$ is arbitrary. □

Proof of Proposition 6.19$_{276}$. First define $h := f - \min f - {}^1\!/_2\left\|f\right\|_{\mathrm{v}}$. Then

$$\max h = \max f - \min f - {}^1\!/_2\left\|f\right\|_{\mathrm{v}} = \left\|f\right\|_{\mathrm{v}} - {}^1\!/_2\left\|f\right\|_{\mathrm{v}} = {}^1\!/_2\left\|f\right\|_{\mathrm{v}}$$

and

$$\min h = \min f - \min f - {}^1\!/_2\left\|f\right\|_{\mathrm{v}} = -{}^1\!/_2\left\|f\right\|_{\mathrm{v}},$$

and therefore,

$$\left\|h\right\| := \max\{|h(x)| : x \in \mathscr{X}\} = {}^1\!/_2\left\|f\right\|_{\mathrm{v}}.$$

Let $\Delta := {}^t\!/_n$, and let $u \in \mathscr{U}_{[0,t]}$ be such that $u = t_0, t_1, \dots, t_n$, where, for all $i \in \{0, 1, \dots, n\}$, $t_i := i\Delta$. Since $\Delta = {}^t\!/_n$, we then have that $t_0 = 0$, $t_n = t$, $\sigma(u) = \Delta$ and, with Φ_u as in Equation (6.6)$_{271}$, $\Phi_u = \prod_{i=1}^n (I + {}^t\!/_n\underline{Q})$. Furthermore, since $n \geq t\left\|\underline{Q}\right\|$, we also know that $\Delta\left\|\underline{Q}\right\| = {}^t\!/_n\left\|\underline{Q}\right\| \leq 1$. Hence, we find that

$$\begin{aligned}\left\|e^{\underline{Q}t}h - \prod_{i=1}^n (I + {}^t\!/_n\underline{Q})h\right\| = \left\|e^{\underline{Q}t}h - \Phi_u h\right\| &\leq \left\|e^{\underline{Q}t} - \Phi_u\right\|\left\|h\right\| \\ &\leq \sigma(u)t\left\|\underline{Q}\right\|^2\left\|h\right\| \\ &= \Delta t\left\|\underline{Q}\right\|^2\frac{\left\|f\right\|_{\mathrm{v}}}{2} \\ &= \frac{\varepsilon}{n}\frac{1}{2\varepsilon}t^2\left\|\underline{Q}\right\|^2\left\|f\right\|_{\mathrm{v}} \leq \varepsilon.\end{aligned}$$

where the first inequality follows from Property N11$_{64}$, the second inequality follows from Lemma 6.38 and the final inequality follows from our lower bound on n. The result is now immediate because $e^{\underline{Q}t}$ and $\prod_{i=1}^n (I + \Delta\underline{Q})$ are both lower transition operators—for the latter, this follows from Propositions 6.6$_{266}$ and 3.33$_{117}$ and the fact that $\Delta\left\|\underline{Q}\right\| \leq 1$—which implies that

$$\left\|e^{\underline{Q}t}h - \prod_{i=1}^n (I + \Delta\underline{Q})h\right\| = \left\|e^{\underline{Q}t}f - \prod_{i=1}^n (I + \Delta\underline{Q})f\right\|$$

because of Property LT6$_{117}$. □

Proof of Proposition 6.20$_{278}$. For notational convenience, let us write $\mathscr{Q} \coloneqq B_r(Q_*)$, and for all $x \in \mathscr{X}$, let $\mathscr{Q}_x \coloneqq \{Q(x,\cdot) : Q \in \mathscr{Q}\}$ denote the set of x-rows of $\mathscr{Q}$.

That $\mathscr{Q}$ is non-empty is clear, since it follows from the fact that $r \geq 0$ that $\|Q_* - Q_*\| = 0 \leq r$, and hence because $Q_* \in \mathscr{R}$ it holds that $Q_* \in \mathscr{Q}$.

Now fix any $Q \in \mathscr{Q}$. It then follows that

$$\|Q\| = \|Q - Q_* + Q_*\| \leq \|Q - Q_*\| + \|Q_*\| \leq r + \|Q_*\|,$$

and hence $\|\mathscr{Q}\| = \sup_{Q \in \mathscr{Q}} \|Q\| \leq \|Q_*\| + r$ which means that $\mathscr{Q}$ is bounded by Definition A.12$_{376}$.

To see that $\mathscr{Q}$ is closed, fix any convergent sequence $\{Q_i\}_{i \in \mathbb{Z}_{>0}}$ and let $Q \coloneqq \lim_{i \to +\infty} Q_i$. According to Proposition A.8$_{376}$, to show that $\mathscr{Q}$ is closed it suffices to show that $Q \in \mathscr{Q}$. By Proposition 4.6$_{151}$, $\mathscr{R}$ is a complete metric space, and because the sequence $\{Q_i\}_{i \in \mathbb{Z}_{>0}}$ is in $\mathscr{Q} \subset \mathscr{R}$, it follows that Q is a rate matrix. So let us show that $\|Q - Q_*\| \leq r$. To this end, fix any $i \in \mathbb{Z}_{>0}$. Then

$$\|Q - Q_*\| = \|Q - Q_i + Q_i - Q_*\| \leq \|Q - Q_i\| + \|Q_i - Q_*\| \leq \|Q - Q_i\| + r.$$

Because this holds for all $i \in \mathbb{Z}_{>0}$, and taking limits, we have $\lim_{i \to +\infty} \|Q - Q_i\| = 0$, which implies that $\|Q - Q_*\| \leq r$, and hence $Q \in \mathscr{Q}$.

Hence we conclude that $\mathscr{Q}$ is closed and, because we already established that it is bounded, $\mathscr{Q}$ is compact by Corollary A.12$_{378}$.

To show that $\mathscr{Q}$ is convex, fix any $Q_1, Q_2 \in \mathscr{Q}$ and any $\lambda \in [0,1]$, and let $Q_\lambda \coloneqq \lambda Q_1 + (1-\lambda) Q_2$. We need to show that $Q_\lambda \in \mathscr{Q}$. Because $Q_1, Q_2 \in \mathscr{Q} \subset \mathscr{R}$, and because $\lambda \in [0,1]$, it follows that $Q_\lambda \in \mathscr{Q}$ by Proposition 4.5$_{151}$. Hence we need to show that $\|Q_\lambda - Q_*\| \leq r$. We have

$$\begin{aligned}
\|Q_\lambda - Q_*\| &= \|Q_\lambda - \lambda Q_* - (1-\lambda) Q_*\| \\
&= \|\lambda(Q_1 - Q_*) + (1-\lambda)(Q_2 - Q_*)\| \\
&\leq \|\lambda(Q_1 - Q_*)\| + \|(1-\lambda)(Q_2 - Q_*)\| \\
&= \lambda \|Q_1 - Q_*\| + (1-\lambda) \|Q_2 - Q_*\| \\
&\leq \lambda r + (1-\lambda) r = r,
\end{aligned}$$

and hence $Q_\lambda \in \mathscr{Q}$, whence $\mathscr{Q}$ is convex.

To see that $\mathscr{Q}$ has separately specified rows, first note that it trivially holds for any $Q \in \mathscr{Q}$ that $Q \in \mathscr{R}$ and that $Q(x,\cdot) \in \mathscr{Q}_x$ for all $x \in \mathscr{X}$. So, according to Definition 5.7$_{193}$, it remains to prove the inclusion in the other direction. To this end, fix any $Q \in \mathscr{R}$ such that $Q(x,\cdot) \in \mathscr{Q}_x$ for all $x \in \mathscr{X}$; we need to prove that $Q \in \mathscr{Q}$. For any $x \in \mathscr{X}$, because

$Q(x,\cdot) \in \mathscr{Q}_x$, there is some $Q_x \in \mathscr{Q}$ such that $Q(x,\cdot) = Q_x(x,\cdot)$, and hence it follows from Equation $(2.7)_{63}$ that

$$\sum_{y\in\mathscr{X}} |Q(x,y) - Q_*(x,y)| = \sum_{y\in\mathscr{X}} |Q_x(x,y) - Q_*(x,y)| \leq \|Q_x - Q_*\| \leq r\,.$$

Because this is true for all $x \in \mathscr{X}$ it follows from Equation $(2.7)_{63}$ that $\|Q - Q_*\| \leq r$, and hence $Q \in \mathscr{Q}$, which implies that $\mathscr{Q}$ has separately specified rows. □

Proof of Proposition 6.21$_{278}$. Let us start by giving the intuition behind the proof. The argument works by constructing a rate matrix $\tilde{Q}$ as a modification of Q_*, and showing that $\underline{Q}g(x) = \tilde{Q}g(x)$ satisfies Equation $(6.12)_{279}$. It will be helpful to understand that each Δ_i, $i \in \{1,\ldots,n\}$ denotes the magnitude of the modification that we make to $Q_*(x,z_i)$, in order to obtain $\tilde{Q}(x,z_i)$. There are a number of constraints that we have to deal with: to ensure that $\tilde{Q}$ stays in $B_r(Q_*)$, we cannot "move away" from Q_* too much. Moreover, to ensure that $\tilde{Q}$ is a rate matrix, it needs to have non-negative off-diagonal entries, and its rows must sum to zero. We make the changes by working from $Q_*(x,z_1)$ to $Q_*(x,z_n)$, that is, we follow the ordering $z_1,\ldots,z_n$. The quantities r_i, $i \in \{1,\ldots,n\}$ denote the *remaining total change* that we still want to make to $Q_*(x,\cdot)$, after modifying $Q_*(x,z_i)$ in a manner that is consistent with these constraints. The zero row-sum constraint will be ensured in $\tilde{Q}(x,z_n)$ and is not kept track of by these r_i.

Before starting the construction, let us establish a number of important properties. Let j be the unique element of $\{1,\ldots,n\}$ such that $z_j = x$. Then $\Delta_j = r_{j-1}$, which means that $r_j = r_{j-1} - \Delta_j = 0$. Therefore, for all $k \in \{1,\ldots,n\}$ with $j < k$, since $Q_*(x,z_k) \geq 0$ due to Property R2$_{150}$ and because $z_k \neq z_j = x$, it follows from a straightforward induction on k that $\Delta_k = \min\{r_{k-1}, Q_*(x,z_k)\} = 0$ and $r_k = r_{k-1} - \Delta_k = 0$. Moreover, for any $i \in \{1,\ldots,n\}$, we have that $r_i = r_{i-1} - \Delta_i$, which means that $\Delta_i = r_{i-1} - r_i$, and therefore, by combining the above properties, it follows that

$$\sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{j} \Delta_i = \sum_{i=1}^{j} r_{i-1} - r_i = r_0 = \frac{r}{2}\,. \tag{6.49}$$

Let us prove by induction that $r_i \geq 0$ and $\Delta_i \geq 0$ for all $i \in \{1,\ldots,j-1\}$. For the induction base, we simply use that $r_0 = r/2 \geq 0$. So now fix any $i \in \{1,\ldots,j-1\}$, and suppose that $r_{i-1} \geq 0$. We will show that then $r_i \geq 0$ and $\Delta_i \geq 0$. Because $i < j$ it holds that $\Delta_i = \min\{r_{i-1}, Q_*(x,z_i)\}$ and, because $z_i \neq z_j = x$ that $Q_*(x,z_i) \geq 0$ due to Property R2$_{150}$. Because by the induction hypothesis $r_{i-1} \geq 0$, it follows that $\Delta_i \geq 0$. Moreover, it holds that $\Delta_i \leq r_{i-1}$, and because $r_i = r_{i-1} - \Delta_i$, and since $r_{i-1} \geq 0$ by the induction hypothesis, it follows that $r_i \geq 0$. This concludes the proof that $r_i \geq 0$ and $\Delta_i \geq 0$ for all $i \in \{1,\dots,j-1\}$. This also implies that $r_{j-1} \geq 0$, and hence that $\Delta_j = r_{j-1} \geq 0$. We already established above that $r_j = 0$ and, moreover, that for all $k \in \{j+1,\dots,n\}$, it holds that $r_k = 0$ and $\Delta_k = 0$, so in summary, we have established that $r_i \geq 0$ and $\Delta_i \geq 0$ for all $i \in \{1,\dots,n\}$.

Moreover, let us similarly show that if $r_k = 0$ for some $k \in \{1,\dots,n-1\}$, then $r_i = 0$ and $\Delta_i = 0$ for all $i \in \{k+1,\dots,n\}$. The argument again proceeds by straightforward induction; suppose that $r_k = 0$ for some $k \in \{1,\dots,n-1\}$. We already established above that $r_i = 0$ and $\Delta_i = 0$ for all $i \in \{j+1,\dots,n\}$, so we can suppose without loss of generality that $k < j$. If $k+1 = j$ then $\Delta_{k+1} = \Delta_j = r_k = 0$, and we already established above that $r_{k+1} = r_j = 0$, so in this case we are done. For the remaining case, suppose that $k+1 < j$. Then $z_{k+1} \neq z_j = x$, so it follows from Property R2150 that $Q_*(x, z_{k+1}) \geq 0$. Moreover, because $\Delta_{k+1} = \min\{r_k, Q_*(x, z_{k+1})\}$, and since $r_k = 0$ by the induction hypothesis, it follows that $\Delta_{k+1} = 0$, and hence that $r_{k+1} = r_k - \Delta_{k+1} = 0$. This concludes the proof that if $r_k = 0$ then also $r_i = 0$ and $\Delta_i = 0$ for all $i \in \{k+1,\dots,n\}$.

Now, let us proceed with the construction. Define $\tilde{Q}(x,\cdot)$ such that, for all $i \in \{1,\dots,n-1\}$,

$$\tilde{Q}(x,z_i) \coloneqq Q_*(x,z_i) - \Delta_i\,,$$

and

$$\tilde{Q}(x,z_n) \coloneqq Q_*(x,z_n) + \frac{r}{2} - \Delta_n\,.$$

It then holds that

$$\begin{aligned}
\sum_{y\in\mathscr{X}} \tilde{Q}(x,y)g(y) &= \sum_{i=1}^{n-1} \bigl(Q_*(x,z_i) - \Delta_i\bigr)g(z_i) + \bigl(Q_*(x,z_n) + \frac{r}{2} - \Delta_n\bigr)g(z_n) \\
&= \sum_{i=1}^{n} Q_*(x,z_i)g(z_i) + \frac{r}{2}g(z_n) - \sum_{i=1}^{n} \Delta_i g(z_i) \\
&= \sum_{i=1}^{n} Q_*(x,z_i)g(z_i) - \sum_{i=1}^{n} \Delta_i\bigl(g(z_i) - g(z_n)\bigr) \\
&= Q_*g(x) - \sum_{i=1}^{n} \Delta_i\bigl(g(z_i) - g(z_n)\bigr)\,,
\end{aligned}$$

where we used Equation (6.49) for the third equality. This establishes that $\sum_{y\in\mathscr{X}} \tilde{Q}(x,y)g(y)$ coincides with the right-hand side of Equation (6.12)279.

For notational convenience, let $\mathscr{Q} \coloneqq B_r(Q_*)$, and let $\mathscr{Q}_x \coloneqq \{Q(x,\cdot) : Q \in \mathscr{Q}\}$ denote the set of x-rows of the elements of $\mathscr{Q}$. We will now show that (i) $\tilde{Q}(x,\cdot) \in \mathscr{Q}_x$, and (ii) that $\underline{Q}g(x) = \sum_{y\in\mathscr{X}} \tilde{Q}(x,y)g(y)$.

First, we note that for any $i \in \{1,\dots,n-1\}$ such that $i \neq j$ it holds that $\Delta_i = \min\{r_{i-1}, Q_*(x,z_i)\}$, and hence

$$\tilde{Q}(x,z_i) = Q_*(x,z_i) - \Delta_i \geq 0\,,$$

because $Q_*(x,z_i) \geq 0$ due to Property R2$_{150}$. Similarly, if $n \neq j$ it holds that $\Delta_n = \min\{r_{n-1}, Q_*(x,z_n)\}$ and hence

$$\tilde{Q}(x,z_n) = Q_*(x,z_n) + \frac{r}{2} - \Delta_n \geq 0\,,$$

because $Q_*(x,z_n) \geq 0$ due to Property R2$_{150}$. Hence we have found that $Q_*(x,z_i) \geq 0$ for all $i \in \{1,\dots,n\}$ with $i \neq j$, whence $\tilde{Q}(x,\cdot)$ also satisfies Property R2$_{150}$. Moreover, it holds that

$$\begin{aligned}\sum_{y\in\mathscr{X}} \tilde{Q}(x,y) &= \sum_{i=1}^{n} \tilde{Q}(x,z_i)\\ &= \sum_{i=1}^{n-1} \big(Q_*(x,z_i) - \Delta_i\big) + Q_*(x,z_n) + \frac{r}{2} - \Delta_n\\ &= \sum_{i=1}^{n} Q_*(x,z_i) + \frac{r}{2} - \sum_{i=1}^{n} \Delta_i = 0\,,\end{aligned}$$

where we used Equation (6.49)$_{318}$ and that Q_* satisfies Property R1$_{150}$. Hence $\tilde{Q}(x,\cdot)$ also satisfies Property R1$_{150}$ and therefore, because we have already established that it satisfies Property R2$_{150}$, it can be interpreted as the x-row of a rate matrix $\tilde{Q}$. Finally, it holds that

$$\begin{aligned}\sum_{y\in\mathscr{X}} \big|Q_*(x,y) - \tilde{Q}(x,y)\big| &= \sum_{i=1}^{n} \big|Q_*(x,z_i) - \tilde{Q}(x,z_i)\big|\\ &= \sum_{i=1}^{n-1} |\Delta_i| + \Big|\frac{r}{2} - \Delta_n\Big| \leq \frac{r}{2} + \sum_{i=1}^{n} |\Delta_i| = r\,, \end{aligned} \tag{6.50}$$

where we used Equation (6.49)$_{318}$ and that $\Delta_i \geq 0$ for all $i \in \{1,\dots,n\}$ for the final equality. Hence, if we now define $\tilde{Q}$ to be a matrix whose x-row is $\tilde{Q}(x,\cdot)$, and whose y-rows are $\tilde{Q}(y,\cdot) := Q_*(y,\cdot)$ for all $y \in \mathscr{X}$ with $y \neq x$, then (i) $\tilde{Q}$ is clearly a rate matrix, and (ii), it holds that

$$\big\|Q_* - \tilde{Q}\big\| = \sum_{y\in\mathscr{X}} \big|Q_*(x,y) - \tilde{Q}(x,y)\big| \leq r\,,$$

where we used Equations (2.7)$_{63}$ and (6.50). Hence it follows that $\tilde{Q} \in B_r(Q_*) = \mathscr{Q}$ and therefore, that $\tilde{Q}(x,\cdot) \in \mathscr{Q}_x$.

We will finally show that $\underline{Q}g(x) = \tilde{Q}g(x)$. Because, as we have just shown, $\tilde{Q} \in \mathscr{Q}$, it clearly holds that $\tilde{Q}g(x) \geq \underline{Q}g(x)$, so let us prove the inequality in the other direction. To this end, we note that due to Proposition 6.20$_{278}$, $\mathscr{Q}$ is a non-empty, compact, and convex set of rate matrices that has separately specified rows. Therefore, it follows from Proposition 6.10$_{268}$ that there is some $Q \in \mathscr{Q}$ such that $\underline{Q}g(x) = Qg(x)$. Let $h := g - \min g = g - g(z_n)$. We note that $h(z_n) = 0$ and that $h(z_i) \geq h(z_{i+1})$ for all $i \in \{1, \ldots, n-1\}$, and hence that $h \geq 0$. Moreover, it follows from Definition 4.4$_{150}$ that

$$\begin{aligned} Qh(x) &= \sum_{y \in \mathscr{X}} Q(x,y)h(y) \\ &= \sum_{y \in \mathscr{X}} Q(x,y)\big(g(y) - \min g\big) = Qg(x) - \min g \sum_{y \in \mathscr{X}} Q(x,y) = Qg(x)\,, \end{aligned}$$

and, similarly, that $\tilde{Q}h(x) = \tilde{Q}g(x)$. Hence, to establish that $\tilde{Q}g(x) \leq Qg(x)$, it suffices to show that $\tilde{Q}h(x) \leq Qh(x)$.

Recall from the beginning of this proof that j is the unique element of $\{1, \ldots, n\}$ such that $z_j = x$, and that $\Delta_j = r_{j-1}$ and hence $r_j = 0$, and that therefore $r_k = 0$ for all $k \in \{j+1, \ldots, n\}$. Now, consider any $k \in \{1, \ldots, n-1\}$ and suppose that $r_k > 0$. This implies that $k \neq j$, and therefore that $\Delta_k = \min\{r_{k-1}, Q_*(x, z_k)\}$. Because $0 < r_k = r_{k-1} - \Delta_k$, this means that $\Delta_k < r_{k-1}$ and hence $\Delta_k = Q_*(x, z_k)$. It therefore follows from the definition of $\tilde{Q}$ that $\tilde{Q}(x, z_k) = Q_*(x, z_k) - \Delta_k = 0$. Moreover, since we already established in the beginning of this proof that $\Delta_k \geq 0$, it follows that also $r_{k-1} = r_k + \Delta_k > 0$.

We now consider several cases. First suppose that $r_{n-1} > 0$. It follows from the above that this implies that $j = n$ and, moreover, that for all $k \in \{1, \ldots, n-1\}$, it holds that $r_k > 0$ and $\tilde{Q}(x, z_k) = 0$. Because we already established that $\sum_{i=1}^n \tilde{Q}(x, z_i) = 0$, it follows that also $\tilde{Q}(x, z_n) = 0$, or in summary, that $\tilde{Q}(x, y) = 0$ for all $y \in \mathscr{X}$. This means that $\tilde{Q}h(x) = \sum_{y \in \mathscr{X}} \tilde{Q}(x,y)h(y) = 0$. Moreover, because Q is a rate matrix and because $j = n$, it follows from Definition 4.4$_{150}$ that $Q(x, z_k) \geq 0$ for all $k \in \{1, \ldots, n-1\}$ and therefore, because $h(z_n) = g(z_n) - \min g = 0$ and $h(z_k) = g(z_k) - \min g \geq 0$ for all $k \in \{1, \ldots, n-1\}$, that

$$Qh(x) = \sum_{i=1}^{n} Q(x, z_i)h(z_i) = \sum_{i=1}^{n-1} Q(x, z_i)h(z_i) \geq 0 = \tilde{Q}h(x)\,.$$

This establishes the desired inequality if $r_{n-1} > 0$.

So, for the other case, suppose that $r_{n-1} = 0$. Then there is a (unique) minimal $k \in \{1, \ldots, n-1\}$ such that $r_k = 0$. As established at the beginning of this proof, this implies that $\Delta_i = 0$ for all $i \in \{k+1, \ldots, n\}$. Moreover, because $r_j = 0$ it follows that $k \leq j$, and hence it follows from our previous reasoning that $\Delta_i = Q_*(x, z_i)$ and $\tilde{Q}(x, z_i) = 0$ for all $i \in \{1, \ldots, k-1\}$. We also note that because $r_k = 0$, it follows that $\Delta_k = r_{k-1}$. Combining these observations, and using Equation (6.49)$_{318}$, it follows that

$$\frac{r}{2} = \sum_{i=1}^{n} \Delta_i = \sum_{i=1}^{k} \Delta_i = r_{k-1} + \sum_{i=1}^{k-1} Q_*(x, z_i)\,,$$

and hence $r_{k-1} = {}^{r}\!/2 - \sum_{i=1}^{k-1} Q_*(x, z_i) = {}^{r}\!/2 - \sum_{i=1}^{k-1} \Delta_i$. This implies that

$$\sum_{i=1}^{n} \Delta_i h(z_i) = \frac{r}{2} h(z_k) + \sum_{i=1}^{k-1} \Delta_i \big(h(z_i) - h(z_k)\big)\,. \tag{6.51}$$

Now, let $v \in \mathscr{L}(\mathscr{X})$ be such that $v(y) := Q(x,y) - Q_*(x,y)$ for all $y \in \mathscr{X}$. Then it holds that

$$\begin{aligned} Qh(x) &= \sum_{y\in\mathscr{X}} Q(x,y)h(y) \\ &= \sum_{y\in\mathscr{X}} \big(Q_*(x,y) + v(y)\big)h(y) = Q_*h(x) + \sum_{y\in\mathscr{X}} v(y)h(y)\,. \end{aligned}$$

Let $\mathscr{N}_- := \big\{i \in \{1,\ldots,n\} : v(z_i) < 0\big\}$, and let $\mathscr{N}_+ := \{1,\ldots,n\} \setminus \mathscr{N}_-$. Then, because $h(z_i) \geq 0$ for all $i \in \{1,\ldots,n\}$, it follows that

$$\begin{aligned} Qh(x) &= Q_*h(x) + \sum_{y\in\mathscr{X}} v(y)h(y) \\ &= Q_*h(x) + \sum_{i\in\mathscr{N}_-} v(z_i)h(z_i) + \sum_{i\in\mathscr{N}_+} v(z_i)h(z_i) \\ &\geq Q_*h(x) + \sum_{i\in\mathscr{N}_-} v(z_i)h(z_i) = Q_*h(x) - \sum_{i\in\mathscr{N}_-} |v(z_i)|\, h(z_i)\,. \end{aligned}$$

Moreover, we note that

$$\begin{aligned} \tilde{Q}h(x) &= \sum_{y\in\mathscr{X}} \tilde{Q}(x,y)h(y) \\ &= \sum_{i=1}^{n-1} \big(Q_*(x,z_i) - \Delta_i\big)h(z_i) + \Big(Q_*(x,z_n) + \frac{r}{2} - \Delta_n\Big)h(z_n) \\ &= Q_*h(x) - \sum_{i=1}^{n} \Delta_i h(z_i)\,, \end{aligned}$$

where we used that $h(z_n) = 0$. So, in order to show that $Qh(x) \geq \tilde{Q}h(x)$, it suffices to show that

$$\sum_{i\in\mathscr{N}_-} |v(z_i)|\, h(z_i) \leq \sum_{i=1}^{n} \Delta_i h(z_i)\,.$$

To this end, we first note that, because Q and Q_* are both rate matrices, it follows from Definition 4.4150 that

$$0 = \sum_{y\in\mathscr{X}} Q(x,y) = \sum_{y\in\mathscr{X}} v(y) + \sum_{y\in\mathscr{X}} Q_*(x,y) = \sum_{y\in\mathscr{X}} v(y)\,.$$

This implies that $\sum_{i\in\mathcal{N}_-}|v(z_i)| = \sum_{i\in\mathcal{N}_+}|v(z_i)|$, which means that $\sum_{i\in\mathcal{N}_-}|v(z_i)| = {}^1\!/\!{}_2\sum_{y\in\mathscr{X}}|v(y)|$. Moreover, because $Q\in\mathscr{Q}=B_r(Q_*)$ it holds that $\|Q-Q_*\|\leq r$ which, using Equation $(2.7)_{63}$, implies that

$$\sum_{y\in\mathscr{X}}|v(y)| = \sum_{y\in\mathscr{X}}|Q(x,y)-Q_*(x,y)| \leq \|Q-Q_*\| \leq r\,.$$

Combining the above, we find that $\sum_{i\in\mathcal{N}_-}|v(z_i)| = {}^1\!/\!{}_2\sum_{y\in\mathscr{X}}|v(y)| \leq {}^r\!/\!{}_2$.

We now have all the pieces to finish the proof. Recall that k is the (unique) minimal element of $\{1,\dots,n-1\}$ such that $r_k=0$. Fix any $i\in\mathcal{N}_-$, and first suppose that $i<k$. Then because $i<k\leq j$, and because Q is a rate matrix, it follows from Definition 4.4_{150} that $Q(x,z_i)\geq 0$, which, using that $v(z_i)<0$, implies that $Q_*(x,z_i)-|v(z_i)|=Q(x,z_i)\geq 0$, or in other words, that $|v(z_i)|\leq Q_*(x,z_i)$. Hence, if $i<k$ then $|v(z_i)|\leq Q_*(x,z_i)=\Delta_i$, and therefore $|v(z_i)|h(z_i)\leq\Delta_i h(z_i)$ because $h\geq 0$. Let $\delta_i \coloneqq \Delta_i-|v(z_i)|$. Then if $i<k$ it holds that $0\leq\delta_i\leq\Delta_i$ and $|v(z_i)|=\Delta_i-\delta_i$, and $\delta_i h(z_i)\geq\delta_i h(z_k)$ because $0\leq h(z_k)\leq h(z_i)$. Conversely, if $k\leq i$ then $h(z_k)\geq h(z_i)$ and hence $|v(z_i)|h(z_k)\geq|v(z_i)|h(z_i)$.

Now, it holds that

$$\begin{aligned}
\sum_{i\in\mathcal{N}_-}|v(z_i)|h(z_i) &= \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_i) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}|v(z_i)|h(z_i)\\
&\leq \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}|v(z_i)|h(z_i)\\
&= \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}(\Delta_i-\delta_i)h(z_i)\\
&= \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_k) - \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\delta_i h(z_i) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\Delta_i h(z_i)\\
&\leq \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_k) - \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\delta_i h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\Delta_i h(z_i)\\
&= \sum_{\substack{i\in\mathcal{N}_-\\:k\leq i}}|v(z_i)|h(z_k) - \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}(\Delta_i-|v(z_i)|)h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\Delta_i h(z_i)\\
&= \sum_{i\in\mathcal{N}_-}|v(z_i)|h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\Delta_i\big(h(z_i)-h(z_k)\big)\\
&\leq \frac{r}{2}h(z_k) + \sum_{\substack{i\in\mathcal{N}_-\\:i<k}}\Delta_i\big(h(z_i)-h(z_k)\big)\\
&\leq \frac{r}{2}h(z_k) + \sum_{i=1}^{k-1}\Delta_i\big(h(z_i)-h(z_k)\big) = \sum_{i=1}^{n}\Delta_i h(z_i)\,,
\end{aligned}$$

where we used that $\sum_{i\in\mathcal{N}_-}|v(z_i)| \leq r/2$ for the third inequality; that $\{i\in\mathcal{N}_- : i<k\}\subseteq\{1,\dots,k-1\}$, that $h(z_i)-h(z_k)\geq 0$ for all $i\in\{1,\dots,k-1\}$, and—as established at the beginning of this proof—that $\Delta_i \geq 0$ for all $i\in\{1,\dots,k-1\}$, for the final inequality; and Equation (6.51)$_{322}$ for the final equality. Hence we find that $\sum_{i\in\mathcal{N}_-}|v(z_i)|h(z_i)\leq\sum_{i=1}^{n}\Delta_i h(z_i)$, which implies that $Qh(x)\geq\tilde{Q}h(x)$, which in turn implies that $\underline{Q}g(x)=\tilde{Q}g(x)$. □

6.C Proofs of Results in Section 6.4

Lemma 6.39. *Consider a non-empty bounded set $\mathcal{Q}$ of rate matrices and let $\underline{Q}$ be the corresponding lower transition rate operator. Then for all $Q\in\mathcal{Q}$, we have that $\|Q\|\leq\|\underline{Q}\|$.*

Proof. Consider any $f\in\mathcal{L}(\mathcal{X})$ such that $\|f\|=1$. It then follows from Definition 6.3$_{267}$ that $Qf\geq\underline{Q}f$ and, due to the linearity of Q, also that $Qf=-Q(-f)\leq-\underline{Q}(-f)$. Hence, we find that $\underline{Q}f\leq Qf\leq-\underline{Q}(-f)$, which implies that

$$\|Qf\|\leq\max\{\|\underline{Q}f\|,\|-\underline{Q}(-f)\|\}=\max\{\|\underline{Q}f\|,\|\underline{Q}(-f)\|\}. \quad (6.52)$$

Since $\|f\|=1$, it follows from Property N11$_{64}$ that $\|\underline{Q}f\|\leq\|\underline{Q}\|$, and similarly, since $\|-f\|=\|f\|=1$, we also find that $\|\underline{Q}(-f)\|\leq\|\underline{Q}\|$. By combining this with Equation (6.52), we find that $\|Qf\|\leq\|\underline{Q}\|$, and since this is true for every $f\in\mathcal{L}(\mathcal{X})$ such that $\|f\|=1$, it now follows from Equation (2.6)$_{63}$ that $\|Q\|\leq\|\underline{Q}\|$. □

Lemma 6.40. *Consider a non-empty and bounded set $\mathcal{Q}$ of rate matrices, let $\underline{Q}$ be the corresponding lower transition rate operator, and consider any $\delta\in\mathbb{R}_{>0}$ such that $\delta\|\underline{Q}\|\leq 1$. Now fix any $n\in\mathbb{Z}_{>0}$ and, for all $i\in\{1,\dots,n\}$, consider some $0\leq\Delta_i\leq\delta$ and $Q_i\in\mathcal{Q}$. Then for any $f\in\mathcal{L}(\mathcal{X})$, it holds that*

$$\prod_{i=1}^{n}(I+\Delta_i Q_i)f\geq\prod_{i=1}^{n}(I+\Delta_i\underline{Q})f.$$

Proof. We provide a proof by induction. For $n=1$, the result follows trivially from Definition 6.3$_{267}$. Consider now any $n>1$ and assume that the result is true for $n-1$. Since Lemma 6.39 implies that $\Delta_1\|Q_1\|\leq\Delta_1\|\underline{Q}\|\leq\delta\|\underline{Q}\|\leq 1$, it then follows from Proposition 4.9$_{153}$ that $I+\Delta_1 Q_1$ is a transition matrix, and therefore, as noted in Section 3.4$_{116}$, also a lower transition operator, which therefore satisfies Property LT5$_{117}$. We now find that

$$\prod_{i=1}^{n}(I+\Delta_i Q_i)f = (I+\Delta_1 Q_1)\prod_{i=2}^{n}(I+\Delta_i Q_i)f$$
$$\geq (I+\Delta_1 Q_1)\prod_{i=2}^{n}(I+\Delta_i \underline{Q})f \geq \prod_{i=1}^{n}(I+\Delta_i \underline{Q})f,$$

where the first inequality follows from the induction hypothesis and Property LT5$_{117}$, and the second inequality follows from Definition 6.3$_{267}$. □

Proof of Proposition 6.22$_{279}$. Fix any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, and let (T^r_{q,x_v}) be its corresponding family of history-dependent transition matrices. By Propositions 4.2$_{149}$ and 6.17$_{273}$, if $t = s$ it holds that $T^t_{t,x_u} f = If = e^{\underline{Q}0} f$, whence the result is then immediate by Proposition 4.3$_{149}$. Hence, for the remainder of this proof, let us assume that $t < s$. Fix any $\varepsilon > 0$ and let $C := (s-t)$. Choose any $\varepsilon_1 > 0$ such that $\varepsilon_1 \|f\| < \varepsilon/2$ and any $\varepsilon_2 > 0$ such that $\varepsilon_2 C \|f\| < \varepsilon/2$; this is clearly always possible.

Due to Theorem 6.16$_{272}$, there is some $\delta \in \mathbb{R}_{>0}$ such that $\delta \|\underline{Q}\| \leq 1$ and

$$(\forall v \in \mathscr{U}_{[t,s]} : \sigma(v) \leq \delta) \left\| e^{\underline{Q}(s-t)} - \Phi_v \right\| \leq \varepsilon_1, \tag{6.53}$$

with Φ_v as in Equation (6.6)$_{271}$. Since $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, it now follows from Proposition 4.23$_{171}$ that there is some $0 < \Delta_1 < \min\{\delta, C\}$ such that

$$(\exists Q_1 \in \overline{\partial}_+ T^t_{t,x_u} \subseteq \mathscr{Q}) \left\| T^{t_1}_{t_0,x_u} - (I+\Delta_1 Q_1) \right\| = \left\| T^{t+\Delta_1}_{t,x_u} - (I+\Delta_1 Q_1) \right\| < \Delta_1 \varepsilon_2,$$

with $t_0 := t$ and $t_1 := t + \Delta_1$. Furthermore, since $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, and because $\Delta_1 < C$ implies that $t_1 = t + \Delta_1 < s$, it follows from Lemma 5.36$_{227}$ that there is some $v \in \mathscr{U}_{[t_1,s]}$ such that $\sigma(v) < \delta$, with $v = t_1, \ldots, t_n$ and $t_n = s$, and such that for all $i \in \{2, \ldots, n\}$, with $\Delta^v_i := t_i - t_{i-1}$, it holds that

$$(\exists Q_i \in \mathscr{Q}) \left\| T^{t_i}_{t_{i-1},x_{u\cup\{t\}}} - (I+\Delta^v_i Q_i) \right\| < \Delta_i \varepsilon_2.$$

For notational convenience, let $\Delta^v_1 := \Delta_1$. Since $\Delta^v_1 = \Delta_1 < \delta$ and, for all $i \in \{2, \ldots, n\}$, $\Delta^v_i \leq \sigma(v) < \delta$, we know that, for all $i \in \{1, \ldots, n\}$, $\Delta^v_i < \delta$ and therefore also, using Lemma 6.39, that $\Delta^v_i \|Q_i\| \leq \delta \|\underline{Q}\| \leq 1$. Therefore, we find that

$$\begin{aligned}
&\left| T^s_{t,x_u} f(x_t) - \left(\prod_{i=1}^n (I + \Delta^v_i Q_i) \right) f(x_t) \right| \\
&\qquad = \left| \left(T^{t_1}_{t_0,x_u} \prod_{i=2}^n T^{t_i}_{t_{i-1},x_{u\cup\{t\}}} \right) f(x_t) - \left(\prod_{i=1}^n (I + \Delta^v_i Q_i) \right) f(x_t) \right| \\
&\qquad \leq \left\| T^{t_1}_{t_0,x_u} \prod_{i=2}^n T^{t_i}_{t_{i-1},x_{u\cup\{t\}}} - \prod_{i=1}^n (I + \Delta^v_i Q_i) \right\| \|f\| \\
&\qquad \leq \left\| T^{t_1}_{t_0,x_u} - (I + \Delta^v_1 Q_1) \right\| \|f\| + \sum_{i=2}^n \left\| T^{t_i}_{t_{i-1},x_{u\cup\{t\}}} - (I + \Delta^v_i Q_i) \right\| \|f\| \\
&\qquad < \sum_{i=1}^n \Delta^v_i \varepsilon_2 \|f\| = C\varepsilon_2 \|f\| < \frac{\varepsilon}{2},
\end{aligned}$$

where the equality follows from Lemma 5.37_{230}, the first inequality follows from the definition of the norm together with Property N11_{64}, and the second inequality follows from Lemma B.5_{393} and Proposition 4.9_{153} together with the fact that $\Delta^v_i \|Q_i\| \leq 1$ for all $i \in \{1, \ldots, n\}$.

