Computing Inferences for Large-Scale Continuous-Time Markov Chains by Combining Lumping with Imprecision

Alexander Erreygers and Jasper De Bock

Ghent University, ELIS, SYSTeMS
{alexander.erreygers, jasper.debock}@ugent.be

Abstract If the state space of a homogeneous continuous-time Markov chain is too large, making inferences—here limited to determining marginal or limit expectations—becomes computationally infeasible. Fortunately, the state space of such a chain is usually too detailed for the inferences we are interested in, in the sense that a less detailed—smaller—state space suffices to unambiguously formalise the inference. However, in general this so-called lumped state space inhibits computing exact inferences because the corresponding dynamics are unknown and/or intractable to obtain. We address this issue by considering an imprecise continuous-time Markov chain. In this way, we are able to provide guaranteed lower and upper bounds for the inferences of interest, without suffering from the curse of dimensionality.

1 Introduction

State space explosion, or the exponential dependency of the size of a finite state space on a system's dimensions, is a frequently encountered inconvenience when constructing mathematical models of systems. In the setting of continuous-time Markov chains (CTMCs), this exponentially increasing number of states has as a consequence that using the model to perform inferences—for the sake of brevity here limited to marginal and limit expectations—about large-scale systems becomes computationally intractable. Fortunately, for many of the inferences we would like to make, a higher-level state description actually suffices, allowing for a reduced state space with considerably fewer states. However, unfortunately, the low-level description and its corresponding larger state space are necessary in order to accurately model the system's dynamics. Therefore, using the reduced state space to make inferences is generally impossible.

In this contribution, we address this problem using imprecise continuous-time Markov chains [5, 11, 16]. In particular, we outline an approach to determine guaranteed lower and upper bounds on marginal and limit expectations using the reduced state space. We introduced a preliminary version of this approach in [7, 15], but the current contribution is—to the best of our knowledge—its first fully general and theoretically justified exposition. Compared to other approaches [3, 8] that also determine lower and upper bounds on expectations, ours has the advantage that it is not restricted to limit expectations. Furthermore, based on our preliminary experiments, our approach seems to produce tighter bounds.

2 Continuous-Time Markov Chains

We are interested in making inferences about a system, more specifically about the state of this system at some future time t, denoted by X_t. The complication is that we are unable to predict the temporal evolution of the state with certainty. Therefore, at all times $t \in \mathbb{R}_{\geq 0}$,[1] the state X_t of the system is a random variable that takes values—generically denoted by x, y or z—in the state space $\mathcal{X}$.

2.1 Homogeneous Continuous-Time Markov Chains

We assume that the stochastic process that models our beliefs about the system, denoted by $(X_t)_{t \in \mathbb{R}_{\geq 0}}$, is a *continuous-time Markov chain* (CTMC) that is *homogeneous*. For a thorough treatment of the terminology and notation concerning CTMCs, we refer to [1,11,13]. Due to length constraints, we here limit ourselves to the bare necessities.

The stochastic process $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ is a CTMC if it satisfies the *Markov property*, which says that for all $t_1, \ldots, t_n, t, \Delta$ in $\mathbb{R}_{\geq 0}$ with $n \in \mathbb{N}$ and $t_1 < \cdots < t_n < t$, and all $x_1, \ldots, x_n, x, y$ in $\mathcal{X}$,

$$P(X_{t+\Delta} = y \mid X_{t_1} = x_1 \ldots, X_{t_n} = x_n, X_t = x) = P(X_{t+\Delta} = y \mid X_t = x). \quad (1)$$

The CTMC $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ is *homogeneous* if for all t, Δ in $\mathbb{R}_{\geq 0}$ and all x, y in $\mathcal{X}$,

$$P(X_{t+\Delta} = y \mid X_t = x) = P(X_\Delta = y \mid X_0 = x). \quad (2)$$

It is well-known that—both in the classical measure-theoretic framework [1] and the full conditional framework [11]—a homogeneous continuous-time Markov chain is uniquely characterised by a triplet $(\mathcal{X}, \pi_0, Q)$, where $\mathcal{X}$ is a state space, π_0 an initial distribution and Q a transition rate matrix.

The state space $\mathcal{X}$ is taken to be a non-empty, finite and—without loss of generality—ordered set. This way, any real-valued function f on $\mathcal{X}$ can be identified with a column vector, the x-component of which is $f(x)$. The set containing all real-valued functions on $\mathcal{X}$ is denoted by $\mathcal{L}(\mathcal{X})$.

The initial distribution π_0 is defined by

$$\pi_0(x) \coloneqq P(X_0 = x) \text{ for all } x \text{ in } \mathcal{X}, \quad (3)$$

and hence is a probability mass function on $\mathcal{X}$. We will (almost) exclusively be concerned with positive (initial) distributions, whom we collect in $\mathcal{D}(\mathcal{X})$ and will identify with row vectors.

[1] We use $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ to denote the set of non-negative real numbers and positive real numbers, respectively. Furthermore, we use $\mathbb{N}$ to denote the natural numbers and write $\mathbb{N}_0$ when including zero.

The transition rate matrix Q is a real-valued $|\mathcal{X}| \times |\mathcal{X}|$ matrix—or equivalently, a linear map from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{X})$—with non-negative off-diagonal entries and rows that sum up to zero. If for any t in $\mathbb{R}_{\geq 0}$ we define the *transition matrix over* t as

$$T_t \coloneqq e^{tQ} = \lim_{n \to +\infty} \left(I + \frac{t}{n} Q \right)^n, \tag{4}$$

then for all t in $\mathbb{R}_{\geq 0}$ and all x, y in $\mathcal{X}$,

$$P(X_t = y \mid X_0 = x) = T_t(x, y). \tag{5}$$

Finally, we denote by E the expectation operator with respect to the homogeneous CTMC $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ in the usual sense. It follows immediately from (3) and (5) that $E(f(X_t)) = \pi_0 T_t f$ for any f in $\mathcal{L}(\mathcal{X})$ and any t in $\mathbb{R}_{\geq 0}$.

2.2 Irreducibility

In order not to be tangled up in edge cases, in the remainder we are only concerned with irreducible transition rate matrices. Many equivalent necessary and sufficient conditions exist; see for instance [13, Theorem 3.2.1]. For the sake of brevity, we here say that a transition rate matrix Q is *irreducible* if, for all t in $\mathbb{R}_{>0}$ and x, y in $\mathcal{X}$, $T_t(x, y) > 0$.

Consider now a homogeneous CTMC that is characterised by $(\mathcal{X}, \pi_0, Q)$. It is then well-known that for any f in $\mathcal{L}(\mathcal{X})$, the limit $\lim_{t \to +\infty} E(f(X_t))$ exists. Even more, since we assume that Q is irreducible, this limit value is the same for all initial distributions π_0 [13, Theorem 3.6.2]! This common limit value, denoted by $E_\infty(f)$, is called the *limit expectation of* f. Furthermore, the irreducibility of Q also implies that there is a unique stationary distribution π_∞ in $\mathcal{D}(\mathcal{X})$ that satisfies the *equilibrium condition* $\pi_\infty Q = 0$. This unique distribution is called the *limit distribution*, as $E_\infty(f) = \pi_\infty f$.

In the remainder of this contribution, a *positive and irreducible CTMC* is any homogeneous CTMC characterised by a positive initial distribution π_0 and an irreducible transition rate matrix Q.

3 Lumping and the Induced (Imprecise) Process

In many practical applications—see for instance [3, 7, 8, 15]—we have a positive and irreducible CTMC that models our system and we want to use this chain to make inferences of the form $E(f(X_t)) = \pi_0 T_t f$ or $E_\infty(f)$. As analytically evaluating the limit in (4) is often infeasible, we usually have to resort to one of the many available numerical methods—see for example [12]—that approximate T_t. However, unfortunately these numerical methods turn out to be computationally intractable when the state space becomes large. Similarly, determining the unique distribution π_∞ that satisfies the equilibrium condition also becomes intractable for large state spaces.

Fortunately, as previously mentioned in Sect. 1, the state space $\mathcal{X}$ is often unnecessarily detailed. Indeed, many interesting inferences can usually still be unambiguously defined using real-valued functions on a less detailed state space that corresponds to a higher-order description of the system, denoted by $\hat{\mathcal{X}}$. However, this provides no immediate solution as the motive behind using the detailed state space $\mathcal{X}$ in the first place is that this allows us to accurately model the (uncertain) dynamics of the system using a homogeneous CTMC; see [3, 7–9, 15] for practical examples. In contrast, the dynamics of the induced stochastic process on the the reduced state space $\hat{\mathcal{X}}$ are often unknown and/or intractable to obtain, which inhibits us from making exact inferences using the induced stochastic process. We now set out to address this by allowing for imprecision.

3.1 Notation and Terminology Concerning Lumping

We assume that the lumped state space $\hat{\mathcal{X}}$ is obtained by *lumping*—sometimes called grouping or aggregating, see [2, 4]—states in $\mathcal{X}$, such that $1 < |\hat{\mathcal{X}}| \leq |\mathcal{X}|$. This lumping is formalised by the surjective *lumping map* $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$, which maps every state x in $\mathcal{X}$ to a state $\Lambda(x) = \hat{x}$ in $\hat{\mathcal{X}}$. In the remainder, we also use the inverse lumping map Γ, which maps every $\hat{x}$ in $\hat{\mathcal{X}}$ to a subset $\Gamma(\hat{x}) \coloneqq \{x \in \mathcal{X} \colon \Lambda(x) = \hat{x}\}$ of $\mathcal{X}$. Given such a lumping map Λ, a function f in $\mathcal{L}(\mathcal{X})$ is *lumpable with respect to* Λ if there is an $\hat{f}$ in $\mathcal{L}(\hat{\mathcal{X}})$ such that $f(x) = \hat{f}(\Lambda(x))$ for all x in $\mathcal{X}$. We use $\mathcal{L}_\Lambda(\mathcal{X}) \subseteq \mathcal{L}(\mathcal{X})$ to denote the set of all real-valued functions on $\mathcal{X}$ that are lumpable with respect to Λ.

As far as our results are concerned, it does not matter in which way the states are lumped. For a given f in $\mathcal{L}(\mathcal{X})$—recall that we are interested in the (limit) expectation of $f(X_t)$—a naive choice is to lump together all states that have the same image under f. However, this is not necessarily a good choice. One reason is that the resulting lumped state space can become very small, for example when f is an indicator, resulting in too much imprecision in the dynamics and/or the inference. Lumping-based methods therefore often let $\hat{\mathcal{X}}$ correspond to a natural higher-level description of the state of the system; see for example [3, 7, 8] for some positive results. An extra benefit of this approach is that the resulting model can be used to determine the (limit) expectation of multiple functions.