Moreover, using Equation $(6.53)_{\curvearrowleft}$, since $\sigma(v) < \delta$, we find that

$$\begin{aligned}
\left| e^{\underline{Q}(s-t)} f(x_t) - \left(\prod_{i=1}^n (I + \Delta^v_i \underline{Q}) \right) f(x_t) \right| &\leq \left\| e^{\underline{Q}(s-t)} - \prod_{i=1}^n (I + \Delta^v_i \underline{Q}) \right\| \|f\| \\
&\leq \varepsilon_1 \|f\| < \frac{\varepsilon}{2}.
\end{aligned}$$

Hence, Lemma 6.40_{324} implies that

$$\begin{aligned}
e^{\underline{Q}(s-t)} f(x_t) &< \left(\prod_{i=1}^n (I + \Delta^v_i \underline{Q}) \right) f(x_t) + \frac{\varepsilon}{2} \\
&\leq \left(\prod_{i=1}^n (I + \Delta^v_i Q_i) \right) f(x_t) + \frac{\varepsilon}{2} < T^s_{t,x_u} f(x_t) + \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this implies that $e^{\underline{Q}(s-t)} f(x_t) \leq T^s_{t,x_u} f(x_t)$. The result now follows because $T^s_{t,x_u} f(x_t) = \mathbb{E}_P[f(X_s) | X_t = x_t, X_u = x_u]$ due to Proposition 4.3_{149}. □

Lemma 6.41. *Let $\mathscr{Q}$ be a non-empty and bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon \in \mathbb{R}_{>0}$, there is some $Q \in \mathscr{Q}$ such that*

$$\left\| \underline{Q} f - Q f \right\| < \varepsilon.$$

Proof. Fix any $x \in \mathscr{X}$. Because $\mathscr{Q}$ is non-empty and bounded, it follows from Definition 6.3$_{267}$ and Lemma 6.8$_{266}$ that $\underline{Q}f(x) = \inf_{Q\in\mathscr{Q}} Qf(x) \in \mathbb{R}$, which implies that there is some $Q_x \in \mathscr{Q}$ such that $\left|\underline{Q}f(x) - Q_x f(x)\right| < \varepsilon$. Now consider the rate matrix Q that is defined, for all $x \in \mathscr{X}$, as $Q(x,\cdot) := Q_x(x,\cdot)$. Then $Q \in \mathscr{Q}$ because $\mathscr{Q}$ has separately specified rows, and it holds that $Qf(x) = Q_x f(x)$ for all $x \in \mathscr{X}$ from the properties of matrix-vector multiplication. Hence, from the definition of the norm, it now follows that

$$\left\|\underline{Q}f - Qf\right\| = \max_{x\in\mathscr{X}} \left|\underline{Q}f(x) - Qf(x)\right| = \max_{x\in\mathscr{X}} \left|\underline{Q}f(x) - Q_x f(x)\right| < \varepsilon\,,$$

which concludes the proof. $\square$

Proof of Proposition 6.23$_{280}$. For any $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ with corresponding transition matrix T_t^s, by Corollary 4.4$_{150}$ and Proposition 6.17$_{273}$, if $t = s$ it holds that $T_t^t f = If = e^{\underline{Q}0} f$, whence the result is then immediate by Corollary 4.4$_{150}$. So, for the remainder of this proof, let us suppose that $t < s$.

Let $C := (s-t)$, choose any $\varepsilon_* > 0$ such that $\varepsilon_* C < \varepsilon$, choose any $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ such that $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 < \varepsilon_*$, choose any $\delta > 0$ such that $\delta \left\|\underline{Q}\right\|^2 \|f\| < \varepsilon_1$ and $\delta \|\mathscr{Q}\|^2 \|f\| < \varepsilon_3$ (this is always possible since $\mathscr{Q}$ is bounded and due to Property LR5$_{266}$), and consider any $u \in \mathscr{U}_{[t,s]}$ such that $\sigma(u) < \delta$, with $u = t_0, \ldots, t_n$ and $n \in \mathbb{Z}_{>0}$.

Now fix any $i \in \{1, \ldots, n\}$ and let $g_i := e^{\underline{Q}(t_n - t_i)} f$. It then follows from Lemma 6.41 that there is some $Q_i \in \mathscr{Q}$ such that $\left\|\underline{Q}g_i - Q_i g_i\right\| < \varepsilon_2$ and, due to Properties N11$_{64}$ and LT4$_{117}$, we also know that $\|g_i\| = \|e^{\underline{Q}(t_n-t_i)} f\| \le \|e^{\underline{Q}(t_n-t_i)}\| \, \|f\| \le \|f\|$. Hence, we find that

$$\begin{aligned}
&\left\|e^{\underline{Q}\Delta_i^u} g_i - e^{Q_i\Delta_i^u} g_i\right\| \\
&\le \left\|e^{\underline{Q}\Delta_i^u} g_i - \left(I + \Delta_i^u \underline{Q}\right) g_i\right\| + \left\|\Delta_i^u \underline{Q} g_i - \Delta_i^u Q_i g_i\right\| + \left\|\left(I + \Delta_i^u Q_i\right) g_i - e^{Q_i\Delta_i^u} g_i\right\| \\
&\le \left\|e^{\underline{Q}\Delta_i^u} - \left(I + \Delta_i^u \underline{Q}\right)\right\| \|g_i\| + \Delta_i^u \left\|\underline{Q} g_i - Q_i g_i\right\| + \left\|I + \Delta_i^u Q_i - e^{Q_i\Delta_i^u}\right\| \|g_i\| \\
&< \left\|e^{\underline{Q}\Delta_i^u} - \left(I + \Delta_i^u \underline{Q}\right)\right\| \|f\| + \Delta_i^u \varepsilon_2 + \left\|I + \Delta_i^u Q_i - e^{Q_i\Delta_i^u}\right\| \|f\| \\
&\le \left(\Delta_i^u\right)^2 \left\|\underline{Q}\right\|^2 \|f\| + \Delta_i^u \varepsilon_2 + \left(\Delta_i^u\right)^2 \|Q_i\|^2 \|f\| \\
&\le \Delta_i^u \big(\delta \left\|\underline{Q}\right\|^2 \|f\| + \varepsilon_2 + \delta \|\mathscr{Q}\|^2 \|f\|\big) < \Delta_i^u (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) < \Delta_i^u \varepsilon_*,
\end{aligned} \tag{6.54}$$

where the second inequality holds because of Property N11$_{64}$, the third inequality holds because $\|g_i\| \le \|f\|$ and $\left\|\underline{Q}g_i - Q_i g_i\right\| < \varepsilon_2$, the fourth inequality holds because of Lemmas B.7$_{394}$ and B.8$_{394}$ and where the fifth inequality holds because $\Delta_i^u \le \sigma(u) < \delta$.

Let Q_0 and Q_{n+1} be two arbitrary elements of $\mathscr{Q}$ and, for all $i \in \{0,\dots,n+1\}$, let $\left(e^{Q_i(q-r)}\right)$ denote the exponential transition matrix system corresponding to Q_i, as in Definition 4.8_{158}. Then, by Proposition 5.10_{192}, there is some $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ with transition matrix system $\left(T_r^q\right)$ given by

$$\left(e^{Q_0(q-r)}\right)_{[0,t_0]} \otimes \left(e^{Q_1(q-r)}\right)_{[t_0,t_1]} \otimes \cdots \otimes \left(e^{Q_n(q-r)}\right)_{[t_{n-1},t_n]} \otimes \left(e^{Q_{n+1}(q-r)}\right)_{[t_n,+\infty)}.$$

For all $i \in \{i,\dots,n\}$, due to Equation $(6.54)_{\curvearrowleft}$ we know that the transition matrices of this process P satisfy

$$\left\|e^{\underline{Q}\Delta_i^u} g_i - T_{t_{i-1}}^{t_i} g_i\right\| = \left\|e^{\underline{Q}\Delta_i^u} g_i - e^{Q_i\Delta_i^u} g_i\right\| < \Delta_i^u \varepsilon^*. \tag{6.55}$$

Furthermore, we also know that

$$\begin{aligned}
&\left\|e^{\underline{Q}(t_n-t_0)} g_n - T_{t_0}^{t_n} g_n\right\| \\
&= \left\|e^{\underline{Q}(t_{n-1}-t_0)} e^{\underline{Q}\Delta_n^u} g_n - T_{t_0}^{t_{n-1}} e^{\underline{Q}\Delta_n^u} g_n + T_{t_0}^{t_{n-1}} \left(e^{\underline{Q}\Delta_n^u} g_n - T_{t_{n-1}}^{t_n} g_n\right)\right\| \\
&\leq \left\|e^{\underline{Q}(t_{n-1}-t_0)} e^{\underline{Q}\Delta_n^u} g_n - T_{t_0}^{t_{n-1}} e^{\underline{Q}\Delta_n^u} g_n\right\| + \left\|T_{t_0}^{t_{n-1}}\right\| \left\|e^{\underline{Q}\Delta_n^u} g_n - T_{t_{n-1}}^{t_n} g_n\right\| \\
&\leq \left\|e^{\underline{Q}(t_{n-1}-t_0)} g_{n-1} - T_{t_0}^{t_{n-1}} g_{n-1}\right\| + \left\|e^{\underline{Q}\Delta_n^u} g_n - T_{t_{n-1}}^{t_n} g_n\right\|,
\end{aligned}$$

using Proposition 6.17_{273} and Equation $(5.1)_{183}$ for the first equality, Property $\mathrm{N11}_{64}$ for the first inequality, and Property $\mathrm{LT4}_{117}$ and the definition of g_{n-1} for the second inequality. Similarly, we also find that

$$\begin{aligned}
\left\|e^{\underline{Q}(t_{n-1}-t_0)} g_{n-1} - T_{t_0}^{t_{n-1}} g_{n-1}\right\| \leq{}& \left\|e^{\underline{Q}(t_{n-2}-t_0)} g_{n-2} - T_{t_0}^{t_{n-2}} g_{n-2}\right\| \\
&+ \left\|e^{\underline{Q}\Delta_{n-1}^u} g_{n-1} - T_{t_{n-2}}^{t_{n-1}} g_{n-1}\right\|.
\end{aligned}$$

By continuing in this way (essentially applying backwards induction) we eventually find, using that $t = t_0$, $s = t_n$, and $g_n = f$ due to Proposition 6.17_{273}, that

$$\begin{aligned}
\left\|e^{\underline{Q}(s-t)} f - T_t^s f\right\| &= \left\|e^{\underline{Q}(t_n-t_0)} g_n - T_{t_0}^{t_n} g_n\right\| \\
&\leq \sum_{i=1}^n \left\|e^{\underline{Q}\Delta_i^u} g_i - T_{t_{i-1}}^{t_i} g_i\right\| \leq \sum_{i=1}^n \Delta_i^u \varepsilon^* = C\varepsilon^* < \varepsilon,
\end{aligned}$$

using Equation (6.55) to establish the second inequality. The result now follows from Corollary 4.4_{150}, which states that $T_t^s f(x_t) = \mathbb{E}_P[f(X_s) | X_t = x_t]$ for all $x_t \in \mathscr{X}$. □

Proof of Corollary 6.24_{280}. Consider any sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ such that $\lim_{i\to+\infty} \varepsilon_i = 0$. Because $\mathscr{Q}$ is compact, it is bounded by Corollary $\mathrm{A.12}_{378}$. Because $\mathscr{Q}$ is also non-empty and has separately specified rows, it follows from Proposition 6.23₂₈₀ that for all $i \in \mathbb{Z}_{>0}$ there is some $P_i \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ such that

$$\left|\mathbb{E}_{P_i}[f(X_s)\,|\,X_t = x_t] - e^{\underline{Q}(s-t)} f(x_t)\right| < \varepsilon_i \;\text{ for all } x_t \in \mathscr{X}\,.$$

For notational convenience, for all $i \in \mathbb{Z}_{>0}$, let $T_i := {}^{P_i}T_t^s$ denote the transition matrix corresponding to P_i. It follows from Corollary 4.4₁₅₀ that $\mathbb{E}_{P_i}[f(X_s)\,|\,X_t = x_t] = T_i f(x_t)$ for all $x_t \in \mathscr{X}$ and $i \in \mathbb{Z}_{>0}$, and hence it holds that

$$\left\|T_i f - e^{\underline{Q}(s-t)} f\right\| = \max_{x \in \mathscr{X}} \left|T_i f(x) - e^{\underline{Q}(s-t)} f(x)\right| < \varepsilon_i \text{ for all } i \in \mathbb{Z}_{>0}\,.$$

Because $\lim_{i \to +\infty} \varepsilon_i = 0$ it follows that also $\lim_{i \to +\infty} \|T_i f - e^{\underline{Q}(s-t)} f\| = 0$.

Now consider the set $\mathscr{T} := \left\{{}^{P}T_t^s \,:\, P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}\right\}$; then for all $i \in \mathbb{Z}_{>0}$ it holds that $T_i \in \mathscr{T}$ because $P_i \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $T_i = {}^{P_i}T_t^s$. Because $\mathscr{Q}$ is non-empty, compact, and convex, it follows from Corollary 5.18₁₉₇ that $\mathscr{T}$ is compact. Due to Corollary A.12₃₇₈ this implies that $\mathscr{T}$ is sequentially compact, and because $T_i \in \mathscr{T}$ for all $i \in \mathbb{Z}_{>0}$ this implies that there is a convergent subsequence $\{T_{i_j}\}_{j \in \mathbb{Z}_{>0}}$ with limit $T_* := \lim_{j \to +\infty} T_{i_j}$ such that $T_* \in \mathscr{T}$. By Lemma A.34₃₉₀ this implies that $\lim_{i \to +\infty} \|T_i f - T_* f\| = 0$, and since we already know that $\lim_{i \to +\infty} \|T_i f - e^{\underline{Q}(s-t)} f\| = 0$, it follows that $\|T_* f - e^{\underline{Q}(s-t)} f\| = 0$. This implies that $T_* f = e^{\underline{Q}(s-t)} f$.

Because $T_* \in \mathscr{T}$, it follows that there is a well-behaved Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ such that $T_* = {}^{P}T_t^s$, and hence it follows from Corollary 4.4₁₅₀ that

$$\mathbb{E}_P[f(X_s)\,|\,X_t = x_t] = T_* f(x_t) = e^{\underline{Q}(s-t)} f(x_t) \;\text{ for all } x_t \in \mathscr{X}\,,$$

which concludes the proof. □

*Proof of Theorem 6.28*₂₈₃. Clearly, it suffices to prove that for any $P \in \mathbb{P}$ that is not well-behaved or not consistent with $\mathscr{Q}_{\underline{Q}}$, there are $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$, $x_t \in \mathscr{X}$ and $f \in \mathscr{L}(\mathscr{X})$ such that

$$\mathbb{E}_P[f(X_s)\,|\,X_t = x_t, X_u = x_u] < e^{\underline{Q}(s-t)} f(x_t). \tag{6.56}$$

We start with the case that P is not well-behaved. Fix any $\varepsilon > 0$ and let $C := \left\|\underline{Q}\right\| + \varepsilon$. It then follows from Proposition 4.2₁₄₉ that there are $t, s \in \mathbb{R}_{\geq 0}$ with $t < s$, $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ such that, with $\Delta := s - t > 0$, it holds that $1/\Delta \left\|T_{t,x_u}^s - I\right\| > C$ and $\Delta \left\|\underline{Q}\right\|^2 < \varepsilon$. Let $Q := 1/\Delta(T_{t,x_u}^s - I)$. Since $\|Q\| > C$, it then follows that there is some $f' \in \mathscr{L}(\mathscr{X})$ such that $\|f'\| = 1$ and $\|Qf'\| > C$, which in turn implies that there is some $x_t \in \mathscr{X}$ such that $|Qf'(x_t)| > C$. If $Qf'(x_t) < 0$, we let $f := f'$, and if $Qf'(x_t) > 0$, we let $f := -f'$. Clearly, this implies that $\|f\| = 1$ and $Qf(x_t) < -C$. From $\|f\| =$

1, it furthermore follows that $\underline{Q}f(x_t) \geq -\left\|\underline{Q}f\right\| \geq -\left\|\underline{Q}\right\|$, and therefore, we find that $Qf(x_t) < -\left\|\underline{Q}\right\| - \varepsilon \leq \underline{Q}f(x_t) - \varepsilon$, which implies that $(I+\Delta Q)f(x_t) \leq (I+\Delta\underline{Q})f(x_t) - \Delta\varepsilon$.

Moreover, we also know that

$$\left|e^{\underline{Q}(s-t)}f(x_t) - (I+\Delta\underline{Q})f(x_t)\right| \leq \left\|e^{\underline{Q}(s-t)} - (I+\Delta\underline{Q})\right\| \leq \Delta^2\left\|\underline{Q}\right\|^2 < \Delta\varepsilon,$$

where we used $\|f\| = 1$ for the first inequality and Lemma B.7$_{394}$ for the second inequality. Hence, it follows that

$$T^s_{t,x_u}f(x_t) = (I+\Delta Q)f(x_t) \leq (I+\Delta\underline{Q})f(x_t) - \Delta\varepsilon < e^{\underline{Q}(s-t)}f(x_t),$$

which, because of Proposition 4.3$_{149}$, implies that Equation (6.56)$_\curvearrowleft$ holds.

Next, we consider the case that P is well-behaved, but not consistent with $\mathscr{Q}_{\underline{Q}}$. In that case, it follows from Definition 5.3$_{189}$ that there are $r \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<r}$ and $x_u \in \mathscr{X}_u$ such that $\overline{\partial}T^r_{r,x_u} \nsubseteq \mathscr{Q}_{\underline{Q}}$. Since we know that $\overline{\partial}T^r_{r,x_u}$ is a non-empty set of rate matrices because of Proposition 4.22$_{169}$, this implies the existence of a rate matrix $Q_* \in \overline{\partial}T^r_{r,x_u}$ such that $Q_* \notin \mathscr{Q}_{\underline{Q}}$. Furthermore, since $Q_* \notin \mathscr{Q}_{\underline{Q}}$, Equation (6.4)$_{268}$ implies that there are $f' \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$ such that $Q_*f'(x) < \underline{Q}f'(x)$. Clearly, this implies that $f' \neq 0$, and therefore, that $\|f'\| > 0$. If we now let $f := {}^1\!/\!_{\|f'\|}f'$, then $\|f\| = 1$, and furthermore, because of the linearity of Q_* and the non-negative homogeneity of $\underline{Q}$, it follows that

$$Q_*f(x) = {}^1\!/\!_{\|f'\|}Q_*f'(x) < {}^1\!/\!_{\|f'\|}\underline{Q}f'(x) = \underline{Q}f(x)\,. \tag{6.57}$$

Consider now any $\varepsilon > 0$ such that $Q_*f(x) \leq \underline{Q}f(x) - 2\varepsilon$; this is clearly possible due to the strict inequality in (6.57). Since $Q_* \in \overline{\partial}T^r_{r,x_u}$, it then follows from Definition 4.11$_{168}$ that there are $t, s \in \mathbb{R}_{\geq 0}$ such that $u < t < s$ and, with $\Delta := s - t > 0$, $\left\|{}^1\!/\!_\Delta(T^s_{t,x_u} - I) - Q_*\right\| \leq \varepsilon$ and $\Delta\left\|\underline{Q}\right\|^2 < \varepsilon$. Let $Q := {}^1\!/\!_\Delta(T^s_{t,x_u} - I)$. Since $\|Q - Q_*\| \leq \varepsilon$ and $\|f\| = 1$, it then follows that $Qf(x) \leq Q_*f(x) + \varepsilon \leq \underline{Q}f(x) - \varepsilon$. The remainder of the argument is now analogous to the argument in the first part of this proof: first, it follows from this inequality that $(I+\Delta Q)f(x) \leq (I+\Delta\underline{Q})f(x) - \Delta\varepsilon$. Moreover, as before, we also know that

$$\left|e^{\underline{Q}(s-t)}f(x) - (I+\Delta\underline{Q})f(x)\right| \leq \left\|e^{\underline{Q}(s-t)} - (I+\Delta\underline{Q})\right\| \leq \Delta^2\left\|\underline{Q}\right\|^2 < \Delta\varepsilon,$$

where we used $\|f\| = 1$ for the first inequality and Lemma B.7$_{394}$ for the second inequality. Hence we find that

$$T^s_{t,x_u}f(x) = (I+\Delta Q)f(x) \leq (I+\Delta\underline{Q})f(x) - \Delta\varepsilon < e^{\underline{Q}(s-t)}f(x),$$

which, using Proposition 4.3$_{149}$, implies that Equation (6.56)$_\curvearrowleft$ holds. □

6.D Proofs of Results in Section 6.5

Proof of Lemma 6.29$_{284}$. If $n=0$ the result follows trivially from Equation (6.16)$_{285}$, so let us assume for the remainder of this proof that $n\geq 1$.

It follows from Equation (6.15)$_{284}$ that, because $g_n = f$,

$$\left|g_{n-1}(x_{w_n}) - \underline{T}_{t_{n-1}}^{t_n} f(x_{w_n})\right| \leq \varepsilon \quad \text{for all } x_{w_n} \in \mathscr{X}_{w_n}\,.$$

This provides the induction base (for $i=n$) in the following induction argument. Suppose that for some $i\in\{2,\dots,n\}$, it holds that

$$\left|g_{i-1}(x_{w_i}) - \underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_i})\right| \leq (n-i+1)\varepsilon \quad \text{for all } x_{w_i} \in \mathscr{X}_{w_i}\,.$$

We will show that then also, for all $x_{w_{i-1}}\in\mathscr{X}_{w_{i-1}}$,

$$\left|g_{i-2}(x_{w_{i-1}}) - \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}})\right| \leq (n-i+2)\varepsilon\,.$$

The induction hypothesis implies that, for all $x_{w_i}\in\mathscr{X}_{w_i}$, it holds that

$$\begin{aligned}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_i}) - (n-i+1)\varepsilon &\leq g_{i-1}(x_{w_i})\\ &\leq \underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_i}) + (n-i+1)\varepsilon\,.\end{aligned}$$

Because $w_i = u\cup\{t_0,\dots,t_{i-1}\}$ and $w_{i-1}=u\cup\{t_0,\dots,t_{i-2}\}$, this implies that for all $x_{w_{i-1}}\in\mathscr{X}_{w_{i-1}}$,

$$\begin{aligned}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot) - (n-i+1)\varepsilon & \qquad (6.58)\\ \leq g_{i-1}(x_{w_{i-1}},\cdot)&\\ \leq \underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot) + (n-i+1)\varepsilon\,, & \qquad (6.59)\end{aligned}$$

where we consider the restrictions (depending on $x_{w_{i-1}}$) of these functions to $\mathscr{L}(\mathscr{X}_{t_{i-1}})$, e.g. $g_{i-1}(x_{w_{i-1}},\cdot)$ is the element corresponding to the t_{i-1}-measurable function $g_{i-1}(x_{w_{i-1}},X_{t_{i-1}})$. So now fix any $x_{w_{i-1}}\in\mathscr{X}_{w_{i-1}}$. Then, using the notation introduced in Section 6.1$_{260}$, and because $\underline{T}_{t_{i-2}}^{t_{i-1}}$ is a lower transition operator, it follows from Proposition 3.32$_{116}$ together with Equation (6.59) that

$$\begin{aligned}\underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) &= \left[\underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}},\cdot)\right](x_{t_{i-2}})\\ &\leq \underline{T}_{t_{i-2}}^{t_{i-1}}\left(\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot) + (n-i+1)\varepsilon\right)(x_{t_{i-2}})\\ &= \left[\underline{T}_{t_{i-2}}^{t_{i-1}}\left(\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot)\right)\right](x_{t_{i-2}}) + (n-i+1)\varepsilon\\ &= \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) + (n-i+1)\varepsilon\,,\end{aligned}$$

and similarly, but using Equation (6.58)$_{\curvearrowleft}$, that

$$\begin{aligned}\underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) &= \left[\underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}},\cdot)\right](x_{t_{i-2}}) \\ &\geq \underline{T}_{t_{i-2}}^{t_{i-1}}\left(\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot) - (n-i+1)\varepsilon\right)(x_{t_{i-2}}) \\ &= \left[\underline{T}_{t_{i-2}}^{t_{i-1}}\left(\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}},\cdot)\right)\right](x_{t_{i-2}}) - (n-i+1)\varepsilon \\ &= \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) - (n-i+1)\varepsilon\,.\end{aligned}$$

Now, Equation (6.15)$_{284}$ implies that

$$\underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) - \varepsilon \leq g_{i-2}(x_{w_{i-1}}) \leq \underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) + \varepsilon\,,$$

and hence, by combining these inequalities, it follows that

$$\begin{aligned}g_{i-2}(x_{w_{i-1}}) &\leq \underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) + \varepsilon \\ &\leq \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) + (n-i+1)\varepsilon + \varepsilon \\ &= \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) + (n-i+2)\varepsilon\,,\end{aligned}$$

and, similarly,

$$\begin{aligned}g_{i-2}(x_{w_{i-1}}) &\geq \underline{T}_{t_{i-2}}^{t_{i-1}} g_{i-1}(x_{w_{i-1}}) - \varepsilon \\ &\geq \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) - (n-i+1)\varepsilon - \varepsilon \\ &= \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}}) - (n-i+2)\varepsilon\,,\end{aligned}$$

which implies that

$$\left|g_{i-2}(x_{w_{i-1}}) - \underline{T}_{t_{i-2}}^{t_{i-1}}\underline{T}_{t_{i-1}}^{t_i}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_{i-1}})\right| \leq (n-i+2)\varepsilon\,.$$

Because this is true for all $x_{w_{i-1}} \in \mathscr{X}_{w_{i-1}}$, this concludes the proof of the induction step. If $n \geq 2$ then the induction argument now implies that, in particular, with $i = 2$, it holds that

$$\left|g_0(x_{w_1}) - \underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_1})\right| \leq n\varepsilon\,, \tag{6.60}$$

for all $x_{w_1} \in \mathscr{X}_{w_1}$. Conversely, if $n = 1$ then this inequality was already established by the induction base; hence this inequality holds in all cases that we consider.

It remains to repeat the above argument one last time to resolve the operator $\underline{T}_{\max u}^{t_0}$. To this end, Equation (6.60) implies that, for all $x_{w_1} \in \mathscr{X}_{w_1}$,

$$\underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_1}) - n\varepsilon \leq g_0(x_{w_1}) \leq \underline{T}_{t_0}^{t_1}\cdots\underline{T}_{t_{n-1}}^{t_n} f(x_{w_1}) + n\varepsilon\,.$$

Because $w_1 = u \cup \{t_0\}$, this implies that for all $x_u \in \mathscr{X}_u$,

$$\underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u, \cdot) - n\varepsilon \leq g_0(x_u, \cdot) \leq \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u, \cdot) + n\varepsilon\,, \tag{6.61}$$

where we consider the restrictions (depending on x_u) of these functions to $\mathscr{L}(\mathscr{X}_{t_0})$, e.g. $g_0(x_u, \cdot)$ is the element corresponding to the t_0-measurable function $g_0(x_u, X_{t_0})$. So now fix any $x_u \in \mathscr{X}_u$. Then, using the notation introduced in Section 6.1$_{260}$, and because $\underline{T}_{\max u}^{t_0}$ is a lower transition operator, it follows from Proposition 3.32$_{116}$ together with Equation (6.61) that

$$\begin{aligned}
\underline{T}_{\max u}^{t_0} g_0(x_u) &= \left[\underline{T}_{\max u}^{t_0} g_0(x_u, \cdot)\right](x_{\max u}) \\
&\leq \underline{T}_{\max u}^{t_0}\left(\left[\underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f\right](x_u, \cdot) + n\varepsilon\right)(x_{\max u}) \\
&= \left[\underline{T}_{\max u}^{t_0}\left(\underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u, \cdot)\right)\right](x_{\max u}) + n\varepsilon \\
&= \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u) + n\varepsilon\,,
\end{aligned}$$

and

$$\begin{aligned}
\underline{T}_{\max u}^{t_0} g_0(x_u) &= \left[\underline{T}_{\max u}^{t_0} g_0(x_u, \cdot)\right](x_{\max u}) \\
&\geq \underline{T}_{\max u}^{t_0}\left(\left[\underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f\right](x_u, \cdot) - n\varepsilon\right)(x_{\max u}) \\
&= \left[\underline{T}_{\max u}^{t_0}\left(\underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u, \cdot)\right)\right](x_{\max u}) - n\varepsilon \\
&= \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u) - n\varepsilon\,.
\end{aligned}$$

Using Equation (6.16)$_{285}$, it follows that

$$\tilde{f}(x_u) \leq \underline{T}_{\max u}^{t_0} g_0(x_u) + \varepsilon \leq \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u) + (n+1)\varepsilon\,,$$

and

$$\tilde{f}(x_u) \geq \underline{T}_{\max u}^{t_0} g_0(x_u) - \varepsilon \geq \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u) - (n+1)\varepsilon\,,$$

from which we obtain

$$\left|\tilde{f}(x_u) - \underline{T}_{\max u}^{t_0} \underline{T}_{t_0}^{t_1} \cdots \underline{T}_{t_{n-1}}^{t_n} f(x_u)\right| \leq (n+1)\varepsilon\,.$$

Because this is true for all $x_u \in \mathscr{X}_u$, this concludes the proof. □

7

Reduction to Discrete-Time Imprecise-Markov Chains

"You're just about drowning me with talk.
What're you getting at? Never mind the in-betweens."

"You've got *to have the in-betweens, or you won't understand."*

Isaac Asimov, "Foundation"

In this final technical chapter, we will establish a strong connection between imprecise-Markov chains in discrete- and continuous time. We will show in Section 7.1$_{\curvearrowright}$ how the parameters $\mathscr{Q}$ and $\mathscr{M}$ of a continuous-time imprecise-Markov chain can induce a discrete-time imprecise-Markov chain with any desired discrete time domain $\mathbb{D}$. We make this identification in such a way that the lower transition operators of this discrete-time model are given by the imprecise exponentials $e^{\underline{Q}t}$ of the lower transition rate operator $\underline{Q}$ corresponding to $\mathscr{Q}$, which means that we can evaluate these lower transition operators numerically using, e.g., Algorithm 1$_{277}$.

The practical use of this construction is illustrated in Section 7.2$_{338}$, where we show that—under some conditions on $\mathscr{Q}$—the lower expectation of a function that only depends on time points in $\mathbb{D}$, is the same for the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and for the induced discrete-time imprecise-Markov chain on $\mathbb{D}$. The immediate practical consequence is this: provided that we have an algorithm to compute this lower expectation for the induced discrete-time model, we can also compute the lower expectation for the continuous-time model. This observation allows us to leverage any number of such algorithms from the literature, which at the time of writing is much more extensive for discrete-time imprecise-Markov chains than for continuous-time imprecise-Markov chains. Moreover, it similarly allows us to exploit existing algorithms for imprecise-probabilistic graphical models, to which—as we already noted in Chapter 383—discrete-time imprecise-Markov chains are related. The results from this chapter can therefore also be interpreted as establishing an analogous connection between continuous-time imprecise-Markov chains and imprecise-probabilistic graphical models, by restricting attention to a finite—or at least countable—number of time points of interest. We illustrate this with an example, where we straightforwardly adopt an efficient algorithm for computing lower expectations of functions satisfying a particular decomposition property, for use with continuous-time imprecise-Markov chains.

We conclude this chapter with Section 7.3345, where—again under some conditions on $\mathscr{Q}$—we provide the induced discrete-time imprecise-Markov chain with an alternative characterisation in terms of sets of stochastic processes. We show that in a precise sense, we can view this induced discrete-time model as the set of restrictions of the elements of the continuous-time imprecise-Markov chain, to the events $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ with time domain $\mathbb{D}$. This provides the connection between these two frameworks also in terms of the sets of processes that constitute our models, in addition to the connection in terms of the corresponding lower expectations discussed earlier.

7.1 Induced Discrete-Time Imprecise-Markov Chains

In this section we show how the parameters $\mathscr{Q}$ and $\mathscr{M}$ of a continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ also induce a discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ on any desired discrete time domain $\mathbb{D}$; specifically in such a way that their lower expectations will agree in a useful sense.

The crucial step is to note that the (generalised) exponentials $e^{\underline{Q}t}$ of the lower transition rate operator $\underline{Q}$ corresponding to $\mathscr{Q}$ are lower transition operators which, as we know from our developments in Section 3.4116, have a dominating set of transition matrices $\mathscr{T}_{e^{\underline{Q}t}}$ that is non-empty, closed, convex, has separately specified rows, and has $e^{\underline{Q}t}$ as its corresponding lower transition operator. Since discrete-time imprecise-Markov chains are parameterised using sets of transition matrices, this allows us to introduce the following definition.

Definition 7.1. *Let $\mathscr{Q}$ be a non-empty and bounded set of rate matrices with corresponding lower transition rate operator $\underline{Q}$, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $\mathbb{D}$ be a discrete time domain with canonical time index τ. We define the discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} := \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$ where, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k := \mathscr{T}_{e^{\underline{Q}(\tau_{k+1}-\tau_k)}}$ is the set of transition matrices that dominate the lower transition operator $e^{\underline{Q}(\tau_{k+1}-\tau_k)}$, and where $\mathscr{M}'$ is given by*

$$\mathscr{M}' := \mathscr{M}\mathscr{T}_{e^{\underline{Q}\tau_0}} = \left\{ \sum_{x\in\mathscr{X}} p(x)T(x,\cdot) \,:\, p \in \mathscr{M}, T \in \mathscr{T}_{e^{\underline{Q}\tau_0}} \right\}, \tag{7.1}$$

with $\mathscr{T}_{e^{\underline{Q}\tau_0}}$ the set of transition matrices that dominate the lower transition operator $e^{\underline{Q}\tau_0}$. We call $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ the discrete-time imprecise-Markov chain with time domain $\mathbb{D}$ induced by $\mathscr{Q}$ and $\mathscr{M}$, *and we use $\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ and $\overline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ to denote its corresponding lower and upper expectations, respectively.*

It follows from Proposition 3.37_{120} that, for all $k \in \mathbb{Z}_{\geq 0}$, the lower transition operator $\underline{T}_k$ corresponding to the set $\mathscr{T}_k = \mathscr{T}_{e^{\underline{Q}(\tau_{k+1}-\tau_k)}}$ in the parameterisation of the discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ is simply the generalised exponential $e^{\underline{Q}(\tau_{k+1}-\tau_k)}$. This implies that we can use e.g. Algorithm 1_{277} to evaluate these lower transition operators $\underline{T}_k$ numerically. By combining this with the results in Section 3.5_{121}—i.e. the fact that these lower transition operators form an alternative representation of the corresponding conditional lower expectation—this allows us to compute the conditional lower expectations for $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$. Moreover, because the lower expectation $\underline{\mathbb{E}}_{\mathscr{M}}$ with respect to the initial model $\mathscr{M}$ of the continuous-time model is solvable by assumption—see the discussion in Section 6.5_{284}—and because we can evaluate the lower transition operator $e^{\underline{Q}\tau_0}$ using e.g. Algorithm 1_{277}, it follows from the following result that we can also efficiently evaluate the lower expectation $\underline{\mathbb{E}}_{\mathscr{M}'}$ with respect to the initial model $\mathscr{M}'$ of the induced discrete-time imprecise-Markov chain, as given by Equation (7.1).