3.2 The Lumped Stochastic Process

Let $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a positive and homogeneous continuous-time Markov chain. Then any lumping map $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ unequivocally induces a *lumped stochastic process* $(\hat{X}_t)_{t\in\mathbb{R}_{\geq 0}}$. It has $\hat{\mathcal{X}}$ as state space and is defined by the relation

$$(\hat{X}_t = \hat{x}) \Leftrightarrow (X_t \in \Gamma(\hat{x})) \text{ for all } t \text{ in } \mathbb{R}_{\geq 0} \text{ and all } \hat{x} \text{ in } \hat{\mathcal{X}}. \tag{6}$$

In some cases, this lumped stochastic process is a homogeneous CTMC, and the inference of interest can then be computed using this reduced CTMC. See for example [2, Theorem 2.3(i)] for a necessary condition and [2, Theorem 2.4] or [4, Theorem 3] for a necessary and sufficient one. However, these conditions are very stringent. Indeed, in general, the lumped stochastic process is not homogeneous nor Markov. For this general case, we are not aware of any previous work that characterises the dynamics of the lumped stochastic process efficiently—i.e., directly from Λ, Q and π_0 and without ever determining T_t.

3.3 The Induced Imprecise Continuous-Time Markov Chain

Nevertheless, that is exactly what we now set out to do. We here only provide an intuitive explanation of our methodology; for a detailed exposition, we refer to Appendix E.

The essential point is that, while we cannot exactly determine the dynamics of the lumped stochastic process $(\hat{X}_t)_{t\in\mathbb{R}_{\geq 0}}$, we can consider a *set of possible stochastic processes*, not necessarily homogeneous and/or Markovian but all with $\hat{\mathcal{X}}$ as state space, that definitely contains the lumped stochastic process $(\hat{X}_t)_{t\in\mathbb{R}_{\geq 0}}$. In the remainder, we will denote this set by $\mathbb{P}_{\pi_0,Q,\Lambda}$. As is indicated by our notation, $\mathbb{P}_{\pi_0,Q,\Lambda}$ is fully characterised by π_0, Q and Λ.

Crucially, it turns out that $\mathbb{P}_{\pi_0,Q,\Lambda}$ takes the form of a so-called *imprecise continuous-time Markov chain*. For a formal definition of general imprecise CTMCs, and an extensive study of their properties, we refer the reader to the work of Krak et. al. [11] and De Bock [5]. For our present purposes, it suffices to know that tight lower and upper bounds on the expectations that correspond to the set of stochastic processes of an imprecise CTMC are relatively easy to obtain. In particular, they can be determined without having to explicitly optimise over this set of processes, thus mitigating the need to actually construct it.

There are many parallels between homogeneous CTMCs and imprecise CTMCs. For instance, the counterpart of a transition rate matrix is a *lower transition rate operator*. For our imprecise CTMC $\mathbb{P}_{\pi_0,Q,\Lambda}$, this lower transition rate operator is $\underline{\hat{Q}}\colon \mathcal{L}(\hat{\mathcal{X}}) \to \mathcal{L}(\hat{\mathcal{X}})\colon g \mapsto \underline{\hat{Q}}g$ where, for every g in $\mathcal{L}(\hat{\mathcal{X}})$, $\underline{\hat{Q}}g$ is defined by

$$[\underline{\hat{Q}}g](\hat{x}) \coloneqq \min\left\{\sum_{\hat{y}\in\hat{\mathcal{X}}} g(\hat{y}) \sum_{y\in\Gamma(\hat{y})} Q(x,y)\colon x\in\Gamma(\hat{x})\right\} \text{ for all } \hat{x} \text{ in } \hat{\mathcal{X}}. \quad (7)$$

Important to mention here is that in case the lumped state space corresponds to some higher-order state description, we often find that executing the optimisation in (7) is fairly straightforward, as is for instance observed in [7, 15].

The counterpart of the transition matrix over t is now the *lower transition operator over* t, denoted by $\underline{\hat{T}}_t\colon \mathcal{L}(\hat{\mathcal{X}}) \to \mathcal{L}(\hat{\mathcal{X}})$ and defined for all g in $\mathcal{L}(\hat{\mathcal{X}})$ by

$$\underline{\hat{T}}_t g \coloneqq \lim_{n\to+\infty}\left(I + \frac{t}{n}\underline{\hat{Q}}\right)^n g, \quad (8)$$

where the n-th power should be interpreted as consecutively applying the operator n times. Note how strikingly (8) resembles (4). Analogous to the precise case, one needs numerical methods—see for instance [6] or [11, Sect. 8.2]—to approximate $\underline{\hat{T}}_t g$ because analytically evaluating the limit in (8) is, at least in general, impossible.

4 Performing Inferences Using The Lumped Process

Everything is now set up to present our main results; see Appendix F for their proofs.

4.1 Guaranteed Bounds On Marginal Expectations

We first turn to marginal expectations. Once we have $\mathbb{P}_{\pi_0,Q,\Lambda}$, the following result is a—not quite immediate—consequence of [11, Corollary 8.3].

Theorem 1. *Consider a positive and irreducible CTMC characterised by $(\mathcal{X}, \pi_0, Q)$ and a lumping map $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$. Let f in $\mathcal{L}(\mathcal{X})$ be lumpable with respect to Λ and let $\hat{f}$ be the corresponding element of $\mathcal{L}(\hat{\mathcal{X}})$. Then for any t in $\mathbb{R}_{\geq 0}$,*

$$\hat{\pi}_0 \underline{\hat{T}}_t \hat{f} \leq E(f(X_t)) = \pi_0 T_t f \leq -\hat{\pi}_0 \underline{\hat{T}}_t(-\hat{f}),$$

where $\hat{\pi}_0$ in $\mathcal{D}(\hat{\mathcal{X}})$ is defined by $\hat{\pi}_0(\hat{x}) \coloneqq \sum_{x \in \Gamma(\hat{x})} \pi_0(x)$ for all $\hat{x}$ in $\hat{\mathcal{X}}$.

This result is highly useful in the setting that was outlined in Sect. 3. Indeed, for large systems we can use Theorem 1 to compute guaranteed lower and upper bounds on marginal expectations that cannot be computed exactly.

4.2 Guaranteed Bounds on Limit Expectations

Our second result provides guaranteed lower and upper bounds on limit expectations. This is extremely useful because the limit expectation is (almost surely) equal to the long-term temporal average due to the ergodic theorem [13, Theorem 3.8.1], and in practice—see for instance [7]—the inference one is interested in is often a long-term temporal average.

Theorem 2. *Consider an irreducible CTMC and a lumping map $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$. Let f in $\mathcal{L}(\mathcal{X})$ be lumpable with respect to Λ and let $\hat{f}$ be the corresponding element of $\mathcal{L}(\hat{\mathcal{X}})$. Then for all n in $\mathbb{N}_0$ and δ in $\mathbb{R}_{>0}$ such that $\delta \max\{|Q(x,x)|\colon x \in \mathcal{X}\} < 1$,*

$$\min(I + \delta \underline{\hat{Q}})^n \hat{f} \leq E_\infty(f) \leq -\min(I + \delta \underline{\hat{Q}})^n(-\hat{f}).$$

Furthermore, for fixed δ, the lower and upper bounds in this expression become monotonously tighter with increasing n, and each converges to a (possibly different) constant as n approaches $+\infty$.

This result can be used to devise an approximation method similar to [7, Algorithm 1]: we fix some value for δ, set $g_0 = \hat{f}$ (or $g_0 = -\hat{f}$) and then repeatedly compute $g_i \coloneqq (I + \delta \underline{\hat{Q}}) g_{i-1} = g_{i-1} + \delta \underline{Q} g_{i-1}$ until we empirically observe convergence of $\min g_i$ (or $-\min g_i$). In general, the lower and upper bounds obtained in this way are dependent on the choice of δ and this choice can therefore influence the tightness of the obtained bounds. Empirically, we have seen that smaller δ tend to yield tighter bounds, at the expense of requiring more iterations—that is, larger n—before empirical convergence.

4.3 Some Preliminary Numerical Results

Due to length constraints, we leave the numerical assessment of Theorem 1 for future work. For an extensive numerical assessment of—the method implied by—Theorem 2, we refer the reader to [7]. We believe that in this contribution, it is more fitting to compare our method to the only existing method—at least the only one that we are aware of—that also uses lumping to provide guaranteed lower and upper bounds on limit expectations. This method was first outlined by Franceschinis and Muntz [8], and then later improved by Buchholz [3]. In order to display the benefit of their methods, they use them to determine bounds on several performance measures for a closed queueing network that consists of a single server in series with multiple parallel servers. We use the method outlined in Sect. 4.2 to also compute bounds on these performance measures, as reported in Table 1. Note that our bounds are tighter than those of [8]. We would very much like to compare our method with the improved method of [3] as well. Unfortunately, the system parameters Buchholz uses do not—as far as we can tell—correspond to the number of states and the values for the performance measures he reports in [3, Fig. 3], thus preventing us from comparing our results.

Table 1. Comparison of the bounds obtained by using Theorem 2 with those obtained by the method presented in [8, Sect. 3.2] for the closed queueing network of [8].

		[8, Tab. 1]		Theorem 2	
	Exact	Lower	Upper	Lower	Upper
Mean queue length	1.2734	1.2507	1.3859	1.2664	1.2802
Throughput	0.9828	0.9676	0.9835	0.9826	0.9831

5 Conclusion

Broadly speaking, the conclusion of this contribution is that imprecise CTMCs are not only a robust uncertainty model—as they were originally intended to be—but also a useful computational tool for determining bounds on inferences for large-scale CTMCs. More concretely, the first important observation of this contribution is that lumping states in a homogeneous CTMC inevitably introduces imprecision, in the sense that we cannot exactly determine the parameters that describe the dynamics of the lumped stochastic process without also explicitly determining the original process. The second is that we can easily characterise a set of processes that definitely contains the lumped process, in the form of an imprecise CTMC. Using this imprecise CTMC, we can then determine guaranteed lower and upper bounds on marginal and limit expectations with respect to the original chain. From a practical point of view, these results are helpful in cases where state space explosion occurs: they allow us to determine guaranteed lower and upper bounds on inferences that we otherwise could not determine at all.

Regarding future work, we envision the following. For starters, a more thorough numerical assessment of the methods outlined in Sect. 4 is necessary. Furthermore, it would be of theoretical as well as practical interest to determine bounds on the *conditional* expectation of a lumpable function, or to consider functions that depend on the state at *multiple* time points. Finally, we are developing a method to determine lower and upper bounds on limit expectations that only requires the solution of a simple linear program.

Acknowledgements

Jasper De Bock's research was partially funded by H2020-MSCA-ITN-2016 UTOPIAE, grant agreement 722734. Furthermore, the authors are grateful to the reviewers for their constructive feedback and useful suggestions.