Lemma 7.1. *Let $\mathscr{Q}$ be a non-empty and bounded set of rate matrices with corresponding lower transition rate operator $\underline{Q}$, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $\mathbb{D}$ be a discrete time domain with canonical time index τ. Let $\mathscr{M}'$ be as in Equation* (7.1), *and let $\underline{\mathbb{E}}_{\mathscr{M}}$ and $\underline{\mathbb{E}}_{\mathscr{M}'}$ be as in Definition 3.14_{114}. Then for all $f \in \mathscr{L}(\mathscr{X})$ it holds that*

$$\underline{\mathbb{E}}_{\mathscr{M}'}[f] = \underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0} f\right].$$

Proof. First fix any $p' \in \mathscr{M}'$. Then by Equation (7.1) there are $p \in \mathscr{M}$ and $T \in \mathscr{T}_{e^{\underline{Q}\tau_0}}$ such that $p'(y) = \sum_{x\in\mathscr{X}} p(x)T(x,y)$ for all $y \in \mathscr{X}$. This implies that

$$\sum_{y\in\mathscr{X}} p'(y)f(y) = \sum_{y\in\mathscr{X}}\sum_{x\in\mathscr{X}} p(x)T(x,y)f(y)$$
$$= \sum_{x\in\mathscr{X}} p(x)Tf(x) \geq \sum_{x\in\mathscr{X}} p(x)e^{\underline{Q}\tau_0}f(x) \geq \underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}f\right],$$

where for the first inequality we used that $p(x) \geq 0$ for all $x \in \mathscr{X}$ because $p \in \mathscr{M}$ is a probability mass function, and that $Tf \geq e^{\underline{Q}\tau_0}f$ since $T \in \mathscr{T}_{e^{\underline{Q}\tau_0}}$; and where for the second inequality we used that $p \in \mathscr{M}$ together with Definition 3.14_{114}. Because this is true for all $p' \in \mathscr{M}'$ it follows from Definition 3.14_{114} that $\underline{\mathbb{E}}_{\mathscr{M}'}[f] \geq \underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}f\right]$.

For the other direction, note that by Propositions 3.36_{119} and 3.37_{120} there is some $T \in \mathscr{T}_{e^{\underline{Q}\tau_0}}$ such that $Tf = e^{\underline{Q}\tau_0}f$, so it follows that $\underline{\mathbb{E}}_{\mathscr{M}}[e^{\underline{Q}\tau_0}f] = \underline{\mathbb{E}}_{\mathscr{M}}[Tf]$. Now fix any $\varepsilon > 0$. By Definition 3.14_{114} it holds that $\underline{\mathbb{E}}_{\mathscr{M}}[f]$ is real-valued, and hence there is some $p \in \mathscr{M}$ such that $|\underline{\mathbb{E}}_{\mathscr{M}}[Tf] - \sum_{x\in\mathscr{X}} p(x)Tf(x)| < \varepsilon$, which implies that $\underline{\mathbb{E}}_{\mathscr{M}}[Tf] > \sum_{x\in\mathscr{X}} p(x)Tf(x) - \varepsilon$. By Equation $(7.1)_{\curvearrowleft}$ it holds that the function $p' : \mathscr{X} \to \mathbb{R}$ that is defined for all $y \in \mathscr{X}$ as $p'(y) := \sum_{x\in\mathscr{X}} p(x)T(x,y)$ is an element of $\mathscr{M}'$. This implies that

$$\underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}f\right] = \underline{\mathbb{E}}_{\mathscr{M}}\left[Tf\right] > \sum_{x\in\mathscr{X}} p(x)Tf(x) - \varepsilon$$
$$= \sum_{x\in\mathscr{X}} p(x)\sum_{y\in\mathscr{X}} T(x,y)f(y) - \varepsilon$$
$$= \sum_{y\in\mathscr{X}} p'(y)f(y) - \varepsilon \geq \underline{\mathbb{E}}_{\mathscr{M}'}[f] - \varepsilon,$$

where we used that $p' \in \mathscr{M}'$ together with Definition 3.14_{114} for the final inequality. Because this is true for all $\varepsilon > 0$ it follows that $\underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}f\right] \geq \underline{\mathbb{E}}_{\mathscr{M}'}[f]$. Since we already established the inequality in the other direction, we conclude that $\underline{\mathbb{E}}_{\mathscr{M}'}[f] = \underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}f\right]$. □

It is worth noting that in the special case that the first time point τ_0 of the discrete time domain $\mathbb{D}$ is equal to zero, then it follows from Proposition 6.17_{273} that $e^{\underline{Q}\tau_0} = e^{\underline{Q}0} = I$, and hence in that case it follows from Lemma $7.1_{\curvearrowleft}$ that then $\underline{\mathbb{E}}_{\mathscr{M}'}[f] = \underline{\mathbb{E}}_{\mathscr{M}}[f]$. Hence if $\tau_0 = 0$ we can simply use the initial model $\mathscr{M}$ of the continuous-time imprecise-Markov chain directly.

7.2 Correspondence of Lower Expectations

In this section we will establish the crucial correspondence between the lower expectations of the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and the discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ with time domain $\mathbb{D}$.

Clearly, we can only obtain such a correspondence in a meaningful way for functions and conditioning events that depend on time points that are included in $\mathbb{D}$; the following definition will therefore be helpful.

Definition 7.2. *Fix any $u \in \mathscr{U}_{\supset\emptyset}$ such that $u = t_0, \ldots, t_n$ with $n \in \mathbb{Z}_{\geq 0}$, and let $\mathbb{D}$ be a discrete time domain with canonical time index τ. We say that $\mathbb{D}$* extends *u, if $t_i = \tau_i$ for all $i \in \{0, \ldots, n\}$.*

The following result tells us that, for functions f that only depend on the state of the system at (finitely many) time points in $\mathbb{D}$, the lower expectations for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and the induced discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ coincide, provided that certain structural assumptions on $\mathscr{Q}$ are satisfied.

Theorem 7.2. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Fix any $u, v \in \mathscr{U}$ with $u < v$ and $v \neq \emptyset$, and let $\mathbb{D}$ be any discrete time domain that extends $u \cup v$. Then for all $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$ and all $x_u \in \mathscr{X}_u$ it holds that*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,|\,X_u = x_u\right] = \underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,|\,X_u = x_u\right].$$

Proof. Let τ denote the canonical time index of $\mathbb{D}$, and let $u = t_0, \ldots, t_{m-1}$ and $v = t_m, \ldots, t_n$ for some $m, n \in \mathbb{Z}_{\geq 0}$ with $m \leq n$ (with $m = 0$ in the case that $u = \emptyset$). Then because $\mathbb{D}$ extends $u \cup v$, it follows that $t_i = \tau_i$ for all $i \in \{0, \ldots, n\}$. Therefore, it follows from Definition 3.1_{85} that $(X_u = x_u)_{\mathbb{D}}$ is a situation with time domain $\mathbb{D}$. Due to Lemma 3.21_{105}, this implies that the lower expectation $\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,|\,X_u = x_u\right]$ is well-defined.

Using Definition 7.1_{337}, it holds that $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$, with $\mathscr{M}'$ as in Equation $(7.1)_{337}$ and, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k = \mathscr{T}_{e^{\underline{Q}(\tau_{k+1}-\tau_k)}}$. Due to Proposition 3.37_{120}, this implies that $\mathscr{T}_k$ is a non-empty, closed, and convex set of transition matrices that has $\underline{T}_k := e^{\underline{Q}(\tau_{k+1}-\tau_k)}$ as its corresponding lower transition operator. Now for all $k \in \mathbb{Z}_{\geq 0}$ let $\underline{T}_{\tau_k}^{\tau_{k+1}}$ denote the lower transition operator corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as in Definition 6.1_{261}. Because $\mathscr{Q}$ is compact, it is bounded by Corollary $\mathrm{A}.12_{378}$. Hence $\mathscr{Q}$ is a non-empty and bounded set of rate matrices that has separately specified rows, and therefore it follows from Proposition 6.26_{281} that for all $k \in \mathbb{Z}_{\geq 0}$ it holds that $\underline{T}_{\tau_k}^{\tau_{k+1}} = e^{\underline{Q}(\tau_{k+1}-\tau_k)} = \underline{T}_k$.

We now consider two cases. First suppose that $u \neq \emptyset$, so $m > 0$. Then, because for all $k \in \mathbb{Z}_{\geq 0}$ it holds that $\mathscr{T}_k$ is non-empty and has separately specified rows, it follows from Proposition 3.44_{123} that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,|\,X_u = x_u\right] &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}\left[f(X_{\tau_{0:n}})\,|\,X_{\tau_{0:(m-1)}} = x_{\tau_{0:(m-1)}}\right] \\ &= \underline{T}_{m-1}\underline{T}_m\cdots\underline{T}_{n-1}f(x_{\tau_{0:(m-1)}}). \qquad (7.2)\end{aligned}$$

Moreover, because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Theorem 6.4$_{263}$ that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right] = \underline{T}_{\tau_{m-1}}^{\tau_m}\underline{T}_{\tau_m}^{\tau_{m+1}}\cdots\underline{T}_{\tau_{n-1}}^{\tau_n}f(x_u)\,. \tag{7.3}$$

Using that $u = \tau_{0:(m-1)}$ and that $\underline{T}_k = \underline{T}_{\tau_k}^{\tau_{k+1}}$ for all $k \in \mathbb{Z}_{\geq 0}$, it follows from Equations (7.2)$_{\curvearrowleft}$ and (7.3) that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,|\,X_u = x_u\right] &= \underline{T}_{m-1}\underline{T}_m\cdots\underline{T}_{n-1}f(x_u)\\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right],\end{aligned}$$

which concludes the proof for the case where $u \neq \emptyset$.

So for the remaining case, suppose that $u = \emptyset$. Then $v = t_0, \dots, t_n$, and because $\mathbb{D}$ extends $u \cup v = v$, this implies that $v = \tau_{0:n}$. Now let $g \in \mathscr{L}(\mathscr{X})$ be defined as $g(x) := \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{\tau_{0:n}})\,|\,X_{\tau_0} = x\right]$ for all $x \in \mathscr{X}$. Then it follows that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right] &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{\tau_{0:n}})\right]\\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{\tau_{0:n}})\,\middle|\,X_{\tau_0}\right]\right]\\ &= \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{\tau_0})\right] = \underline{\mathbb{E}}_{\mathscr{M}}\left[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_{\tau_0})\,\middle|\,X_0\right]\right]\\ &= \underline{\mathbb{E}}_{\mathscr{M}}\left[e^{\underline{Q}\tau_0}g(X_0)\right],\end{aligned} \tag{7.4}$$

where for the second equality we used Theorem 5.32$_{208}$, which we can do because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows; for the third equality we used the definition of g; for the fourth equality we used Theorem 6.33$_{289}$, which we can do because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows; and for the final equality we used Corollary 6.25$_{280}$, which we can do because $\mathscr{Q}$ is non-empty and bounded (since it is compact, due to Corollary A.12$_{378}$) and has separately specified rows.

Similarly, because $\mathscr{T}_k$ is non-empty and has separately specified rows for all $k \in \mathbb{Z}_{\geq 0}$, it follows from Corollary 3.31$_{115}$ that

$$\begin{aligned}\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}\left[f(X_{u\cup v})\,\middle|\,X_u = x_u\right] &= \underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}\left[f(X_{\tau_{0:n}})\right]\\ &= \underline{\mathbb{E}}_{\mathscr{M}'}\left[\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}\left[f(X_{\tau_{0:n}})\,\middle|\,X_{\tau_0}\right]\right].\end{aligned} \tag{7.5}$$

Because $\mathscr{Q}$ is non-empty and bounded (since it is compact, due to Corollary A.12$_{378}$), it follows from Lemma 7.1$_{337}$ and Equations (7.4) and (7.5) that it remains to establish that, for all $x \in \mathscr{X}$,

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}[f(X_{\tau_{0:n}})\,|\,X_{\tau_0} = x] = g(x) = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_{\tau_{0:n}})\,|\,X_{\tau_0} = x]\,.$$

This readily follows from an argument that is completely analogous to that used in the first part of this proof—with $\{\tau_0\}$ in place of u—and therefore we omit it here for brevity. □

The crucial observation that makes Theorem 7.2339 applicable in practice, is that we can *first* choose a function f for which we know the time points on which it depends, and then obtain an induced discrete-time imprecise-Markov chain with a discrete time domain $\mathbb{D}$ that extends these time points. That is, we can tailor the discrete-time model to the inference problem that we are trying to solve. This "reduction" to a discrete-time model may then be used to translate known (ideally efficient) inference algorithms from the literature, to compute the corresponding inference for the original continuous-time model. Such algorithms will typically be expressed using the local models—i.e., the lower transition operators—of the induced discrete-time imprecise-Markov chain. As we already noted in Section 7.1336, these lower transition operators simply correspond to the imprecise exponentials $e^{\underline{Q}(\tau_{k+1}-\tau_k)}$, and we can therefore use the machinery that we developed in Chapter 6259 to evaluate them.

In summary, once we have an algorithm to compute inferences for a discrete-time imprecise-Markov chain that is expressed in terms of lower transition operators, then we can apply this algorithm to the induced model $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$, using Algorithm 1277 to evaluate the lower transition operators numerically. It then follows from Theorem 7.2339 that the computed inference for the discrete-time model coincides with the quantity of interest for the continuous-time model.

What follows is an example application of this strategy, to devise an algorithm that can compute lower expectations of a specific class of functions; we will show that this resulting algorithm is much more efficient than our general Algorithm 2286 discussed in Chapter 6259.

Example 7.1. Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Consider any $u \in \mathscr{U}_{\supset\emptyset}$ with $u = t_0, \ldots, t_n$, $n \in \mathbb{Z}_{\geq 0}$, and, for all $i \in \{0, \ldots, n\}$, consider any $f_i \in \mathscr{L}(\mathscr{X}_{t_i})$. Let $f \in \mathscr{L}(\mathscr{X}_u)$ be defined, for all $x_u \in \mathscr{X}_u$, as $f(x_u) := \prod_{i=0}^n f_i(x_{t_i})$. Suppose that we are interested in computing the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)]$. Because the structural assumptions on $\mathscr{Q}$ are satisfied, we know from our developments in Section 6.5284 that this can be done by combining Theorem 6.33289 with Algorithm 2286. Because the function f depends on $n+1$ time points, we also know from our discussion in Section 6.5284 that this approach will have a time complexity that is exponential in $n+1$. However, the function f has a lot of structure; it fully factorises into functions $f_0, \ldots, f_n$ that each only depend on a single time point t_i, $i \in \{0,\dots,n\}$. This suggests that it may be possible to exploit this structure for the development of more efficient algorithms.

It is the purpose of this example to illustrate how we can leverage existing results from the literature, where this problem has already been investigated, without having to translate such results explicitly to the continuous-time setting. So, let $\mathbb{D}$ be any discrete time domain that extends u, as in Definition 7.2339, and let $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$ be the discrete-time imprecise-Markov chain with time domain $\mathbb{D}$ that is induced by $\mathscr{Q}$ and $\mathscr{M}$. Then Theorem 7.2339 tells us that to obtain $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)]$, it suffices to compute $\underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}[f(X_u)]$.

Inferences of this type have been extensively studied in the literature on (discrete-time) imprecise-Markov chains and (more generally) imprecise-probabilistic graphical models. As it turns out, the factorising structure of this function f can be used to devise an algorithm with a runtime complexity that is only *linear* in n. This result has been described in the literature in multiple contexts and to varying degrees of generality. In the specific context of discrete-time imprecise-Markov chains, it is for example reported as a special case of the core algorithm described in [107], although that work makes the restricting assumption that the family $(\mathscr{T}_k)$ of transition matrices describing the discrete-time imprecise-Markov chain is constant in time, i.e. that $\mathscr{T}_k = \mathscr{T}_0$ for all $k \in \mathbb{Z}_{\geq 0}$. In the more general context of imprecise-probabilistic graphical models—specifically credal networks under epistemic irrelevance [16]—it can be obtained as a special case of the results in [16, Section 7.5.4], and should look recognisable to readers who are familiar with [21]. The same technique has also been employed to derive efficient inference algorithms for discrete-time *hidden* imprecise-Markov chains in [18]. The ideas behind the results in this chapter—which at the time were not formalised—were also central to the application of these existing methods for the development of a similar inference algorithm for continuous-time hidden imprecise-Markov chains [60].

So let us now summarise this algorithm using our current notation and terminology. As should hopefully be clear from the discussion above, this algorithm is effectively well-known in the literature and we are not attempting to claim it as our own. Moreover, because we are including it largely for illustrative purposes, we will omit the full derivation and proof of correctness; we simply refer to the above references for the technical details.

Let τ denote the canonical time index of $\mathbb{D}$ and, for all $k \in \mathbb{Z}_{\geq 0}$, let $\underline{T}_k = e^{\underline{Q}(\tau_{k+1}-\tau_k)}$ denote the lower transition operator corresponding to $\mathscr{T}_k$. The algorithm works by constructing two finite sequences $\underline{\Upsilon}_0,\dots,\underline{\Upsilon}_n$ and $\overline{\Upsilon}_0,\dots,\overline{\Upsilon}_n$ in $\mathscr{L}(\mathscr{X})$, which are computed in a backwards recursive manner—i.e. using dynamic programming—as $\underline{\Upsilon}_n := \overline{\Upsilon}_n := f_n$

and, for all $i \in \{0, \dots, n-1\}$ and all $x \in \mathscr{X}$,

$$\underline{\Upsilon}_i(x) := \begin{cases} f_i(x)\underline{T}_i\big(\underline{\Upsilon}_{i+1}\big)(x) & \text{if } f_i(x) \geq 0 \\ -f_i(x)\underline{T}_i\big(-\overline{\Upsilon}_{i+1}\big)(x) & \text{otherwise} \end{cases} \tag{7.6}$$

and

$$\overline{\Upsilon}_i(x) := \begin{cases} -f_i(x)\underline{T}_i\big(-\overline{\Upsilon}_{i+1}\big)(x) & \text{if } f_i(x) \geq 0 \\ f_i(x)\underline{T}_i\big(\underline{\Upsilon}_{i+1}\big)(x) & \text{otherwise.} \end{cases} \tag{7.7}$$

It can then be shown—see the above discussion of the related literature—that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}[f(X_u)] = \underline{\mathbb{E}}_{\mathscr{M}'}[\underline{\Upsilon}_0(X_{\tau_0})], \tag{7.8}$$

which relies on the fact that, due to Proposition 3.37_{120}, $\mathscr{T}_k$ has separately specified rows for all $k \in \mathbb{Z}_{\geq 0}$.

The crucial point here is that what we obtain from the literature are the backwards-recursive relations in Equations (7.6) and (7.7), as well as their relation to the inference of interest in Equation (7.8). The heavy lifting[1] of exploiting the factorising structure of f is done there, and what remains for our purposes is only to compute these (hopefully simpler) quantities.

So let us consider the remaining computational aspects of this particular example. The trick will be to use the existing numerical method to evaluate the lower transition operators $\underline{T}_i$, so as to obtain a workable algorithmic method to evaluate the quantities in Equations (7.6), (7.7), and (7.8). We first need to compute $2n$ functions $\underline{\Upsilon}_0, \dots, \underline{\Upsilon}_{n-1}$ and $\overline{\Upsilon}_0, \dots, \overline{\Upsilon}_{n-1}$; the functions $\underline{\Upsilon}_n = \overline{\Upsilon}_n = f_n$ are essentially obtained for free. So fix any $i \in \{0, \dots, n-1\}$. Then according to Equations (7.6) and (7.7), if we can evaluate $\underline{T}_i = e^{\underline{Q}(\tau_{i+1}-\tau_i)}$ in both $\underline{\Upsilon}_{i+1}$ and in $-\overline{\Upsilon}_{i+1}$, then computing $\underline{\Upsilon}_i(x)$ and $\overline{\Upsilon}_i(x)$ for any $x \in \mathscr{X}$ is a matter of straightforward multiplication of two scalar quantities based on the sign of $f_i(x)$. Therefore, the functions $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$ can be computed by employing the machinery described in Section $6.3.3_{275}$, in particular using Algorithm 1_{277}.[2]

After obtaining $\underline{\Upsilon}_0$ in this manner, according to Equation (7.8), it remains to evaluate $\underline{\mathbb{E}}_{\mathscr{M}'}[\underline{\Upsilon}_0(X_{\tau_0})]$. Due to Lemma 7.1_{337}, it holds that

[1] For the case in this example, proving the correctness of this approach is actually relatively straightforward once the solution is pointed out, but we envision that the strategy illustrated here will be applied to much more complicated problems.

[2] One should be careful with any numerical errors that are introduced in the approximation of these quantities, because they can be amplified by the magnitude of $f_i(x)$ when multiplying, e.g., $f_i(x)$ and $e^{\underline{Q}(\tau_{i+1}-\tau_i)}\underline{\Upsilon}_{i+1}(x)$ in computing $\underline{\Upsilon}_i(x)$. Specifically, one should rescale the desired numerical error ε to $\varepsilon/\|f_i\|$ for any numerical algorithm that is used to compute $e^{\underline{Q}(\tau_{i+1}-\tau_i)}\underline{\Upsilon}_{i+1}$ and $-e^{\underline{Q}(\tau_{i+1}-\tau_i)}\big(-\overline{\Upsilon}_{i+1}\big)$.

$\underline{\mathbb{E}}_{\mathscr{M}'}[\underline{\Upsilon}_0(X_{\tau_0})] = \underline{\mathbb{E}}_{\mathscr{M}}[e^{\underline{Q}\tau_0}\underline{\Upsilon}_0]$. Therefore, we need one final application of Algorithm 1_{277}—or an alternative numerical method—to evaluate $e^{\underline{Q}\tau_0}$ in $\underline{\Upsilon}_0$. What remains is then to compute $\underline{\mathbb{E}}_{\mathscr{M}}$ in $e^{\underline{Q}\tau_0}\underline{\Upsilon}_0$, which is solvable by assumption, as explained in Section 6.5_{284}. This concludes the discussion of the computational approach to evaluate $\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}[f(X_u)]$.

It should be clear from the above discussion that this method scales linearly in n; every $\underline{\Upsilon}_i$ and $\overline{\Upsilon}_i$ is a function in $\mathscr{L}(\mathscr{X})$, and there is no combinatorial explosion leading to exponential growth of the number of quantities that need to be computed. Since by Theorem 7.2_{339} it holds that

$$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}[f(X_u)] = \underline{\mathbb{E}}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}[f(X_u)] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u)]\,,$$

this gives a method with linear time complexity—in n—for computing the lower expectation of f for the imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. ◊

At this point, it is probably useful to mention that factorising functions of the form discussed in Example 7.1_{341} lie at the heart of even more advanced inference algorithms for imprecise-probabilistic models. For instance, for any $u \in \mathscr{U}_{\supset\emptyset}$, with $u = t_0, \dots, t_n$ and $n \in \mathbb{Z}_{\geq 0}$, any $x_u \in \mathscr{X}_u$, *any* $t \in \mathbb{R}_{\geq 0}$, and any $f \in \mathscr{L}(\mathscr{X})$, we may be interested in computing the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u]$. This deceptively simple looking expression is, however, entirely non-trivial to compute whenever $t < \max u$ and $t \notin u$, because in that case the conditional events $(X_t = x, X_u = x_u)$, $x \in \mathscr{X}$, are *not* in the domain $\mathscr{C}^{\mathrm{SP}}$, and hence most of the usual machinery from this dissertation cannot be used; see e.g. the preconditions in Proposition 2.23_{73}, which crucially relies on the assumption that $t \in u \cup \mathbb{R}_{>u}$.

An in-depth discussion of the following concepts is outside the scope of what we want to present here, but we believe that some pointers may nevertheless be useful. In order to obtain $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u]$, one can use Walley's *generalised Bayes's rule* [114, Theorem 6.4.1], provided that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_{x_u}(X_u)] > 0$; as explained in Section 5.4_{198}, this condition means that the lower probability $\underline{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}(X_u = x_u)$ is strictly positive.

Let us expand on the computational aspects of this approach; we would first need to verify that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_{x_u}(X_u)] > 0$. It is immediately clear that, for any $y_u \in \mathscr{X}_u$, it holds that $\mathbb{I}_{x_u}(y_u) = \prod_{i=0}^{n} \mathbb{I}_{x_{t_i}}(y_{t_i})$. In other words, the indicator $\mathbb{I}_{x_u}$ is a function that factorises in a way that allows us to apply the algorithm from Example 7.1_{341}. This gives us an efficient computational method to verify this precondition.

So now suppose that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_{x_u}(X_u)] = \underline{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}(X_u = x_u) > 0$. This implies that for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it holds that $P(X_u = x_u) > 0$. It can be shown that this, in turn, implies that the value $E[f(X_t)\,|\,X_u = x_u]$ of any coherent conditional prevision E corresponding to P—see Definition 2.4_{53}—whose domain includes the pair $(f(X_t), X_u = x_u) \in \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ is uniquely determined by P; essentially using Bayes's rule, which is applicable since the conditioning event has strictly positive probability.

Due to Definition 2.5$_{54}$, this implies that the pair $(f(X_t), X_u = x_u)$ is in the domain $\mathscr{D}_P$ of the conditional expectation $\mathbb{E}_P$ of P. Because this holds for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, it follows from Definition 5.8$_{198}$ that the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u]$ is in fact well-defined whenever $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[\mathbb{I}_{x_u}(X_u)] > 0$. Nevertheless, as we already noted, it cannot always readily be computed using the existing machinery from this work. The solution is offered by Walley's generalised Bayes's rule [114, Theorem 6.4.1] which essentially states that, with $\mu \in \mathbb{R}$, it holds that

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u] = \mu \quad \Leftrightarrow \quad \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[\mathbb{I}_{x_u}(X_u)\big(f(X_t) - \mu\big)\big] = 0\,. \tag{7.9}$$

In words, this means that the numerical value μ of the inference of interest $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u]$, is the unique value for which the unconditional lower expectation on the right-hand side of this expression equals zero. Moreover, the lower expectation on the right-hand side of Equation (7.9) is clearly of a function that factorises over the time points in $u \cup \{t\}$, in the sense that for all $y_{u\cup\{t\}} \in \mathscr{X}_{u\cup\{t\}}$ it holds that

$$\mathbb{I}_{x_u}(y_u)\big(f(y_t) - \mu\big) = \left(\prod_{i=0}^{n} \mathbb{I}_{x_{t_i}}(y_{t_i})\right) \cdot \big(f(y_t) - \mu\big)\,.$$

Therefore, for any $\mu \in \mathbb{R}$, we can use the algorithm from Example 7.1$_{341}$ to efficiently compute $G(\mu) := \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\big[\mathbb{I}_{x_u}(X_u)\big(f(X_t) - \mu\big)\big]$. It follows from Equation (7.9) that, in order to compute $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_t)\,|\,X_u = x_u]$, it remains to find the (unique) value of μ such that $G(\mu) = 0$. As a function of μ, this G is very well-behaved [16, Section 2.7.3]: it is (Lipschitz) continuous, concave, and non-increasing, and the value μ for which $G(\mu) = 0$ is guaranteed to lie between $\min f$ and $\max f$. Therefore, finding this value of μ can be done using a straightforward bisection method, which in this context is often called *Lavine's algorithm* [12]. See also [16, Section 2.7.3] for some notes on how to account for numerical errors in the computation of $G(\mu)$, and see e.g. [21,115] for root-finding methods that have been specifically tailored for this problem.

7.3 Correspondence of Sets of Processes

We have already established in Section 7.2$_{338}$ that the lower expectations of the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and the discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ with time domain $\mathbb{D}$ that is induced by $\mathscr{Q}$ and $\mathscr{M}$, have the same lower expectations for functions that depend only on (finitely many) time points in $\mathbb{D}$. From a practical point of view, as we have seen with Example 7.1$_{341}$, this connection suffices to leverage existing inference algorithms from the literature, for use with continuous-time imprecise-Markov chains.

It is the purpose of this final section to strengthen the connection between these two models. Specifically, we will show how we can obtain from a continuous-time stochastic process P—which is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{R}_{\geq 0}}$—a discrete-time stochastic process $P|_{\mathbb{D}}$—a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$—that can be interpreted as the restriction of P to the events that depend only on time points in $\mathbb{D}$. This uses the embedding $\mathbb{D} \subset \mathbb{R}_{\geq 0}$, and is ultimately the motivation for introducing discrete time domains at the level of generality in Definition 2.7$_{59}$. We will establish in Theorem 7.13$_{361}$ below, that—under some structural assumptions on $\mathscr{Q}$—the induced model $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ is the set of restrictions $P|_{\mathbb{D}}$ induced by the elements $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$.

We have several reasons for including this result. Throughout this dissertation we have taken *sets* of stochastic processes as the fundamental representation of our imprecise-probabilistic models. Hence, it makes sense to also investigate the connection between these sets of processes, in addition to the correspondence between the lower expectations established in Section 7.2$_{338}$. A more subtle point is that Theorem 7.2$_{339}$ only holds for u-measurable functions, that is, functions that depend on the state at finitely-many time points. By deriving the connection also in terms of the sets of processes, we hope to pave the way for future work that extends these results.

Finally, we find this result aesthetically satisfying; the construction of the induced discrete-time model is done in a fairly roundabout way—through the sets of transition matrices that dominate the generalised exponential $e^{\underline{Q}t}$ of the lower transition rate operator $\underline{Q}$ corresponding to $\mathscr{Q}$—which makes the corresponding lower expectation easy to evaluate using these generalised exponentials $e^{\underline{Q}t}$ (see Example 7.1$_{341}$), but it does not seem immediately obvious that every discrete-time process in this set is accounted for by a corresponding continuous-time process in the original model. Theorem 7.13$_{361}$, however, resolves this issue: for every continuous-time process $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ there is a corresponding discrete-time process $P|_{\mathbb{D}} \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ and, conversely, for every discrete-time process $P \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$, there is a continuous-time process $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that $P = P_*|_{\mathbb{D}}$.

Let us begin by setting up the definition of the restriction of a continuous-time process P to a discrete-time process $P|_{\mathbb{D}}$. In order to do this, we need to establish a correspondence between the domains of discrete-time and continuous-time stochastic processes. Unfortunately, as the following example makes clear, due to the level of generality at which we introduced the outcome spaces $\Omega_{\mathbb{D}}$ and $\Omega_{\mathbb{R}_{\geq 0}}$, we cannot work with the paths in these spaces directly.

Example 7.2. Let $\mathscr{X}$ contain at least two states, let $\mathbb{D}$ be a discrete time domain with canonical time index τ, and let $\Omega_{\mathbb{D}}$ be the set of all paths with time domain $\mathbb{D}$ that *eventually* become constant, i.e. such that for all $\omega \in \Omega_{\mathbb{D}}$ there is some $n \in \mathbb{Z}_{\geq 0}$ such that for all $m \in \mathbb{Z}_{\geq 0}$ with $n \leq m$ it holds that $\omega(\tau_n) = \omega(\tau_m)$. It is easy to see that $\Omega_{\mathbb{D}}$ satisfies Equation (2.8)$_{65}$, so this is a valid outcome space for a discrete-time stochastic process. Let $\Omega_{\mathbb{R}_{\geq 0}}$ be the set of all paths with time domain $\mathbb{R}_{\geq 0}$ that *never* become constant on $\mathbb{D}$, i.e. such that for all $\omega \in \Omega_{\mathbb{R}_{\geq 0}}$ and all $n \in \mathbb{Z}_{\geq 0}$ there is some $m \in \mathbb{Z}_{\geq 0}$ with $n \leq m$ and $\omega(\tau_n) \neq \omega(\tau_m)$. This $\Omega_{\mathbb{R}_{\geq 0}}$ also satisfies Equation (2.8)$_{65}$, and we can take it as the outcome space for a continuous-time stochastic process. For any $\omega \in \Omega_{\mathbb{R}_{\geq 0}}$, let $\omega|_{\mathbb{D}} : \mathbb{D} \to \mathscr{X} : t \mapsto \omega(t)$ denote the restriction of ω to $\mathbb{D}$.

In order to establish a correspondence between discrete-time and continuous-time processes, what we would like to do is identify, for any discrete-time event $A \subseteq \Omega_{\mathbb{D}}$, a corresponding continuous-time event $A' \subseteq \Omega_{\mathbb{R}_{\geq 0}}$. The intuitive identification that we would like to make is

$$A' := \left\{ \omega \in \Omega_{\mathbb{R}_{\geq 0}} \,:\, \omega|_{\mathbb{D}} \in A \right\},$$

which is to say, we want to let A' be the set of all continuous-time paths that agree with the event A on the time points in $\mathbb{D}$. Due to our choice of $\Omega_{\mathbb{D}}$ and $\Omega_{\mathbb{R}_{\geq 0}}$ in this example, it follows that, for any $\omega \in \Omega_{\mathbb{R}_{\geq 0}}$, $\omega|_{\mathbb{D}}$ will never become constant, and hence $\omega|_{\mathbb{D}} \notin \Omega_{\mathbb{D}}$. It follows that $A' = \emptyset$ regardless of the choice of A. In other words, no information about A is preserved in this identification of A', so this does not provide us with a useful correspondence between the domains of such processes. ◊

The fact that we cannot work with these outcome spaces directly does not really influence our results, but unfortunately it complicates the analysis somewhat; we will be forced to explicitly work with the algebraic structure of the domains of stochastic processes.

Working with this algebraic structure is easy when the events in question are situations: for any $n \in \mathbb{Z}_{\geq 0}$ and, with $u = \tau_{0:n}$, any $x_u \in \mathscr{X}_u$, the events $(X_u = x_u)_{\mathbb{D}}$ and $(X_u = x_u)_{\mathbb{R}_{\geq 0}}$ essentially carry the same information. What we need next is a way to similarly map between events with a more complicated structure; for this we will rely on Lemma 3.3$_{86}$. The following result provides the required conditions to do this uniquely; essentially it tells us that, if a discrete-time event A has two different representations in terms of unions of situations, then the two corresponding unions of the analogous continuous-time events are also representations of the same (continuous-time) event.

Proposition 7.3. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ. Consider any $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$, and suppose that there are $n, m \in \mathbb{Z}_{\geq 0}$ and*

$S \subseteq \mathscr{X}_{\tau_{0:n}}$ and $V \subseteq \mathscr{X}_{\tau_{0:m}}$ such that

$$\bigcup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}} = A = \bigcup_{x_{\tau_{0:m}} \in V} (X_{\tau_{0:m}} = x_{\tau_{0:m}})_{\mathbb{D}} . \tag{7.10}$$

Then also $\cup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}} = \cup_{x_{\tau_{0:m}} \in V} (X_{\tau_{0:m}} = x_{\tau_{0:m}})_{\mathbb{R}_{\geq 0}}$.

Proof. It clearly suffices to show that

$$\bigcup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}} \subseteq \bigcup_{x_{\tau_{0:m}} \in V} (X_{\tau_{0:m}} = x_{\tau_{0:m}})_{\mathbb{R}_{\geq 0}} .$$

To this end, fix any $\omega \in \cup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}}$. Then there is some $x_{\tau_{0:n}} \in S$ such that $\omega \in (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}}$, or equivalently, that $\omega(\tau_i) = x_{\tau_i}$ for all $i \in \{0, \ldots, n\}$.

Now let $k := \max\{n, m\}$. Due to Equation (2.8)$_{65}$, there is some $\tilde{\omega} \in \Omega_{\mathbb{D}}$ such that $\tilde{\omega}(\tau_i) = \omega(\tau_i)$ for all $i \in \{0, \ldots, k\}$. Because $n \leq k$, this implies that $\tilde{\omega}(\tau_i) = \omega(\tau_i) = x_{\tau_i}$ for all $i \in \{0, \ldots, n\}$, which means that $\tilde{\omega} \in (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}}$. Due to Equation (7.10), this implies that $\tilde{\omega} \in A$ and, again by Equation (7.10), there is therefore some $y_{0:m} \in V$ such that also $\tilde{\omega} \in (X_{\tau_{0:m}} = y_{\tau_{0:m}})_{\mathbb{D}}$. This means that $\tilde{\omega}(\tau_i) = y_{\tau_i}$ for all $i \in \{0, \ldots, m\}$ and, because $m \leq k$, also $\omega(\tau_i) = \tilde{\omega}(\tau_i) = y_{\tau_i}$ for all $i \in \{0, \ldots, m\}$. Hence, indeed, $\omega \in (X_{\tau_{0:m}} = y_{\tau_{0:m}})_{\mathbb{R}_{\geq 0}}$. □

To obtain the identification that we are after, we note that for any discrete-time event $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ there are $n \in \mathbb{Z}_{\geq 0}$ and $S \subseteq \mathscr{X}_{\tau_{0:n}}$ such that $A = \cup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{D}}$, due to Lemma 3.3$_{86}$. Hence, for this A we can define a unique continuous-time event $A\rceil_{\mathbb{R}_{\geq 0}}$ as

$$A\rceil_{\mathbb{R}_{\geq 0}} := \cup_{x_{\tau_{0:n}} \in S} (X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}} . \tag{7.11}$$

In particular, it follows from Proposition 7.3$_{\curvearrowleft}$ that this $A\rceil_{\mathbb{R}_{\geq 0}}$ is uniquely determined by A and independent of the choice of n and S.

Let us first establish that this identification does not have the same issue that we observed previously:

Example 7.3. Let $\mathscr{X}$, $\mathbb{D}$, τ, $\Omega_{\mathbb{D}}$, and $\Omega_{\mathbb{R}_{\geq 0}}$ be as in Example 7.2$_{346}$, and choose any $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ such that $A \neq \emptyset$. Then, as we know from Example 7.2$_{346}$, it holds that $\{\omega \in \Omega_{\mathbb{R}_{\geq 0}} : \omega|_{\mathbb{D}} \in A\} = \emptyset$. We want to show that $A\rceil_{\mathbb{R}_{\geq 0}} \neq \emptyset$. Due to Lemma 3.3$_{86}$, there are $n \in \mathbb{Z}_{\geq 0}$ and, with $u := \tau_{0:n}$, $S \subseteq \mathscr{X}_u$ such that $A = \cup_{x_u \in S} (X_u = x_u)_{\mathbb{D}}$ and, due to Equation (7.11), $A\rceil_{\mathbb{R}_{\geq 0}} = \cup_{x_u \in S} (X_u = x_u)_{\mathbb{R}_{\geq 0}}$. Because $A \neq \emptyset$, there is some $\omega \in A$, which means that there is some $x_u \in S$ such that $\omega \in (X_u = x_u)_{\mathbb{D}}$. Due to Equation (2.8)$_{65}$, there is some $\omega' \in \Omega_{\mathbb{R}_{\geq 0}}$ such that $\omega'|_u = x_u$, which implies that $\omega' \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$ and hence it follows that $\omega' \in A\rceil_{\mathbb{R}_{\geq 0}}$. ◊

Moreover, as the following result makes clear, this definition yields the intuitive identification that we were after in Example 7.2346, provided that the outcome spaces in question are "compatible".

Proposition 7.4. *Let $\mathbb{D}$ be a discrete time domain, and consider the outcome space $\Omega_{\mathbb{D}} = \{\omega|_{\mathbb{D}} : \omega \in \Omega_{\mathbb{R}_{\geq 0}}\}$ with time domain $\mathbb{D}$ that corresponds to the restrictions of the elements of $\Omega_{\mathbb{R}_{\geq 0}}$ to $\mathbb{D}$. Then for all $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$ it holds that*

$$A\rceil_{\mathbb{R}_{\geq 0}} = \{\omega \in \Omega_{\mathbb{R}_{\geq 0}} : \omega|_{\mathbb{D}} \in A\}.$$

Proof. Let τ denote the canonical time index of $\mathbb{D}$, and let $A' := \{\omega \in \Omega_{\mathbb{R}_{\geq 0}} : \omega|_{\mathbb{D}} \in A\}$. Due to Lemma 3.386, there are $n \in \mathbb{Z}_{\geq 0}$ and, with $u := \tau_{0:n}$, $S \subseteq \mathscr{X}_u$ such that $A = \cup_{x_u \in S}(X_u = x_u)_{\mathbb{D}}$ and $A\rceil_{\mathbb{R}_{\geq 0}} = \cup_{x_u \in S}(X_u = x_u)_{\mathbb{R}_{\geq 0}}$, due to Equation (7.11).

Now first fix any $\omega \in A\rceil_{\mathbb{R}_{\geq 0}}$. Then there is some $x_u \in S$ such that $\omega \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$, which implies that $\omega|_u = x_u$. Because $u \subset \mathbb{D}$ it follows that $(\omega|_{\mathbb{D}})|_u = x_u$. Hence, and because $\Omega_{\mathbb{D}}$ corresponds to the restrictions of the elements of $\Omega_{\mathbb{R}_{\geq 0}}$ to $\mathbb{D}$, it follows that $\omega|_{\mathbb{D}} \in (X_u = x_u)_{\mathbb{D}}$. This implies that $\omega|_{\mathbb{D}} \in A$, and because $\omega \in A\rceil_{\mathbb{R}_{\geq 0}} \subseteq \Omega_{\mathbb{R}_{\geq 0}}$, it follows that $\omega \in A'$. Since $\omega \in A\rceil_{\mathbb{R}_{\geq 0}}$ is arbitrary we conclude that $A\rceil_{\mathbb{R}_{\geq 0}} \subseteq A'$.

For the other direction, fix any $\omega \in A'$. Then $\omega|_{\mathbb{D}} \in A$, and therefore there is some $x_u \in S$ such that $\omega|_{\mathbb{D}} \in (X_u = x_u)_{\mathbb{D}}$. This implies that $(\omega|_{\mathbb{D}})|_u = x_u$, which in turn implies that $\omega|_u = x_u$. Hence it follows that $\omega \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$, which implies that $\omega \in A\rceil_{\mathbb{R}_{\geq 0}}$. Because $\omega \in A'$ is arbitrary we conclude that also $A' \subseteq A\rceil_{\mathbb{R}_{\geq 0}}$. ☐

Let us next establish that Equation (7.11) allows us to map from the discrete-time domain of *conditional* events, to the continuous-time one.

Lemma 7.5. *Let $\mathbb{D}$ be a discrete-time domain and consider any conditional event $(A, X_u = x_u)_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$. Then $\left(A\rceil_{\mathbb{R}_{\geq 0}}, X_u = x_u\right)_{\mathbb{R}_{\geq 0}} \in \mathscr{C}_{\mathbb{R}_{\geq 0}}^{\mathrm{SP}}$.*

Proof. Let τ denote the canonical time index of $\mathbb{D}$. By Lemma 2.1968 it holds that $A \in \mathscr{A}_{\emptyset}^{\mathbb{D}}$. Hence, due to Equation (7.11), there are $n \in \mathbb{Z}_{\geq 0}$ and $S \subseteq \mathscr{X}_{\tau_{0:n}}$ such that $A\rceil_{\mathbb{R}_{\geq 0}} = \cup_{x_{\tau_{0:n}} \in S}(X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}}$.

Because $(A, X_u = x_u)_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$ it follows from Lemma 3.286 that the event $(X_u = x_u)_{\mathbb{D}}$ is a situation, which means that there is some $m \in \mathbb{Z}_{\geq 0}$ such that $u = \tau_{0:(m-1)}$. It follows that for all $i \in \{0, \ldots, n\}$ it holds that $\tau_i \in u \cup \mathbb{R}_{>u}$, which implies that $(X_{\tau_i} = x)_{\mathbb{R}_{\geq 0}} \in \mathscr{E}_u^{\mathbb{R}_{\geq 0}} \subseteq \mathscr{A}_u^{\mathbb{R}_{\geq 0}}$ for all $x \in \mathscr{X}$. Because $\mathscr{A}_u^{\mathbb{R}_{\geq 0}}$ is an algebra, this set is closed under finite intersections of its elements, and hence it follows that for all $x_{\tau_{0:n}} \in S$ it holds that

$$(X_{\tau_{0:n}} = x_{\tau_{0:n}})_{\mathbb{R}_{\geq 0}} \in \mathscr{A}_u^{\mathbb{R}_{\geq 0}}.$$

Since $\mathscr{A}_u^{\mathbb{R}_{\geq 0}}$ is an algebra, it is also closed under finite unions of its elements, and hence it follows that $A\rceil_{\mathbb{R}_{\geq 0}} \in \mathscr{A}_u^{\mathbb{R}_{\geq 0}}$. We therefore find that $\left(A\rceil_{\mathbb{R}_{\geq 0}}, X_u = x_u\right)_{\mathbb{R}_{\geq 0}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{R}_{\geq 0}}$ by Definition 2.1067. □

We are now finally ready to define the restriction of a continuous-time stochastic process.

Definition 7.3. *Let $P \in \mathbb{P}$ be a continuous-time stochastic process, and let $\mathbb{D}$ be a discrete time domain. The* restriction of P to $\mathbb{D}$ *is the map $P|_{\mathbb{D}} : \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}} \to \mathbb{R}$ that is defined for all $(A, X_u = x_u)_{\mathbb{D}} \in \mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$ as*

$$P|_{\mathbb{D}}\left(A \,\middle|\, X_u = x_u\right) := P\left(A\rceil_{\mathbb{R}_{\geq 0}} \,\middle|\, X_u = x_u\right). \tag{7.12}$$

We note that the right-hand side of Equation (7.12) is well-defined due to Lemma 7.5↶. Let us next show that this map $P|_{\mathbb{D}}$ is a discrete-time stochastic process with time domain $\mathbb{D}$. We will first need the following technical result.

Lemma 7.6. *Let $\mathbb{D}$ be a discrete time domain with canonical time index τ, fix any $m, n \in \mathbb{Z}_{\geq 0}$ such that $m-1 \leq n$, and let $u := \tau_{0:(m-1)}$. Let $\omega \in \Omega_{\mathbb{D}}$ and $\omega' \in \Omega_{\mathbb{R}_{\geq 0}}$ be such that $\omega|_{\tau_{0:n}} = \omega'|_{\tau_{0:n}}$. Then for all $x_u \in \mathscr{X}_u$ it holds that $\omega \in (X_u = x_u)_{\mathbb{D}}$ if and only if $\omega' \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$.*

Proof. If $m = 0$ it holds that $u = \emptyset$, and therefore, as noted in Section 2.364, it then holds that $(X_u = x_u)_{\mathbb{D}} = \Omega_{\mathbb{D}}$ and $(X_u = x_u)_{\mathbb{R}_{\geq 0}} = \Omega_{\mathbb{R}_{\geq 0}}$, whence the claim follows trivially. On the other hand, if $m > 0$ then $u \neq \emptyset$. In that case, first suppose that $\omega' \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$, which implies that $\omega'|_u = x_u$. Because $m - 1 \leq n$, it holds that $\omega(\tau_i) = \omega'(\tau_i) = x_{\tau_i}$ for all $i \in \{0, \ldots, m-1\}$, which implies that also $\omega|_u = x_u$, and hence that $\omega \in (X_u = x_u)_{\mathbb{D}}$. Completely analogously, if $\omega \in (X_u = x_u)_{\mathbb{D}}$ then $\omega|_u = x_u$, which implies that $\omega'(\tau_i) = \omega(\tau_i) = x_{\tau_i}$ for all $i \in \{0, \ldots, m-1\}$ since $m - 1 \leq n$, whence $\omega' \in (X_u = x_u)_{\mathbb{R}_{\geq 0}}$. □

Proposition 7.7. *Let $P \in \mathbb{P}$ be a continuous-time stochastic process, and let $\mathbb{D}$ be a discrete time domain. Then the restriction $P|_{\mathbb{D}}$ of P to $\mathbb{D}$, is a discrete-time stochastic process with time domain $\mathbb{D}$.*

Proof. By Definitions 2.1268 and 2.1369, in order to establish that $P|_{\mathbb{D}}$ is a stochastic process with time domain $\mathbb{D}$, we need to establish that $P|_{\mathbb{D}}$ is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$. By Definition 7.3, $P|_{\mathbb{D}}$ clearly has the correct domain $\mathscr{C}^{\mathrm{SP}}_{\mathbb{D}}$.

Moreover, because P is a stochastic process, it follows from Definition 2.1268 that P is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$. By Definition 2.248, this implies that P is a real-valued map. Due to Definition 7.3, this implies that $P|_{\mathbb{D}}$ is also a real-valued map.

Hence we have established that $P|_{\mathbb{D}}$ is a real-valued map on $\mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$. By Definition 2.2₄₈, in order to establish that $P|_{\mathbb{D}}$ is a coherent conditional probability, we need to show that it satisfies the defining coherence condition. So choose any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1,\dots,n\}$, any $\lambda_i \in \mathbb{R}$ and $(A_i, C_i)_{\mathbb{D}} \in \mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$. According to Definition 2.2₄₈, we need to show that

$$\max\left\{\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P|_{\mathbb{D}}(A_i \,|\, C_i) - \mathbb{I}_{A_i}(\omega)\big) \,\middle|\, \omega \in C_0\right\} \geq 0\,, \tag{7.13}$$

with $C_0 := \cup_{i=1}^{n} C_i$.

Let τ denote the canonical time index of $\mathbb{D}$, and fix any $i \in \{1,\dots,n\}$. By Lemma 3.2₈₆ it holds that $C_i \in \mathscr{S}_{\mathbb{D}}$, which implies that there is some $m_i \in \mathbb{Z}_{\geq 0}$ and, with $u_i := \tau_{0:(m_i-1)}$, some $x_{u_i}^{(i)} \in \mathscr{X}_{u_i}$, such that $C_i = (X_{u_i} = x_{u_i}^{(i)})_{\mathbb{D}}$. Moreover, following Lemma 3.3₈₆, there are $n_i \in \mathbb{Z}_{\geq 0}$ and, with $v_i := \tau_{0:n_i}$, $S_i \subseteq \mathscr{X}_{v_i}$ such that $A_i = \cup_{x_{v_i} \in S_i}(X_{v_i} = x_{v_i})_{\mathbb{D}}$. Let $C_i' := (X_{u_i} = x_{u_i}^{(i)})_{\mathbb{R}_{\geq 0}}$ and let $A_i' := A_i \!\upharpoonright_{\mathbb{R}_{\geq 0}} = \cup_{x_{v_i} \in S_i}(X_{v_i} = x_{v_i})_{\mathbb{R}_{\geq 0}}$ as in Equation (7.11)₃₄₈. By Definition 7.3 it then holds that

$$P|_{\mathbb{D}}(A_i \,|\, C_i) = P\big(A_i' \,\big|\, C_i'\big)\,. \tag{7.14}$$

We already established that P is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$, and hence it follows from Definition 2.2₄₈ that

$$\max\left\{\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i'}(\omega)\big(P(A_i' \,|\, C_i') - \mathbb{I}_{A_i'}(\omega)\big) \,\middle|\, \omega \in C_0'\right\} \geq 0\,,$$

with $C_0' := \cup_{i=1}^{n} C_i'$. This implies that there is some $\omega' \in C_0'$ such that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i'}(\omega')\big(P(A_i' \,|\, C_i') - \mathbb{I}_{A_i'}(\omega')\big) \geq 0\,. \tag{7.15}$$

Now let $N := \max_{i \in \{1,\dots,n\}} \max\{m_i - 1, n_i\}$; then $N \geq 0$ since $n_i \in \mathbb{Z}_{\geq 0}$ for all $i \in \{1,\dots,n\}$. Due to Equation (2.8)₆₅, there is some $\omega \in \Omega_{\mathbb{D}}$ such that $\omega(\tau_i) = \omega'(\tau_i)$ for all $i \in \{0,\dots,N\}$.

Now fix any $i \in \{1,\dots,n\}$. Because it holds that $C_i = (X_{u_i} = x_{u_i}^{(i)})_{\mathbb{D}}$ and $C_i' = (X_{u_i} = x_{u_i}^{(i)})_{\mathbb{R}_{\geq 0}}$, and since $u_i = \tau_{0:(m_i-1)}$ and $m_i - 1 \leq N$, it follows from Lemma 7.6 that $\omega \in C_i$ if and only if $\omega' \in C_i'$. In turn, this implies that $\mathbb{I}_{C_i}(\omega) = \mathbb{I}_{C_i'}(\omega')$.

Next, note that $A_i = \cup_{x_{v_i} \in S_i}(X_{v_i} = x_{v_i})_{\mathbb{D}}$ and $A_i' = \cup_{x_{v_i} \in S_i}(X_{v_i} = x_{v_i})_{\mathbb{R}_{\geq 0}}$. Since $v_i = \tau_{0:n_i}$ and $n_i \leq N$, it follows from Lemma 7.6 that $\omega \in A_i$ if and only if $\omega' \in A_i'$. This also implies that $\mathbb{I}_{A_i}(\omega) = \mathbb{I}_{A_i'}(\omega')$.

Hence in summary, we have established that $\mathbb{I}_{C_i}(\omega) = \mathbb{I}_{C_i'}(\omega')$ and $\mathbb{I}_{A_i}(\omega) = \mathbb{I}_{A_i'}(\omega')$ for all $i \in \{1,\dots,n\}$. By combining this with Equations $(7.14)_{\curvearrowleft}$ and $(7.15)_{\curvearrowleft}$, we find that

$$0 \leq \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i'}(\omega')\big(P(A_i' \,|\, C_i') - \mathbb{I}_{A_i'}(\omega')\big) = \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P|_{\mathbb{D}}(A_i \,|\, C_i) - \mathbb{I}_{A_i}(\omega)\big)\,. \tag{7.16}$$

Finally, since $\omega' \in C_0'$, there is some $i \in \{1,\dots,n\}$ such that $\omega' \in C_i'$. We already established that this implies that $\omega \in C_i$, whence it follows that $\omega \in C_0$. Therefore, it follows from Equation (7.16) that the inequality in Equation $(7.13)_{\curvearrowleft}$ indeed holds. This implies that $P|_{\mathbb{D}}$ is a coherent conditional probability on $\mathscr{C}_{\mathbb{D}}^{\mathrm{SP}}$, or in other words, as established in the beginning of this proof, that it is a discrete-time stochastic process with time domain $\mathbb{D}$. ☐

Our aim in the remainder of this section will now be to establish that, under some conditions on $\mathscr{Q}$, it holds that

$$\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \left\{ P|_{\mathbb{D}} \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \right\}, \tag{7.17}$$

which reveals the interpretation of the induced discrete-time imprecise-Markov chain $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ as the restriction of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ to the events that only deal with the time points in $\mathbb{D}$. To this end, we first give two technical results which help to further clarify the connection between $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$.

Lemma 7.8. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows and corresponding lower transition rate operator $\underline{Q}$, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Fix any $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, let $\mathscr{T}_{e^{\underline{Q}(s-t)}}$ be the set of transition matrices that dominate $e^{\underline{Q}(s-t)}$, and let ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$ be the set of transition matrices corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Then it holds that $\mathscr{T}_{e^{\underline{Q}(s-t)}} = {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s$.*

Proof. Let $\underline{T}_t^s$ be the lower transition operator corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, as in Definition 6.1_{261}. Because $\mathscr{Q}$ is compact it is bounded by Corollary $\mathrm{A}.12_{378}$. Hence $\mathscr{Q}$ is a non-empty and bounded set of rate matrices that has separately specified rows, and so it follows from Proposition 6.26_{281}that $\underline{T}_t^s = e^{\underline{Q}(s-t)}$. Moreover, because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Theorem 6.2_{261} that ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}_t^s = \mathscr{T}_{\underline{T}_t^s} = \mathscr{T}_{e^{\underline{Q}(s-t)}}$. ☐

Lemma 7.9. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows and corresponding lower transition rate operator $\underline{Q}$, and let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$. Fix any $t \in \mathbb{R}_{\geq 0}$, and let*

$$\mathscr{M}' := \mathscr{M}\mathscr{T}_{e^{\underline{Q}t}} = \left\{ \sum_{x \in \mathscr{X}} p(x)T(x,\cdot) \,:\, p \in \mathscr{M}, T \in \mathscr{T}_{e^{\underline{Q}t}} \right\},$$

where $\mathscr{T}_{e^{\mathcal{Q}t}}$ is the set of transition matrices that dominate $e^{\mathcal{Q}t}$. Moreover, let

$$\mathscr{M}_t := \left\{ p : \mathscr{X} \to \mathbb{R} : x \mapsto P(X_t = x) \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}} \right\} .$$

Then $\mathscr{M}' = \mathscr{M}_t$.

Proof. We first note that, because $\mathcal{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 7.8 that $\mathscr{T}_{e^{\mathcal{Q}t}} = {}^{\mathcal{Q}}_{\mathscr{M}}\mathscr{T}^t_0$ is the set of transition matrices corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$.

Now, for the first direction, fix any $P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$. Then the transition matrix ${}^{P}T^t_0$ corresponding to P is in ${}^{\mathcal{Q}}_{\mathscr{M}}\mathscr{T}^t_0 = \mathscr{T}_{e^{\mathcal{Q}t}}$, due to Equation $(5.11)_{197}$. Moreover, because $P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$ it holds that $P \sim \mathscr{M}$, which implies that there is some $p \in \mathscr{M}$ such that $p(x) = P(X_0 = x)$ for all $x \in \mathscr{X}$. Therefore, the map $q : \mathscr{X} \to \mathbb{R}$ that is defined for all $x \in \mathscr{X}$ as $q(x) := \sum_{y \in \mathscr{X}} p(y)\, {}^{P}T^t_0(y,x)$ is an element of $\mathscr{M}'$. Moreover, for all $x \in \mathscr{X}$ it holds that

$$\begin{aligned} q(x) &= \sum_{y \in \mathscr{X}} p(y)\, {}^{P}T^t_0(y,x) \\ &= \sum_{y \in \mathscr{X}} P(X_0 = y) P(X_t = x \,|\, X_0 = y) \\ &= \sum_{y \in \mathscr{X}} P(X_t = x, X_0 = y) = P(X_t = x) , \end{aligned}$$

where for the second equality we used the choice of p and the definition of the transition matrix ${}^{P}T^t_0$, for the third equality we used Property $\mathrm{F4}_{47}$, and for the final equality we used Property $\mathrm{F3}_{47}$. Hence we conclude that the map $\mathscr{X} \to \mathbb{R} : x \mapsto P(X_t = x)$ is in $\mathscr{M}'$. Because $P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$ is arbitrary it follows that $\mathscr{M}_t \subseteq \mathscr{M}'$.

For the other direction, fix any $q \in \mathscr{M}'$. This implies that there are $p \in \mathscr{M}$ and $T \in \mathscr{T}_{e^{\mathcal{Q}t}}$ such that $q(x) = \sum_{y \in \mathscr{X}} p(y) T(y,x)$ for all $x \in \mathscr{X}$. Because $p \in \mathscr{M}$, and since $\mathscr{M}$ and $\mathcal{Q}$ are non-empty, it follows from Proposition 6.30_{287} that there is some $P_\emptyset \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$ such that $P_\emptyset(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$. Moreover, because $T \in \mathscr{T}_{e^{\mathcal{Q}t}} = {}^{\mathcal{Q}}_{\mathscr{M}}\mathscr{T}^t_0$, it follows from Equation $(5.11)_{197}$ that there is some $P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$ with corresponding transition matrix ${}^{P}T^t_0 = T$. Because $\mathcal{Q}$ is non-empty, convex, and has separately specified rows, it follows from Theorem 5.11_{193}—with $u = \{0\}$ and $P_{x_u} = P$ for all $x_u \in \mathscr{X}_u$—that there is some $P' \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathscr{M}}$ such that, for all $x \in \mathscr{X}$,

$$P'(X_0 = x) = P_\emptyset(X_0 = x) = p(x) ,$$

and, for all $x, y \in \mathscr{X}$,

$$P'(X_t = y \,|\, X_0 = x) = P(X_t = y \,|\, X_0 = x) = {}^{P}T^t_0(x,y) .$$

It follows that, for all $x \in \mathscr{X}$,

$$P'(X_t = x) = \sum_{y\in\mathscr{X}} P'(X_t = x, X_0 = y)$$
$$= \sum_{y\in\mathscr{X}} P'(X_t = x \,|\, X_0 = y)P'(X_0 = x) = \sum_{y\in\mathscr{X}} {}^{P}T_0^t(y,x)p(x) = q(x)\,,$$

where we used Properties F3$_{47}$ and F4$_{47}$. Hence it follows that $q \in \mathscr{M}_t$, and because $q \in \mathscr{M}'$ is arbitrary, that $\mathscr{M}' \subseteq \mathscr{M}_t$. □

We will now prove Equation (7.17)$_{352}$ in two lemmas, which each take care of one direction of the inclusion.

Lemma 7.10. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $\mathbb{D}$ be a discrete-time domain. Then for all $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it holds that $P|_{\mathbb{D}} \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$.*

Proof. We know from Proposition 7.7$_{350}$ that $P|_{\mathbb{D}}$ is a discrete-time stochastic process with time domain $\mathbb{D}$, which means that $P \in \mathbb{P}^{\mathbb{D}}$ according to Definition 2.11$_{68}$.

Let τ denote the canonical time index of $\mathbb{D}$, and let $\underline{Q}$ denote the lower transition rate operator corresponding to $\mathscr{Q}$. We will first show that $P|_{\mathbb{D}} \sim \mathscr{M}'$, where $\mathscr{M}' = \mathscr{M}\mathscr{T}_{e^{\underline{Q}\tau_0}}$ as in Definition 7.1$_{337}$. From Definition 7.3$_{350}$ and Equation (7.11)$_{348}$ it follows that, for all $x \in \mathscr{X}$,

$$P|_{\mathbb{D}}(X_{\tau_0} = x) = P(X_{\tau_0} = x)\,. \tag{7.18}$$

Therefore, and because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 7.9$_{352}$ that there is some $p \in \mathscr{M}'$ such that $P|_{\mathbb{D}}(X_{\tau_0} = x) = P(X_{\tau_0} = x) = p(x)$ for all $x \in \mathscr{X}$. Hence, by the definition given in Section 3.3.1$_{102}$, we find that $P|_{\mathbb{D}} \sim \mathscr{M}'$.

Therefore, according to Definition 3.11$_{104}$, in order to show that $P|_{\mathbb{D}} \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$, it remains to show that $P|_{\mathbb{D}} \sim (\mathscr{T}_k)$ where, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k = \mathscr{T}_{e^{\underline{Q}(\tau_{k+1}-\tau_k)}}$ is the set of transition matrices that dominate $e^{\underline{Q}(\tau_{k+1}-\tau_k)}$, as in Definition 7.1$_{337}$. Let (T_{k,x_u}) denote the family of history-dependent transition matrices corresponding to $P|_{\mathbb{D}}$, as in Definition 3.8$_{101}$. Then, according to Definition 3.9$_{102}$, we need to show that $T_{k,x_u} \in \mathscr{T}_k$ for all $k \in \mathbb{Z}_{\geq 0}$ and all $x_u \in \mathscr{X}_u$, with $u = \tau_{0:(k-1)}$.

So fix any $k \in \mathbb{Z}_{\geq 0}$ and any $x_u \in \mathscr{X}_u$, with $u = \tau_{0:(k-1)}$. By Definition 3.8$_{101}$ it then follows that, for all $x, y \in \mathscr{X}$, it holds that

$$T_{k,x_u}(x,y) = P|_{\mathbb{D}}(X_{\tau_{k+1}} = y \,|\, X_{\tau_k} = x, X_u = x_u)$$
$$= P(X_{\tau_{k+1}} = y \,|\, X_{\tau_k} = x, X_u = x_u) = {}^{P}T_{\tau_k,x_u}^{\tau_{k+1}}(x,y)\,,$$

where for the second equality we used Definition 7.3$_{350}$, and where ${}^{P}T^{\tau_{k+1}}_{\tau_k,x_u}$ is the history-dependent transition matrix corresponding to P, as in Definition 4.2$_{148}$. Since this is true for all $x,y \in \mathscr{X}$, it follows that $T_{k,x_u} = {}^{P}T^{\tau_{k+1}}_{\tau_k,x_u}$. Because $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it follows from Equation (5.11)$_{197}$ that ${}^{P}T^{\tau_{k+1}}_{\tau_k,x_u} \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{\tau_{k+1}}_{\tau_k}$ which, using the previous equality, implies that

$$T_{k,x_u} = {}^{P}T^{\tau_{k+1}}_{\tau_k,x_u} \in {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{\tau_{k+1}}_{\tau_k} = \mathscr{T}_{e^{\mathscr{Q}(\tau_{k+1}-\tau_k)}} = \mathscr{T}_k\,,$$

where for the second equality we used Lemma 7.8$_{352}$ and that $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows. Since this is true for all $x_u \in \mathscr{X}_u$ and all $k \in \mathbb{Z}_{\geq 0}$, it follows from Definition 3.9$_{102}$ that $P|_{\mathbb{D}} \sim (\mathscr{T}_k)$. Hence, since we already know that $P|_{\mathbb{D}} \in \mathbb{P}^{\mathbb{D}}$ and $P|_{\mathbb{D}} \sim \mathscr{M}'$, by Definition 3.11$_{104}$ it holds that $P|_{\mathbb{D}} \in \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'} = \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$. □

Proving the inclusion in the other direction is, unfortunately, somewhat more involved. We need the following auxiliary property.

Lemma 7.11. *Choose any conditional event $(A,C) \in \mathscr{C}^{\mathrm{SP}}$. Then there are $u \in \mathscr{U}$ and $x_u \in \mathscr{X}_u$ such that $C = (X_u = x_u)$. Moreover, there are $v \in \mathscr{U}$ and $S \subseteq \mathscr{X}_v$ such that $v \subset u \cup \mathbb{R}_{>u}$ and $A = \cup_{y_v \in S}(X_v = y_v)$.*

Proof. Because $(A,C) \in \mathscr{C}^{\mathrm{SP}}$ it follows from Definition 2.10$_{67}$ that $C = (X_u = x_u)$ for some $u \in \mathscr{U}$ and $x_u \in \mathscr{X}_u$ and, moreover, that $A \in \mathscr{A}_u$. The second claim is now immediate from Proposition 2.18$_{66}$. □

Lemma 7.12. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $\mathbb{D}$ be a discrete-time domain. Then for all $P \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ there is some $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that $P = P_*|_{\mathbb{D}}$.*

Proof. The proof works by constructing a continuous-time stochastic process $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ that satisfies $P_*|_{\mathbb{D}} = P$, where the construction essentially uses an induction argument. In order to obtain this P_*, we first build a sequence $\{P_n\}_{n \in \mathbb{Z}_{\geq 0}}$ in $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. This sequence is constructed iteratively; let us start by finding P_0.

Let τ denote the canonical time index of $\mathbb{D}$. By Definition 7.1$_{337}$ it holds that $\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$, where $\mathscr{M}' = \mathscr{M}\mathscr{T}_{e^{\mathscr{Q}\tau_0}}$ and, for all $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{T}_k = \mathscr{T}_{e^{\mathscr{Q}(\tau_{k+1}-\tau_k)}}$ is the set of transition matrices that dominate $e^{\mathscr{Q}(\tau_{k+1}-\tau_k)}$. Because $P \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}}$ it therefore holds that $P \sim \mathscr{M}'$, which by the definition in Section 3.3.1$_{102}$ means that there is some $q \in \mathscr{M}'$ such that $P(X_{\tau_0} = x) = q(x)$ for all $x \in \mathscr{X}$. Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 7.9$_{352}$ that there is some $P_0 \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that

$$P(X_{\tau_0} = x) = q(x) = P_0(X_{\tau_0} = x) \quad \text{for all } x \in \mathscr{X}\,. \tag{7.19}$$

This P_0 will be the first element of the sequence $\{P_n\}_{n\in\mathbb{Z}_{\geq 0}}$, and the remainder of the sequence will now be constructed iteratively. So, fix any $n \in \mathbb{Z}_{\geq 0}$ and consider P_n; we will now construct P_{n+1}, as follows.

Let $v := \tau_{0:(n-1)}$, so that $v = \emptyset$ if $n = 0$, and let $u := \tau_{0:n}$, so that $u = v \cup \{\tau_n\}$. Then, because $P \in \mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_k),\mathscr{M}'}$ it holds that $P \sim (\mathscr{T}_k)$. By Definition 3.9_{102} this implies that for all $x_v \in \mathscr{X}_v$ there is some $T_{x_v} \in \mathscr{T}_n$ such that $T_{x_v} = {}^{P}T_{n,x_v}$, where ${}^{P}T_{n,x_v}$ is the history-dependent transition matrix corresponding to P, as in Definition 3.8_{101}.

Because $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows, it follows from Lemma 7.8_{352} that $\mathscr{T}_n = {}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{\tau_{n+1}}_{\tau_n}$, where ${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{\tau_{n+1}}_{\tau_n}$ is the set of history-dependent transition matrices corresponding to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Because $T_{x_v} \in \mathscr{T}_n$, this implies, together with Lemma 5.42_{243}—since $\mathscr{Q}$ is non-empty, compact, convex, and has separately specified rows—and Definition 4.3_{150}, that for all $x_{\tau_n} \in \mathscr{X}_{\tau_n}$, there is a Markov chain $P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ that satisfies, for all $y \in \mathscr{X}$,

$$\begin{aligned}
P_{x_u}(X_{\tau_{n+1}} = y \,|\, X_{\tau_n} = x_{\tau_n}) &= T_{x_v}(x_{\tau_n}, y) \\
&= {}^{P}T_{n,x_v}(x_{\tau_n}, y) \\
&= P(X_{\tau_{n+1}} = y \,|\, X_{\tau_n} = x_{\tau_n}, X_v = x_v) \\
&= P(X_{\tau_{n+1}} = y \,|\, X_u = x_u)\,. \qquad (7.20)
\end{aligned}$$

By repeating this selection for all $x_v \in \mathscr{X}_v$ and all $x_{\tau_n} \in \mathscr{X}_{\tau_n}$, we obtain a (possibly different) Markov chain $P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ for every $x_u \in \mathscr{X}_u$. We will now construct P_{n+1} by combining P_n with this collection of Markov chains, as follows.

First let

$$\mathscr{C}^n_{\emptyset} := \{(A, X_v = x_v) \in \mathscr{C}^{\mathrm{SP}} \colon v \in \mathscr{U}_{<\tau_n} \text{ and } A \in \langle\{(X_t = x) \colon x \in \mathscr{X}, t \in [0,\tau_n]\}\rangle\}\,. \quad (7.21)$$

Because $u = \tau_{0:n}$ and hence $\max u = \tau_n$, it follows that $\mathscr{C}^n_{\emptyset}$ is equal to the set $\mathscr{C}_{\emptyset}$ defined in Equation $(5.29)_{215}$—this simply serves to make the dependence on n notationally explicit. Because $\mathscr{Q}$ is non-empty, convex, and has separately specified rows, because $\mathscr{M}$ is non-empty, and because $P_n, P_{x_u} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ for all $x_u \in \mathscr{X}_u$, it follows from Lemma 5.35_{215} that there is some $P_{n+1} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ such that for all $(A,C) \in \mathscr{C}^n_{\emptyset}$ it holds that

$$P_{n+1}(A \,|\, C) = P_n(A \,|\, C)\,, \qquad (7.22)$$

and, for all $x_u \in \mathscr{X}_u$ and $A \in \mathscr{A}_u$—with $\mathscr{A}_u$ as in Section 2.3_{64}—it holds that

$$P_{n+1}(A \,|\, X_u = x_u) = P_{x_u}(A \,|\, X_u = x_u)\,. \qquad (7.23)$$

This concludes the construction of P_{n+1}.

In order to construct P_*, we define, for all $(A,C) \in \mathscr{C}^{\mathrm{SP}}$,

$$P_*(A\,|\,C) := \lim_{n\to+\infty} P_n(A\,|\,C)\,. \tag{7.24}$$

We will now show that this limit is well-defined, by proving an important property: as we are about to show, for all $(A,C) \in \mathscr{C}^{\mathrm{SP}}$ there is some $N \in \mathbb{Z}_{\geq 0}$ such that $P_n(A\,|\,C) = P_N(A\,|\,C)$ for all $n \geq N$. That is, the processes P_n eventually all agree on the probability of any event in their domain. Put differently, this essentially means that the sequence $\{P_n\}_{n\in\mathbb{Z}_{\geq 0}}$ converges pointwise, but specifically in a way that these pointwise limits are each reached after finitely many steps in the sequence (and of course N can depend on (A,C)); P_* is defined as this pointwise limit.

To establish this property, for any $u,v \in \mathscr{U}$ let first

$$N_{u,v} := \min\{n \in \mathbb{Z}_{\geq 0} \,:\, u < \tau_n,\ v \leq \tau_n\}\,.$$

In other words, $N_{u,v}$ is the smallest time point in $\mathbb{D}$ that is *strictly* greater than $\max u$, and at least as great as $\max v$. Next, consider any $u,v \in \mathscr{U}$ such that $v \subset u \cup \mathbb{R}_{>u}$ and consider any $x_u \in \mathscr{X}_u$ and $y_v \in \mathscr{X}_v$. Then it follows from Definition 2.10$_{67}$ that $(X_v = y_v, X_u = x_u) \in \mathscr{C}^{\mathrm{SP}}$. Let $n \in \mathbb{Z}_{\geq 0}$ be such that $n \geq N_{u,v}$. Then $u \in \mathscr{U}_{<\tau_n}$ and $v \subset [0,\tau_n]$, and therefore,

$$(X_v = y_v) \in \langle\{(X_t = x)\colon x \in \mathscr{X}, t \in [0,\tau_n]\}\rangle\,.$$

Hence it follows from Equation (7.21) that $(X_v = y_v, X_u = x_u) \in \mathscr{C}^n_{\emptyset}$ and therefore, by Equation (7.22), that

$$P_{n+1}(X_v = y_v\,|\,X_u = x_u) = P_n(X_v = y_v\,|\,X_u = x_u)\,.$$

Because this is true for all $n \in \mathbb{Z}_{\geq 0}$ such that $n \geq N_{u,v}$, it follows that for all $k \in \mathbb{Z}_{\geq 0}$ it holds that

$$P_{N_{u,v}+k}(X_v = y_v\,|\,X_u = x_u) = P_{N_{u,v}}(X_v = y_v\,|\,X_u = x_u)\,. \tag{7.25}$$

Having established this property, it will be straightforward to show that P_* is well-defined. Fix any $(A,C) \in \mathscr{C}^{\mathrm{SP}}$. Then it follows from Lemma 7.11$_{355}$ that there are $u,v \in \mathscr{U}$, $x_u \in \mathscr{X}_u$, and $S \subseteq \mathscr{X}_v$ such that $C = (X_u = x_u)$, $v \subset u \cup \mathbb{R}_{>u}$, and $A = \cup_{y_v\in S}(X_v = y_v)$. Moreover, it follows from Equation (7.25) that, for all $y_v \in S$ and all $k \in \mathbb{Z}_{\geq 0}$ it holds that

$$P_{N_{u,v}+k}(X_v = y_v\,|\,C) = P_{N_{u,v}}(X_v = y_v\,|\,C)\,,$$

which, using Property F3$_{47}$ and that $P_{N_{u,v}+k}$ and $P_{N_{u,v}}$ are coherent conditional probabilities, implies that

$$\begin{aligned}P_{N_{u,v}+k}(A\,|\,C) &= \sum_{y_v\in S} P_{N_{u,v}+k}(X_v = y_v\,|\,C)\\ &= \sum_{y_v\in S} P_{N_{u,v}}(X_v = y_v\,|\,C) = P_{N_{u,v}}(A\,|\,C)\,.\end{aligned} \tag{7.26}$$

Because the $k \in \mathbb{Z}_{\geq 0}$ in Equation $(7.26)_{\curvearrowleft}$ is arbitrary, it follows that

$$\begin{aligned} P_*(A \,|\, C) &= \lim_{n \to +\infty} P_n(A \,|\, C) \\ &= P_{N_{u,v}}(A \,|\, C) = \sum_{y_v \in S} P_{N_{u,v}}(X_v = y_v \,|\, X_u = x_u)\,. \end{aligned} \tag{7.27}$$

It will be helpful to establish some of the intuition behind this statement. Essentially we have noted that every conditional event (A,C) depends on a finite number of time points $u \cup v$ (due to Lemma 7.11_{355}), and that the sequence $\{P_n(A|C)\}_{n \in \mathbb{Z}_{\geq 0}}$ is constant for $n \geq N_{u,v}$, where $N_{u,v}$ indexes the first (discrete) time point $\tau_{N_{u,v}}$ after the time points $u \cup v$ (and whether it comes strictly after depends on how u and v are related). Put differently, for any time point $\tau_n \in \mathbb{D}$, the probabilities that P_* assigns to events that depend only on time points *before* τ_n, are already given by P_n (and sometimes by P_{n-1}, depending on the specific time points). This will be useful in the remainder of the proof because, once we have determined an event of interest, we now know that we can assess its probabilities by considering a particular element of the sequence $\{P_n\}_{n \in \mathbb{Z}_{\geq 0}}$.