References

1. Anderson, W.J.: Continuous-Time Markov Chains. Springer-Verlag (1991). doi:10.1007/978-1-4612-3038-0
2. Ball, F., Yeo, G.F.: Lumpability and marginalisability for continuous-time Markov chains. Journal of Applied Probability **30**(3), 518–528 (1993). doi:10.2307/3214762
3. Buchholz, P.: An improved method for bounding stationary measures of finite Markov processes. Performance Evaluation **62**(1), 349–365 (2005). doi:10.1016/j.peva.2005.07.002
4. Burke, C.J., Rosenblatt, M.: A Markovian function of a Markov chain. The Annals of Mathematical Statistics **29**(4), 1112–1122 (1958). doi:10.1214/aoms/1177706444
5. De Bock, J.: The limit behaviour of imprecise continuous-time Markov chains. Journal of Nonlinear Science **27**(1), 159–196 (2017). doi:10.1007/s00332-016-9328-3
6. Erreygers, A., De Bock, J.: Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In: Proceedings of ISIPTA'17. pp. 145–156. PMLR (2017), extended pre-print: arXiv:1702.07150 [math.PR]
7. Erreygers, A., Rottondi, C., Verticale, G., De Bock, J.: Imprecise Markov models for scalable and robust performance evaluation of flexi-grid spectrum allocation policies (2018), submitted, arXiv:1801.05700 [cs:NI]
8. Franceschinis, G., Muntz, R.R.: Bounds for quasi-lumpable Markov chains. Performance Evaluation **20**(1), 223–243 (1994). doi:10.1016/0166-5316(94)90015-9
9. Ganguly, A., Petrov, T., Koeppl, H.: Markov chain aggregation and its applications to combinatorial reaction networks. Journal of Mathematical Biology **69**(3), 767–797 (2014). doi:10.1007/s00285-013-0738-7
10. Hermans, F., de Cooman, G.: Characterisation of ergodic upper transition operators. International Journal of Approximate Reasoning **53**(4), 573–583 (2012). doi:10.1016/j.ijar.2011.12.008
11. Krak, T., De Bock, J., Siebes, A.: Imprecise continuous-time Markov chains. International Journal of Approximate Reasoning **88**, 452–528 (2017). doi:10.1016/j.ijar.2017.06.012
12. Moler, C., Van Loan, C.: Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Review **45**(1), 3–49 (2003). doi:10.1137/S00361445024180

13. Norris, J.R.: Markov chains. Cambridge University Press (1997). doi:10.1017/CBO9780511810633
14. Regazzini, E.: Finitely additive conditional probabilities. Rendiconti del Seminario Matematico e Fisico di Milano **55**(1), 69–89 (1985). doi:10.1007/BF02924866
15. Rottondi, C., Erreygers, A., Verticale, G., De Bock, J.: Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains. In: Proceedings of DRCN 2017. pp. 39–46. VDE Verlag (2017)
16. Škulj, D.: Efficient computation of the bounds of continuous time imprecise Markov chains. Applied Mathematics and Computation **250**, 165–180 (2015). doi:10.1016/j.amc.2014.10.092
17. Stewart, W.J.: Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling. Princeton University Press (2009)
18. Škulj, D., Hable, R.: Coefficients of ergodicity for Markov chains with uncertain parameters. Metrika **76**(1), 107–133 (2013). doi:10.1007/s00184-011-0378-0

A Some Notation

We often rely on results from Krak et al. [11] when proving our results. In order to facilitate the use of these results, we try to adhere to the notation used in [11] as much as possible. For instance, throughout this appendix we will denote the original CTMC by P instead of by $(X_t)_{t\in\mathbb{R}_{\geq 0}}$, which we previously used in the main text.

A.1 Sequences of Time Points

A finite number of time points $t_1, \dots, t_n$ in $\mathbb{R}_{\geq 0}$ is always taken to be increasing, in the sense that $t_1 < \cdots < t_n$. Following Krak et al. [11], we collect all such sequences—including the empty sequence $\emptyset$—in the set $\mathcal{U}$, and denote a generic element of this set by u. Furthermore, we denote the set of all time sequences without the empty sequence by $\mathcal{U}_\emptyset$, and for all t in $\mathbb{R}_{\geq 0}$ use $\mathcal{U}_{<t}$ (or $\mathcal{U}_{\emptyset,<t}$) to denote the set of all (non-empty) time sequences of which the last time point strictly precedes t. Moreover, for any sequence $u = t_1, \dots, t_n$ in $\mathcal{U}$, we define $\mathcal{X}_u \coloneqq \prod_{i=1}^n \mathcal{X}$ and we use x_u to elegantly denote a generic n-tuple $(x_{t_1}, \dots, x_{t_n})$ in $\mathcal{X}_u$. For the empty sequence $\emptyset$, we have that $x_\emptyset$ is equal to the empty tuple $\diamond$ and that $\mathcal{X}_\emptyset = \{\diamond\}$.

We will sometimes need to concatenate two increasing sequences of finite time points, for instance u and v in $\mathcal{U}$. Since u and v can be identified with sets, we let $u \cup v$ denote their concatenation, taken to be their ordered union. Finally, for any sequence $u = t_0, \dots, t_n$ in $\mathcal{U}$, we let $\max u \coloneqq \max\{t_i \colon i \in \{1, \dots, n\}\}$, which, due to our convention that u is increasing, is equal to t_n. If u is the empty sequence, then statements like "$\max u < \cdot$" are taken to be vacuously true.

A.2 Indicators, Operators and Norms

Consider any non-empty finite set S, and collect all real-valued functions on S in $\mathcal{L}(S)$. An often-used type of function in $\mathcal{L}(S)$ is the *indicator* of some subset

$A \subseteq S$, denoted by $\mathbb{1}_A$ and defined by $\mathbb{1}_A(x) \coloneqq 1$ if x is an element of S and $\mathbb{1}_A(x) \coloneqq 0$ otherwise. In order not to obfuscate the notation too much, for all x in S, we write $\mathbb{1}_x$ instead of $\mathbb{1}_{\{x\}}$.

We now turn to operators on $\mathcal{L}(S)$. Let M be an operator with domain $\mathcal{L}(S)$ and range $\mathcal{L}(S)$. Then M is *non-negatively homogeneous* if, for all f in $\mathcal{L}(S)$ and all λ in $\mathbb{R}_{\geq 0}$, $M(\lambda f) = \lambda M f$. Such operators play an important role in (imprecise) CTMCs. Examples of non-negatively homogeneous operators that we have seen so far are I, Q, T_t, $\underline{\hat{Q}}$ and $\underline{\hat{T}}_t$. If furthermore $M(f+g) = Mf + Mg$ for all f, g in $\mathcal{L}(S)$, then M is *linear*. It is well-known that linear operators can be represented by matrices; previously encountered examples are I, Q and T_t.

We bestow $\mathcal{L}(S)$ with the maximum norm:

$$\|f\| \coloneqq \max\{|f(x)| \colon x \in S\} \text{ for all } f \text{ in } \mathcal{L}(S).$$

The maximum norm on $\mathcal{L}(S)$ induces an operator norm for non-negatively homogeneous operators:

$$\|M\| \coloneqq \sup\{\|Mf\| \colon f \in \mathcal{L}(S), \|f\| = 1\}.$$

Finally, we turn to transition rate matrices, i.e., matrices with non-negative off-diagonal elements and rows that sum to zero. We use $\mathcal{R}(S)$ to denote the set of all transition rate matrices that map $\mathcal{L}(S)$ to $\mathcal{L}(S)$. It is well-known that, for all Q in $\mathcal{R}(S)$,

$$\|Q\| = 2\max\{|Q(s,s)| \colon s \in S\} = 2\max\{-[Q\mathbb{1}_s](s) \colon s \in S\}. \tag{9}$$

B Extra Material for Sect. 2

Recall from Sect. 2.1 that we can consider stochastic processes in both the classical measure-theoretic framework and the full conditional framework. For the former, we refer to [1, Sect. 1.1] and references therein. Since the latter is the approach that is introduced and followed by Krak et. al. [11], it will be the approach that we will follow here. Therefore, we here briefly recall the notation, terminology and results from [11, Sects. 4 and 5] that we need in the remainder: we discuss coherent conditional probabilities in Sect. B.1, explain how stochastic processes are coherent conditional probabilities with a specific domain in Sect. B.2 and treat the special case of homogeneous CTMCs in Sect.B.3.

B.1 Coherent Conditional Probabilities

Fix some non-empty set S called the *outcome space*. For this outcome space S, we let $\mathcal{E}(S)$ denote the set of all subsets of S, and furthermore let $\mathcal{E}(S)_\emptyset \coloneqq \mathcal{E}(S) \setminus \emptyset$. The following definition is one of the most elementary and essential ones that we will need throughout the remainder.

Definition 1. Let S be a non-empty set and P a real-valued map from $\mathcal{C} \subseteq \mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$ to $\mathbb{R}$. Then P is a *coherent conditional probability* if, for all n in $\mathbb{N}$, (A_1, C_1), …, (A_n, C_n) in $\mathcal{C}$ and $\lambda_1, \dots, \lambda_n$ in $\mathbb{R}$,

$$\max\left\{ \sum_{i=1}^{n} \lambda_i \mathbb{1}_{C_i}(s)(P(A_i \mid C_i) - \mathbb{1}_{A_i}(s)) \colon s \in \cup_{i=1}^{n} C_i \right\} \geq 0.$$

Lemma 3 (Theorem 4 in [14]). *Let S be a non-empty set. If P is a coherent conditional probability on $\mathcal{C} \subseteq \mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$, then for any $\mathcal{C}^{\star}$ such that $\mathcal{C} \subseteq \mathcal{C}^{\star} \subseteq \mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$, P can be extended to a coherent conditional probability $P^{\star}$ on $\mathcal{C}^{\star}$, in the sense that $P^{*}(A \mid C) = P(A \mid C)$ for all $(A, C) \in \mathcal{C}$.*

Lemma 4 (Corollary 4.3 in [11]). *Let S be a non-empty set. Then P is a coherent conditional probability on $\mathcal{C} \subseteq \mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$ if and only if it can be extended to a coherent conditional probability on $\mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$.*

Definition 1 might seem rather abstract on first encounter, but our motivation for using it is the following result.

Lemma 5 ((5)–(8) in [14]). *Let S be a non-empty set. If P is a coherent conditional probability on $\mathcal{C} \subseteq \mathcal{E}(S) \times \mathcal{E}(S)_{\emptyset}$, then for all (A, C), (B, C) and (D, C) in $\mathcal{C}$ such that $(A, D \cap C)$ is in $\mathcal{C}$,*

(i) $P(A \mid C) \geq 0$;
(ii) $P(A \mid C) = 1$ *if* $C \subseteq A$;
(iii) $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$ *if* $A \cap B = \emptyset$;
(iv) $P(A \cap D \mid C) = P(A \mid D \cap C) P(D \mid C)$.

Lemma 5 states that a coherent conditional probability satisfies the standard laws of (conditional) probability on its domain: properties (i)–(iii) state that $P(\cdot \mid C)$ is a probability measure, while (iv) is Bayes' rule.

B.2 Stochastic Processes

Fix some finite state space $\mathcal{X}$. We are then uncertain about what the actual path $\omega \colon \mathbb{R}_{\geq 0} \to \mathcal{X}$ of the system will be. We therefore consider a set of paths Ω, which contains all "feasible" paths. The only thing that is required of Ω is that

$$(\forall u \in \mathcal{U}_{\emptyset})(\forall x_u \in \mathcal{X}_u)(\exists \omega \in \Omega)\ \omega(t) = x_t \text{ for all } t \text{ in } u. \tag{10}$$

For all t in $\mathbb{R}_{\geq 0}$ and x in $\mathcal{X}$, we then define the basic event

$$(X_t = x) \coloneqq \{\omega \in \Omega \colon \omega(t) = x\}.$$

Similarly, for all u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, we let

$$(X_u = x_u) \coloneqq \bigcap_{t \in u} (X_t = x_t).$$

We follow the convention that an empty intersection in expressions similar to the one above correspond to Ω; hence $(X_u = x_u) = \Omega$ if u is the empty sequence $\emptyset$.

For all u in $\mathcal{U}$, the set of elementary events

$$\mathcal{E}_u \coloneqq \{(X_t = x)\colon t \in u \cup [\max u, +\infty), x \in \mathcal{X}\}$$

induces an algebra $\mathcal{A}_u \coloneqq \langle \mathcal{E}_u \rangle$. We use these algebras to define the domain

$$\mathcal{C}^{\mathrm{SP}} \coloneqq \{(A_u, X_u = x_u)\colon u \in \mathcal{U}, x_u \in \mathcal{X}_u, A_u \in \mathcal{A}_u\},$$

and consider maps of the form

$$P\colon \mathcal{C}^{\mathrm{SP}} \to \mathbb{R}\colon (A_u, X_u = x_u) \mapsto P(A_u \mid X_u = x_u),$$

where—in order to not to unnecessarily clutter our notation—we leave out the conditioning event if it is $(X_\emptyset = x_\emptyset) = \Omega$:

$$P(A) \coloneqq P(A \mid X_\emptyset = x_\emptyset) = P(A \mid \Omega) \quad \text{for any } A \text{ in } \mathcal{A}_\emptyset.$$

Definition 2 (Definition 4.3 in [11]). A real-valued map P on $\mathcal{C}^{\mathrm{SP}}$ is a *stochastic process* if it is a coherent conditional probability on $\mathcal{C}^{\mathrm{SP}}$.

It immediately follows from Lemma 5 that a stochastic process P satisfies the laws of (conditional) probability. Because these laws are so well-known, we will frequently use them without explicitly referring to Lemma 5.

B.3 Precise (Homogeneous) Continuous-Time Markov Chains As Special Cases

The following is a more formal definition of the terms introduced in Sect. 2.1.

Definition 3. A stochastic process $P\colon \mathcal{C}^{\mathrm{SP}} \to \mathbb{R}$ is a *continuous-time Markov chain* (CTMC) if, for all t, Δ in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$, x, y in $\mathcal{X}$ and x_u in $\mathcal{X}_u$,

$$P(X_{t+\Delta} = y \mid X_u = x_u, X_t = x) = P(X_{t+\Delta} = y \mid X_t = x).$$

This CTMC P is *homogeneous* if furthermore

$$P(X_{t+\Delta} = y \mid X_t = x) = P(X_\Delta = y \mid X_0 = x)$$

for all t, Δ in $\mathbb{R}_{\geq 0}$ and x, y in $\mathcal{X}$.

Our statement in Sect. 2.1 that a homogeneous CTMC is uniquely characterised by the triplet $(\mathcal{X}, \pi_0, Q)$ is justified due to the following result.

Proposition 6 (Theorem 5.2 in [11]). *Let $\mathcal{X}$ be a state space, π_0 a distribution on $\mathcal{X}$ and Q a transition rate matrix. Then there is a unique homogeneous CTCM P such that (i) $P(X_0 = x) = \pi_0(x)$ for all x in $\mathcal{X}$ and (ii) $P(X_{t+\Delta} = y \mid X_t = x) = T_\Delta(x, y)$ for all x, y in $\mathcal{X}$ and t, Δ in $\mathbb{R}_{\geq 0}$.*

In this appendix, a *positive and irreducible CTMC* is any stochastic process P that is a homogeneous CTMC and that is (uniquely) characterised by a positive initial distribution π_0 and an irreducible transition rate matrix Q.

B.4 Irreducibility

An easy to check necessary and sufficient condition for irreducibility is based on the accessibility relation $\cdot \rightsquigarrow \cdot$ [13]. We say that a state x is *accessible* from a state y (or that y *leads* to x) if there is a sequence $y = x_0, x_1 \ldots, x_n = x$ in $\mathcal{X}$ such that $Q(x_{i-1}, x_i) > 0$ for all i in $\{1, \ldots, n\}$.

Proposition 7 (Theorem 3.2.1 in [13]). *The transition rate matrix Q is irreducible if and only if every state is accessible from any other state. More formally, this condition reads*

$$\mathcal{X}_{\text{top}} \coloneqq \{x \in \mathcal{X} \colon (\forall y \in \mathcal{X}) y \rightsquigarrow x\} = \mathcal{X}.$$

The following lemma is our main reason for assuming that the CTMC has a positive initial distribution and an irreducible transition rate matrix.

Lemma 8. *If P is a positive and irreducible CTMC, then for any $u = t_1, \ldots, t_n$ in $\mathcal{U}_\emptyset$ and $x_u = (x_1, \ldots, x_n)$ in $\mathcal{X}_u$,*

$$P(X_u = x_u) = P(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) > 0.$$

Proof. Repeated application of Bayes' rule and the Markov property (1) yields

$$\begin{aligned} &P(X_u = x_u) = P(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) \\ &\quad = \sum_{x_0 \in \mathcal{X}} P(X_0 = x_0) P(X_{t_1} = x_1 \mid X_0 = x_0) \prod_{i=2}^{n} P(X_{t_i} = x_i \mid X_{t_{i-1}} = x_{i-1}). \end{aligned}$$

We now use (3) and (5) to obtain

$$P(X_u = x_u) = \sum_{x_0 \in \mathcal{X}} \pi_0(x_0) T_{t_1}(x_0, x_1) \prod_{i=2}^{n} T_{(t_i - t_{i-1})}(x_{i-1}, x_i). \tag{11}$$

As Q is irreducible, $T_t(x, y)$ is positive for all t in $\mathbb{R}_{>0}$ and all x, y in $\mathcal{X}$. Hence, all terms in the product on the right hand side of (11) are positive. We now distinguish two cases: $t_1 > 0$ and $t_1 = 0$. In the first case, it again follows from the irreducibility that all $T_{t_1}(x_0, x_1)$ are positive. In the second case, $T_{t_1}(x_0, x_1)$ is zero if $x_0 \neq x_1$ and 1 if $x_0 = x_1$. Furthermore, $\pi_0(x_0) > 0$ for all x_0 in $\mathcal{X}$ by assumption. The stated now follows by observing that at least one of the terms in the sum is a product of positive real numbers and therefore positive itself, and that the other terms are non-negative. □

We will also need the following properties. The first one is essentially well-known, but we could not immediately find a good reference for it.

Lemma 9. *Let Q be a transition rate matrix and δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$. Then $(I + \delta Q)$ is a transition matrix. If Q is furthermore irreducible, then $(I + \delta Q)$ is aperiodic and irreducible in the sense of [13].*

Proof. Fix any δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$. It can then be immediately verified—see for instance [17, p. 289]—that T is a transition matrix.

That T is irreducible follows from the irreducibility of Q. Recall from Proposition 7 that the irreducibility of Q implies that for any x, y in $\mathcal{X}$ such that $x \neq y$, there is a sequence $y = x_0, \ldots, x_n$ in $\mathcal{X}$ such that $Q(x_{i-1}, x_i) > 0$ for all i in $\{1, \ldots, n\}$. Clearly, this implies that, for all i in $\{1, \ldots, n\}$,

$$T(x_{i-1}, x_i) = (I + \delta Q)(x_{i-1}, x_i) = I(x_{i-1}, x_i) + \delta Q(x_{i-1}, x_i) > 0.$$

From [13, Theorem 1.2.1], it follows that y leads to x with respect to T. Since any two distinct states are communicating with respect to T, we conclude that T is irreducible.

That T is aperiodic follows from [17, p. 304]. □

Lemma 10. *If Q is an irreducible transition rate matrix, then for all f in $\mathcal{L}(\mathcal{X})$, δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$ and n in $\mathbb{N}_0$,*

$$\min(I + \delta Q)^n f \leq \pi_\infty f,$$

where π_∞ is the stationary distribution of Q.

Proof. Fix any δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$, and let $T \coloneqq I + \delta Q$. We recall from Lemma 9 that T is an aperiodic and irreducible transition matrix. Furthermore, from the equilibrium condition $\pi_\infty Q = 0$ it follows that

$$\pi_\infty T = \pi_\infty(I + \delta Q) = \pi_\infty + \delta\pi_\infty Q = \pi_\infty.$$

Since π_∞ is an invariant distribution for the aperiodic and irreducible transition matrix T, it follows from [13, Theorem 1.8.3] that $\lim_{n\to+\infty}[T^n f](x) = \pi_\infty f$ for all f in $\mathcal{L}(\mathcal{X})$ and x in $\mathcal{X}$.

Fix now any f in $\mathcal{L}(\mathcal{X})$ and consider the sequence

$$\{\min(I + \delta Q)^n f\}_{n\in\mathbb{N}_0} = \{\min T^n f\}_{n\in\mathbb{N}_0}.$$

From the previous, we know that this sequence converges to $\pi_\infty f$ in the limit for $n \to +\infty$. Since T is a transition matrix (a matrix with non-negative elements and rows that sum to 1), it clearly holds that $\min g \leq \min Tg$ for all g in $\mathcal{L}(\mathcal{X})$. It now follows from repeated application of this inequality that the sequence $\{\min T^n f\}_{n\in\mathbb{N}_0}$ is non-decreasing, which proves the stated. □

C Imprecise Continuous-Time Markov Chains: A Brief Summary

In this supplementary section, we briefly introduce the notation, terminology and results concerning imprecise CTMCs [5,11,16] that we will need in the remainder.

C.1 Sets of Consistent Processes and Lower Expectations

In general, the main idea behind imprecise CTMCs is to consider a set of stochastic process instead of a single stochastic process. In particular, Krak et. al. [11] focus on three nested sets of processes, all characterised by a non-empty set of initial distributions $\mathcal{M}$ and a non-empty bounded set of transition rate matrices $\mathcal{Q} \subseteq \mathcal{R}(\mathcal{X})$. More specifically, they collect in $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$ all stochastic processes that are: (i) well-behaved, a technical condition [11, Definition 4.4]; (ii) consistent with $\mathcal{Q}$, in the sense that at all times the "instantaneous transition rate matrix" is contained in $\mathcal{Q}$ [11, Definition 6.1]; and (iii) consistent with $\mathcal{M}$, in the sense that $\mathcal{M}$ contains the initial distribution [11, Definition 6.2]. Similarly, $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ ($\mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}}$) contains all well-behaved (homogeneous) Markov processes that are consistent with $\mathcal{Q}$ and $\mathcal{M}$. These sets are clearly nested, in the sense that

$$\mathbb{P}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}} \subseteq \mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}} \subseteq \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}. \tag{12}$$

Using these sets of consistent stochastic processes, Krak et. al. [11] construct lower (and conjugate upper) expectations as follows. For any non-empty set of initial distributions $\mathcal{M}$ and non-empty bounded set of transition rate matrices $\mathcal{Q}$, they let

$$\underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(\cdot \mid \cdot) \coloneqq \inf\{E_P(\cdot \mid \cdot) \colon P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}\},$$

where E_P denotes the expectation with respect to the process P, and similarly for $\underline{E}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$ and $\underline{E}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}}$. It is now intuitively clear from (12) that

$$\underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(\cdot \mid \cdot) \leq \underline{E}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}(\cdot \mid \cdot) \leq \underline{E}^{\mathrm{WHM}}_{\mathcal{Q},\mathcal{M}}(\cdot \mid \cdot).$$

C.2 Lower Transition (Rate) Operators

With any non-empty bounded set of transition rate matrices $\mathcal{Q}$, we associate the operator $\underline{Q} \colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X}) \colon f \mapsto \underline{Q}f$ where, for all f in $\mathcal{L}(\mathcal{X})$,

$$[\underline{Q}f](x) \coloneqq \inf\{[Qf](x) \colon Q \in \mathcal{Q}\} \text{ for all } x \text{ in } \mathcal{X}. \tag{13}$$

This operator $\underline{Q}$ is called the *lower envelope* of $\mathcal{Q}$. By [11, Proposition 7.5], we know that it is a *lower transition rate operator* [11, Definition 7.2], a specific type of non-homogeneous operator that is a non-linear generalisation of the concept of a transition rate matrix. Hence, it should not come as a surprise that there is an equivalent of (4). Indeed, for any t in $\mathbb{R}_{\geq 0}$, one defines the *lower transition operator over* t as

$$\underline{T}_t \coloneqq \lim_{n \to +\infty} \left(I + \frac{t}{n}\underline{Q}\right)^n, \tag{14}$$

where the n-th power should be interpreted as n consecutive applications.