Moving on, Equation (7.27) clearly also shows that P_* is well-defined. So, at this point we have identified P_*, and we now need to show that (i) P_* is an element of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ and (ii) that $P_*|_{\mathbb{D}} = P$.

In order to establish that $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, we need to show that it is a well-behaved stochastic process that is consistent with both $\mathscr{Q}$ and $\mathscr{M}$. Let us start by showing that it is a stochastic process, or in other words, using Definition 2.11_{68}, that P_* is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$. Because, for all $n \in \mathbb{Z}_{\geq 0}$, P_n is a stochastic process, it follows from Definition 2.11_{68} that P_n is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$, which by Definition 2.2_{48} means that P_n is a real-valued map. Using Equation (7.27), this implies that P_* is also a real-valued map, and it is clearly defined on $\mathscr{C}^{\mathrm{SP}}$. Let us next establish that P_* satisfies the crucial coherence condition. So, fix any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1,\dots,n\}$, any $\lambda_i \in \mathbb{R}$ and $(A_i,C_i) \in \mathscr{C}^{\mathrm{SP}}$. By Definition 2.2_{48}, we need to show that

$$\max\left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega)\big(P_*(A_i \,|\, C_i) - \mathbb{I}_{A_i}(\omega)\big) \,\middle|\, \omega \in C_0 \right\} \geq 0\,, \tag{7.28}$$

with $C_0 := \cup_{i=1}^{n} C_i$. For all $i \in \{1,\dots,n\}$, let $N_i := N_{(A_i,C_i)}$ be such that $P_*(A_i \,|\, C_i) = P_{N_i+k}(A_i \,|\, C_i)$ for all $k \in \mathbb{Z}_{\geq 0}$; this N_i exists by Equations $(7.26)_{\curvearrowleft}$ and (7.27). Let $N := \max_{i=1}^{n} N_i$; then, for all $i \in \{1,\dots,n\}$ it holds that $N \geq N_i = N_{(A_i,C_i)}$, which due to Equations $(7.26)_{\curvearrowleft}$ and (7.27) implies that $P_*(A_i \,|\, C_i) = P_{N_i}(A_i \,|\, C_i) = P_N(A_i \,|\, C_i)$. Because we know that

P_N is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$, it holds that

$$\max\left\{\sum_{i=1}^{n}\lambda_i\mathbb{I}_{C_i}(\omega)\big(P_N(A_i\,|\,C_i)-\mathbb{I}_{A_i}(\omega)\big)\,\middle|\,\omega\in C_0\right\}\geq 0\,,$$

which, because $P_*(A_i\,|\,C_i)=P_N(A_i\,|\,C_i)$ for all $i\in\{1,\ldots,n\}$, immediately implies that the inequality in Equation (7.28) must also hold. Therefore, it follows from Definition 2.2$_{48}$ that P_* is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$, or in other words, that it is a continuous-time stochastic process.

Let us show next that $P_*\sim\mathscr{M}$. With $u=\emptyset$ and $v=\{0\}$ we have that $N_{u,v}=0$, and hence it follows from Equations (7.25)$_{357}$ and (7.27) that $P_*(X_0=x)=P_{N_{u,v}}(X_0=x)=P_0(X_0=x)$ for all $x\in\mathscr{X}$. Because $P_0\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, this implies that there is some $q\in\mathscr{M}$ such that $P_*(X_0=x)=P_0(X_0=x)=q(x)$, whence $P_*\sim\mathscr{M}$.

In order to prove that P_* is well-behaved and consistent with $\mathscr{Q}$, it will be helpful to first establish a connection between the history-dependent transition matrices ${}^{P_*}T^{s}_{t,x_u}$ and ${}^{P_n}T^{s}_{t,x_u}$ corresponding to P_* and the processes P_n, $n\in\mathbb{Z}_{\geq 0}$, respectively. To this end, fix any $r\in\mathbb{R}_{\geq 0}$, and let $N_r=\min\{n\in\mathbb{Z}_{\geq 0}\,:\,r<\tau_n\}$. Then for all $t,s\in\mathbb{R}_{\geq 0}$ with $t\leq s\leq r$, and all $u\in\mathscr{U}_{<t}$, it holds that $N_{u\cup\{t\},s}\leq N_r$, and hence it follows from Equations (7.26)$_{357}$ and (7.27) that for all $x_u\in\mathscr{X}_u$ and $x,y\in\mathscr{X}$ it holds that

$$P_*(X_s=y\,|\,X_t=x,X_u=x_u)=P_{N_r}(X_s=y\,|\,X_t=x,X_u=x_u)\,.$$

It therefore follows from Definition 4.2$_{148}$ that ${}^{P_*}T^{s}_{t,x_u}={}^{P_{N_r}}T^{s}_{t,x_u}$.

Using this property, let us first show that P_* is well-behaved. To this end, fix any $t\in\mathbb{R}_{\geq 0}$, any $u\in\mathscr{U}_{<t}$, and any $x_u\in\mathscr{X}_u$. Fix an arbitrary $\delta\in\mathbb{R}_{>0}$. Then, due to the above, with $r=t+\delta$, it holds for all $\Delta\in\mathbb{R}_{>0}$ with $\Delta\leq\delta$, that ${}^{P_*}T^{t+\Delta}_{t,x_u}={}^{P_{N_r}}T^{t+\Delta}_{t,x_u}$ and, if $t-\Delta\geq 0$ and $t-\Delta>u$, also ${}^{P_*}T^{t}_{t-\Delta,x_u}={}^{P_{N_r}}T^{t}_{t-\Delta,x_u}$. Because $P_{N_r}\in\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ we know that P_{N_r} is well-behaved, and therefore it follows from Proposition 4.2$_{149}$ that

$$\limsup_{\Delta\to 0^+}\frac{1}{\Delta}\left\|{}^{P_*}T^{t+\Delta}_{t,x_u}-I\right\|=\limsup_{\Delta\to 0^+}\frac{1}{\Delta}\left\|{}^{P_{N_r}}T^{t+\Delta}_{t,x_u}-I\right\|<+\infty\,.\tag{7.29}$$

Similarly, if $t\neq 0$ it follows from Proposition 4.2$_{149}$, from the well-behavedness of P_{N_r}, and from the choice of N_r that also

$$\limsup_{\Delta\to 0^+}\frac{1}{\Delta}\left\|{}^{P_*}T^{t}_{t-\Delta,x_u}-I\right\|=\limsup_{\Delta\to 0^+}\frac{1}{\Delta}\left\|{}^{P_{N_r}}T^{t}_{t-\Delta,x_u}-I\right\|<+\infty\,.\tag{7.30}$$

Because Equations (7.29) and (7.30) hold for all $x_u\in\mathscr{X}_u$, all $u\in\mathscr{U}_{<t}$, and all $t\in\mathbb{R}_{\geq 0}$, it follows from Proposition 4.2$_{149}$ that P_* is well-behaved.

That $P_*\sim\mathscr{Q}$ is established analogously. Fix any $t\in\mathbb{R}_{\geq 0}$, any $u\in\mathscr{U}_{<t}$, and any $x_u\in\mathscr{X}_u$; according to Definition 5.3$_{189}$ we need to show that

$\overline{\partial}\,{}^{P_*}T^t_{t,x_u} \subseteq \mathscr{Q}$. Choose an arbitrary $\delta \in \mathbb{R}_{>0}$. Then with $r = t + \delta$, it holds for all $\Delta \in \mathbb{R}_{>0}$ with $\Delta \leq \delta$ that ${}^{P_*}T^{t+\Delta}_{t,x_u} = {}^{P_{N_r}}T^{t+\Delta}_{t,x_u}$ and, if $t - \Delta \geq 0$ and $t - \Delta > u$, also ${}^{P_*}T^t_{t-\Delta,x_u} = {}^{P_{N_r}}T^t_{t-\Delta,x_u}$. It immediately follows from Definition 4.11$_{168}$ that this implies that $\overline{\partial}\,{}^{P_*}T^t_{t,x_u} = \overline{\partial}\,{}^{P_{N_r}}T^t_{t,x_u}$. Because $P_{N_r} \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ it follows that

$$\overline{\partial}\,{}^{P_*}T^t_{t,x_u} = \overline{\partial}\,{}^{P_{N_r}}T^t_{t,x_u} \subseteq \mathscr{Q}\,,$$

and since this is true for all $x_u \in \mathscr{X}_u$, all $u \in \mathscr{U}_{<t}$, and all $t \in \mathbb{R}_{\geq 0}$, we conclude that $P_* \sim \mathscr{Q}$.

In summary, we have established that P_* is a stochastic process that is well-behaved and consistent with both $\mathscr{M}$ and $\mathscr{Q}$, so it follows from Definition 5.5$_{189}$ that $P_* \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Our next and final goal will be to establish that the restriction $P_*|_{\mathbb{D}}$ of P_* to $\mathbb{D}$ coincides with P.

To this end, we first need some additional properties of P_* that specifically deal with the time points in $\mathbb{D}$. First, with $u = \emptyset$ and $v = \{0\}$ it holds that $N_{u,v} = 0$, and hence it follows from Equations (7.25)$_{357}$ and (7.27)$_{358}$ that for all $x \in \mathscr{X}$ it holds that

$$P_*(X_{\tau_0} = x) = \lim_{n\to+\infty} P_n(X_{\tau_0} = x) = P_0(X_{\tau_0} = x) = P(X_{\tau_0} = x)\,, \tag{7.31}$$

where for the final equality we used Equation (7.19)$_{355}$. In words, this establishes a correspondence between P_* and P for the state at time τ_0.

Next, fix any $n \in \mathbb{Z}_{\geq 0}$. Then with $u = \tau_{0:n}$ and $v = \{t_{n+1}\}$ it holds that $N_{u,v} = n+1$. Hence, it follows from Equations (7.25)$_{357}$ and (7.27)$_{358}$ that for all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$ and all $y \in \mathscr{X}$ it holds that

$$\begin{aligned}
P_*(X_{\tau_{n+1}} = y \,|\, X_u = x_u) &= P_{n+1}(X_{\tau_{n+1}} = y \,|\, X_u = x_u) \\
&= P_{x_u}(X_{\tau_{n+1}} = y \,|\, X_u = x_u) \\
&= P_{x_u}(X_{\tau_{n+1}} = y \,|\, X_{\tau_n} = x_{\tau_n}) = P(X_{\tau_{n+1}} = y \,|\, X_u = x_u)\,,
\end{aligned}$$

where for the second equality we used the definition of P_{n+1} in Equation (7.23)$_{356}$ and the fact that $(X_{\tau_{n+1}} = y) \in \mathscr{A}_u$; for the third equality we used that P_{x_u} is a Markov chain, i.e. we used the Markov property; and for the final equality we used Equation (7.20)$_{356}$. So, we have found that for all $n \in \mathbb{Z}_{\geq 0}$, all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, and all $y \in \mathscr{X}$, it holds that

$$P_*(X_{\tau_{n+1}} = y \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) = P(X_{\tau_{n+1}} = y \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}})\,, \tag{7.32}$$

which establishes a correspondence between P_* and P for these specific conditional probabilities.

We now have all the pieces to complete the proof. Let $P_*|_{\mathbb{D}}$ be the restriction of P_* to $\mathbb{D}$, as in Definition 7.3$_{350}$. Then it follows from Equations (7.11)$_{348}$ and (7.31) that for all $x \in \mathscr{X}$ it holds that

$$P_*|_{\mathbb{D}}(X_{\tau_0} = x) = P_*(X_{\tau_0} = x) = P(X_{\tau_0} = x)\,. \tag{7.33}$$

Moreover, for all $n \in \mathbb{Z}_{\geq 0}$, all $x_{\tau_{0:n}} \in \mathscr{X}_{\tau_{0:n}}$, and all $y \in \mathscr{X}$, it follows from Equations (7.11)$_{348}$ and (7.32) that

$$\begin{aligned} P_*|_{\mathbb{D}}(X_{\tau_{n+1}} = y \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) &= P_*(X_{\tau_{n+1}} = y \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) \\ &= P(X_{\tau_{n+1}} = y \,|\, X_{\tau_{0:n}} = x_{\tau_{0:n}}) . \end{aligned} \tag{7.34}$$

Because P and $P_*|_{\mathbb{D}}$ are both elements of $\mathbb{P}^{\mathbb{D}}$, it follows from Equations (7.33) and (7.34) and Corollary 3.6$_{88}$ that $P = P_*|_{\mathbb{D}}$. ☐

Hence we obtain the following result.

Theorem 7.13. *Let $\mathscr{Q}$ be a non-empty, compact, and convex set of rate matrices that has separately specified rows, let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, and let $\mathbb{D}$ be a discrete-time domain. Then*

$$\mathbb{P}^{\mathbb{D}}_{\mathscr{Q},\mathscr{M}} = \left\{ P|_{\mathbb{D}} \,\middle|\, P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \right\} .$$

Proof. This is immediate from Lemmas 7.10$_{354}$ and 7.12$_{355}$. ☐

In summary, in this chapter we have established an important connection between discrete-time and continuous-time imprecise-Markov chains: under some regularity conditions on the parameters, we can "restrict" a given continuous-time imprecise-Markov chain to any desired discrete time domain, in order to obtain a discrete-time imprecise-Markov chain. Furthermore, for functions that depend only on the state at (finitely many) time points in this discrete time domain, the lower expectations for these two models coincide.

Consequently, we might say that there does not seem to be a *practical* reason to consider the technical minutiae of continuous-time imprecise-Markov chains, provided that one is only interested in inferences that depend on finitely-many time points.[3] That is, although continuous-time models may be easier or more natural to conceptualise for certain applications, the numerical methods from Chapter 6$_{259}$ and the results from this current chapter allow one to reduce such finitary inferences to a discrete-time setting that, perhaps, is more intuitive or easier to work with.

It should nevertheless be noted that this connection is not entirely one-to-one. For instance, while we have shown that continuous-time models can induce discrete-time ones, the converse is not necessarily true: essentially, there are discrete-time stochastic processes that cannot be extended to continuous-time ones. Moreover, we suspect that in

[3]Of course, we have spent a large part of this dissertation on exploring these technical minutiae in great detail, before we could reach this conclusion.

(future) work that deals with inferences that depend on infinitely many time points—in particular, where these time points are dense in (parts of) the continuous time domain—one will have to deal with the technical details of continuous-time stochastic processes explicitly; and we expect that some of our results will prove essential to achieve this.

8

Conclusions

"The wind blew southward, through knotted forests, over shimmering plains and toward lands unexplored."

Robert Jordan, "A Memory of Light"

In this dissertation we have developed a theory of continuous-time imprecise-Markov chains. These are imprecise-probabilistic generalisations of continuous-time Markov chains that can be used when numerical parameters are only partially specified and/or when structural assumptions like Markovianity and time-homogeneity are unwarranted. Inferences computed with these models are lower- and upper expectations of quantities of interest, which can essentially be interpreted as (tight) lower and upper bounds on the traditional probabilistic expectation of the same quantities of interest, taken with respect to the probabilistic model that is the subject of such parameter- and structural uncertainties.

The basis of our theory is a formalisation of (precise) stochastic processes using *full conditional probabilities* and *coherence*, which is notably different from the measure-theoretic formalisation that is more typical in the literature. Similarly, the corresponding notion of conditional expectations that we use derives from the related notion of *coherent conditional previsions*, and does not rely on measure-theoretic constructions like (Lebesgue-)integration. The main inferences we have concerned ourselves with are (lower- and upper) expectations of *u-measurable functions*. Using more classical terminology, these can be seen as functions that are *simple* with respect to the domain of definition of our stochastic processes. Despite the conceptual and foundational differences, as we have seen, expectations of such functions essentially agree with their typical measure-theoretic characterisations. Similarly, we have shown how continuous-time homogeneous Markov chains—under our definition—share many crucial characteristics with their more typical measure-theoretic constructions; they are uniquely characterised by the specification of an *initial distribution* and a *transition rate matrix*, and their induced family of transition matrices—which characterises its corresponding conditional expectation operator—corresponds to the *semigroup of transition matrices* generated by this transition rate matrix.

The reason for using this alternative formalisation is therefore not so much one of consequences—in the sense that we expect our results to agree to a large extent with what could be obtained had we used a measure-theoretic foundation—as it is methodological and philosophical. Here, "methodological" is taken to mean that the notion of coherence, and in particular the existence of coherent *extensions* from a chosen domain to a larger one, has provided us with a powerful tool that has been the workhorse of many of our proofs, especially where we needed to prove the existence of processes with certain characteristics. And the reason is "philosophical" in the sense that (i) the formalisation provides our theory with a clear behavioural and subjectivist interpretation, and (ii) the axiomatisation does not impose *sigma*-additivity, but only finite additivity. While we are not vehemently opposed to using sigma-additivity, the point is that we have managed to obtain all of our results *without* assuming this sometimes controversial axiom,[1] so we see no real reason to adopt it here. If we had to point to one difference that we *do* expect to come out of a similar treatment when using a measure-theoretic basis, then it would be that we have managed to obtain uniqueness results that might not hold in the more traditional framework. In particular, we have not had to deal with results that hold "almost surely", that is, that are only true up to modification of processes on a null-set. We believe that this makes the theory somewhat more elegant, but we leave it to the reader to appreciate the real or potential relevance of this distinction.

We also believe that the reason that we did not need to invoke sigma-additivity, and did not encounter substantial differences from what would be expected under an alternative measure-theoretic formalisation, is that in this dissertation we have remained solidly in the

[1] The success of the measure-theoretic school notwithstanding.

realm of problems that are, in a definite sense, *finitary*. In particular, we have only dealt with systems with finitely many states, with algebras of events that are only closed with respect to finitely many algebraic operations, and with functions that depend on the state of the system at finitely many time points. Therefore it does not seem surprising that we did not have to rely on notions like sigma-additivity, which is fundamentally infinitary. On the other hand, we also believe that the most interesting and important future work lies in extending and generalising our results to settings that go beyond such finite domains. We expect that in doing so, fundamental problems may crop up that cannot be solved without either imposing further regularity conditions such as sigma-additivity or continuity under monotone convergence; or by breaking away from what is expected under measure-theoretic formalisations and ending up with substantially different (more conservative) results.

Moving on, using our formalisation of stochastic processes we have introduced three distinct definitions of what we call a *continuous-time imprecise-Markov chain*. All three of these types of models are parameterised using a *set* $\mathscr{M}$ of initial distributions and a *set* $\mathscr{Q}$ of transition rate matrices. They are all *sets* of stochastic processes which are *consistent* with $\mathscr{M}$ and $\mathscr{Q}$; this means that the initial distribution of each of their elements is in $\mathscr{M}$ and, roughly speaking, that the dynamical behaviour of their elements can be described using the transition rate matrices in $\mathscr{Q}$. Moreover, all three models contain only *well-behaved* stochastic processes, which is a condition to which we adhered to prevent the consideration of processes whose behaviour is overly pathological. The difference between these types of continuous-time imprecise-Markov chains is in the structural properties of their elements: they correspond to sets of homogeneous Markov chains, sets of—not necessarily homogeneous—Markov chains, and sets of general—not necessarily homogeneous or Markovian—stochastic processes.

The reason that we call all three models *imprecise-Markov chains* is that they all satisfy an *imprecise-Markov property*; albeit, for the third of these models, only under some conditions on $\mathscr{Q}$. This means in particular that the corresponding lower (and hence also upper) expectations for these models are history-independent in a manner that is entirely analogous to the classical Markov property.

We have argued that the most straightforward of these models—$\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$, the set of all continuous-time homogeneous Markov chains that are well-behaved and consistent with $\mathscr{M}$ and $\mathscr{Q}$—is conceptually simple, but essentially does not admit sufficiently many degrees of freedom to make working with it practically feasible. In particular, the non-linear optimisation problem that is central to the computation of its lower (and hence upper) expectations cannot really be simplified.

This is analogous to the difficulty of working with the conceptually related *discrete-time* imprecise-Markov chains that consist of discrete-time homogeneous Markov chains, for which this issue has long been acknowledged in the literature [69].

On the other end of the complexity scale, we have the model $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, which consists of *all* well-behaved stochastic processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$. Its elements need not be time-homogeneous or even Markovian, and their dynamic behaviour can depend almost arbitrarily on their history, provided that the well-behavedness and consistency conditions are satisfied. While conceptually the most complicated of our three models, this is also the one for which we were able to prove the most powerful properties. Notably, Theorems 5.11_{193}, 5.21_{198}, 5.32_{208}, 6.2_{261}, 6.4_{263}, 6.28_{283}, 6.33_{289}, 7.2_{339}, 7.13_{361} and Algorithm 2_{286} are all fundamental and crucial results for the analysis and practical use of continuous-time imprecise-Markov chains, and which we have *only* been able to prove for this most imprecise of our models, $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. We deem the results that are listed above to be amongst the most important in this dissertation.

Finally, in between these two extremes lies $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$: the set of all well-behaved continuous-time Markov chains that are consistent with $\mathscr{M}$ and $\mathscr{Q}$, but which need not be homogeneous. We have seen that for this model we *can* develop fairly efficient algorithms to compute lower (and hence upper) expectations of functions that depend on the state of the underlying system at a single point in time, and that these coincide with such inferences for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. However, we have also established that for functions that depend on the state of the system at more than one time point, computational methods like Algorithm 2_{286} do *not* work (in general) for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$.

One thing that is important to reiterate is that these three definitions are nested subsets of each other, in the sense that $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. Consequently, even if the model that one would like to work with is, say, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, the lower and upper expectations computed for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ are guaranteed to be conservative, or "cautious", approximations to the lower and upper expectations for the intended model $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$. As such, while some "precision" may be lost by working with this more general model, it has the benefit of being computationally more tractable, while still *guaranteeing* that its lower and upper expectations are robust bounds on the inference of interest.

In the same vein, many of our results for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ require certain structural properties of the set $\mathscr{Q}$; in particular, we have often seen that this set needs to be non-empty, compact (i.e. closed and bounded), convex, and/or have separately specified rows. Two of these properties, *viz.* non-emptiness and boundedness, will probably be trivially satisfied in practice; non-emptiness because it gives us a model to talk about in the first place, and boundedness because physical systems arguably do not evolve arbitrarily quickly. The other conditions—closedness, convexity, and having separately specified rows—are all, in a sense, *closure* properties. We mean this in the sense that given any set $\mathscr{Q}$ that is non-empty and bounded, we can construct a set $\mathscr{Q}' \supseteq \mathscr{Q}$ that is non-empty, compact, convex, and has separately specified rows: simply first construct the closed convex hull of $\mathscr{Q}$, and then close that set under all row-wise combinations of its elements, in the sense of Definition 5.7$_{193}$. Then $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}',\mathscr{M}}$ and, hence, any inferences computed from the model induced by $\mathscr{Q}'$ are guaranteed to be conservative bounds on the intended inferences of interest with respect to $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$.

In summary, we conclude that the continuous-time imprecise-Markov chain $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is the most analytically and computationally tractable one out of the three definitions that we introduced. We believe that the properties of this model make it the most promising one to focus on for future theoretical work, and the most practical to use from a computational point of view. Perhaps unsurprisingly, this is analogous to the developments in the literature on discrete-time imprecise-Markov chains, where the introduction of the analogous model by De Cooman and Hermans [20] opened up many avenues for theoretical and algorithmic development. Our results in Chapter 7$_{335}$ allow us to translate to the continuous-time setting many of these existing algorithms from the literature—we believe with minimal future effort.

So let us now end on a historical note. It was Kolmogorov who initially extended Markov's ideas to the continuous-time setting [113]. At the time, Kolmogorov observed:

> *"[...] it is* a matter of indifference *which of the two following assumptions is made: either the time variable t runs through all real values, or only through the integers. The classical understanding of Markov chains corresponds to the second possibility."* [56, emphasis ours]

We are happy to repeat this conclusion for imprecise-Markov chains.

A

Analysis in Finite-Dimensional Normed Vector Spaces

"Hmm"

Geralt of Rivia, The Witcher

This appendix contains some general results about the analysis of sets, limits, and sequences in (finite-dimensional) vector spaces, and on which we rely throughout this dissertation. We start this appendix by stating some of the key definitions and properties in a fairly general setting, and then discuss the relevant special cases in Sections A.1 $_{379}$–A.3 $_{383}$. In particular, we discuss there how the space $\mathscr{L}(\mathscr{X})$ of real-valued functions on $\mathscr{X}$, its dual space of real linear functionals on $\mathscr{L}(\mathscr{X})$, and the space of linear maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, are all special cases to which we can apply the machinery developed in the first part of this appendix.

None of the results that we present here is new, although we do provide some explicit proofs for statements where we could not easily find a reference stating the exact result. Throughout, we largely base ourselves on the developments in References [9, 51, 100].

Finally, a word of encouragement for the weary reader: the somewhat abstract generality in which we state these results, is merely aimed at matching them more easily with results in the literature, when we use these properties in our proofs. Unfortunately, this belies the simplicity that stems from the fact that, as we shall see, the finite dimensionality of the spaces we are working with ensures that we can think about them as simply being (at least isomorphic to) the space $\mathbb{R}^n$. Hence, the core results and intuition should, hopefully, be familiar to readers with some background in linear algebra and real analysis. With that out of the way, let us start with some definitions.

Definition A.1 ([9, Chapter 1, Definition 1.4]). *A* real vector space $\mathbb{V}$ *is a vector space over $\mathbb{R}$, i.e. a non-empty set $\mathbb{V}$ together with two operations, $\mathbb{V} \times \mathbb{V} \to \mathbb{V} : (v,w) \mapsto v+w$ (called addition) and $\mathbb{R} \times \mathbb{V} \to \mathbb{V} : (\alpha, v) \mapsto \alpha v$ (called scalar multiplication), which satisfy the following axioms:*

V1: $v+w = w+v$ *for all* $v,w \in \mathbb{V}$*;*

V2: $v+(w+u) = (v+w)+u$ *for all* $v,w,u \in \mathbb{V}$*;*

V3: *there is some* $0 \in \mathbb{V}$ *(called the* origin*) such that* $0+v = v$ *for all* $v \in \mathbb{V}$*;*

V4: *for all* $v \in \mathbb{V}$*, there is some* $-v \in \mathbb{V}$ *such that* $v+(-v) = 0$*;*

V5: $(\alpha\beta)v = \alpha(\beta v)$ *for all* $\alpha, \beta \in \mathbb{R}$ *and* $v \in \mathbb{V}$*;*

V6: $\alpha(v+w) = \alpha v + \alpha w$ *for all* $\alpha \in \mathbb{R}$ *and* $v,w \in \mathbb{V}$*;*

V7: $(\alpha+\beta)v = \alpha v + \beta v$ *for all* $\alpha, \beta \in \mathbb{R}$ *and* $v \in \mathbb{V}$*;*

V8: $1v = v$ *for all* $v \in \mathbb{V}$*.*

For any real vector space $\mathbb{V}$ and any non-empty $\mathscr{V} \subseteq \mathbb{V}$, we say that $\mathscr{V}$ is a *subspace* of $\mathbb{V}$, if $\mathscr{V}$ is a real vector space under the same addition and scalar multiplication operations as $\mathbb{V}$ [9, Chapter 1, Definition 1.12]. Moreover, for any subset $\mathscr{V} \subseteq \mathbb{V}$, we define the *linear span* [9, Chapter 1, Section 1] of $\mathscr{V}$ as

$$\operatorname{span}(\mathscr{V}) := \bigcap_{\mathscr{W} \in \boldsymbol{\mathcal{W}}} \mathscr{W},$$

where

$$\boldsymbol{\mathcal{W}} := \left\{ \mathscr{W} \subseteq \mathbb{V} \,\middle|\, \mathscr{V} \subseteq \mathscr{W},\ \mathscr{W} \text{ is a subspace of } \mathbb{V} \right\}.$$

Then $\operatorname{span}(\mathscr{V})$ is the smallest subspace of $\mathbb{V}$ that includes $\mathscr{V}$ [9, Chapter 1, Section 1]. However, this definition is fairly abstract, so the following property may be more helpful:

Proposition A.1 ([9, Chapter 1, Theorem 1.15]). *Let $\mathbb{V}$ be a real vector space and let $\mathscr{V} \subseteq \mathbb{V}$. Then*

$$\operatorname{span}(\mathscr{V}) = \left\{ \sum_{i=1}^{n} \alpha_i v_i \,\middle|\, n \in \mathbb{Z}_{\geq 0},\ \forall i \in \{1, \ldots, n\} : v_i \in \mathscr{V},\ \alpha_i \in \mathbb{R} \right\}.$$

Definition A.2 ([9, Chapter 1, Definition 2.3]). *Let $\mathbb{V}$ be a real vector space and let $\mathscr{V} \subseteq \mathbb{V}$. Then $\mathscr{V}$ is* linearly dependent *if, for some $n \in \mathbb{Z}_{>0}$, there are $\{v_1, \ldots, v_n\} \subseteq \mathscr{V}$ and non-zero $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, such that $\sum_{i=1}^{n} \alpha_i v_i = 0$. Conversely, $\mathscr{V}$ is called* linearly independent *if it is not linearly dependent.*

Definition A.3 ([9, Chapter 1, Definition 2.5]). *Let $\mathbb{V}$ be a real vector space. Then a set $\mathscr{V} \subseteq \mathbb{V}$ is called a* basis *of $\mathbb{V}$, if $\mathscr{V}$ is linearly independent, and* $\mathrm{span}(\mathscr{V}) = \mathbb{V}$.

In general, a real vector space $\mathbb{V}$ has many different bases. However, as the next result shows, they all have the same *cardinality*. The cardinality $|\mathscr{V}|$ of a set $\mathscr{V}$ essentially measures how many elements it contains; if $\mathscr{V}$ is finite then $|\mathscr{V}|$ is the number of elements it contains, and otherwise it suffices for our purposes to set $|\mathscr{V}| := +\infty$. [1]

Proposition A.2 ([9, Chapter 1, Theorem 2.12]). *Let $\mathbb{V}$ be a real vector space, and suppose that $\mathscr{V}_1 \subseteq \mathbb{V}$ and $\mathscr{V}_2 \subseteq \mathbb{V}$ are both bases of $\mathbb{V}$. Then $|\mathscr{V}_1| = |\mathscr{V}_2|$.*

This gives rise to the *dimension* of a real vector space, as follows.

Definition A.4 ([9, Chapter 1, Section 2]). *Let $\mathbb{V}$ be a real vector space. Then its dimension* $\dim(\mathbb{V})$ *is defined as the (common) cardinality of any basis of $\mathbb{V}$. $\mathbb{V}$ is called* finite-dimensional *if* $\dim(\mathbb{V}) = n$ *for some $n \in \mathbb{Z}_{\geq 0}$.*

At this point it is probably worth noting the canonical example of finite-dimensional real vector spaces: the space $\mathbb{R}^n$, with $n \in \mathbb{Z}_{>0}$. In fact, every n-dimensional real vector space is isomorphic to $\mathbb{R}^n$ [9, Chapter 1, Theorem 3.15], so for the sake of intuition we may proceed under this interpretation.

Let us now move the discussion to *normed* vector spaces.

Definition A.5 ([9, Chapter 4, Definition 1.1 and 1.2]). *A* normed vector space *is a real vector space $\mathbb{V}$ together with a map $\|\cdot\| : \mathbb{V} \to \mathbb{R}$ (called a* norm*) that satisfies*

1. $\|v\| > 0$ *for all $v \in \mathbb{V}$ such that $v \neq 0$;*
2. $\|\alpha v\| = |\alpha| \, \|v\|$ *for all $\alpha \in \mathbb{R}$ and $v \in \mathbb{V}$;*
3. $\|v + w\| \leq \|v\| + \|w\|$ *for all $v, w \in \mathbb{V}$.*

[1] Proposition A.2 holds for bases with finitely many elements as well as for bases with infinitely elements, but this requires some subtleties of the cardinality of sets that we do not deal with here; we refer to Reference [9, Chapter 1, Section 2] for details.

One norm that is often encountered is the *Euclidean* norm $\|\cdot\|_2$ on $\mathbb{R}^n$, $n \in \mathbb{Z}_{>0}$, which is defined as

$$\|v\|_2 := \sqrt{v^\top v} \quad \text{for all } v \in \mathbb{R}^n ,$$

where, for all $v \in \mathbb{R}^n$, $v^\top v$ denotes the dot product of v with itself. However, in this work we typically use other norms; see Sections 2.2.3$_{62}$ and A.1$_{379}$–A.3$_{383}$ for details.

As is well-known, normed spaces are a special case of *metric spaces*. Let us start with the general definition.

Definition A.6 ([100, Definition 1.2.1])**.** *A* metric space *is a non-empty set* $\mathbb{X}$ *together with a map* $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ *(called a* metric*), that satisfies, for all* $x, y, z \in \mathbb{X}$*,*

1. $d(x,y) \geq 0$*;*
2. $d(x,y) = 0$ *if and only if* $x = y$*;*
3. $d(x,y) = d(y,x)$*;*
4. $d(x,z) \leq d(x,y) + d(y,z)$*.*

As mentioned, any normed vector space $\mathbb{V}$ is also a metric space, with the metric $d : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ induced by the norm on $\mathbb{V}$, as

$$d(v,w) := \|v - w\| \quad \text{for all } v, w \in \mathbb{V}. \tag{A.1}$$

This metric induces a topology on $\mathbb{V}$ (specifically, the metric topology), with respect to which we can define the convergence of a sequence $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathbb{V}$. Let us start with some general definitions; we use [94, Definition 1.5] for what follows.

A *topology* τ on a set $\mathbb{X}$ is a collection of subsets of $\mathbb{X}$ such that $\mathbb{X} \in \tau$, $\emptyset \in \tau$, and such that τ is closed under finite intersections and arbitrary unions. The ordered tuple $(\mathbb{X}, \tau)$ is called a *topological space*, and the elements of τ are called the *open* sets of $\mathbb{X}$ (with respect to τ). A subset of $\mathbb{X}$ is called *closed* (with respect to τ), if its complement in $\mathbb{X}$ is open (with respect to τ).

A *neighbourhood* of $x \in \mathbb{X}$ is an element of τ that contains x. A topological space $(\mathbb{X}, \tau)$ is called a *Hausdorff space* if for all $x, y \in \mathbb{X}$ such that $x \neq y$, there are neighbourhoods N_x, N_y of x and y, respectively, such that $N_x \cap N_y = \emptyset$; τ is then called a *Hausdorff topology* (on $\mathbb{X}$).

The reason that this is interesting is that, following [94, Definition 1.5], in a Hausdorff space $(\mathbb{X}, \tau)$ one may define the convergence of a sequence $\{x_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathbb{X}$ to $x_* \in \mathbb{X}$, and write $\lim_{i \to +\infty} x_i = x_*$, if for every neighbourhood N_{x_*} of x_*, there is some $n \in \mathbb{Z}_{>0}$ such that for all $k \in \mathbb{Z}_{>0}$

with $k > n$, it holds that $x_k \in N_{x_*}$; we then say that $\{x_i\}_{i\in\mathbb{Z}_{>0}}$ *converges* to x_* with respect to τ. Specifically, it is easily seen that the Hausdorff character of the space guarantees that this limit is unique.

Let us now show that for normed vector spaces, and for finite-dimensional vector spaces in particular, this topological notion of convergence coincides with the usual notion of metric convergence.

Definition A.7 ([94, Section 1.2])**.** *Let $\mathbb{V}$ be a normed vector space with norm $\|\cdot\|$. For any $v \in \mathbb{V}$ and $r \in \mathbb{R}_{>0}$, the* open ball *in $\mathbb{V}$ with center v and radius r, is the set $B_r(v) := \{w \in \mathbb{V} : \|v - w\| < r\}$. The* metric topology τ_m *on $\mathbb{V}$ induced by the norm $\|\cdot\|$, is the collection of subsets of $\mathbb{V}$ that can be written as a (possibly empty) union of open balls in $\mathbb{V}$.*

Lemma A.3. *Let $\mathbb{V}$ be a normed vector space with norm $\|\cdot\|$, and let τ_m be the metric topology on $\mathbb{V}$ that is induced by $\|\cdot\|$. Then $(\mathbb{V}, \tau_m)$ is a Hausdorff space.*

Proof. Fix any $v, w \in \mathbb{V}$ such that $v \neq w$, and let $\delta := \|v - w\|$; then $\delta > 0$ by Definition A.5$_{371}$. Let $\delta' := \delta/2$. By Definition A.7, the open balls $B_{\delta'}(v)$ and $B_{\delta'}(w)$ are neighbourhoods of v and w, respectively. To show that $(\mathbb{V}, \tau_m)$ is Hausdorff, it suffices to show that $B_{\delta'}(v) \cap B_{\delta'}(w) = \emptyset$.