Almost everything has now been set up to state Proposition 11, which is the main result from imprecise CTMCs that we will need; we just have to introduce the following definition.

Definition 4 (Definition 7.3 in [11]). A non-empty set of transition rate matrices $\mathcal{Q} \subseteq \mathcal{R}(\mathcal{X})$ has *separately specified rows* if for any $|\mathcal{X}|$-tuple $(Q_x)_{x\in\mathcal{X}}$ with entries that are all elements of $\mathcal{Q}$, there is a $Q^\star$ in $\mathcal{Q}$ such that

$$Q^\star(x,y) = Q_x(x,y) \text{ for all } x,y \in \mathcal{X}.$$

Proposition 11 (Corollary 8.3 in [11]). *Let $\mathcal{M}$ be a non-empty set of initial distributions and $\mathcal{Q}$ a non-empty and bounded set of transition rate matrices that has separately specified rows. If $\underline{Q}$ is the corresponding lower transition rate operator (13), then for any t,Δ in $\overline{\mathbb{R}}_{\geq 0}$, u in $\mathcal{U}_{<t}$, x in $\mathcal{X}$, x_u in $\mathcal{X}_u$ and f in $\mathcal{L}(\mathcal{X})$,*

$$\begin{aligned}\underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(f(X_{t+\Delta}) \mid X_u = x_u, X_t = x) = \underline{E}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}(f(X_{t+\Delta}) \mid X_u = x_u, X_t = x)& \\ = [\underline{T}_\Delta f](x).& \quad (15)\end{aligned}$$

This result justifies calling $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$ (and $\mathbb{P}^{\mathrm{WM}}_{\mathcal{Q},\mathcal{M}}$) an imprecise CTMC, as (15) is an imprecise version of the Markov property (1). Even more, the imprecise CTMC $\mathbb{P}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}$ also satisfies an imprecise version of the law of iterated expectation.

Proposition 12 (Theorem 6.5 in [11]). *If $\mathcal{M}$ is a non-empty set of initial distributions and $\mathcal{Q}$ a non-empty, bounded and convex set of transition rate matrices, then for any u,v,w in $\mathcal{U}$ with $\max u < \min v$ and $\max v < \min w$, x_u in $\mathcal{X}_u$ and f in $\mathcal{L}(\mathcal{X}_{u\cup v\cup w})$,*

$$\begin{aligned}&\underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(f(X_u,X_v,X_w) \mid X_u = x_u) \\ &\qquad = \underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(\underline{E}^{\mathrm{W}}_{\mathcal{Q},\mathcal{M}}(f(X_u,X_v,X_w) \mid X_u = x_u, X_v) \mid X_u = x_u).\end{aligned}$$

We conclude this section with a strengthened version of [11, LR5].

Lemma 13. *Let $\mathcal{Q}$ be a non-empty and bounded set of transition rate matrices with associated lower transition rate operator $\underline{Q}$. Then $\|\underline{Q}\| = \sup\{\|Q\| \colon Q \in \mathcal{Q}\}$, such that $\|Q\| \leq \|\underline{Q}\|$ for all Q in $\mathcal{Q}$.*

Proof. Recall from (9) that, for all Q in $\mathcal{Q}$,

$$\|Q\| = 2\max\{|Q(x,x)| \colon x \in \mathcal{X}\} = 2\max\{-[Q\mathbb{1}_x](x) \colon x \in \mathcal{X}\}.$$

Moreover, by [6, Proposition 4],

$$\|\underline{Q}\| = 2\max\big\{-[\underline{Q}\mathbb{1}_x](x) \colon x \in \mathcal{X}\big\}.$$

Using (13) and executing some straightforward manipulations yields

$$\begin{aligned}\|\underline{Q}\| &= 2\max\{-\inf\{[Q\mathbb{1}_x](x) \colon Q \in \mathcal{Q}\} \colon x \in \mathcal{X}\} \\ &= 2\max\{\sup\{-[Q\mathbb{1}_x](x) \colon Q \in \mathcal{Q}\} \colon x \in \mathcal{X}\} \\ &= \sup\{2\max\{-[Q\mathbb{1}_x](x) \colon x \in \mathcal{X}\} \colon Q \in \mathcal{Q}\} = \sup\{\|Q\| \colon Q \in \mathcal{Q}\}.\end{aligned}$$

The stated now follows immediately from the final equality. ◻

C.3 Irreducibility

Just like precise CTMCs, their imprecise counterparts also have some nice ergodic properties. For a detailed exposition of these properties, we refer the interested reader to our previous work [5, 6]. We here only mention the definitions and results that we will need in the remainder.

Let $\mathcal{Q}$ be a non-empty and bounded set of lower transition rate operators. As previously mentioned in Appendix C.2, the corresponding lower envelope $\underline{Q}$ is a lower transition rate operator. For such a lower transition rate operator, the imprecise counterpart of the accessibility relation $\cdot \rightsquigarrow \cdot$ is the upper reachability relation [5, Definition 7]. We say that a state x is *upper reachable* from the state y, denoted by $y \rightarrowtail x$, if there is a sequence $y = x_0, \dots, x_n = x$ in $\mathcal{X}$ such that $-[\underline{Q}(-\mathbb{1}_{x_i})](x_{i-1}) > 0$ for all i in $\{1, \dots, n\}$.

Definition 5. Let $\mathcal{Q}$ be an non-empty bounded set of transition rate matrices. The corresponding lower transition rate operator $\underline{Q}$ is *irreducible* if any state is upper reachable from any other state, that is, if

$$\mathcal{X}_{\mathrm{top}} \coloneqq \{x \in \mathcal{X} \colon (\forall y \in \mathcal{X}) y \rightarrowtail x\} = \mathcal{X}.$$

It now follows from [5, Theorem 19] that if $\underline{Q}$ is irreducible, then $\underline{Q}$ is *ergodic*, meaning that, for all f in $\mathcal{L}(\mathcal{X})$, $\underline{T}_t f$ converges to a constant function in the limit for $t \to +\infty$ [5, Definition 6]. For all t in $\mathbb{R}_{>0}$ and x, y in $\mathcal{X}$, this also implies that $-[\underline{T}_t(-\mathbb{1}_x)](y) > 0$ [5, Proposition 17], which is similar to the definition of irreducibility in the precise case. Note also the similarity between Proposition 7 and Definition 5, which justifies the use of the term irreducible. Furthermore, the following property holds.

Corollary 14. *Let $\mathcal{Q}$ be a non-empty bounded set of transition rate matrices. If the corresponding lower envelope $\underline{Q}$ is irreducible, then for any f in $\mathcal{L}(\mathcal{X})$ and δ in $\mathbb{R}_{>0}$ such that $\delta\|\underline{Q}\| < 2$, $(I + \delta\underline{Q})^n f$ converges to a constant function in the limit for $n \to +\infty$: there is some f_δ in $\mathbb{R}$ such that $\lim_{n\to+\infty}(I + \delta\underline{Q})^n f = f_\delta \mathbb{1}_{\mathcal{X}}$. Moreover, $\{\min(I + \delta\underline{Q})^n f\}_{n\in\mathbb{N}}$ is a non-decreasing sequence that converges to f_δ.*

Proof. Fix any δ in $\mathbb{R}_{>0}$ such that $\delta\|\underline{Q}\| < 2$ and let $\underline{T} \coloneqq (I + \delta\underline{Q})$. Then by [6, Proposition 3], $\underline{T}$ is a *lower transition operator* (see [11, Definition 7.1] or [6, Definition 1]). Furthermore, since $\underline{Q}$ is irreducible and hence ergodic, it follows from [6, Theorem 8] and either [10, Proposition 7] or [18, Theorem 21] that the lower transition operator $\underline{T}$ is also *ergodic*, meaning that, for all f in $\mathcal{L}(\mathcal{X})$, $\lim_{n\to+\infty} \underline{T}^n f = \lim_{n\to+\infty}(I + \delta\underline{Q})^n f$ exists and is a constant function, here denoted by $f_\delta \mathbb{1}_{\mathcal{X}}$. Finally, the non-decreasing character of the sequence in the statement can be verified by repeatedly applying [11, Definition 7.1(LT1)]; that the sequence converges to f_δ follows immediately from the previous. □

D The Lumped Stochastic Process

Before diving in head first, we first extend the inverse lumping map Γ to tuples of state assignments. For any u in $\mathcal{U}$, similar to what we did in Section A.1, we let $\hat{x}_u$ denote an element of $\hat{\mathcal{X}}_u \coloneqq \prod_{t \in u} \hat{\mathcal{X}}$. The domain of the inverse lumping map Γ can then be trivially extended to $\hat{\mathcal{X}}_u$ as follows: we let $\Gamma(\hat{x}_\emptyset) \coloneqq x_\emptyset$ and for all u in $\mathcal{U}_\emptyset$ and all $\hat{x}_u$ in $\hat{\mathcal{X}}_u$, we let

$$\Gamma(\hat{x}_u) \coloneqq \{x_u \in \mathcal{X}_u \colon (\forall t \in u) \Lambda(x_t) = \hat{x}_t\}.$$

D.1 A Formal Definition of the Lumped Stochastic Process

In order to define the lumped process rigorously, we need a more formal construction than that given in the main text (6). To that end, we now consider a positive and irreducible CTMC P that is characterised by $(\mathcal{X}, \pi_0, Q)$ and a lumping map $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$.