So fix any $u \in B_{\delta'}(v)$. Then, using Definition A.5$_{371}$, it holds that

$$\|v - w\| = \|v - u + u - w\| \leq \|v - u\| + \|u - w\| ,$$

which, since $\|v - u\| < \delta'$ because $u \in B_{\delta'}(v)$, implies that

$$\delta' = \delta - \delta' < \|v - w\| - \|v - u\| \leq \|u - w\| .$$

This means that $u \notin B_{\delta'}(w)$. Because this is true for any $u \in B_{\delta'}(v)$, it follows that $B_{\delta'}(v) \cap B_{\delta'}(w) = \emptyset$. □

Definition A.8 ([75, Definition 2.2.1])**.** *Let $(\mathbb{V}, \tau)$ be a topological space, where $\mathbb{V}$ is a real vector space, for which the vector space operations are continuous with respect to τ.*[2] *Then $(\mathbb{V}, \tau)$ is called a* topological vector space, *and τ is called a* vector topology.

Proposition A.4 ([75, Theorem 2.2.3])**.** *Let $\mathbb{V}$ be a normed vector space with norm $\|\cdot\|$, and let τ_m be the metric topology on $\mathbb{V}$ that is induced by $\|\cdot\|$. Then $(\mathbb{V}, \tau_m)$ is a topological vector space, and τ_m is a vector topology.*

Proposition A.5 ([94, Section 1.19])**.** *Any finite-dimensional normed vector space $\mathbb{V}$ has a unique Hausdorff vector topology τ_v.*

[2] Such that the preimage of any open set, under these operations, is itself open; for vector addition, this preimage should be open in the product topology. See also [75, Section 2.1] or [94, Section 1.4] for further information.

Corollary A.6. *Let $\mathbb{V}$ be a finite-dimensional normed vector space. Then the metric topology τ_m induced by the norm $\|\cdot\|$ is the unique Hausdorff vector topology τ_v on $\mathbb{V}$.*

Proof. By Lemma A.3↶ and Proposition A.4↶, $(\mathbb{V}, \tau_m)$ is a Hausdorff topological vector space, i.e. τ_m is a Hausdorff vector topology on $\mathbb{V}$. Because $\mathbb{V}$ is finite-dimensional, by Proposition A.5↶ there is a unique Hausdorff vector topology τ_v on $\mathbb{V}$, whence $\tau_m = \tau_v$. □

Hence, if $\mathbb{V}$ is a finite-dimensional normed vector space, then due to the above results, and because, as we mentioned earlier, $\mathbb{V}$ is then also isomorphic to $\mathbb{R}^n$, we can intuitively treat any analysis in $\mathbb{V}$ *as if* we were working in the Euclidean space $\mathbb{R}^n$ equipped with the norm $\|\cdot\|_2$; this may be helpful to provide some intuition.

Because we are working in normed/metric spaces, we can of course use notions of convergence that are much more intuitive than the abstract topological one discussed above. Let us state the following well-known definitions.

Definition A.9 ([100, Definition 1.4.1])**.** *Let $\mathbb{X}$ be a metric space. Then a sequence $\{x_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathbb{X}$ is called a* Cauchy sequence *if for all $\varepsilon > 0$, there is some $n \in \mathbb{Z}_{>0}$ such that, for all $k, \ell \in \mathbb{Z}_{>0}$ with $k > n$ and $\ell > n$, it holds that $d(x_k, x_\ell) < \varepsilon$.*

Definition A.10. *Let $\mathbb{X}$ be a metric space. Then $\mathbb{X}$ is said to be* complete *if for every Cauchy sequence $\{x_i\}_{i\in\mathbb{Z}_{>0}}$ in $\mathbb{X}$, there is some $x_* \in \mathbb{X}$ such that $\lim_{i\to+\infty} d(x_i, x_*) = 0$.*

In words, this means that in a complete metric space $\mathbb{X}$, every Cauchy sequence $\{x_i\}_{i\in\mathbb{Z}_{>0}}$ has a limit x_* that also belongs to $\mathbb{X}$.

The following property is now crucial.

Proposition A.7 ([9, Chapter 4, Corollary 4.6])**.** *Let $\mathbb{V}$ be a finite-dimensional normed vector space. Then $\mathbb{V}$ is a complete metric space under the metric induced by its norm.*

A complete normed vector space is also called a *Banach space* [94, Section 1.2]; the above result therefore implies that any finite-dimensional normed vector space is a Banach space.

Summarising the above results, we can use the following equivalent notions of convergence when working with finite-dimensional normed vector spaces (and Banach spaces more generally).

Definition A.11. *Let $\mathbb{V}$ be a Banach space with norm $\|\cdot\|$, let d be the metric induced by this norm, and let τ_m be its metric topology. Let $\{v_i\}_{i\in\mathbb{Z}_{>0}}$ be any sequence in $\mathbb{V}$. Then the following are equivalent:*

1. *for some* $v_* \in \mathbb{V}$, $\lim_{i\to+\infty} v_i = v_*$ *with respect to* τ_m*;*

2. *for some* $v_* \in \mathbb{V}$, $\lim_{i\to+\infty} d(v_i, v_*) = 0$*;*

3. *for some* $v_* \in \mathbb{V}$, $\lim_{i\to+\infty} \|v_i - v_*\| = 0$*;*

4. $\{v_i\}_{i\in\mathbb{Z}_{>0}}$ *is a Cauchy sequence with respect to d.*

If any (and therefore all) of the above properties hold, there is a (unique) element $v_* \in \mathbb{V}$ *that satisfies Properties 1, 2, and 3, and we say that the sequence* $\{v_i\}_{i\in\mathbb{Z}_{>0}}$ converges *to* v_*.

Proof. We will first show that properties 1, 2, and 3 are equivalent.

1 implies 2: Suppose that $\lim_{i\to+\infty} v_i = v_*$ with respect to τ_m. Then, by Definition A.7$_{373}$, for any $\varepsilon > 0$ the open ball $B_\varepsilon(v_*)$ is a neighbourhood of v_*, whence there is some $n \in \mathbb{Z}_{>0}$ such that for all $k > n$, it holds that $v_k \in B_\varepsilon(v_*)$. This means that $d(v_k, v_*) < \varepsilon$ by the definition of the open ball. Because this is true for all $\varepsilon > 0$, it follows that $\lim_{i\to+\infty} d(v_i, v_*) = 0$.

2 implies 3: Suppose that $\lim_{i\to+\infty} d(v_i, v_*) = 0$. Since $d(v_i, v_*) = \|v_i - v_*\|$ for all $i \in \mathbb{Z}_{>0}$ it immediately follows that $\lim_{i\to+\infty} \|v_i - v_*\| = 0$.

3 implies 1: Suppose that $\lim_{i\to+\infty} \|v_i - v_*\| = 0$. Let N_{v_*} be a neighbourhood of v_*. By Definition A.7$_{373}$, N_{v_*} can be written as a (possibly empty) union of open balls in $\mathbb{V}$. However, this union cannot be empty because that would imply that $N_{v_*} = \emptyset$, which is impossible since $v_* \in N_{v_*}$ by the definition of a neighbourhood. Hence, there is some $v \in \mathbb{V}$ and $r \in \mathbb{R}$, such that $v_* \in B_r(v) \subseteq N_{v_*}$. Let $\varepsilon := r - \|v_* - v\|$; then $\varepsilon > 0$ because $\|v_* - v\| < r$ since $v_* \in B_r(v)$. Because $\lim_{i\to+\infty} \|v_i - v_*\| = 0$, there is some $n \in \mathbb{Z}_{\geq 0}$ such that, for all $k > n$, it holds that $\|v_k - v_*\| < \varepsilon$. Using Definition A.5$_{371}$, this means that also

$$\|v_k - v\| \leq \|v_k - v_*\| + \|v_* - v\| < \varepsilon + \|v_* - v\| = r\,,$$

which implies that $v_k \in B_r(v) \subseteq N_{v_*}$. Because this is true for any neighbourhood of v_*, it follows that $\lim_{i\to+\infty} v_i = v_*$ with respect to τ_m.

Let us next show that if there is some $v_* \in \mathbb{V}$ that satisfies property 2 (and hence, as we have already established, also properties 1 and 3), that this v_* is then unique. So suppose that there are $v_*, w_* \in \mathbb{V}$ such that $\lim_{i\to+\infty} d(v_i, v_*) = 0$ and $\lim_{i\to+\infty} d(v_i, w_*) = 0$; we will show that $v_* = w_*$. By Definition A.6$_{372}$ it holds that $d(v_*, w_*) \leq d(v_*, v_i) + d(v_i, w_*)$ for all $i \in \mathbb{Z}_{>0}$. Because $\lim_{i\to+\infty} d(v_i, v_*) = 0$ and $\lim_{i\to+\infty} d(v_i, w_*) = 0$ this implies that $d(v_*, w_*) = 0$ which, by Definition A.6$_{372}$, in turn implies that $v_* = w_*$.

To complete the proof we need to show that property 4 holds if and only if one (and therefore all) of the other properties hold.

2 implies 4: Suppose that $\lim_{i\to+\infty} d(v_i, v_*) = 0$. Fix any $\varepsilon > 0$. Then there is some $n \in \mathbb{Z}_{>0}$ such that, for all $k > n$, it holds that $d(v_k, v_*) < \varepsilon/2$.

Using Definition A.6$_{372}$, it follows that for all $k, \ell \in \mathbb{Z}_{>0}$ such that $k > n$ and $\ell > n$, it holds that $d(v_k, v_\ell) \leq d(v_k, v_*) + d(v_*, v_\ell) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. Because this is true for all $\varepsilon > 0$, it follows from Definition A.9$_{374}$ that $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ is a Cauchy sequence.

4$_{\curvearrowleft}$ implies 2$_{\curvearrowleft}$: Suppose that $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ is a Cauchy sequence. Because $\mathbb{V}$ is a Banach space, by definition it is a complete metric space under the metric d. Hence, by Definition A.10$_{374}$, there is some $v_* \in \mathbb{V}$ such that $\lim_{i \to +\infty} d(v_i, v_*) = 0$. □

We will next need some properties of subsets of normed vector spaces. First, for any normed vector space $\mathbb{V}$ with norm $\|\cdot\|$, and any $\mathscr{V} \subseteq \mathbb{V}$, we define

$$\|\mathscr{V}\| := \sup\big\{\|v\| \,:\, v \in \mathscr{V}\big\}\,.$$

We use the following terminology throughout this dissertation.

Definition A.12. *Let $\mathbb{V}$ be a finite-dimensional normed vector space, and let $\mathscr{V} \subseteq \mathbb{V}$. We then say that $\mathscr{V}$ is, respectively,*

S1: non-empty *if $\mathscr{V} \neq \emptyset$;*

S2: convex *if $\lambda v + (1-\lambda) w \in \mathscr{V}$ for all $\lambda \in [0,1]$ and $v, w \in \mathscr{V}$;*

S3: open *if $\mathscr{V}$ is open with respect to the metric topology on $\mathbb{V}$;*

S4: closed *if $\mathscr{V}$ is closed with respect to the metric topology on $\mathbb{V}$;*

S5: bounded *if $\|\mathscr{V}\| < +\infty$;*

S6: compact *if $\mathscr{V}$ is compact in the metric topology on $\mathbb{V}$.*

The terminology of Properties S1 and S2 should be obvious. Properties S3 and S4 are perhaps a bit less transparent. We repeat that $\mathscr{V}$ is closed if and only if its complement $\mathscr{V}^c := \mathbb{V} \setminus \mathscr{V}$ is open [94, Definition 1.5]. However, this alternative (and well-known) characterisation may be more helpful (which uses that any normed vector space is a metric space):

Proposition A.8 ([51, Proposition 1.41]). *Let $\mathbb{V}$ be a normed vector space, and let $\mathscr{V} \subseteq \mathbb{V}$. Then $\mathscr{V}$ is closed if and only if, for every convergent sequence $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathscr{V}$ with $v_* := \lim_{i \to +\infty} v_i$, it holds that $v_* \in \mathscr{V}$.*

The following alternative characterisation of boundedness shows that Property S5 coincides with the way that Reference [9, Chapter 4, Definition 1.11] defines it.

Proposition A.9. *Let $\mathbb{V}$ be a normed vector space and, for any $r \in \mathbb{R}_{>0}$ and $v \in \mathbb{V}$, let $B_r(v) := \{w \in \mathbb{V} : \|v - w\| < r\}$ denote the open ball in $\mathbb{V}$ with center v and radius r. Let $\mathscr{V} \subseteq \mathbb{V}$. Then $\mathscr{V}$ is bounded if and only if $\mathscr{V} \subseteq B_r(0)$ for some $r \in \mathbb{R}_{>0}$, where $0 \in \mathbb{V}$ denotes the origin of $\mathbb{V}$.*

Proof. First note that, for any $v \in \mathbb{V}$, it follows from Definition A.1$_{370}$ that $v - 0 = v$, and hence by Definition A.5$_{371}$ that $\|v - 0\| = \|v\|$.

Now suppose that $\mathscr{V}$ is bounded. Then due to Property S5, there is some $r \in \mathbb{R}_{>0}$ such that $\|\mathscr{V}\| < r$. For any $v \in \mathscr{V}$ it therefore holds that

$$\|v - 0\| = \|v\| \leq \|\mathscr{V}\| < r\,,$$

where we used the definition of $\|\mathscr{V}\|$ for the first inequality. This means that $v \in B_r(0)$ and hence, because $v \in \mathscr{V}$ was arbitrary, we conclude that $\mathscr{V} \subseteq B_r(0)$.

For the other direction, suppose that there is some $r \in \mathbb{R}_{>0}$ such that $\mathscr{V} \subseteq B_r(0)$. Then, using the definition of $\|\mathscr{V}\|$, it holds that

$$\|\mathscr{V}\| = \sup_{v \in \mathscr{V}} \|v\| \leq \sup_{v \in B_r(0)} \|v\| = \sup_{v \in B_r(0)} \|v - 0\| \leq r\,,$$

using the definition of $B_r(0)$ for the final inequality; it follows that $\mathscr{V}$ satisfies Property S5. ☐

For Property S6, we first note that a set $\mathscr{V}$ is *compact* (in a specific topology) if every open cover of $\mathscr{V}$ contains a finite subcover of $\mathscr{V}$; a collection $\mathbb{C}$ of subsets of $\mathbb{V}$ is an open cover of $\mathscr{V}$, if every element of $\mathbb{C}$ is open, and $\mathscr{V} \subseteq \cup_{\mathscr{W} \in \mathbb{C}} \mathscr{W}$ [100, Definition 5.1.1]. Moreover, a set $\mathscr{V} \subseteq \mathbb{V}$ is said to be *sequentially compact*, if every sequence $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ in $\mathscr{V}$ contains a convergent *subsequence* $\{v_{i_k}\}_{k \in \mathbb{Z}_{>0}}$ for which $\lim_{k \to +\infty} v_{i_k} =: v_* \in \mathscr{V}$. Now, as shown by the following result, the notions of compactness and sequential compactness coincide for metric spaces.

Proposition A.10 ([51, Theorem 1.62])**.** *Let $\mathbb{X}$ be a metric space, and let $\mathscr{V} \subseteq \mathbb{X}$. Then $\mathscr{V}$ is compact if and only if $\mathscr{V}$ is sequentially compact.*

Moreover, the following result provides an alternative characterisation of sequential compactness in finite-dimensional spaces, which generalises the well-known Bolzano-Weierstrass theorem, and which will be helpful in the sequel.

Proposition A.11 ([9, Chapter 4, Corollary 3.28])**.** *Let $\mathbb{V}$ be a finite-dimensional normed vector space, and let $\mathscr{V} \subseteq \mathbb{V}$. Then $\mathscr{V}$ is sequentially compact if and only if $\mathscr{V}$ is closed and bounded.*

The following summarizes the above; these are important properties of which we make frequent use in this dissertation.

Corollary A.12. *Let $\mathbb{V}$ be a finite-dimensional normed vector space, and let $\mathcal{V} \subseteq \mathbb{V}$. Then the following properties are equivalent:*

1. *$\mathcal{V}$ is closed and bounded;*
2. *$\mathcal{V}$ is sequentially compact;*
3. *$\mathcal{V}$ is compact.*

Proof. By Proposition A.11↶, $\mathcal{V}$ is sequentially compact if and only if it is closed and bounded. Because $\mathbb{V}$ is a metric space with the metric induced by its norm, by Proposition A.10↶ $\mathcal{V}$ is compact if and only if it is sequentially compact. Hence $\mathcal{V}$ is compact if and only if it is closed and bounded. □

Lemma A.13. *Let $\mathbb{V}$ be a normed vector space and, for all $r \in \mathbb{R}_{\geq 0}$, let $\overline{B}_r(0) := \{w \in \mathbb{V} : \|w\| \leq r\}$ be the closed ball in $\mathbb{V}$ with center 0 and radius r. Then $\overline{B}_r(0)$ is closed, and $\|\overline{B}_r(0)\| \leq r$.*

Proof. The set $\overline{B}_r(0)$ is closed by [75, Proposition B.10]. Note that $\|0 - 0\| = \|0\| = 0 \leq r$, and hence $0 \in \overline{B}_r(0)$; in particular, $\overline{B}_r(0)$ is non-empty. Hence, it follows that

$$\|\overline{B}_r(0)\| = \sup_{w \in \overline{B}_r(0)} \|w\| = \sup_{w \in \overline{B}_r(0)} \|w - 0\| \leq r,$$

using the definition and non-emptiness of $\overline{B}_r(0)$ for the final inequality. □

Corollary A.14. *Let $\mathbb{V}$ be a finite-dimensional normed vector space and let $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ be a bounded sequence in $\mathbb{V}$, i.e. such that there is an $r \in \mathbb{R}_{>0}$ for which $\|v_i\| < r$ for all $i \in \mathbb{Z}_{>0}$. Then $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ contains a convergent subsequence $\{v_{i_k}\}_{k \in \mathbb{Z}_{>0}}$. Moreover, with $v_* := \lim_{k \to +\infty} v_{i_k}$, it holds that $\|v_*\| \leq r$.*

Proof. Because the sequence $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ is bounded, there is some $r \in \mathbb{R}_{>0}$ such that $\|v_i\| < r$ for all $i \in \mathbb{Z}_{>0}$. This implies that $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ is contained in the origin-centered closed ball $\overline{B}_r(0)$ in $\mathbb{V}$ with radius r. By Lemma A.13, $\overline{B}_r(0)$ is closed and, since $r \in \mathbb{R}_{>0}$, bounded. By Corollary A.12, this implies that $\overline{B}_r(0)$ is sequentially compact, which in turn implies that the sequence $\{v_i\}_{i \in \mathbb{Z}_{>0}}$ in $\overline{B}_r(0)$ contains a convergent subsequence $\{v_{i_k}\}_{k \in \mathbb{Z}_{>0}}$ such that, with $v_* := \lim_{k \to +\infty} v_{i_k}$, $v_* \in \overline{B}_r(0)$. Hence it holds that $\|v_*\| \leq r$. □

A.1 The Space $\mathscr{L}(\mathscr{X})$

We now consider the space $\mathscr{L}(\mathscr{X})$ of real-valued functions on the finite set $\mathscr{X}$, that we introduce in Section 2.2.3$_{62}$. It is straightforward to verify that $\mathscr{L}(\mathscr{X})$ is a real vector space—that it satisfies Definition A.1$_{370}$—under the normal operations of addition and scalar multiplication, i.e. $(f+g)(x) := f(x)+g(x)$ and $(\alpha f)(x) := \alpha f(x)$ for all $x \in \mathscr{X}$, $f, g \in \mathscr{L}(\mathscr{X})$, and $\alpha \in \mathbb{R}$.

Now, for any $x \in \mathscr{X}$ consider the *indicator* $\mathbb{I}_x \in \mathscr{L}(\mathscr{X})$ of x, which is defined for all $y \in \mathscr{X}$ such that $\mathbb{I}_x(y) := 1$ if $x = y$, and $\mathbb{I}_x(y) := 0$, otherwise. Let us now prove the following.

Lemma A.15. *For any $f \in \mathscr{L}(\mathscr{X})$ it holds that $f = \sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x$.*

Proof. For any $y \in \mathscr{X}$ it holds that

$$\left(\sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x\right)(y) = \sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x(y) = f(y)\,,$$

using the definition of vector addition for the first equality, and the definition of the indicators $\mathbb{I}_x$ for the second equality. □

Proposition A.16. *The set $\mathscr{B} := \{\mathbb{I}_x : x \in \mathscr{X}\}$ is a basis of $\mathscr{L}(\mathscr{X})$.*

Proof. According to Definition A.3$_{371}$, we need to show that $\mathscr{B}$ is linearly independent, and that $\operatorname{span}(\mathscr{B}) = \mathscr{L}(\mathscr{X})$.

To show that $\mathscr{B}$ is linearly independent, consider any $\{f_1, \dots, f_n\} \subseteq \mathscr{B}$, $n \in \mathbb{Z}_{\geq 0}$, and any non-zero $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. For any $i \in \{1, \dots, n\}$, because $f_i \in \mathscr{B}$, there is some $x_i \in \mathscr{X}$ such that $f_i = \mathbb{I}_{x_i}$. Then

$$\left(\sum_{i=1}^{n} \alpha_i f_i\right)(x_1) = \sum_{i=1}^{n} \alpha_i f_i(x_1) = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{x_i}(x_1) = \alpha_1 \neq 0\,,$$

where we used the definition of vector addition for the first equality, the identification $f_i = \mathbb{I}_{x_i}$ for the second equality, the definition of $\mathbb{I}_{x_i}$ for the third equality, and the fact that $\alpha_i \neq 0$ for the inequality. Hence, we conclude that $\sum_{i=1}^{n} \alpha_i f_i \neq 0$. Because this is true for all $\{f_1, \dots, f_n\} \subseteq \mathscr{B}$ and non-zero $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, it follows from Definition A.2$_{371}$ that $\mathscr{B}$ is linearly independent.

To show that $\operatorname{span}(\mathscr{B}) = \mathscr{L}(\mathscr{X})$, we first note that, from the definition of the linear span, it holds that $\operatorname{span}(\mathscr{B}) \subseteq \mathscr{L}(\mathscr{X})$. So, it suffices to prove the inclusion in the other direction. To this end, fix any $f \in \mathscr{L}(\mathscr{X})$. Then, by Lemma A.15, it holds that $f = \sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x$, and because $\mathscr{X}$ is finite, it follows from Proposition A.1$_{370}$ that $f \in \operatorname{span}(\mathscr{B})$. Because this is true for all $f \in \mathscr{L}(\mathscr{X})$, it follows that $\mathscr{L}(\mathscr{X}) \subseteq \operatorname{span}(\mathscr{B})$, whence $\operatorname{span}(\mathscr{B}) = \mathscr{L}(\mathscr{X})$. □

Corollary A.17. *The space $\mathscr{L}(\mathscr{X})$ is a finite-dimensional real vector space. In particular,* $\dim\bigl(\mathscr{L}(\mathscr{X})\bigr) = |\mathscr{X}|$.

Proof. From the above discussion, we know that $\mathscr{L}(\mathscr{X})$ is a real vector space. By Proposition A.16$_{\frown}$, the set $\{\mathbb{I}_x : x \in \mathscr{X}\}$ is a basis of $\mathscr{L}(\mathscr{X})$. Hence, it follows from Definition A.4$_{371}$ that $\dim\bigl(\mathscr{L}(\mathscr{X})\bigr) = |\mathscr{X}|$. Therefore, and because $\mathscr{X}$ is finite, $\mathscr{L}(\mathscr{X})$ is finite-dimensional. □

As discussed in Section 2.2.3$_{62}$, throughout this dissertation we equip $\mathscr{L}(\mathscr{X})$ with the supremum norm, whence it follows that $\mathscr{L}(\mathscr{X})$ is then a finite-dimensional normed vector space by Definition A.5$_{371}$ (c.f. Proposition 2.16$_{63}$).

A.2 The Dual Space of $\mathscr{L}(\mathscr{X})$

Having discussed above how the space $\mathscr{L}(\mathscr{X})$ is a finite-dimensional normed vector space, let us next consider the space of real linear functionals on $\mathscr{L}(\mathscr{X})$; these are maps $\phi^\top : \mathscr{L}(\mathscr{X}) \to \mathbb{R} : f \mapsto \phi^\top f$ that satisfy $\phi^\top(f+g) = \phi^\top f + \phi^\top g$ and $\phi^\top(\lambda f) = \lambda\phi^\top f$ for all $f, g \in \mathscr{L}(\mathscr{X})$ and $\lambda \in \mathbb{R}$. These functionals are collected in the space $\mathscr{L}(\mathscr{X})^\top$, which is the *dual space* [51, Definition 5.54] of $\mathscr{L}(\mathscr{X})$.[3]

Proposition A.18 ([51, Section 5.6])**.** *The space $\mathscr{L}(\mathscr{X})^\top$ is a finite-dimensional real vector space under the usual operations of addition and scalar multiplication. In particular,* $\dim\bigl(\mathscr{L}(\mathscr{X})^\top\bigr) = \dim\bigl(\mathscr{L}(\mathscr{X})\bigr)$.

This introduces the dual space at a level of abstraction at which we can later apply some general results, but for the sake of one's intuition it is useful to note that, if $\mathscr{L}(\mathscr{X})$ is interpreted as the space of $|\mathscr{X}|$-dimensional real *column* vectors, then $\mathscr{L}(\mathscr{X})^\top$ is simply the space of $|\mathscr{X}|$-dimensional real *row* vectors; moreover, for any $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$ and $f \in \mathscr{L}(\mathscr{X})$, the value $\phi^\top f$ is simply the dot product of these vectors. This also explains our choice of notation.

To make this more explicit, consider the following result.

Proposition A.19. *For any $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$, let ϕ be the unique element of $\mathscr{L}(\mathscr{X})$ such that $\phi(x) := \phi^\top \mathbb{I}_x$ for all $x \in \mathscr{X}$, where $\mathbb{I}_x$ is the indicator of x. Then, for all $f \in \mathscr{L}(\mathscr{X})$ it holds that*

$$\phi^\top f = \sum_{x \in \mathscr{X}} \phi(x) f(x)\,.$$

[3] As in [51, Definition 5.54], in general one may make a distinction between the *algebraic* and *topological* dual spaces of a linear space; however, since $\mathscr{L}(\mathscr{X})$ is finite-dimensional by Corollary A.17, this distinction disappears [51, Section 5.6].

Proof. By Lemma A.15379, for any $f \in \mathscr{L}(\mathscr{X})$ we can write $f = \sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x$. Hence, it holds that

$$\phi^\top f = \phi^\top \sum_{x\in\mathscr{X}} f(x)\mathbb{I}_x = \sum_{x\in\mathscr{X}} \phi^\top\big(f(x)\mathbb{I}_x\big) = \sum_{x\in\mathscr{X}} f(x)\phi^\top\mathbb{I}_x = \sum_{x\in\mathscr{X}} f(x)\phi(x)\,,$$

using the linear character of $\phi^\top$ for the second and third equalities. □

Conversely, for any $\phi \in \mathscr{L}(\mathscr{X})$ we can define a unique element $\phi^\top$ of $\mathscr{L}(\mathscr{X})^\top$, for all $f \in \mathscr{L}(\mathscr{X})$, as $\phi^\top f := \sum_{x\in\mathscr{X}} \phi(x)f(x)$. Thus the spaces $\mathscr{L}(\mathscr{X})$ and $\mathscr{L}(\mathscr{X})^\top$ can be treated as essentially the same; formally, there is a linear isomorphism between the two [51, Section 5.6].

However, these spaces do differ in the norms with which we equip them; while $\mathscr{L}(\mathscr{X})$ received the supremum norm, we equip $\mathscr{L}(\mathscr{X})^\top$ with the induced dual norm $\lVert\cdot\rVert_*$, defined for all $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$ as

$$\lVert\phi^\top\rVert_* := \sup\Big\{\lvert\phi^\top f\rvert \,:\, f \in \mathscr{L}(\mathscr{X}), \lVert f\rVert = 1\Big\}.$$

Using Proposition A.18, this turns $\mathscr{L}(\mathscr{X})^\top$ into a finite-dimensional normed vector space, and all the relevant results from the first section of this appendix therefore apply to it.

It is worth noting that, as discussed in the first section of this appendix, the fact that $\mathscr{L}(\mathscr{X})^\top$ is finite-dimensional ensures that its (metric) topology is independent of the choice of norm. Hence in particular, we could have also equipped it with the supremum norm, which would make its identification with $\mathscr{L}(\mathscr{X})$ complete. However, for practical reasons, we will find the dual norm to be slightly easier to work with.

To that end, we do need the following result, which provides a more convenient expression of the dual norm; essentially, this shows that the dual norm $\lVert\cdot\rVert_*$ (for the dual of a space equipped with the supremum norm) is the ℓ_1 norm.

Proposition A.20. *For any $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$, let ϕ be the unique element of $\mathscr{L}(\mathscr{X})$ such that $\phi(x) := \phi^\top\mathbb{I}_x$ for all $x \in \mathscr{X}$, where $\mathbb{I}_x$ is the indicator of x. Then it holds that*

$$\lVert\phi^\top\rVert_* = \sum_{x\in\mathscr{X}} \lvert\phi(x)\rvert\,.$$

Proof. Let $f \in \mathscr{L}(\mathscr{X})$ be such that $\lVert f\rVert = 1$. Then, using Proposition A.19 it holds that

$$\begin{aligned}\lvert\phi^\top f\rvert = \left\lvert\sum_{x\in\mathscr{X}} f(x)\phi(x)\right\rvert &\leq \sum_{x\in\mathscr{X}} \lvert f(x)\phi(x)\rvert \\ &= \sum_{x\in\mathscr{X}} \lvert f(x)\rvert\,\lvert\phi(x)\rvert \leq \sum_{x\in\mathscr{X}} \lVert f\rVert\,\lvert\phi(x)\rvert = \sum_{x\in\mathscr{X}} \lvert\phi(x)\rvert\,,\end{aligned}$$

using the definition of the supremum norm for the second inequality, and the assumption that $\|f\| = 1$ for the final equality. Because this is true for all $f \in \mathscr{L}(\mathscr{X})$ with $\|f\| = 1$, it follows that $\|\phi^\top\|_* \leq \sum_{x \in \mathscr{X}} |\phi(x)|$.

To get the inequality in the other direction, let $f \in \mathscr{L}(\mathscr{X})$ be defined, for all $x \in \mathscr{X}$, such that $f(x) := 1$ if $\phi(x) \geq 0$, and $f(x) := -1$, otherwise. Then $\|f\| = 1$, and, because $f(x)\phi(x) = |\phi(x)| \geq 0$ for all $x \in \mathscr{X}$, it holds that

$$\left|\phi^\top f\right| = \left|\sum_{x \in \mathscr{X}} f(x)\phi(x)\right| = \sum_{x \in \mathscr{X}} |\phi(x)| \, ,$$

using Proposition A.19$_{380}$ for the first equality. Hence it follows that also $\sum_{x \in \mathscr{X}} |\phi(x)| \leq \|\phi^\top\|_*$. $\square$

Moreover, we need the notion of the *double dual* $\mathscr{L}(\mathscr{X})^{\top\top}$, which is the dual space of $\mathscr{L}(\mathscr{X})^\top$: the space of all real-valued linear functionals that map $\mathscr{L}(\mathscr{X})^\top$ into $\mathbb{R}$. Because $\mathscr{L}(\mathscr{X})$ is finite-dimensional, this turns out to just coincide with $\mathscr{L}(\mathscr{X})$.

Lemma A.21 ([42, Chapter 1, Section 16]). *The space $\mathscr{L}(\mathscr{X})$ is* reflexive, *which means that $\mathscr{L}(\mathscr{X})^{\top\top} = \mathscr{L}(\mathscr{X})$. Moreover, let $f \in \mathscr{L}(\mathscr{X})^{\top\top} = \mathscr{L}(\mathscr{X})$. Then, for all $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$, f maps $\phi^\top$ into $\phi^\top f$.*

Intuitively, this says that the transpose $v^{\top\top}$ of a transposed vector $v^\top$, is just the vector v itself. When this space is equipped with the induced dual norm $\|\cdot\|_{**}$, i.e.

$$\|f\|_{**} := \sup\left\{\left|\phi^\top f\right| \,:\, \phi^\top \in \mathscr{L}(\mathscr{X})^\top,\ \left\|\phi^\top\right\|_* = 1\right\},$$

it can be shown that $\|\cdot\|_{**} = \|\cdot\|_\infty$, so $\mathscr{L}(\mathscr{X})^{\top\top}$ equipped with the dual norm recovers $\mathscr{L}(\mathscr{X})$ as a normed vector space.

Usually when we reference convergence in $\mathscr{L}(\mathscr{X})^\top$, we mean it with respect to the (dual) norm, i.e. as in Definition A.11$_{374}$. However, in general, one may consider other notions of convergence. Specifically, for technical reasons we need to introduce the following:

Definition A.13 ([75, Section 2.6]). *A sequence $\{\phi_i^\top\}_{i \in \mathbb{Z}_{>0}}$ in $\mathscr{L}(\mathscr{X})^\top$* weak⋆-converges *to $\phi_*^\top \in \mathscr{L}(\mathscr{X})^\top$ if $\lim_{i \to +\infty} \phi_i^\top f = \phi_*^\top f$ for all $f \in \mathscr{L}(\mathscr{X})$.*

This notion of convergence has an associated topology:

Definition A.14 ([75, Section 2.6]). *The* weak⋆-topology *$\tau_{w\star}$ on $\mathscr{L}(\mathscr{X})^\top$ is the smallest collection of subsets of $\mathscr{L}(\mathscr{X})^\top$ for which all sequences $\{\phi_i^\top\}_{i \in \mathbb{Z}_{>0}}$ that weak⋆-converge to $\phi_*^\top$, converge to $\phi_*^\top$ with respect to $\tau_{w\star}$.*

It turns out that, in our current finite-dimensional setting, this topology is just the one that we were already familiar with:

Proposition A.22. *Let $\mathscr{L}(\mathscr{X})^\top$ be equipped with the dual norm $\|\cdot\|_*$, let τ_m be its metric topology, and let $\tau_{w\star}$ be its weak⋆ topology. Then $\tau_m = \tau_{w\star}$.*

Proof. By [75, Corollary 2.6.3], the weak⋆ and metric topologies of the dual space of a normed vector space $\mathbb{V}$ are the same, if and only if $\mathbb{V}$ is finite-dimensional. Since $\mathscr{L}(\mathscr{X})^\top$ is the dual space of $\mathscr{L}(\mathscr{X})$, which is a finite-dimensional normed vector space by Corollary A.17$_{380}$, the result is immediate. □

This allows us to state the following technical result, which we need in this dissertation:

Corollary A.23. *A subset $\mathscr{V} \subseteq \mathscr{L}(\mathscr{X})^\top$ is weak⋆-compact if and only if it is compact in the metric topology induced by its norm $\|\cdot\|_*$.*

Proof. Immediate from Proposition A.22 and the fact that compactness is a topological property. □

A.3 Linear Maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$

Let us conclude this appendix by considering linear maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$; these are maps $T : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto Tf$ such that for all $f, g \in \mathscr{L}(\mathscr{X})$ and all $\lambda \in \mathbb{R}$ it holds that $T(f+g) = Tf + Tg$ and $T(\lambda f) = \lambda Tf$. We collect all these maps in the space $\mathbb{M}$. It is clear that $\mathbb{M}$ is a real vector space under the usual operations of addition and scalar multiplication.

For any $T \in \mathbb{M}$, we can consider its *matrix representation* $\tilde{T} : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ with entries $\tilde{T}(x,y) := T\mathbb{I}_y(x)$ for all $x, y \in \mathscr{X}$, where $\mathbb{I}_y$ is the indicator of y. As is well-known, this matrix representation is the dual of the linear map T, in the sense that one can always obtain one from the other; we already established the first direction above. The following result shows that, given a matrix representation, one can always recover a corresponding linear map in $\mathbb{M}$:

Lemma A.24. *Consider a function $\tilde{T} : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ and define the map $T : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) : f \mapsto Tf$, for all $f \in \mathscr{L}(\mathscr{X})$, as*

$$Tf(x) := \sum_{y \in \mathscr{X}} \tilde{T}(x,y) f(y) \quad \text{for all } x \in \mathscr{X}. \tag{A.2}$$

Then $T \in \mathbb{M}$, and $\tilde{T}$ is the matrix representation of T.

Proof. Let us first show that $T \in \mathbb{M}$. It is clear that T is defined on $\mathscr{L}(\mathscr{X})$, and because $\mathscr{X}$ is finite, it follows from (A.2)$_{\curvearrowleft}$ that $Tf(x)$ is real-valued for any $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$. Therefore, we conclude that T maps $\mathscr{L}(\mathscr{X})$ into $\mathscr{L}(\mathscr{X})$.

To establish the linearity of T, fix any $x \in \mathscr{X}$. It follows from Equation (A.2)$_{\curvearrowleft}$ that, for all $f, g \in \mathscr{L}(\mathscr{X})$, it holds that

$$\begin{aligned} T(f+g)(x) &= \sum_{y\in\mathscr{X}} \tilde{T}(x,y)\big(f(y)+g(y)\big) \\ &= \sum_{y\in\mathscr{X}} \tilde{T}(x,y)f(y) + \sum_{y\in\mathscr{X}} \tilde{T}(x,y)g(y) = Tf(x)+Tg(x)\,. \end{aligned}$$

Moreover, for any $f \in \mathscr{L}(\mathscr{X})$ and $\lambda \in \mathbb{R}$, it follows from Equation (A.2)$_{\curvearrowleft}$ that

$$T(\lambda f)(x) = \sum_{y\in\mathscr{X}} \tilde{T}(x,y)\lambda f(y) = \lambda \sum_{y\in\mathscr{X}} \tilde{T}(x,y)f(y) = \lambda Tf(x)\,.$$

Because $x \in \mathscr{X}$ is arbitrary, we conclude from the above that $T(f+g) = Tf+Tg$ and $T(\lambda f) = \lambda Tf$ for all $f, g \in \mathscr{L}(\mathscr{X})$ and $\lambda \in \mathbb{R}$, and hence that $T \in \mathbb{M}$. It remains to establish that $\tilde{T}$ is the matrix representation of T. To this end, fix any $x, y \in \mathscr{X}$. Then it follows from Equation (A.2)$_{\curvearrowleft}$ and the definition of the indicator $\mathbb{I}_y$ that

$$T\mathbb{I}_y(x) = \sum_{z\in\mathscr{X}} \tilde{T}(x,z)\mathbb{I}_y(z) = \tilde{T}(x,y)\,,$$

which concludes the proof. □

Moreover, this mapping between elements of $\mathbb{M}$ and their matrix representation is unique; that a given element of $\mathbb{M}$ uniquely determines its corresponding matrix representation is immediate from the definition. The following result takes care of the other direction:

Proposition A.25. *For any $T, S \in \mathbb{M}$ whose matrix representations $\tilde{T}$ and $\tilde{S}$ satisfy $\tilde{T}(x,y) = \tilde{S}(x,y)$ for all $x, y \in \mathscr{X}$, it holds that $T = S$.*

Proof. For any $f \in \mathscr{L}(\mathscr{X})$ it follows from Lemma A.15$_{379}$ that $f = \sum_{y\in\mathscr{X}} f(y)\mathbb{I}_y$ and therefore, using the linearity of T, it follows that for all $x \in \mathscr{X}$ it holds that

$$Tf(x) = T\left(\sum_{y\in\mathscr{X}} f(y)\mathbb{I}_y\right)(x) = \sum_{y\in\mathscr{X}} f(y)T\mathbb{I}_y(x) = \sum_{y\in\mathscr{X}} \tilde{T}(x,y)f(y)\,,$$

using the definition of the matrix representation $\tilde{T}$ of T. Similarly, we find that $Sf(x) = \sum_{y\in\mathscr{X}} \tilde{S}(x,y)f(y)$. Because $\tilde{T}(x,y) = \tilde{S}(x,y)$ for all $y \in \mathscr{X}$

it follows that

$$Tf(x) = \sum_{y\in\mathscr{X}} \tilde{T}(x,y)f(y) = \sum_{y\in\mathscr{X}} \tilde{S}(x,y)f(y) = Sf(x)\,,$$

which, since $x \in \mathscr{X}$ is arbitrary, implies that $Tf = Sf$. Since $f \in \mathscr{L}(\mathscr{X})$ is arbitrary it follows that $T = S$. $\square$

Due to the above, we will also simply refer to the elements of $\mathbb{M}$ as *matrices*, and we will denote the entries of the matrix representation of $T \in \mathbb{M}$ simply as $T(x,y)$ with $x,y \in \mathscr{X}$, without further notational distinction. As such, it follows from Lemma A.24383 that for any $T \in \mathbb{M}$, any $f \in \mathscr{L}(\mathscr{X})$, and any $x \in \mathscr{X}$, it holds that

$$Tf(x) = \sum_{y\in\mathscr{X}} T(x,y)f(y)\,,$$

which is an identity that we will use often throughout this work. We emphasize that, the technical rigour of the current discussion notwithstanding, this says that Tf is essentially just the matrix-vector product of T with the vector $f \in \mathscr{L}(\mathscr{X})$. The main difference with the "usual" linear-algebraic approach is that we can make the above identification without fixing an ordering of $\mathscr{X}$.

Analogously, we note that for any $T,S \in \mathbb{M}$, their composition TS is again an element of $\mathbb{M}$—since the composition of two linear maps is clearly itself linear—that satisfies, for all $x,y \in \mathscr{X}$, that

$$TS(x,y) = TS\mathbb{I}_y(x) = T\big(S\mathbb{I}_y\big)(x) = \sum_{z\in\mathscr{X}} T(x,z)S\mathbb{I}_y(z) = \sum_{z\in\mathscr{X}} T(x,z)S(z,y)\,,$$

which simply expresses the composition TS in terms of the familiar matrix product.

Next, let us note that for any $T \in \mathbb{M}$ and any $x \in \mathscr{X}$, we can consider the functional $T(\cdot)(x) : f \mapsto Tf(x)$ on $\mathscr{L}(\mathscr{X})$. By the linearity of T, this functional is also linear, and it is therefore an element of the dual space $\mathscr{L}(\mathscr{X})^\top$ of $\mathscr{L}(\mathscr{X})$ (see Section A.2380). Therefore, by Proposition A.19380, this linear functional has a representation $\phi \in \mathscr{L}(\mathscr{X})$ where $\phi(y) := T\mathbb{I}_y(x)$ for all $y \in \mathscr{X}$, so that, for all $f \in \mathscr{L}(\mathscr{X})$,

$$Tf(x) = \sum_{y\in\mathscr{X}} \phi(y)f(y) = \sum_{y\in\mathscr{X}} T\mathbb{I}_y(x)f(y) = \sum_{y\in\mathscr{X}} T(x,y)f(y)\,,$$

where for the final step we used the definition of the matrix representation of T. We also note that it holds that $\phi(y) = T(x,y)$ for all $y \in \mathscr{X}$, and hence, the x-row $T(x,\cdot)$ of T is essentially the representation of the linear functional $T(\cdot)(x)$. Therefore, we will in the sequel identify this functional with the x-row of T, so that we write

$$T(x,\cdot)f := Tf(x) = \sum_{y\in\mathscr{X}} T(x,y)f(y) \text{ for all } f \in \mathscr{L}(\mathscr{X}).$$

This is, effectively, the same kind of identification that we make when identifying the linear map $T : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ with its matrix representation, and should not be a cause for confusion. Moreover, when we write this linear functional $T(x,\cdot) \in \mathscr{L}(\mathscr{X})^\top$ as it is applied to the function $f \in \mathscr{L}(\mathscr{X})$, this notation coincides with the intuitive interpretation of $T(x,\cdot)f$ being the dot product of the x-row of T with f.

In summary, we have seen in the above discussion that we can construct linear functionals on $\mathscr{L}(\mathscr{X})$, from matrices $T \in \mathbb{M}$. However, the construction in the other direction will also be useful.

Proposition A.26. *For all $x \in \mathscr{X}$ choose some $\phi_x^\top \in \mathscr{L}(\mathscr{X})^\top$, and let $T : f \mapsto Tf$ be the unique map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ that is defined, for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{X}$, as $Tf(x) := \phi_x^\top f$. Then T is a linear map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, and its matrix representation satisfies $T(x,y) = \phi_x^\top \mathbb{I}_y$ for all $x,y \in \mathscr{X}$.*

Proof. That T has $\mathscr{L}(\mathscr{X})$ as its codomain follows from the fact that, by definition, each $\phi_x^\top \in \mathscr{L}(\mathscr{X})^\top$, $x \in \mathscr{X}$, has $\mathbb{R}$ as its codomain; thus $Tf(x)$ is real-valued for all $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$. That T is a linear map follows from the fact that each $\phi_x^\top \in \mathscr{L}(\mathscr{X})^\top$ is a linear map.

It remains to prove the statement about the matrix representation. As noted in the introduction to this section, for all $x,y \in \mathscr{X}$ it holds that $T(x,y) = T\mathbb{I}_y(x)$, and hence it follows that $T(x,y) = T\mathbb{I}_y(x) = \phi_x^\top \mathbb{I}_y$ by the definition of Tf. □

The space of matrix representations of the elements of $\mathbb{M}$ is itself a real vector space under the usual operations of addition and scalar multiplication. In particular, these spaces are isomorphic, in the following sense.

Proposition A.27. *For any $T,S \in \mathbb{M}$ and any $\lambda \in \mathbb{R}$ it holds that*

$$(T+S)(x,y) = T(x,y) + S(x,y) \text{ and } (\lambda T)(x,y) = \lambda T(x,y) \quad \text{for all } x,y \in \mathscr{X}.$$

Proof. For any $x,y \in \mathscr{X}$ it holds that

$$(T+S)(x,y) = (T+S)\mathbb{I}_y(x) = T\mathbb{I}_y(x) + S\mathbb{I}_y(x) = T(x,y) + S(x,y).$$

Moreover, for any $\lambda \in \mathbb{R}$ and any $x,y \in \mathscr{X}$ it holds that

$$(\lambda T)(x,y) = \lambda T\mathbb{I}_y(x) = \lambda T(x,y),$$

which concludes the proof. □

In words, the above result establishes that the matrix representation of a sum of elements of $\mathbb{M}$, is simply the sum of the matrix representations of these elements; and similarly for scalar multiplications. Moreover, we have the following result:

Lemma A.28. *The (unique) origin of* $\mathbb{M}$ *is the matrix* O *with entries* $O(x,y) = 0$ *for all* $x,y \in \mathscr{X}$.

Proof. Let $O \in \mathbb{M}$ be defined such that $O(x,y) = 0$ for all $x,y \in \mathscr{X}$. To show that O is the origin of $\mathbb{M}$, we need to show that $O+T = T$ for all $T \in \mathbb{M}$. To this end, fix any $T \in \mathbb{M}$ and any $f \in \mathscr{L}(\mathscr{X})$. Then for all $x \in \mathscr{X}$ it follows from Proposition A.27 that

$$\begin{aligned}\big((O+T)f\big)(x) &= \sum_{y\in\mathscr{X}} (O+T)(x,y)f(y) \\ &= \sum_{y\in\mathscr{X}} O(x,y)f(y) + \sum_{y\in\mathscr{X}} T(x,y)f(y) \\ &= \sum_{y\in\mathscr{X}} T(x,y)f(y) = Tf(x)\,,\end{aligned}$$

where for the third equality we used that $O(x,y) = 0$ for all $y \in \mathscr{X}$. Because this is true for all $x \in \mathscr{X}$ it follows that $(O+T)f = Tf$, which in turn implies that $O+T = T$ because $f \in \mathscr{L}(\mathscr{X})$ is arbitrary. By Definition A.1$_{370}$ this implies that O is the origin of $\mathbb{M}$.

To see that this is the unique origin, consider any $O' \in \mathbb{M}$ such that $O'+T = T$ for all $T \in \mathbb{M}$. Then because we already know that O is an origin of $\mathbb{M}$, it follows that $O = O'+O = O'$. □

Next, for any $x,y \in \mathscr{X}$, consider the (unique) matrix $B_{x,y} \in \mathbb{M}$ that satisfies, for all $x',y' \in \mathscr{X}$, that $B_{x,y}(x',y') := 1$ if $x = x'$ and $y = y'$, and $B_{x,y}(x',y') := 0$ otherwise. Essentially, $B_{x,y}$ is just the matrix whose x,y-entry is one and that otherwise contains only zeroes. Clearly, these matrices allow us to represent any other matrix:

Lemma A.29. *For any* $T \in \mathbb{M}$, *it holds that*

$$T = \sum_{x,y\in\mathscr{X}} T(x,y)B_{x,y}\,.$$

Proof. Fix any $x',y' \in \mathscr{X}$. Then, by the definition of $B_{x,y}$, with $x,y \in \mathscr{X}$, it holds that

$$\begin{aligned}T(x',y') &= T(x',y')B_{x',y'}(x',y') \\ &= \sum_{x,y\in\mathscr{X}} T(x,y)B_{x,y}(x',y') = \left(\sum_{x,y\in\mathscr{X}} T(x,y)B_{x,y}\right)(x',y')\,,\end{aligned}$$

where for the last step we used Proposition A.27$_{386}$. Because this is true for all $x', y' \in \mathscr{X}$ it follows from Proposition A.25$_{384}$ that $T = \sum_{x,y \in \mathscr{X}} T(x,y) B_{x,y}$. ☐

Hence, we have the following result.

Proposition A.30. *The set* $\mathscr{B} := \{B_{x,y} : x,y \in \mathscr{X}\}$ *is a basis of* $\mathbb{M}$.

Proof. According to Definition A.3$_{371}$, we need to show that $\mathscr{B}$ is linearly independent, and that $\operatorname{span}(\mathscr{B}) = \mathbb{M}$.

To show that $\mathscr{B}$ is linearly independent, choose any $\{v_1, \ldots, v_n\} \subseteq \mathscr{B}$, $n \in \mathbb{Z}_{\geq 0}$, and any non-zero $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. For any $i \in \{1, \ldots, n\}$, because $v_i \in \mathscr{B}$, there are $x_i, y_i \in \mathscr{X}$ such that $v_i = B_{x_i,y_i}$. Then it follows from Proposition A.27$_{386}$ that

$$\left(\sum_{i=1}^{n} \alpha_i v_i\right)(x_1, y_1) = \sum_{i=1}^{n} \alpha_i v_i(x_1, y_1) = \sum_{i=1}^{n} \alpha_i B_{x_i,y_i}(x_1, y_1) = \alpha_1 \neq 0,$$

where we used the definition of B_{x_i,y_i} for the final equality, and that α_1 is non-zero for the inequality. Due to Lemma A.28$_{\curvearrowleft}$ this implies that $\sum_{i=1}^{n} \alpha_i v_i$ is not the origin of $\mathbb{M}$ which, by Definition A.2$_{371}$, implies that $\mathscr{B}$ is linearly independent.

To show that $\operatorname{span}(\mathscr{B}) = \mathbb{M}$, we first note that, from the definition of the linear span, it holds that $\operatorname{span}(\mathscr{B}) \subseteq \mathbb{M}$. So, it suffices to prove the inclusion in the other direction. To this end, fix any $T \in \mathbb{M}$. Then it follows from Lemma A.29$_{\curvearrowleft}$ that $T = \sum_{x,y \in \mathscr{X}} T(x,y) B_{x,y}$. Therefore, and because $\mathscr{X}$ is finite, it follows from Proposition A.1$_{370}$ that $T \in \operatorname{span}(\mathscr{B})$. Because this is true for all $T \in \mathbb{M}$ it follows that $\mathbb{M} \subseteq \operatorname{span}(\mathscr{B})$ and hence, that $\operatorname{span}(\mathscr{B}) = \mathbb{M}$. ☐

The following result should therefore not be surprising:

Corollary A.31. *The space* $\mathbb{M}$ *is a finite-dimensional real vector space. In particular,* $\dim(\mathbb{M}) = |\mathscr{X}|^2$.

Proof. We have already established that $\mathbb{M}$ is a real vector space. By Proposition A.30, the set $\{B_{x,y} : x,y \in \mathscr{X}\}$ is a basis of $\mathbb{M}$. Because this basis contains $|\mathscr{X}|^2$ elements, it follows from Definition A.4$_{371}$ that $\dim(\mathbb{M}) = |\mathscr{X}|^2$. Therefore, and because $\mathscr{X}$ is finite, $\mathbb{M}$ is finite-dimensional. ☐

Of course, this result should be intuitively clear; since $\mathbb{M}$ is isomorphic to the space $\mathbb{R}^{\mathscr{X} \times \mathscr{X}}$ of matrix representations $(x,y) \mapsto T(x,y)$ of elements $T \in \mathbb{M}$, it makes sense that $\mathbb{M}$ should have exactly $|\mathscr{X}|^2$ dimensions.

Moving on, as discussed in Section 2.2.3$_{62}$, in this work we equip $\mathbb{M}$ with the induced operator norm $\|\cdot\|$, which is defined for all $T \in \mathbb{M}$ as

$$\|T\| := \sup\big\{\|Tf\| \,:\, f \in \mathscr{L}(\mathscr{X}),\ \|f\| = 1\big\}. \tag{A.3}$$

This turns $\mathbb{M}$ into a finite-dimensional normed vector space by Definition A.5$_{371}$ (c.f. Proposition 2.16$_{63}$), and therefore all the relevant results from the first section of this appendix apply. In particular, therefore, $\mathbb{M}$ is a Banach space.

Let us consider some properties of the operator norm $\|\cdot\|$ on $\mathbb{M}$.

Proposition A.32. *For any $T \in \mathbb{M}$, it holds that*

$$\|T\| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)|\,.$$

Proof. For any $x \in \mathscr{X}$, define $f_x \in \mathscr{L}(\mathscr{X})$ such that, for all $y \in \mathscr{X}$, $f_x(y) := 1$ if $T(x,y) \geq 0$, and $f_x(y) := -1$ otherwise. Then $T(x,y)f_x(y) = |T(x,y)|$ for all $x,y \in \mathscr{X}$. Therefore, it follows that for all $x \in \mathscr{X}$ it holds that

$$Tf_x(x) = \sum_{y\in\mathscr{X}} T(x,y)f_x(y) = \sum_{y\in\mathscr{X}} |T(x,y)|\,.$$

Moreover, it follows from the definition of the supremum norm on $\mathscr{L}(\mathscr{X})$ that $Tf_x(x) \leq \|Tf_x\|$. We also note that $\|f_x\| = 1$ for all $x \in \mathscr{X}$. By combining the above results with Equation (A.3) it follows that

$$\max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)| \leq \max_{x\in\mathscr{X}} \|Tf_x\| \leq \|T\|\,.$$

For the other direction, consider any $f \in \mathscr{L}(\mathscr{X})$ such that $\|f\| = 1$, and fix any $x \in \mathscr{X}$. Then, for all $y \in \mathscr{L}(\mathscr{X})$, it holds that $|f(y)| \leq 1$ because $\|f\| = 1$. Hence it follows that, for all $x,y \in \mathscr{X}$,

$$|T(x,y)f(y)| = |T(x,y)|\,|f(y)| \leq |T(x,y)|\,.$$

This implies that, for all $x \in \mathscr{X}$,

$$|Tf(x)| = \left|\sum_{y\in\mathscr{X}} T(x,y)f(y)\right| \leq \sum_{y\in\mathscr{X}} |T(x,y)f(y)| \leq \sum_{y\in\mathscr{X}} |T(x,y)| = Tf_x(x)\,,$$

and because this is true for all $x \in \mathscr{X}$, it follows that

$$\|Tf\| = \max_{x\in\mathscr{X}} |Tf(x)| \leq \max_{x\in\mathscr{X}} Tf_x(x) = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)|\,.$$

Because this is true for all $f \in \mathscr{L}(\mathscr{X})$ with $\|f\| = 1$, it follows from Equation (A.3) that

$$\|T\| \leq \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)|\,,$$

which concludes the proof. $\square$

Due to this result, and using the interpretation of the x-rows of T as linear functionals on $\mathscr{L}(\mathscr{X})$ together with Proposition A.20[381], it follows that

$$\|T\| = \max_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} |T(x,y)| = \max_{x\in\mathscr{X}} \|T(x,\cdot)\|_* \,. \tag{A.4}$$

This shows the relation between the induced operator norm on $\mathbb{M}$, and the dual norm $\|\cdot\|_*$ on $\mathscr{L}(\mathscr{X})^\top$. This also implies the following result.

Proposition A.33. *For any matrix $T \in \mathbb{M}$ and any $x \in \mathscr{X}$, it holds that $\|T(x,\cdot)\|_* \leq \|T\|$.*

Finally, the next (well-known) result will be helpful; it states that convergence of a sequence of matrices with respect to the operator norm (i.e. as in Definition A.11[374]), implies its convergence in the strong operator topology [51, Definition 5.45]; that is, the elementwise convergence as these matrices are applied to the elements of $\mathscr{L}(\mathscr{X})$.

Lemma A.34 ([51, Theorem 5.45])**.** *Let $\{T_i\}_{i\in\mathbb{Z}_{>0}}$ be a convergent sequence in $\mathbb{M}$ with $T_* := \lim_{i\to+\infty} T_i$. Then $T_* f = \lim_{i\to+\infty} T_i f$ for all $f \in \mathscr{L}(\mathscr{X})$.*

B

Norm Inequalities

"Tell me one last thing.
Is this real? Or has this been happening inside my head?"

"Of course it is happening inside your head,
but why on earth should that mean that it is not real?"

J. K. Rowling, "Harry Potter and the Deathly Hallows"

This appendix contains some technical inequalities that we will need throughout the dissertation. In particular, they are inequalities between—and bounds on—norms of particular operators. We specifically consider transition matrices T (Definition 3.5$_{91}$), rate matrices Q (Definition 4.4$_{150}$), matrix exponentials e^{Qt}, with $t \in \mathbb{R}_{\geq 0}$ (Definition 4.5$_{154}$), and their generalised counterparts: lower transition operators $\underline{T}$ (Definition 3.15$_{116}$), lower transition rate operators $\underline{Q}$ (Definition 6.2$_{265}$), and the generalised exponentials $e^{\underline{Q}t}$ (Theorem 6.16$_{272}$), respectively.

From a technical point of view, we largely restrict ourselves to proving these inequalities for the general version of these operators. Unfortunately, on a chronological reading of this dissertation one may not be aware of the properties (or indeed, the existence) of these more general objects, so for convenience we repeat the relevant results explicitly for the case of linear operators, i.e., when we are dealing with matrices. The following results provide the required relations that make this possible without having to fully duplicate the proofs.

Lemma B.1. *Any transition matrix T is a lower transition operator.*

Proof. This is proved in the main text of Section 3.4_{116}. □

Lemma B.2. *Any rate matrix Q is a lower transition rate operator.*

Proof. This is proved in the main text of Section 6.2_{265}. □

Lemma B.3. *Let Q be a rate matrix, and let $\underline{Q}$ be the lower transition rate operator such that $\underline{Q} := Q$ (this is possible by Lemma B.2). Fix any $t \in \mathbb{R}_{\geq 0}$, let e^{Qt} denote the matrix exponential as in Definition 4.5_{154}, and let $e^{\underline{Q}t}$ denote the generalised exponential as in Theorem 6.16_{272}. Then $e^{Qt} = e^{\underline{Q}t}$.*

Proof. This is proved in the main text of Section $6.3.2_{273}$. □

With these relations out of the way, let us now move on to the actual statements of the norm inequalities which are the subject of this appendix.

Lemma B.4. *For any $n \in \mathbb{Z}_{>0}$, let $\underline{T}_1, \ldots, \underline{T}_n$ and $\underline{S}_1, \ldots, \underline{S}_n$ be two finite sequences of lower transition operators. Then*

$$\left\| \prod_{i=1}^{n} \underline{T}_i - \prod_{i=1}^{n} \underline{S}_i \right\| \leq \sum_{i=1}^{n} \left\| \underline{T}_i - \underline{S}_i \right\|. \tag{B.1}$$

Proof. We provide a proof by induction. Clearly, Equation (B.1) holds for $n = 1$. Suppose that it holds for $n - 1$. We show that it then also holds for n.

$$\begin{aligned}
\left\| \prod_{i=1}^{n} \underline{T}_i - \prod_{i=1}^{n} \underline{S}_i \right\| &= \left\| \prod_{i=1}^{n} \underline{T}_i - \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n + \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n - \prod_{i=1}^{n} \underline{S}_i \right\| \\
&\leq \left\| \prod_{i=1}^{n} \underline{T}_i - \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n \right\| + \left\| \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n - \prod_{i=1}^{n} \underline{S}_i \right\| \\
&= \left\| \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{T}_n - \left(\prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n \right\| + \left\| \left(\prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right) \underline{S}_n \right\| \\
&\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \left\| \prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right\| \left\| \underline{S}_n \right\| \\
&\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \left\| \prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right\| \\
&\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \sum_{i=1}^{n-1} \left\| \underline{T}_i - \underline{S}_i \right\| = \sum_{i=1}^{n} \left\| \underline{T}_i - \underline{S}_i \right\| .
\end{aligned}$$

Here, in the second inequality, we applied Proposition 3.33_{117} and properties $\mathrm{N10}_{64}$ and $\mathrm{LT7}_{117}$. In the third inequality, we used property $\mathrm{LT4}_{117}$. In the final inequality, we used the induction hypothesis. □

Lemma B.5. *For any $n \in \mathbb{Z}_{>0}$, let $T_1,\dots,T_n$ and $S_1,\dots,S_n$ be two finite sequences of transition matrices. Then $\|\prod_{i=1}^n T_i - \prod_{i=1}^n S_i\| \leq \sum_{i=1}^n \|T_i - S_i\|$.*

Proof. This follows immediately from Lemmas B.1 and B.4. □

Lemma B.6. *Let $\underline{Q}$ be a lower transition rate operator, fix any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1,\dots,n\}$, any $\Delta_i \geq 0$ such that $\Delta_i \left\|\underline{Q}\right\| \leq 1$. Let $\Delta := \sum_{i=1}^n \Delta_i$. Then*

$$\left\|\prod_{i=1}^n (I+\Delta_i\underline{Q}) - (I+\Delta\underline{Q})\right\| \leq \Delta^2 \left\|\underline{Q}\right\|^2.$$

Proof. We provide a proof by induction. For $n=1$, the result is trivial. So consider the case $n \geq 2$ and assume that the result is true for $n-1$.

For all $i \in \{2,\dots,n\}$, since $\Delta_i \left\|\underline{Q}\right\| \leq 1$, it follows from Proposition 6.6_{266} that I and $(I+\Delta_i\underline{Q})$ are lower transition operators. Therefore,

$$\begin{aligned}
&\left\|\prod_{i=1}^n (I+\Delta_i\underline{Q}) - (I+\Delta\underline{Q})\right\| \\
&= \left\|\prod_{i=2}^n (I+\Delta_i\underline{Q}) + \Delta_1\underline{Q}\prod_{i=2}^n (I+\Delta_i\underline{Q}) - \left(I+\sum_{i=2}^n \Delta_i\underline{Q}\right) - \Delta_1\underline{Q}\right\| \\
&\leq \left\|\prod_{i=2}^n (I+\Delta_i\underline{Q}) - \left(I+\sum_{i=2}^n \Delta_i\underline{Q}\right)\right\| + \left\|\Delta_1\underline{Q}\prod_{i=2}^n (I+\Delta_i\underline{Q}) - \Delta_1\underline{Q}\right\| \\
&\leq \left(\sum_{i=2}^n \Delta_i\right)^2 \left\|\underline{Q}\right\|^2 + 2\Delta_1 \left\|\underline{Q}\right\| \left\|\prod_{i=2}^n (I+\Delta_i\underline{Q}) - I\right\| \\
&\leq \left(\sum_{i=2}^n \Delta_i\right)^2 \left\|\underline{Q}\right\|^2 + 2\Delta_1 \left\|\underline{Q}\right\| \sum_{i=2}^n \left\|(I+\Delta_i\underline{Q}) - I\right\| \\
&= \left(\sum_{i=2}^n \Delta_i\right)^2 \left\|\underline{Q}\right\|^2 + \left(2\Delta_1 \sum_{i=2}^n \Delta_i\right) \left\|\underline{Q}\right\|^2 \leq \left(\Delta_1 + \sum_{i=2}^n \Delta_i\right)^2 \left\|\underline{Q}\right\|^2 = \Delta^2 \left\|\underline{Q}\right\|^2,
\end{aligned}$$

where the second inequality follows from the induction hypothesis and property $\mathrm{LR6}_{266}$, and the third inequality follows from Lemma B.4. □

Lemma B.7. *Let $\underline{Q}$ be a lower transition rate operator and consider any $\Delta \in \mathbb{R}_{\geq 0}$. Then,*

$$\left\| e^{\underline{Q}\Delta} - (I + \Delta \underline{Q}) \right\| \leq \Delta^2 \left\| \underline{Q} \right\|^2.$$

Proof. Fix any $\varepsilon > 0$. Because of Theorem 6.16_{272}, there is some $u \in \mathscr{U}_{[0,\Delta]}$ such that $\sigma(u) \left\| \underline{Q} \right\| \leq 1$ and $\left\| e^{\underline{Q}\Delta} - \Phi_u \right\| \leq \varepsilon$, with Φ_u as in Equation $(6.6)_{271}$. By combining this with Lemma B.6, it follows that

$$\left\| e^{\underline{Q}\Delta} - (I + \Delta \underline{Q}) \right\| \leq \left\| e^{\underline{Q}\Delta} - \Phi_u \right\| + \left\| \Phi_u - (I + \Delta \underline{Q}) \right\| \leq \varepsilon + \Delta^2 \left\| \underline{Q} \right\|^2.$$

The result is now immediate since $\varepsilon > 0$ is arbitrary. □

Lemma B.8. *Let Q be a rate matrix and consider any $\Delta \in \mathbb{R}_{\geq 0}$. Then,*

$$\left\| e^{Q\Delta} - (I + \Delta Q) \right\| \leq \Delta^2 \left\| Q \right\|^2.$$

Proof. This follows immediately from Lemmas B.2_{392}, B.3_{392} and B.7. □

Lemma B.9. *Let $\underline{Q}$ be a lower transition rate operator, and consider any $\Delta \in \mathbb{R}_{\geq 0}$. Then,*

$$\left\| e^{\underline{Q}\Delta} - I \right\| \leq \Delta \left\| \underline{Q} \right\|.$$

Proof. Fix any $n \in \mathbb{Z}_{>0}$ and let $\Delta_n := \Delta/n$. Proposition 6.17_{273} then implies that

$$e^{\underline{Q}\Delta} = e^{\underline{Q}n\Delta_n} = e^{\underline{Q}\Delta_n} \cdots e^{\underline{Q}\Delta_n} = (e^{\underline{Q}\Delta_n})^n,$$

and therefore, it follows from Lemma B.4_{392} and the fact that $e^{\underline{Q}\Delta_n}$ is a lower transition operator, that

$$\left\| e^{\underline{Q}\Delta} - I \right\| = \left\| (e^{\underline{Q}\Delta_n})^n - I^n \right\| \leq n \left\| e^{\underline{Q}\Delta_n} - I \right\| \leq n \left\| e^{\underline{Q}\Delta_n} - (I + \Delta_n \underline{Q}) \right\| + n\Delta_n \left\| \underline{Q} \right\|,$$

which, when combined with Lemma B.7, implies that

$$\left\| e^{\underline{Q}\Delta} - I \right\| \leq n\Delta_n^2 \left\| \underline{Q} \right\|^2 + n\Delta_n \left\| \underline{Q} \right\| = \frac{1}{n}\Delta^2 \left\| \underline{Q} \right\|^2 + \Delta \left\| \underline{Q} \right\|.$$

Since $n \in \mathbb{Z}_{>0}$ is arbitrary, the result is now immediate. □

Lemma B.10. *Let Q be a rate matrix and consider any $\Delta \in \mathbb{R}_{\geq 0}$. Then,*

$$\left\| e^{Q\Delta} - I \right\| \leq \Delta \left\| Q \right\|.$$

Proof. This follows immediately from Lemmas B.2_{392}, B.3_{392} and B.9. □

Lemma B.11. *Let Q_1, Q_2 be two rate matrices and consider any $\Delta \in \mathbb{R}_{\geq 0}$. Then $\left\|e^{Q_1\Delta} - e^{Q_2\Delta}\right\| \leq \Delta \|Q_1 - Q_2\|$.*

Proof. Consider any $n \in \mathbb{Z}_{>0}$. It then follows from Proposition 4.11_{154} and Lemma $B.5_{393}$ that

$$\left\|e^{Q_1\Delta} - e^{Q_2\Delta}\right\| = \left\|\prod_{k=1}^{n} e^{Q_1\frac{\Delta}{n}} - \prod_{k=1}^{n} e^{Q_2\frac{\Delta}{n}}\right\| \leq n\left\|e^{Q_1\frac{\Delta}{n}} - e^{Q_2\frac{\Delta}{n}}\right\|.$$

which, since we know from Lemma B.8 that

$$\begin{aligned}
&\left\|e^{Q_1\frac{\Delta}{n}} - e^{Q_2\frac{\Delta}{n}}\right\| \\
&\qquad\leq \left\|e^{Q_1\frac{\Delta}{n}} - (I + \Delta/nQ_1)\right\| + \left\|\frac{\Delta}{n}(Q_1 - Q_2)\right\| + \left\|(I + \Delta/nQ_2) - e^{Q_2\Delta/n}\right\| \\
&\qquad\leq \frac{\Delta^2}{n^2}\|Q_1\|^2 + \frac{\Delta}{n}\|Q_1 - Q_2\| + \frac{\Delta^2}{n^2}\|Q_2\|^2,
\end{aligned}$$

implies that

$$\left\|e^{Q_1\Delta} - e^{Q_2\Delta}\right\| \leq \frac{\Delta^2}{n}\|Q_1\|^2 + \Delta\|Q_1 - Q_2\| + \frac{\Delta^2}{n}\|Q_2\|^2.$$

Since $n \in \mathbb{Z}_{>0}$ is arbitrary, the result is now immediate. □

Lemma B.12. *Consider a non-empty and bounded set of rate matrices $\mathcal{Q}$, any $n \in \mathbb{Z}_{>0}$ and, for all $i \in \{1, \ldots, n\}$, any $\Delta_i \in \mathbb{R}_{\geq 0}$ and $Q_i \in \mathcal{Q}$ such that $\Delta_i \|Q_i\| \leq 1$. Let $\Delta := \sum_{i=1}^{n} \Delta_i$. Then*

$$\left\|\prod_{i=1}^{n}(I + \Delta_i Q_i) - \left(I + \sum_{i=1}^{n} \Delta_i Q_i\right)\right\| \leq \Delta^2 \|\mathcal{Q}\|^2.$$

Proof. We provide a proof by induction. For $n = 1$, the result is trivial. So consider the case $n \geq 2$ and assume that the result is true for $n - 1$.

For all $i \in \{2, \ldots, n\}$, since $\Delta_i \|Q_i\| \leq 1$, it follows from Proposition 4.9[153] that $(I + \Delta_i Q_i)$ and I are transition matrices. Therefore,

$$\begin{aligned}
&\left\| \prod_{i=1}^{n}(I+\Delta_i Q_i) - \left(I + \sum_{i=1}^{n}\Delta_i Q_i\right)\right\| \\
&= \left\| \prod_{i=2}^{n}(I+\Delta_i Q_i) + \Delta_1 Q_1 \prod_{i=2}^{n}(I-\Delta_i Q_i) - \left(I + \sum_{i=2}^{n}\Delta_i Q_i\right) - \Delta_1 Q_1\right\| \\
&\leq \left\| \prod_{i=2}^{n}(I+\Delta_i Q_i) - \left(I + \sum_{i=2}^{n}\Delta_i Q_i\right)\right\| + \left\| \Delta_1 Q_1 \prod_{i=2}^{n}(I-\Delta_i Q_i) - \Delta_1 Q_1\right\| \\
&\leq \left(\sum_{i=2}^{n}\Delta_i\right)^2 \|\mathscr{Q}\|^2 + \Delta_1 \|Q_1\| \left\| \prod_{i=2}^{n}(I+\Delta_i Q_i) - I\right\| \\
&\leq \left(\sum_{i=2}^{n}\Delta_i\right)^2 \|\mathscr{Q}\|^2 + \Delta_1 \|Q_1\| \sum_{i=2}^{n} \|(I+\Delta_i Q_i) - I\| \\
&= \left(\sum_{i=2}^{n}\Delta_i\right)^2 \|\mathscr{Q}\|^2 + \Delta_1 \|Q_1\| \sum_{i=2}^{n} \Delta_i \|Q_i\| \\
&\leq \left(\sum_{i=2}^{n}\Delta_i\right)^2 \|\mathscr{Q}\|^2 + \left(\Delta_1 \sum_{i=2}^{n}\Delta_i\right) \|\mathscr{Q}\|^2 \leq \left(\Delta_1 + \sum_{i=2}^{n}\Delta_i\right)^2 \|\mathscr{Q}\|^2 = \Delta^2 \|\mathscr{Q}\|^2,
\end{aligned}$$

where the second inequality follows from the induction hypothesis and the third inequality follows from Lemma B.5[393]. □

List of Symbols and Terminology

This section aggregates the most important symbols and terminology used throughout this dissertation and, for symbols, provides a short description of their meaning. For ease of use, we present this collection using several lists, ordered by topic, and provide references to the location in this work where the symbol or terminology is first introduced.