For starters, we first need to specify the set of "feasible" paths for the lumped process. A natural way is to map Ω, the set of "feasible" paths on $\mathcal{X}$, to a set of paths on $\hat{\mathcal{X}}$ using Λ:

$$\hat{\Omega} \coloneqq \{\Lambda \circ \omega \colon \omega \in \Omega\}, \tag{16}$$

where $\Lambda \circ \omega$ denotes the function composition of $\omega \colon \mathbb{R}_{\geq 0} \to \mathcal{X}$ and $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$, given by $\Lambda \circ \omega \colon \mathbb{R}_{\geq 0} \to \hat{\mathcal{X}} \colon t \mapsto (\Lambda \circ \omega)(t) \coloneqq \Lambda(\omega(t))$. Note that because Ω satisfies (10), $\hat{\Omega}$ clearly satisfies a lumped version of (10):

$$(\forall u \in \mathcal{U}_\emptyset)(\forall \hat{x}_u \in \hat{\mathcal{X}}_u)(\exists \hat{\omega} \in \hat{\Omega})\ \hat{\omega}(t) = \hat{x}_t \text{ for all } t \text{ in } u.$$

For any t in $\mathbb{R}_{\geq 0}$ and any $\hat{x}$ in $\hat{\mathcal{X}}$, we can now consider the elementary event

$$(\hat{X}_t = \hat{x}) \coloneqq \{\hat{\omega} \in \hat{\Omega} \colon \hat{\omega}(t) = \hat{x}\}.$$

As before, for any u in $\mathcal{U}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$, we also let

$$(\hat{X}_u = \hat{x}_u) \coloneqq \bigcap_{t \in u} (\hat{X}_t = \hat{x}_t),$$

where $(\hat{X}_\emptyset = \hat{x}_\emptyset) = \hat{\Omega}$. For any u in $\mathcal{U}$, the set of elementary elements

$$\hat{\mathcal{E}}_u \coloneqq \{(\hat{X}_t = \hat{x}) \colon t \in u \cup [\max u, +\infty), \hat{x} \in \hat{\mathcal{X}}\}$$

induces the algebra $\hat{\mathcal{A}}_u \coloneqq \langle \hat{\mathcal{E}}_u \rangle$. The domain of the lumped stochastic process $\hat{P}$ should hence be

$$\hat{\mathcal{C}}^{\mathrm{SP}} \coloneqq \{(\hat{A}_u, \hat{X}_u = \hat{x}_u) \colon u \in \mathcal{U}, \hat{x}_u \in \hat{\mathcal{X}}_u, \hat{A}_u \in \hat{\mathcal{A}}_u\}.$$

We have now introduced almost all concepts to formally define the lumped stochastic process $\hat{P}$. The sole remaining concept that we need is another inverse derived from Λ, this time from $\hat{\Omega}$ to Ω. To that end, we consider the map Γ_Ω that maps any subset $\hat{A}$ of $\hat{\Omega}$ to

$$\Gamma_\Omega(\hat{A}) \coloneqq \{\omega \in \Omega \colon \Lambda \circ \omega \in \hat{A}\}, \tag{17}$$

which is a subset of Ω. Note that Γ_Ω is indeed an inverse, as clearly

$$\{\Lambda \circ \omega \colon \omega \in \Gamma_\Omega(\hat{A})\} = \hat{A}. \tag{18}$$

Fix any u in $\mathcal{U}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$. Then some straightforward manipulations—similar to those used in the proof of Lemma 15—yield

$$\Gamma_\Omega(\hat{X}_u = \hat{x}_u) = \bigcup_{x_u \in \Gamma(\hat{x}_u)} (X_u = x_u) =: (X_u \in \Gamma(\hat{x}_u)), \tag{19}$$

More generally, we find the following.

Lemma 15. *If $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$ is a lumping map, then for all u in $\mathcal{U}$ and $\hat{A}_u$ in $\hat{\mathcal{A}}_u$, $\Gamma_\Omega(\hat{A}_u)$ is an element of $\mathcal{A}_u$.*

Proof. First, we observe that from (16) and (17) it follows that $\Gamma_\Omega(\hat{A}) = \emptyset$ if and only if $\hat{A} = \emptyset$. Next, we fix some u in $\mathcal{U}$ and some $\hat{A}_u$ in $\hat{\mathcal{A}}_u$ such that $\hat{A}_u \neq \emptyset$. Because $\hat{\mathcal{A}}_u$ is an algebra generated by the elementary events in $\hat{\mathcal{E}}_u$—see for instance also [11, Proof of Lemma C.3]—there is some time sequence v in $\mathcal{U}$ and a non-empty set of tuples $\hat{S} \subseteq \hat{\mathcal{X}}_{u \cup v}$ such that $\max u < \min v$ and

$$\hat{A}_u = \bigcup_{\hat{z}_{u \cup v} \in \hat{S}} (\hat{X}_v = \hat{z}_v).$$

If we let $w := u \cup v$, then

$$\begin{aligned}
\Gamma_\Omega(\hat{A}_u) &= \left\{ \omega \in \Omega \colon \Lambda \circ \omega \in \bigcup_{\hat{z}_w \in \hat{S}} (\hat{X}_w = \hat{z}_w) \right\} \\
&= \bigcup_{\hat{z}_w \in \hat{S}} \{\omega \in \Omega \colon \Lambda \circ \omega \in (\hat{X}_w = \hat{z}_w)\}.
\end{aligned}$$

Using the definition of $(\hat{X}_w = \hat{z}_w)$ and (16), we write this as

$$\begin{aligned}
\Gamma_\Omega(\hat{A}_u) &= \bigcup_{\hat{z}_w \in \hat{S}} \{\omega \in \Omega \colon (\forall t \in w) \Lambda(\omega(t)) = \hat{z}_t\} \\
&= \bigcup_{\hat{z}_w \in \hat{S}} \left(\bigcap_{t \in w} \{\omega \in \Omega \colon \Lambda(\omega(t)) = \hat{z}_t\} \right) \\
&= \bigcup_{\hat{z}_w \in \hat{S}} \left(\bigcap_{t \in w} \left[\bigcup_{z_t \in \Gamma(\hat{z}_t)} (X_t = z_t) \right] \right).
\end{aligned}$$

It is now immediately clear that $\Gamma_\Omega(\hat{A}_u)$ is an element of $\mathcal{A}_u$. ☐

The inverse Γ_Ω naturally suggests a sensible formal definition of the lumped stochastic process $\hat{P} \colon \hat{\mathcal{C}}^{\mathrm{SP}} \to \mathbb{R}$ where, for all $(\hat{A}_u, \hat{X}_u = \hat{x}_u)$ in $\hat{\mathcal{C}}^{\mathrm{SP}}$,

$$\hat{P}(\hat{A}_u \mid \hat{X}_u = \hat{x}_u) := \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(\Gamma_\Omega(\hat{A}_u) \mid X_u = x_u) P(X_u = x_u)}{\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = z_u)}. \tag{20}$$

That this is a proper definition follows from Lemmas 8 and 15: we know that $\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = x_u) > 0$ by the former and that $\Gamma_\Omega(\hat{A}_u)$ is in $\mathcal{A}_u$ by the latter, which in turn implies that $(\Gamma_\Omega(\hat{A}_u), X_u = x_u)$ is in $\mathcal{C}^{\mathrm{SP}}$. Note that if the conditioning event is $(\hat{X}_\emptyset = \hat{x}_\emptyset)$, then (20) reduces to

$$\hat{P}(\hat{A}_\emptyset) \coloneqq \hat{P}(\hat{A}_\emptyset \mid \hat{X}_\emptyset = \hat{x}_\emptyset) = P(\Gamma_\Omega(\hat{A}_\emptyset)). \tag{21}$$

One intuitively expects that this definition yields a stochastic process, and this intuition is verified by the following result.

Theorem 16. *If P is a positive and irreducible CTMC and $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then $\hat{P} \colon \hat{\mathcal{C}}^{\mathrm{SP}} \to \mathbb{R}$, as defined by (20), is a stochastic process.*

Proof. To prove the stated, we take a little detour. First, we combine Definition 3, Definition 2 and Lemma 3 to see that P can be extended to a coherent conditional probability $P^\star$ on $\mathcal{E}(\Omega) \times \mathcal{E}(\Omega)_\emptyset$. We take any such coherent extension $P^\star$, and use it to construct the real-valued map $\hat{P}^\star$ on $\mathcal{E}(\hat{\Omega}) \times \mathcal{E}(\hat{\Omega})_\emptyset$, defined by

$$\hat{P}^\star(\hat{A} \mid \hat{C}) \coloneqq P^\star(\Gamma_\Omega(\hat{A}) \mid \Gamma_\Omega(\hat{C})) \quad \text{for all } (\hat{A}, \hat{C}) \in \mathcal{E}(\hat{\Omega}) \times \mathcal{E}(\hat{\Omega})_\emptyset. \tag{22}$$

We will show that $\hat{P}$ is a stochastic process by verifying that $\hat{P}^\star$ is its (coherent) extension to $\mathcal{E}(\hat{\Omega}) \times \mathcal{E}(\hat{\Omega})_\emptyset$, after which we can simply invoke Lemma 4.

To that end, we first verify that $\hat{P}^\star$ is a coherent conditional probability. Therefore, we fix any n in $\mathbb{N}$, $(\hat{A}_1, \hat{C}_1)$, …, $(\hat{A}_n, \hat{C}_n)$ in $\mathcal{E}(\hat{\Omega}) \times \mathcal{E}(\hat{\Omega})_\emptyset$ and λ_1, …, λ_n in $\mathbb{R}$ and show that $\max S \geq 0$, where

$$S \coloneqq \left\{ \sum_{i=1}^n \lambda_i \mathbb{1}_{\hat{C}_i}(\hat{\omega}) \Big(\hat{P}^\star(\hat{A}_i \mid \hat{C}_i) - \mathbb{1}_{\hat{A}_i}(\hat{\omega}) \Big) \colon \hat{\omega} \in \bigcup_{i=1}^n \hat{C}_i \right\}.$$

Substituting (22) yields

$$S = \left\{ \sum_{i=1}^n \lambda_i \mathbb{1}_{\hat{C}_i}(\hat{\omega}) \Big(P^\star(\Gamma_\Omega(\hat{A}_i) \mid \Gamma_\Omega(\hat{C}_i)) - \mathbb{1}_{\hat{A}_i}(\hat{\omega}) \Big) \colon \hat{\omega} \in \bigcup_{i=1}^n \hat{C}_i \right\}.$$

Furthermore, using (17) and (18) yields

$$S = \left\{ \sum_{i=1}^n \lambda_i \mathbb{1}_{\hat{C}_i}(\Lambda \circ \omega) \Big(P^\star(\Gamma_\Omega(\hat{A}_i) \mid \Gamma_\Omega(\hat{C}_i)) - \mathbb{1}_{\hat{A}_i}(\Lambda \circ \omega) \Big) \colon \omega \in \bigcup_{i=1}^n \Gamma_\Omega(\hat{C}_i) \right\}.$$

Observe that for all ω in Ω and $\hat{A} \subseteq \hat{\Omega}$,

$$\mathbb{1}_{\hat{A}}(\Lambda \circ \omega) = \begin{cases} 1 & \text{if } \Lambda \circ \omega \in \hat{A} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } \omega \in \Gamma_\Omega(\hat{A}) \\ 0 & \text{otherwise} \end{cases} = \mathbb{1}_{\Gamma_\Omega(\hat{A})}(\omega), \tag{23}$$

where the second equality follows immediately from (17). We substitute (23) in our expression for S, to yield

$$S = \left\{ \sum_{i=1}^n \lambda_i \mathbb{1}_{C_i}(\omega) (P^\star(A_i \mid C_i) - \mathbb{1}_{A_i}(\omega)) \colon \omega \in \bigcup_{i=1}^n C_i \right\},$$

where, for all i in $\{1, \ldots, n\}$, we let $A_i \coloneqq \mathit{\Gamma}_{\Omega}(\hat{A}_i)$ and $C_i \coloneqq \mathit{\Gamma}_{\Omega}(\hat{C}_i)$. Because $P^\star$ is a coherent conditional probability on $\mathcal{E}(\Omega) \times \mathcal{E}(\Omega)_\emptyset$, it follows from Definition 1 that $\max S \geq 0$.