Basic Notation

Symbol	Meaning	Location
$:=$	Definition	-
$\mathbb{Z}_{>0}$, $\mathbb{Z}_{\geq 0}$	The positive and non-negative integers	Sec. 1.6 43
$\mathbb{R}$, $\mathbb{Q}$	The real numbers and rational numbers	Sec. 1.6 43
$\mathbb{R}_{\geq c}$, $\mathbb{R}_{>c}$, $\mathbb{R}_{<c}$	For $c \in \mathbb{R}$, the reals that are at least/greater than/less than c; the non-negative/positive/negative reals when $c = 0$	Sec. 1.6 43
$\{a_i\}_{i \in \mathbb{Z}_{>0}}$	Sequence of quantities indexed by $i \in \mathbb{Z}_{>0}$	Sec. 1.6 43
$\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c$	Limit statement that $\lim_{i \to +\infty} a_i = c$	Sec. 1.6 43
$\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c^+$	Limit from above, i.e. $a_i \geq c$ for all $i \in \mathbb{Z}_{>0}$	Sec. 1.6 43
$\{a_i\}_{i \in \mathbb{Z}_{>0}} \to c^-$	Limit from below, i.e. $a_i \leq c$ for all $i \in \mathbb{Z}_{>0}$	Sec. 1.6 43
$\mathbb{I}_A$	Indicator of a set A; for any superset $C \supseteq A$ of A, and any $a \in C$, we let $\mathbb{I}_A(a) := 1$ if $a \in A$, and $\mathbb{I}_A(a) := 0$, otherwise	Sec. 1.6 43

Probabilities and Conditional Expectation

Symbol	Meaning	Location
Ω	Abstract outcome space	Sec. 2.1 46
$\mathscr{E}(\Omega)$	Set of all events; set of all subsets of Ω	Sec. 2.1 46
$\mathscr{E}(\Omega)_{\supset\emptyset}$	Set of all non-empty events; $\mathscr{E}(\Omega) \setminus \{\emptyset\}$	Sec. 2.1 46
P	Coherent conditional probability	Def. 2.2 48
$P(A)$	Probability of $A \in \mathscr{E}(\Omega)$	Def. 2.1 47
$P(A \mid C)$	Probability of $A \in \mathscr{E}(\Omega)$, conditional on $C \in \mathscr{E}(\Omega)_{\supset\emptyset}$	Def. 2.1 47
$\mathbb{B}$	Set of all functions on Ω that are bounded	Sec. 2.1.1 51
E	Coherent conditional prevision on $\mathscr{D} \subseteq \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$	Def. 2.3 52
$\mathscr{D}_{\mathscr{C}}$	Set of pairs $(\mathbb{I}_A, C) \in \mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ such that $(A, C) \in \mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\supset\emptyset}$	Sec. 2.1.1 51
$\mathbb{E}_P$	Conditional expectation that corresponds to P; coherent conditional prevision	Def. 2.5 54
$\mathscr{D}_P$	Largest subset of $\mathbb{B} \times \mathscr{E}(\Omega)_{\supset\emptyset}$ on which there is a unique coherent conditional prevision $\mathbb{E}_P$ corresponding to P	Def. 2.5 54

Time Points

Symbol	Meaning	Location
$\mathbb{H}$	General time domain	Def. 2.6 59
$\mathbb{D}$	Discrete time domain	Def. 2.7 59
$\mathscr{U}^{\mathbb{H}}$	Set of all finite sequences of time points in $\mathbb{H}$; each $u \in \mathscr{U}^{\mathbb{H}}$ is ordered and $u \subset \mathbb{H}$	Sec. 2.2.1 58
$\mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$	Subset of $\mathscr{U}^{\mathbb{H}}$: $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$ is non-empty	Sec. 2.2.1 58
$\mathscr{U}^{\mathbb{H}}_{<t}$	Subset of $\mathscr{U}^{\mathbb{H}}$: $u \in \mathscr{U}^{\mathbb{H}}_{<t}$ satisfies $u \subset \mathbb{R}_{<t}$	Sec. 2.2.1 58
$\mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$	Subset of $\mathscr{U}^{\mathbb{R}_{\geq 0}}_{\supset\emptyset}$: $u \in \mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$ partitions $[t,s]$	Sec. 2.2.1 58
Δ^u_i	Sequential difference $\Delta^u_i := t_i - t_{i-1}$ for $u \in \mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$ with $u = t_0, \dots, t_n$	Sec. 2.2.1 58
$\sigma(u)$	Maximum Δ^u_i for $u = t_0, \dots, t_n \in \mathscr{U}^{\mathbb{R}_{\geq 0}}_{[t,s]}$, that is $\sigma(u) := \max\{\Delta^u_i : i \in \{1, \dots, n\}\}$	Sec. 2.2.1 58

States, Functions, and Operators

Symbol	Meaning	Location
$\mathscr{X}$	Generic state space; non-empty finite set	Sec. 2.2.2$_{61}$
x	State; generic element of $\mathscr{X}$	Sec. 2.2.2$_{61}$
$\mathscr{X}_t$	State space at explicit time point $t \in \mathbb{H}$	Sec. 2.2.2$_{61}$
x_t	State at time t; generic element of $\mathscr{X}_t$	Sec. 2.2.2$_{61}$
$\mathscr{X}_u$	Joint state space at time points $u \in \mathscr{U}^{\mathbb{H}}$	Sec. 2.2.2$_{61}$
x_u	Joint state at time points $u \in \mathscr{U}^{\mathbb{H}}$; generic element of $\mathscr{X}_u$	Sec. 2.2.2$_{61}$
$\mathscr{L}(\mathscr{X})$	Set of real-valued functions on $\mathscr{X}$	Sec. 2.2.3$_{62}$
$\mathscr{L}(\mathscr{X}_u)$	Set of real-valued functions on $\mathscr{X}_u$, with $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$	Sec. 2.2.3$_{62}$
$\mathscr{L}(\mathscr{X})^\top$	The dual space of $\mathscr{L}(\mathscr{X})$; set of all real-valued linear functionals on $\mathscr{L}(\mathscr{X})$	App. A.2$_{380}$
$\mathbb{M}$	Set of all linear maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$; any $T \in \mathbb{M}$ is interchangeably called a matrix	Sec. 2.2.3$_{62}$
$T(x,y)$	The x,y-entry of the matrix representation of $T \in \mathbb{M}$, with $x,y \in \mathscr{X}$	App. A.3$_{383}$
$T(x,\cdot)$	The x-row of the matrix $T \in \mathbb{M}$; also a linear functional in $\mathscr{L}(\mathscr{X})^\top$	App. A.3$_{383}$
$\|f\|$	Supremum norm of $f \in \mathscr{L}(\mathscr{X}_u)$, with $u \in \mathscr{U}^{\mathbb{H}}_{\supset\emptyset}$	Sec. 2.2.3$_{62}$
$\|\phi^\top\|_*$	Induced dual norm of $\phi^\top \in \mathscr{L}(\mathscr{X})^\top$	App. A.2$_{380}$
$\|T\|$	Induced operator norm of non-negatively homogeneous operator $T : \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$	Sec. 2.2.3$_{62}$
$\|\mathscr{V}\|$	Supremum of norms $\|v\|$ where each $v \in \mathscr{V}$ is an element of a normed vector space	App. A$_{369}$

Stochastic Processes

Symbol	Meaning	Location
$\Omega_{\mathbb{H}}$	Outcome space of stochastic process with time domain $\mathbb{H}$; satisfies Eq. (2.8)$_{65}$	Sec. 2.3$_{64}$
$\omega \in \Omega_{\mathbb{H}}$	A *path*; generic realisation of stochastic process with time domain $\mathbb{H}$; function $\omega : \mathbb{H} \to \mathscr{X}$	Sec. 2.3$_{64}$
$\omega\|_u$	Path $\omega \in \Omega_{\mathbb{H}}$ restricted to $u \in \mathscr{U}^{\mathbb{H}}$	Sec. 2.3$_{64}$
$\langle \mathscr{E} \rangle$	Algebra generated by set $\mathscr{E} \subseteq \mathscr{E}(\Omega_{\mathbb{H}})$	Sec. 2.3$_{64}$
$(X_t = x)_{\mathbb{H}}$	Elementary event with $t \in \mathbb{H}$	Sec. 2.3$_{64}$
$\mathscr{E}_u^{\mathbb{H}}$	Set of elementary events whose time point either follows or belongs to $u \in \mathscr{U}^{\mathbb{H}}$	Sec. 2.3$_{64}$
$\mathscr{A}_u^{\mathbb{H}}$	Algebra generated by $\mathscr{E}_u^{\mathbb{H}}$	Sec. 2.3$_{64}$
$(X_u = x_u)_{\mathbb{H}}$	Conjunction of elementary events on time points $u \in \mathscr{U}^{\mathbb{H}}$	Sec. 2.3$_{64}$
$(A, X_u = x_u)_{\mathbb{H}}$	Shorthand for the pair $(A, (X_u = x_u)_{\mathbb{H}})$; element of $\mathscr{E}(\Omega_{\mathbb{H}}) \times \mathscr{E}(\Omega_{\mathbb{H}})_{\supset \emptyset}$; conditional event with $A \in \mathscr{A}_u^{\mathbb{H}}$, $x_u \in \mathscr{X}_u$, and $u \in \mathscr{U}^{\mathbb{H}}$	Sec. 2.3$_{64}$
$\mathscr{C}_{\mathbb{H}}^{\mathrm{SP}}$	Domain of stochastic process with time domain $\mathbb{H}$; subset of $\mathscr{E}(\Omega_{\mathbb{H}}) \times \mathscr{E}(\Omega_{\mathbb{H}})_{\supset \emptyset}$	Sec. 2.3$_{64}$
P	Stochastic process (with time domain $\mathbb{H}$); coherent conditional probability on $\mathscr{C}_{\mathbb{H}}^{\mathrm{SP}}$	Def. 2.12$_{68}$
$\mathbb{P}^{\mathbb{H}}$	Set of all stochastic processes with time domain $\mathbb{H}$	Def. 2.12$_{68}$
$\mathscr{C}^{\mathrm{SP}}, \mathbb{P}, \ldots$	For any of the above symbols, if the time domain $\mathbb{H}$ is not explicitly given, it is usually implied that $\mathbb{H} = \mathbb{R}_{\geq 0}$	Sec. 2.3$_{64}$
$f(X_u)$	u-measurable function corresponding to $f \in \mathscr{L}(\mathscr{X}_u)$, with $u \in \mathscr{U}_{\supset \emptyset}^{\mathbb{H}}$	Def. 2.15$_{72}$

Discrete-Time Stochastic Processes

Symbol	Meaning	Location
$\tau_{0:n}$	Shorthand for the sequence $\tau_0, \dots, \tau_n$, where τ is the canonical time index of a discrete time domain $\mathbb{D}$. Moreover, $\tau_{0:(-1)} = \emptyset$ by convention	Sec. 3.1$_{85}$
$\mathscr{S}_{\mathbb{D}}$	The set of *situations* with time domain $\mathbb{D}$	Def. 3.1$_{85}$
$P\vert_{\mathbb{D}}$	Discrete-time process with time domain $\mathbb{D}$ that is the restriction of *continuous-time* stochastic process P	Def. 7.3$_{350}$

Transition Matrices and Lower Transition Operators

Symbol	Meaning	Location
T	Transition matrix; matrix that is row-stochastic	Def. 3.5$_{91}$
$\mathbb{T}$	Set of all transition matrices	Def. 3.5$_{91}$
$\underline{T}$	Lower transition operator	Def. 3.15$_{116}$
$\underline{\mathbb{T}}$	Set of all lower transition operators	Def. 3.15$_{116}$
$\mathscr{T}_{\underline{T}}$	Set of transition matrices that dominate $\underline{T}$	Def. 3.17$_{120}$

Terminology	Location
Set $\mathscr{T} \subseteq \mathbb{T}$ with separately specified rows	Def. 3.13$_{111}$

Discrete-Time (Imprecise-)Markov Chains

Symbol	Meaning	Location
$P(X_{\tau_{n+1}} = x_{\tau_{n+1}} \vert X_{\tau_n} = x_{\tau_n})$	Transition probabilities of discrete-time Markov chain	Def. 3.3$_{89}$

$\mathbb{P}^{\mathbb{D},\mathrm{M}}$	Set of all discrete-time Markov chains with time domain $\mathbb{D}$	Def. 3.3$_{89}$
$\mathbb{P}^{\mathbb{D},\mathrm{HM}}$	Set of all *homogeneous* discrete-time Markov chains with time domain $\mathbb{D}$	Def. 3.4$_{91}$
(T_n)	Family $(T_n)_{n\in\mathbb{Z}_{\geq 0}}$ of transition matrices corresponding to a discrete-time Markov chain	Def. 3.7$_{94}$
(T_{n,x_u})	Family of *history-dependent* transition matrices T_{n,x_u}, with $n \in \mathbb{Z}_{\geq 0}$, $u = \tau_{0:(n-1)}$, and $x_u \in \mathscr{X}_u$, corresponding to a stochastic process $P \in \mathbb{P}^{\mathbb{D}}$, where τ is the canonical time index of $\mathbb{D}$	Def. 3.8$_{101}$
$(\mathscr{T}_n)$	Family of sets of transition matrices with $n \in \mathbb{Z}_{\geq 0}$	Sec. 3.3$_{101}$
$P \sim (\mathscr{T}_n)$	Consistency of process P with $(\mathscr{T}_n)$	Def. 3.9$_{102}$
$\mathscr{M}$	Set of probability mass functions on $\mathscr{X}$	Sec. 3.3$_{101}$
$P \sim \mathscr{M}$	Consistency of process P with $\mathscr{M}$	Sec. 3.3$_{101}$
$\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$	Imprecise-Markov chain with time domain $\mathbb{D}$	Def. 3.11$_{104}$
$\underline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$	Lower expectation for $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$	Def. 3.12$_{105}$
$\overline{\mathbb{E}}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$	Upper expectation for $\mathbb{P}^{\mathbb{D}}_{(\mathscr{T}_n),\mathscr{M}}$	Def. 3.12$_{105}$

Transition Rate Matrices, Lower Transition Rate Operators, and their Exponentials

Symbol	Meaning	Location
Q	A generic (transition) rate matrix	Def. 4.4$_{150}$
$\mathscr{R}$	The set of all (transition) rate matrices	Def. 4.4$_{150}$
$\mathscr{Q}$	A set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$	Sec. 4.3$_{150}$
e^{Qt}	The *matrix exponential* of the matrix Qt, with $Q \in \mathscr{R}$ and $t \in \mathbb{R}_{\geq 0}$	Def. 4.5$_{154}$
$\underline{Q}$	Lower transition rate operator; often the *lower envelope* of a set $\mathscr{Q}$ of rate matrices	Def. 6.2$_{265}$
$\mathscr{Q}_{\underline{Q}}$	Set of rate matrices that dominate $\underline{Q}$	Def. 6.4$_{268}$
$e^{\underline{Q}t}$	(Generalised) exponential of $\underline{Q}t$, with $\underline{Q}$ a lower transition rate operator and $t \in \mathbb{R}_{\geq 0}$; a lower transition operator	Thm 6.16$_{272}$

Terminology	Location
Set $\mathscr{Q} \subseteq \mathscr{R}$ with separately specified rows	Def. 5.7$_{193}$

Semigroups and (Restricted) Transition Matrix Systems

Symbol	Meaning	Location
(e^{Qt})	Semigroup of transition matrices generated by $Q \in \mathscr{R}$; family of transition matrices e^{Qt} with $t \in \mathbb{R}_{\geq 0}$	Sec. 4.4$_{156}$
(T_t^s)	*Transition matrix system*, a two-parameter family of transition matrices T_t^s, with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, that satisfies specific properties	Def. 4.6$_{156}$
$\boldsymbol{\mathcal{T}}$	Set of all transition matrix systems	Def. 4.6$_{156}$
$(e^{Q(s-t)})$	*Exponential transition matrix system* corresponding to $Q \in \mathscr{R}$	Def. 4.8$_{158}$
$(T_t^s)_\mathbf{I}$	Transition matrix system *restricted* to interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$	Sec. 4.5$_{158}$
$\boldsymbol{\mathcal{T}}_\mathbf{I}$	Set of all restricted transition matrix systems on $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$	Sec. 4.5$_{158}$
$(T_t^s)_\mathbf{I} \otimes (S_t^s)_\mathbf{J}$	*Concatenation* of two restricted transition matrix systems $(T_t^s)_\mathbf{I}$ and $(S_t^s)_\mathbf{J}$	Def. 4.9$_{160}$
d	Metric on $\boldsymbol{\mathcal{T}}_\mathbf{I}$	Eq. (4.15)$_{162}$
$(e^{\underline{Q}t})$	Semigroup of lower transition operators generated by $\underline{Q}$; family of lower transition operators $e^{\underline{Q}t}$ with $t \in \mathbb{R}_{\geq 0}$	Sec. 6.3.1$_{270}$

Continuous-Time Stochastic Processes

Symbol	Meaning	Location
$\mathbb{P}^\mathrm{W}$	Set of *well-behaved* stochastic processes	Def. 4.1$_{145}$
(T_{t,x_u}^s)	Family of history-dependent transition matrices corresponding to stochastic process, with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$	Def. 4.2$_{148}$
(T_t^s)	Family of transition matrices corresponding to stochastic process, with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$	Def. 4.3$_{150}$

$\partial_+ T^t_{t,x_u}$	Right-sided partial derivative of history-dependent transition matrix T^t_{t,x_u}	Def. 4.10$_{167}$
$\partial_- T^t_{t,x_u}$	Left-sided partial derivative of history-dependent transition matrix T^t_{t,x_u}	Def. 4.10$_{167}$
$\partial T^t_{t,x_u}$	Partial derivative of history-dependent transition matrix T^t_{t,x_u}	Def. 4.10$_{167}$
$\overline{\partial}_+ T^t_{t,x_u}$	Right-sided *outer* partial derivative of history-dependent transition matrix T^t_{t,x_u}; subset of $\mathscr{R}$	Def. 4.11$_{168}$
$\overline{\partial}_- T^t_{t,x_u}$	Left-sided *outer* partial derivative of history-dependent transition matrix T^t_{t,x_u}; subset of $\mathscr{R}$	Def. 4.11$_{168}$
$\overline{\partial} T^t_{t,x_u}$	*Outer* partial derivative of history-dependent transition matrix T^t_{t,x_u}; subset of $\mathscr{R}$	Def. 4.11$_{168}$
$(\overline{\partial}_+ T^t_{t,x_u})$	Family of right-sided outer partial derivatives, with $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$	Sec. 4.6$_{166}$
$(\overline{\partial}_- T^t_{t,x_u})$	Family of left-sided outer partial derivatives, with $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$	Sec. 4.6$_{166}$
$(\overline{\partial} T^t_{t,x_u})$	Family of outer partial derivatives, with $t \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$, and $x_u \in \mathscr{X}_u$	Sec. 4.6$_{166}$

Continuous-Time (Imprecise-)Markov Chains

Symbol	Meaning	Location
$\mathbb{P}^{\mathrm{M}}$	Set of all Markov chains	Def. 5.1$_{182}$
$\mathbb{P}^{\mathrm{WM}}$	Set of all well-behaved Markov chains	Def. 5.1$_{182}$
$\mathbb{P}^{\mathrm{HM}}$	Set of all homogeneous Markov chains	Def. 5.2$_{185}$
$\mathbb{P}^{\mathrm{WHM}}$	Set of all well-behaved homogeneous Markov chains	Def. 5.2$_{185}$
Q_P	Unique transition rate matrix corresponding to a well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$	Sec. 5.1.1$_{185}$
$P \sim \mathscr{Q}$	Consistency of process P with $\mathscr{Q}$	Def. 5.3$_{189}$
$\mathscr{M}$	Set of probability mass functions on $\mathscr{X}$	Sec. 5.2$_{188}$
$P \sim \mathscr{M}$	Consistency of process P with $\mathscr{M}$	Def. 5.4$_{189}$
$\mathscr{P}_{\mathscr{Q},\mathscr{M}}$	Subset of $\mathscr{P} \subseteq \mathbb{P}$ consistent with $\mathscr{Q}$ and $\mathscr{M}$	Def. 5.5$_{189}$

$\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Continuous-time imprecise-Markov chain	Def. 5.6$_{190}$
$\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Continuous-time imprecise-Markov chain	Def. 5.6$_{190}$
$\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$	Continuous-time imprecise-Markov chain	Def. 5.6$_{190}$
$\mathcal{T}^{\mathscr{Q}}_{[a,b]}$	Set of restricted transition matrix systems on $[a,b]$ induced by $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Eq. (5.9)$_{195}$
${}^{\mathscr{Q}}_{\mathscr{M}}\mathscr{T}^{s}_{t}$	Set of (history-dependent) transition matrices induced by $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Eq. (5.11)$_{197}$
$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Lower expectation for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\overline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Upper expectation for $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Lower expectation for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\overline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Upper expectation for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$	Lower expectation for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\overline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$	Upper expectation for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$	Def. 5.8$_{198}$
$\underline{\mathbb{E}}_{\mathscr{M}}$	Lower expectation for the initial model described by $\mathscr{M}$	Def. 6.5$_{287}$
$(\underline{T}^{s}_{t})$	Family of lower transition operators $\underline{T}^{s}_{t}$ induced by set $\mathscr{P}$ of stochastic processes, with $t,s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$	Def. 6.1$_{261}$

Bibliography

[1] Odd O. Aalen and Søren Johansen. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5(3):141–150, 1978.

[2] Søren Asmussen. *Applied Probability and Queues*. Springer-Verlag New York, 2003.

[3] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[4] Patrizia Berti, Eugenio Regazzini, and Pietro Rigo. Coherent statistical inference and Bayes theorem. *The Annals of Statistics*, 19(1):366–381, 1991.

[5] Patrizia Berti and Pietro Rigo. On coherent conditional probabilities and disintegrations. *Annals of Mathematics and Artificial Intelligence*, 35(1-4):71–82, 2002.

[6] François Besnard and Lina Bertling. An approach for condition-based maintenance optimization applied to wind turbine blades. *Sustainable Energy, IEEE Transactions on*, 1(2):77–83, 2010.

[7] Gunter Bolch, Stefan Greiner, Hermann de Meer, and Kishor S. Trivedi. *Queueing Networks and Markov Chains: modeling and performance evaluation with computer science applications*. John Wiley & Sons, 2006.

[8] Jan van den Brand. A deterministic linear program solver in current matrix multiplication time. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 259–278, 2020.

[9] William C. Brown. *A Second Course in Linear Algebra*. John Wiley & Sons, 1988.

[10] Marcilia A. Campos, Graçaliz P. Dimuro, Antônio C. da Rocha Costa, and Vladik Kreinovich. Computing 2-step predictions for interval-valued finite stationary Markov chains. *Technical report utep-cs-03-20a, University of Texas at El Paso*, 2003.

[11] Gary A. Churchill. Stochastic models for heterogeneous DNA sequences. *Bulletin of Mathematical Biology*, 51(1):79–94, 1989.

[12] Fabio G. Cozman. Alternatives to Lavine's algorithm for calculation of posterior bounds given convex sets of distributions. *Technical report, CMU-RI-TR-97-38, Robotics Institute, Carnegie Mellon University*, 1997.

[13] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[14] Fabio G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[15] José C. F. da Rocha and Fabio G. Cozman. Evidence propagation in credal networks: an exact algorithm based on separately specified sets of probability. In *Advances in Artificial Intelligence (Proceedings of SBIA 2002)*, pages 376–385. Springer, 2002.

[16] Jasper De Bock. *Credal Networks under Epistemic Irrelevance: Theory and Algorithms*. PhD thesis, Ghent University, 2015.

[17] Jasper De Bock. The limit behaviour of continuous-time imprecise Markov chains. *Journal of Nonlinear Science*, 27(1):159–196, 2017.

[18] Jasper De Bock and Gert de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. *Journal of Artificial Intelligence Research*, 50:189–233, 2014.

[19] Gert de Cooman, Jasper De Bock, and Stavros Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76:18–46, 2016.

[20] Gert de Cooman and Filip Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, 2008.

[21] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9):1029–1052, 2010.

[22] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009.

[23] Gert de Cooman and Enrique Miranda. Lower previsions induced by filter maps. *Journal of Mathematical Analysis and Applications*, 410:101–116, 2014.

[24] Bruno de Finetti. *Theory of Probability*. New York, 1974.

[25] Robert Denk, Michael Kupper, and Max Nendel. Kolmogorov-type and general extension results for nonlinear expectations. *Banach Journal of Mathematical Analysis*, 12(3):515–540, 2018.

[26] Robert Denk, Michael Kupper, and Max Nendel. A semigroup approach to nonlinear Lévy processes. *Stochastic Processes and their Applications*, 130(3):1616–1642, 2020.

[27] Sébastien Destercke and Gert de Cooman. Relating epistemic irrelevance to event trees. In *Soft Methods for Handling Variability and Imprecision*, pages 66–73. Springer, 2008.

[28] Joseph L. Doob. *Stochastic Processes*. Wiley Classics Library Edition. John Wiley & Sons, 1953.

[29] Lester E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3(1):89–99, 1975.

[30] Stephen W. Duffy, Hsiu-Hsi Chen, Laszlo Tabar, and Nicholas E. Day. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in Medicine*, 14(14):1531–1543, 1995.

[31] Robert J. Elliott and John van der Hoek. Default times in a continuous time Markov chain economy. *Applied Mathematical Finance*, 20(5):450–460, 2013.

[32] Alexander Erreygers and Jasper De Bock. Imprecise continuous-time Markov chains: efficient computational methods with guaranteed error bounds. In *Proceedings of Machine Learning Research, Vol. 62 (Proceedings of ISIPTA 2017)*, pages 145–156. SIPTA, 2017.

[33] Alexander Erreygers and Jasper De Bock. Bounding inferences for large-scale continuous-time Markov chains: A new approach based on lumping and imprecise Markov chains. *International Journal of Approximate Reasoning*, 115:96–133, 2019.

[34] Alexander Erreygers and Jasper De Bock. First steps towards an imprecise Poisson process. *Proceedings of Machine Learning Research, Vol. 103 (Proceedings of ISIPTA 2019)*, pages 175–184, 2019.

[35] Alexander Erreygers, Cristina Rottondi, Giacomo Verticale, and Jasper De Bock. Imprecise Markov models for scalable and robust performance evaluation of flexi-grid spectrum allocation policies. *IEEE Transactions on Communications*, 66:5401–5414, 2018.

[36] Swapna S. Gokhale, Michael R. Lyu, and Kishor S. Trivedi. Analysis of software fault removal policies using a non-homogeneous continuous time Markov chain. *Software Quality Journal*, 12(3):211–230, 2004.

[37] Alexandre Goldsztejn and Arnold Neumaier. On the exponentiation of interval matrices. *Reliable Computing*, 20:52–72, 2014.

[38] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 3rd edition, 2001.

[39] Charles M. Grinstead and James L. Snell. *Introduction to Probability*. American Mathematical Society, 2012.

[40] Xianping Guo and Onésimo Hernández-Lerma. *Continuous-time Markov decision processes: theory and applications*. Springer, 2009.

[41] Xianping Guo and Onésimo Hernández-Lerma. Continuous-time controlled Markov chains. *Annals of Applied Probability*, 13(1):363–388, 2003.

[42] Paul R. Halmos. *Finite-Dimensional Vector Spaces*. Springer-Verlag, 1974.

[43] Darald J. Hartfiel. On the solutions to $x'(t) = A(t)x(t)$ over all $A(t)$, where $P \leq A(t) \leq Q$. *Journal of Mathematical Analysis and Applications*, 108(1):230–240, 1985.

[44] Darald J. Hartfiel. Sequential limits in Markov set-chains. *Journal of Applied Probability*, 28(4):910–913, 1991.

[45] Darald J. Hartfiel. *Markov Set-Chains*. Springer, 1998.

[46] Darald J. Hartfiel and Eugene Seneta. On the theory of Markov set-chains. *Advances in Applied Probability*, 26(4):947–964, 1994.

[47] Filip Hermans. *An operational approach to graphical uncertainty modelling*. PhD thesis, Ghent University, 2012.

[48] Filip Hermans and Damjan Škulj. Stochastic processes. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, chapter 11. Wiley, 2014.

[49] Ronald A. Howard. *Dynamic Probabilistic Systems. Volume 1: Markov Models*. John Wiley & Sons, 1971.

[50] Peter J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[51] John K. Hunter and Bruno Nachtergaele. *Applied Analysis*. World Scientific Publishing Company, 2001.

[52] Christopher H. Jackson, Linda D. Sharples, Simon G. Thompson, Stephen W. Duffy, and Elisabeth Couto. Multistate Markov models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(2):193–209, 2003.

[53] Jean T. Johnson and Glenn R. Luecke. Nonhomogeneous, continuous-time Markov chains defined by series of proportional intensity matrices. *Stochastic Processes and their Applications*, 32(1):171–181, 1989.

[54] Olav Kallenberg. *Foundations of Modern Probability*. Springer Science & Business Media, 2006.

[55] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, 1976.

[56] Andrey N. Kolmogorov. On the theory of Markov chains. Translated from the Russian by G. Lindquist. In A. N. Shiryayev, editor, *Selected Works of A. N. Kolmorogov. Volume II: Probability Theory and Mathematical Statistics*, chapter 21. Springer Science & Business Media, 1992.

[57] Igor O. Kozine and Lev V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8(2):97–113, 2002.

[58] Thomas Krak. Computing expected hitting times for imprecise Markov chains. *Proceedings of UQOP 2020 (forthcoming)*, 2020.

[59] Thomas Krak. An introduction to imprecise Markov chains. In Massimiliano Vasile, editor, *Optimization Under Uncertainty with Applications to Aerospace Engineering*, chapter 5. Springer Nature, 2021.

[60] Thomas Krak, Jasper De Bock, and Arno Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains. In *Proceedings of Machine Learning Research, Vol. 62 (Proceedings of ISIPTA 2017)*, pages 193–204. SIPTA, 2017.

[61] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017.

[62] Thomas Krak, Alexander Erreygers, and Jasper De Bock. An imprecise probabilistic estimator for the transition rate matrix of a continuous-time Markov chain. *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*, pages 124–132, 2018.

[63] Thomas Krak and Ad Feelders. Exceptional model mining with tree-constrained gradient ascent. *Proceedings of SIAM ICDM 2015*, pages 487–495, 2015.

[64] Thomas Krak, Natan T'Joens, and Jasper De Bock. Hitting times and probabilities for imprecise Markov chains. *Proceedings of Machine Learning Research, Vol. 103 (Proceedings of ISIPTA 2019)*, pages 265–275, 2019.

[65] Thomas Krak and Linda C. van der Gaag. Knowledge-based bias correction – a case study in veterinary decision support. *Frontiers in Artificial Intelligence and Applications (Proceedings of ECAI 2014)*, pages 489–494, 2014.

[66] Daniel Krpelík, Frank P. A. Coolen, and Louis J. M. Aslett. On robust Markov analysis for reliability assessment of complex systems using imprecise Markov chains. In *2019 International Conference on Information and Digital Technologies (IDT)*, pages 245–252. IEEE, 2019.

[67] Philippe Lemey, Marc Suchard, and Andrew Rambaut. Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm. *PLoS Currents Influenza*, 1, 2009.

[68] Thomas M. Liggett. *Continuous time Markov processes: an introduction.* American Mathematical Society, 2010.

[69] Stavros Lopatatzidis. *Robust Modelling and Optimisation in Stochastic Processes Using Imprecise Probabilities, with an Application to Queueing Theory.* PhD thesis, Ghent University, 2016.

[70] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. Computational methods for imprecise continuous-time birth-death processes: a preliminary study of flipping times. *Poster abstract in Proceedings of ISIPTA 2015*, page 344, 2015.

[71] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. Computing lower and upper expected first-passage and return times in imprecise birth–death chains. *International Journal of Approximate Reasoning*, 80:137–173, 2017.

[72] Stavros Lopatatzidis, Jasper De Bock, Gert de Cooman, Stijn Vuyst, and Joris Walraevens. Robust queueing theory: an initial study using imprecise probabilities. *Queueing Systems*, 82(1):75–101, 2015.

[73] Andrey A. Markov. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. Translated from Russian to English. In Ronald A. Howard, editor, *Dynamic Probabilistic Systems. Volume 1: Markov models*. John Wiley & Sons, 1971.

[74] Andrey A. Markov. An example of statistical investigation of the text *Eugene Onegin* concerning the connection of samples in chains. Translated from Russian to German to English. *Science in Context*, 19(4):591–600, 2006.

[75] Robert E. Megginson. *An Introduction to Banach Space Theory*. Springer, 1998.

[76] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45(1):3–49, 2003.

[77] James Munkres. *Topology*. Prentice Hall, 2nd edition, 2000.

[78] Max Nendel. Markov chains under nonlinear expectation. *Mathematical Finance*, 31:474–507, 2021.

[79] Max Nendel. On nonlinear expectations and Markov chains under model uncertainty. *International Journal of Approximate Reasoning*, 130:226–245, 2021.

[80] Jacques Neveu. *Mathematical foundations of the calculus of probability*. Holden-Day, 1965.

[81] Makiko Nisio. On a non-linear semi-group attached to stochastic optimal control. *Publications of the Research Institute for Mathematical Sciences*, 12:513–537, 1976.

[82] James R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[83] Edward P. Oppenheimer and Anthony N. Michel. Application of interval analysis techniques to linear systems. II. The interval matrix exponential function. *IEEE Transactions on Circuits and Systems*, 35(10):1230–1242, 1988.

[84] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the Web. Technical report, 1999-0120, Stanford University, 1999.

[85] Shige Peng. Nonlinear expectations and nonlinear Markov chains. *Chinese Annals of Mathematics*, 26(02):159–184, 2005.

[86] Frank P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, chapter 7, pages 156–198. 1926.

[87] Eugenio Regazzini. Finitely additive conditional probabilities. *Rendiconti del Seminario Matematico e Fisico di Milano*, 55(1):69–89, 1985.

[88] Andy Rindos, Steven Woolet, Ioannis Viniotis, and Kishor Trivedi. Exact methods for the transient analysis of nonhomogeneous continuous time Markov chains. In *Computations with Markov chains*, pages 121–133. Springer, 1995.

[89] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer-Verlag, 1998.

[90] L. C. G. Rogers and David Williams. *Diffusions, Markov Processes, and Martingales. Volume 1. Foundations*. Cambridge University Press, 2000.

[91] Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels. *Stochastic Processes for Insurance and Finance*. John Wiley & Sons, 2009.

[92] Cristina Rottondi, Alexander Erreygers, Giacomo Verticale, and Jasper De Bock. Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains. In *Proceedings of DRCN 2017*, pages 39–46, 2017.

[93] Halsey L. Royden and Patrick M. Fitzpatrick. *Real Analysis*. Prentice Hall, 4th edition, 2010.

[94] Walter Rudin. *Functional Analysis*. McGraw-Hill, 2nd edition, 1991.

[95] Jörn Sass and Ulrich G. Haussmann. Optimizing the terminal wealth under partial information: the drift process as a continuous time Markov chain. *Finance and Stochastics*, 8(4):553–577, 2004.

[96] Richard Serfozo. *Basics of Applied Stochastic Processes*. Springer, 2009.

[97] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, 2001.

[98] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance*. Wiley, 2019.

[99] Claude E. Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[100] Satish Shirali and Harkrishan L. Vasudeva. *Metric Spaces*. Springer-Verlag London, 1st edition, 2006.

[101] Damjan Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50:1314–1329, 2009.

[102] Damjan Škulj. Interval matrix differential equations. *arXiv preprint arXiv:1204.0467*, 2012.

[103] Damjan Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250(C):165–180, 2015.

[104] Damjan Škulj. Computing bounds for imprecise continuous-time Markov chains using normal cones. *arXiv preprint arXiv:2012.01029*, 2020.

[105] Terence Tao. *An Introduction to Measure Theory*. American Mathematical Society Providence, 2011.

[106] Natan T'Joens and Jasper De Bock. Limit behaviour of upper and lower expected time averages in discrete-time imprecise Markov chains. In *Communications in Computer and Information Science, Vol. 1238 (Proceedings of IPMU 2020, Part II)*, pages 224–238. Springer, 2020.

[107] Natan T'Joens, Thomas Krak, Jasper De Bock, and Gert de Cooman. A recursive algorithm for computing inferences in imprecise Markov chains. *Lecture Notes in Artificial Intelligence, Vol. 11726 (Proceedings of ECSQARU 2019)*, pages 455–465, 2019.

[108] Matthias Troffaes, Thomas Krak, and Henna Bains. Two-state imprecise Markov chains for statistical modelling of two-state non-Markovian processes. *Proceedings of Machine Learning Research, Vol. 103 (Proceedings of ISIPTA 2019)*, pages 394–404, 2019.

[109] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[110] Matthias C. M. Troffaes, Jacob Gledhill, Damjan Škulj, and Simon Blake. Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. *Proceedings of ISIPTA 2015*, pages 287–294, 2015.

[111] Charles F. Van Loan. A study of the matrix exponential. *Numerical Analysis Report No. 10, University of Manchester*, Reissued as MIMS EPrint 2006.397, Manchester Institute for Mathematical Sciences, University of Manchester, 2006.

[112] Paolo Vicig, Marco Zaffalon, and Fabio G. Cozman. Notes on "Notes on conditional previsions". *International Journal of Approximate Reasoning*, 44(3):358–365, 2007.

[113] Philipp Von Hilgers and Amy N. Langville. The five greatest applications of Markov chains. *Proceedings of the Markov Anniversary Meeting*, pages 155–158, 2006.

[114] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[115] Peter Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83(1):1–58, 1996.

[116] Dazhi Wang and Kishor S. Trivedi. Reliability analysis of phased-mission system with independent component repairs. *IEEE Transactions on Reliability*, 56(3):540–551, 2007.

[117] Peter M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, 1975. Revised journal version: [118].

[118] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

Dynamic Uncertainty. Amy Genser, 2021, mulberry paper and acrylic on sintra board.
Private collection.