Next, we verify that $\hat{P}^\star$ coincides with $\hat{P}$ on $\hat{\mathcal{C}}^{\mathrm{SP}}$. To that end, we fix any $(\hat{A}_u, \hat{X}_u = \hat{x}_u)$ in $\hat{\mathcal{C}}^{\mathrm{SP}}$. Then by (22),

$$\hat{P}^\star(\hat{A}_u \mid \hat{X}_u = \hat{x}_u) = P^\star(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \mid \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)).$$

As $P^\star$ is a coherent conditional probability on $\mathcal{E}(\mathcal{X}) \times \mathcal{E}(\mathcal{X})_\emptyset$, it follows from Lemma 5(iv) that

$$\begin{aligned} P^\star(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \mid \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) P^\star(\mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) \\ = P^\star(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)), \end{aligned} \tag{24}$$

where $P^\star(\mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) \coloneqq P^\star(\mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u) \mid \Omega)$.

Recall from (19) that

$$\mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u) = \bigcup_{z_u \in \mathit{\Gamma}(\hat{x}_u)} (X_u = z_u),$$

which clearly is an element of $\mathcal{A}_\emptyset$. Consequently, $(\cup_{z_u \in \mathit{\Gamma}(\hat{x}_u)} (X_u = z_u), \Omega)$ is an element of $\mathcal{C}^{\mathrm{SP}}$. Since furthermore $P^\star$ is an extension of P, we find that

$$P^\star(\mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) = P(\cup_{z_u \in \mathit{\Gamma}(\hat{x}_u)} (X_u = z_u)) = \sum_{z_u \in \mathit{\Gamma}(\hat{x}_u)} P(X_u = z_u). \tag{25}$$

Recall from Lemma 15 that $\mathit{\Gamma}_{\Omega}(\hat{A}_u)$ is an element of $\mathcal{A}_u$, such that $\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)$ is an element of $\mathcal{A}_u$ as well. Consequently, we now find that

$$\begin{aligned} P^\star(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) &= P(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) \\ = P(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap (\cup_{x_u \in \hat{x}_u} (X_u = x_u))) &= \sum_{x_u \in \hat{x}_u} P(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \cap (X_u = x_u)) \\ &= \sum_{x_u \in \hat{x}_u} P(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \mid X_u = x_u) P(X_u = x_u). \end{aligned} \tag{26}$$

Since we know from Lemma 8 that $\sum_{z_u \in \mathit{\Gamma}(\hat{x}_u)} P(X_u = z_u) > 0$, substituting (25) and (26) in (24) yields

$$\begin{aligned} \hat{P}^\star(\hat{A}_u \mid \hat{X}_u = \hat{x}_u) &= P^\star(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \mid \mathit{\Gamma}_{\Omega}(\hat{X}_u = \hat{x}_u)) \\ &= \frac{\sum_{x_u \in \mathit{\Gamma}(\hat{x}_u)} P(\mathit{\Gamma}_{\Omega}(\hat{A}_u) \mid X_u = x_u) P(X_u = x_u)}{\sum_{z_u \in \mathit{\Gamma}(\hat{x}_u)} P(X_u = z_u)} \\ &= \hat{P}(\hat{A}_u \mid \hat{X}_u = \hat{x}_u). \end{aligned}$$

Hence, the coherent conditional probability $\hat{P}^\star$ on $\mathcal{E}(\hat{\mathcal{X}}) \times \mathcal{E}(\hat{\mathcal{X}})_\emptyset$ coincides with the real-valued map $\hat{P}$ on $\hat{\mathcal{C}}^{\mathrm{SP}}$. It now follows from Lemma 4 that $\hat{P}$ is a coherent conditional probability on $\hat{\mathcal{C}}^{\mathrm{SP}}$, such that it is a stochastic process by Definition 2. $\square$

Lemma 17. *If P is a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$, $\hat{y}$ in $\hat{\mathcal{X}}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$,*

$$\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u) = \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} \sum_{y \in \Gamma(\hat{y})} P(X_t = y \mid X_u = x_u) P(X_u = x_u)}{\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = z_u)}. \tag{27}$$

Proof. Follows immediately from (19) and (20). □

D.2 The Instantaneous Transition Rate Matrix of the Lumped Stochastic Process

For any t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$, we consider the real-valued map $\pi_{(u,\hat{x}_u,t)}$ that maps any x in $\mathcal{X}$ to

$$\pi_{(u,\hat{x}_u,t)}(x) \coloneqq \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = x)}{\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = z_u)}. \tag{28}$$

We use this notation in the following result, which provides the main motivation for seeing the lumped process as belonging to an imprecise CTMC that is consistent with a specific set of transition rate matrices.

Proposition 18. *If P is a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then, for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$, $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$,*

$$\lim_{\Delta \to 0^+} \frac{1}{\Delta} \Big(\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y}) \Big) = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} Q(x,y), \tag{29}$$

which is a convex combination of terms $\sum_{y \in \Gamma(\hat{y})} Q(x,y)$ with x in $\Gamma(\hat{x})$, and, if $t \neq 0$, also

$$\lim_{\Delta \to 0^+} \frac{1}{\Delta} \Big(\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_{t-\Delta} = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y}) \Big) = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} Q(x,y). \tag{30}$$

Before proving this result, we first state and, if necessary, prove three intermediary technical results.

Lemma 19. *Let P be a positive and irreducible CTMC. If t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$, then $\pi_{(u,\hat{x}_u,t)}$ is a positive distribution on $\mathcal{X}$.*

Proof. For any x in $\mathcal{X}$,

$$\pi_{(u,\hat{x}_u,t)}(x) = \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = x)}{\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = z_u)}$$

is a well-defined (in the sense that we do not divide by zero) positive real number due to Lemma 8. Hence, since

$$\begin{aligned}
\sum_{x \in \mathcal{X}} \pi_{(u,\hat{x}_u,t)}(x) &= \sum_{x \in \mathcal{X}} \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = x)}{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u)} \\
&= \sum_{x \in \mathcal{X}} \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = x)}{\sum_{z \in \mathcal{X}} \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = z)} = 1,
\end{aligned}$$

$\pi_{(u,\hat{x}_u,t)}$ is a positive distribution on $\mathcal{X}$. □

Lemma 20 (Theorem 2.1.1 in [13]). *Let Q be a transition rate matrix. Then for all t in $\mathbb{R}_{\geq 0}$ and all x, y in $\mathcal{X}$,*

$$\lim_{\Delta \to 0^+} \frac{T_{t+\Delta}(x,y) - T_t(x,y)}{\Delta} = [QT_t](x,y)$$

and, if $t \neq 0$,

$$\lim_{\Delta \to 0^+} \frac{T_t(x,y) - T_{t-\Delta}(x,y)}{\Delta} = [QT_t](x,y).$$

Lemma 21. *If P is a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then, for all t in $\mathbb{R}_{>0}$, u in $\mathcal{U}_{<t}$, $\hat{x}$ in $\hat{\mathcal{X}}$ and x in $\mathcal{X}$,*

$$\lim_{\Delta \to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(x) = \pi_{(u,\hat{x}_u,t)}(x).$$

Proof. Fix some Δ in $\mathbb{R}_{>0}$ such that $\Delta \leq t$ and $\max u < t - \Delta$. Then by (28),

$$\pi_{(u,\hat{x}_u,t-\Delta)}(x) = \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_{t-\Delta} = x)}{\sum_{z_u \in \Gamma(\hat{x}_u)} P(X_u = z_u)}.$$

For notational simplicity, we distinguish between two cases.

First, we assume that $u = t_1, \dots, t_n$ is not the empty sequence $\emptyset$. Then using (1), (2) and (5), we can write the numerator as

$$\begin{aligned}
\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_{t-\Delta} = x) &= \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u) P(X_{t-\Delta} = x | X_u = x_u) \\
&= \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u) P(X_{t-\Delta} = x \mid X_{t_n} = x_{t_n}) \\
&= \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u) P(X_{t-\Delta-t_n} = x \mid X_0 = x_{t_n}) \\
&= \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u) T_{t-\Delta-t_n}(x_{t_n}, x).
\end{aligned}$$

Since it follows from Lemma 20 that $\lim_{\Delta\to 0^+} T_{t-\Delta-t_n}(y,x) = T_{t-t_n}(y,x)$ for all y in $\mathcal{X}$, we find that

$$\lim_{\Delta\to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(x) = \frac{\sum_{x_u\in\Gamma(\hat{x}_u)} \lim_{\Delta\to 0^+} P(X_u = x_u) T_{t-\Delta-t_n}(x_{t_n}, x)}{\sum_{z_u\in\Gamma(\hat{x}_u)} P(X_u = z_u)}$$
$$= \frac{\sum_{x_u\in\Gamma(\hat{x}_u)} P(X_u = x_u) T_{t-t_n}(x_{t_n}, x)}{\sum_{z_u\in\Gamma(\hat{x}_u)} P(X_u = z_u)}.$$

Executing the same manipulations as before in reverse order yields

$$\lim_{\Delta\to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(x) = \pi_{(u,\hat{x}_u,t)}(x).$$

Next, we assume that u is the empty sequence $\emptyset$. Then some straightforward manipulations yield

$$\pi_{(u,\hat{x}_u,t-\Delta)}(x) = \frac{P(X_\emptyset = x_\emptyset, X_{t-\Delta} = x)}{P(X_\emptyset = x_\emptyset)} = P(X_{t-\Delta} = x)$$
$$= \sum_{y\in\mathcal{X}} P(X_0 = y) P(X_{t-\Delta} = x \mid X_0 = y) = \sum_{y\in\mathcal{X}} \pi_0(y) T_{t-\Delta}(y,x).$$

Again, it now follows from Lemma 20 that

$$\lim_{\Delta\to 0^+} \pi_{(\emptyset,\hat{x}_u,t-\Delta)}(x) = \sum_{y\in\mathcal{X}} \pi_0(y) \lim_{\Delta\to 0^+} T_{t-\Delta}(y,x) = \sum_{y\in\mathcal{X}} \pi_0(y) T_t(y,x),$$

such that

$$\lim_{\Delta\to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(x) = \pi_{(u,\hat{x}_u,t)}(x)$$

□

Lemma 22. *Let P be a positive and irreducible CTMC, $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map and $\hat{P}$ the corresponding lumped stochastic process. Fix any t, Δ in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$, $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$. Then*

$$\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) = \sum_{x\in\Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z\in\Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y\in\Gamma(\hat{y})} T_\Delta(x,y), \quad (31)$$

which is a convex combination of terms $\sum_{y\in\Gamma(\hat{y})} T_\Delta(x,y)$ with x in $\Gamma(\hat{x})$. If moreover $\Delta \leq t$ and $\max u < t - \Delta$, then

$$\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_{t-\Delta} = \hat{x}) = \sum_{x\in\Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t-\Delta)}(x)}{\sum_{z\in\Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t-\Delta)}(z)} \sum_{y\in\Gamma(\hat{y})} T_\Delta(x,y), \quad (32)$$

Proof. We only prove the first equality because the proof of the second equality is largely analoguous. To that end, we recall that by (27),

$$\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) = \frac{\sum_{x_u \in \Gamma(\hat{x}_u)} \sum_{x \in \Gamma(\hat{x})} \sum_{y \in \Gamma(\hat{y})} P(X_u = x_u, X_t = x, X_{t+\Delta} = y)}{\sum_{z_u \in \Gamma(\hat{x}_u)} \sum_{z \in \Gamma(\hat{x})} P(X_u = z_u, X_t = z)}.$$

After applying (1)–(5) and reordering the sums, we end up with

$$\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) = \frac{\sum_{x \in \Gamma(\hat{x})} \sum_{y \in \Gamma(\hat{y})} T_\Delta(x, y) \sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u, X_t = x)}{\sum_{z_u \in \Gamma(\hat{x}_u)} \sum_{z \in \Gamma(\hat{x})} P(X_u = z_u, X_t = z)}.$$

It is a matter of straightforward verification that if we divide both the numerator and the denominator in this final expression by $\sum_{x_u \in \Gamma(\hat{x}_u)} P(X_u = x_u)$, then the obtained expression is indeed equal to (31). Finally, verifying that (31) is a convex combination is trivial because $\pi_{(u,\hat{x}_u,t)}$ is a positive distribution on $\mathcal{X}$ by Lemma 19. □

Proof of Proposition 18. We start by proving (31), i.e., the limit from the right. To that end, we fix any Δ in $\mathbb{R}_{>0}$ and recall that by Lemma 22,

$$\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} T_\Delta(x, y),$$

where $\pi_{(u,\hat{x}_u,t)}$ is a positive distribution on $\mathcal{X}$ by Lemma 19. Subtracting $\mathbb{1}_{\hat{x}}(\hat{y})$ from both sides of the equality and dividing both sides of the equality by Δ yields

$$\frac{\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} = \frac{1}{\Delta}\left(\left(\sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} T_\Delta(x, y)\right) - \mathbb{1}_{\hat{x}}(\hat{y})\right).$$

Recall that the sum for x ranging over $\Gamma(\hat{x})$ is a convex combination of the terms $\sum_{y \in \Gamma(\hat{y})} T_\Delta(x, y)$, such that we can rewrite this equality as

$$\frac{\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} = \frac{1}{\Delta}\left(\sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \left(\left(\sum_{y \in \Gamma(\hat{y})} T_\Delta(x, y)\right) - \mathbb{1}_{\hat{x}}(\hat{y})\right)\right).$$

Furthermore, it clearly holds that $\mathbb{1}_{\hat{x}}(\hat{y}) = \sum_{y \in \Gamma(\hat{y})} \mathbb{1}_x(y)$, such that

$$\begin{aligned}\frac{\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} \\ = \frac{1}{\Delta}\left(\sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} (T_\Delta(x,y) - \mathbb{1}_x(y))\right).\end{aligned}$$

Since $T_0(x,y) = I(x,y) = \mathbb{1}_x(y)$, it follows from Lemma 20 that

$$\begin{aligned}\lim_{\Delta \to 0^+} \frac{\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} \\ = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} \lim_{\Delta \to 0^+} \frac{T_\Delta(x,y) - \mathbb{1}_x(y)}{\Delta} \\ = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} Q(x,y).\end{aligned}$$

Next, we prove (30), i.e., the limit from the left. To that end, we use Lemma 22 and execute similar manipulations as in the first part of the proof, to yield

$$\begin{aligned}\lim_{\Delta \to 0^+} \frac{\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_{t-\Delta} = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} \\ = \lim_{\Delta \to 0^+} \frac{1}{\Delta}\left(\sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t-\Delta)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t-\Delta)}(z)} \sum_{y \in \Gamma(\hat{y})} (T_\Delta(x,y) - \mathbb{1}_x(y))\right).\end{aligned}$$

It now follows from Lemmas 20 and 21 that

$$\begin{aligned}\lim_{\Delta \to 0^+} \frac{\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_{t-\Delta} = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} \\ = \sum_{x \in \Gamma(\hat{x})} \frac{\lim_{\Delta \to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(x)}{\sum_{z \in \Gamma(\hat{x})} \lim_{\Delta \to 0^+} \pi_{(u,\hat{x}_u,t-\Delta)}(z)} \sum_{y \in \Gamma(\hat{y})} \lim_{\Delta \to 0^+} \frac{T_\Delta(x,y) - \mathbb{1}_x(y)}{\Delta} \\ = \sum_{x \in \Gamma(\hat{x})} \frac{\pi_{(u,\hat{x}_u,t)}(x)}{\sum_{z \in \Gamma(\hat{x})} \pi_{(u,\hat{x}_u,t)}(z)} \sum_{y \in \Gamma(\hat{y})} Q(x,y).\end{aligned}$$

□

The following corollary essentially allows us to use the results from [11].

Corollary 23. *If P is a positive and irreducible CTMC and $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then the corresponding lumped process $\hat{P}$ is* well-behaved *[11, Definition 4.4], in the sense that, for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{<t}$, x, y in $\hat{\mathcal{X}}$ and $\hat{x}_u$ in $\hat{\mathcal{X}}_u$,*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta}\left|\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})\right| < +\infty$$

and, if $t \neq 0$*,*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left| \hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}, \hat{X}_{t-\Delta} = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y}) \right| < +\infty.$$

Proof. Follows immediately from Proposition 18. □

E The Induced Imprecise Continuous-Time Markov Chains

Everything is now set up to characterise the imprecise CTMC induced by lumping. Recall from Appendix C that such an imprecise CTMC is fully characterised by a non-empty bounded set of transition rate matrices and a non-empty set of initial distributions. Therefore, we first focus on the set of *lumped transition rate matrices.*

E.1 The Set of Lumped Transition Rate Matrices

Let Q be an irreducible transition rate matrix and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map. Then for any π in $\mathcal{D}(\mathcal{X})$, the matrix $\hat{Q}_\pi\colon \mathcal{L}(\hat{\mathcal{X}}) \to \mathcal{L}(\hat{\mathcal{X}})$ is defined by

$$\hat{Q}_\pi(\hat{x}, \hat{y}) \coloneqq \sum_{x \in \Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z \in \Gamma(\hat{x})} \pi(z)} \sum_{y \in \Gamma(\hat{y})} Q(x, y) \text{ for all } \hat{x}, \hat{y} \text{ in } \hat{\mathcal{X}}. \tag{33}$$

Lemma 24. *If Q is an irreducible transition rate matrix on $\mathcal{X}$, $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map and π an element of $\mathcal{D}(\mathcal{X})$, then $\hat{Q}_\pi$, defined by (33), is an irreducible transition rate matrix.*

Proof. We first verify that $\hat{Q}_\pi$ is indeed a transition rate matrix on $\hat{\mathcal{X}}$. To that end, we observe that $\hat{Q}_\pi$ is a real-valued $|\hat{\mathcal{X}}| \times |\hat{\mathcal{X}}|$ matrix. We also need to verify that $\hat{Q}_\pi$ has non-negative off-diagonal elements and rows that sum up to zero. Note that for any $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$ such that $\hat{x} \neq \hat{y}$, $\hat{Q}_\pi(\hat{x}, \hat{y})$ is a convex combination of non-negative real numbers, and hence the off-diagonal elements are non-negative. Also note that for any $\hat{x}$ in $\hat{\mathcal{X}}$,

$$\begin{aligned}
\sum_{\hat{y} \in \hat{\mathcal{X}}} \hat{Q}_\pi(\hat{x}, \hat{y}) &= \sum_{\hat{y} \in \hat{\mathcal{X}}} \sum_{x \in \Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z \in \Gamma(\hat{x})} \pi(z)} \sum_{y \in \Gamma(\hat{y})} Q(x, y) \\
&= \sum_{x \in \Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z \in \Gamma(\hat{x})} \pi(z)} \sum_{\hat{y} \in \hat{\mathcal{X}}} \sum_{y \in \Gamma(\hat{y})} Q(x, y) \\
&= \sum_{x \in \Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z \in \Gamma(\hat{x})} \pi(z)} \sum_{y \in \mathcal{X}} Q(x, y) = 0,
\end{aligned}$$

where the second and third equality follow from manipulations of finite sums and the last equality holds because Q is a transition rate matrix.

Next, we prove that $\hat{Q}_\pi$ is irreducible. To that end, we fix any two $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$ such that $\hat{x} \neq \hat{y}$. Fix now any x in $\Gamma(\hat{x})$ and any y in $\Gamma(\hat{y})$. Then as Q is irreducible, it follows from Proposition 7 that there is a sequence $x_0, \dots, x_n$ in $\mathcal{X}$ such that $x_0 = x$, $x_n = y$ and $Q(x_{i-1}, x_i) > 0$ for all $i = 1, \dots, n$. If for all i in $\{0, \dots, n\}$ we let $\hat{x}_i \coloneqq \Lambda(x_i)$, then $\hat{x}_0, \dots, \hat{x}_n$ is obviously a sequence in $\hat{\mathcal{X}}$ such that $\hat{x}_0 = \hat{x}$ and $\hat{x}_n = \hat{y}$. It may occur for several indices j in $\{0, \dots, n-1\}$ that there are consecutive entries $\hat{x}_j, \hat{x}_{j+1}, \dots$ that are all equal to $\hat{x}_j$. For each of those indices j we delete these consecutive entries $\hat{x}_{j+1}, \dots$ from the sequence; this way, we end up with the shorter sequence $\hat{x}_{i_0}, \dots, \hat{x}_{i_m}$ in $\hat{\mathcal{X}}$, where $\{i_0, \dots, i_m\}$ is an increasing subsequence of $\{1, \dots, n\}$. Note that by construction $\hat{x}_{i_0} = \hat{x}$, $\hat{x}_{i_m} = \hat{y}$ and $\hat{x}_{i_{(k-1)}} \neq \hat{x}_{i_k}$ for all k in $\{1, \dots, m\}$. Fix now any k in $\{1, \dots, m\}$. While it does not necessarily hold that $Q(x_{i_{(k-1)}}, x_{i_k}) > 0$, we have removed the consecutive entries in such a way that $Q(x_{i_k - 1}, x_{i_k}) > 0$. Because clearly $x_{i_k - 1} \in \Gamma(\hat{x}_{i_{(k-1)}})$ and $x_{i_k} \in \Gamma(\hat{x}_{i_k})$, it now follows that

$$\hat{Q}_\pi(\hat{x}_{i_{(k-1)}}, \hat{x}_{i_k}) = \sum_{x \in \Gamma(\hat{x}_{i_{(k-1)}})} \frac{\pi(x)}{\sum_{z \in \Gamma(\hat{x}_{i_{(k-1)}})} \pi(z)} \sum_{y \in \Gamma(\hat{x}_{i_k})} Q(x, y) > 0.$$

Since this is true for any $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$ such that $\hat{x} \neq \hat{y}$, it follows from Proposition 7 that $\hat{Q}_\pi$ is irreducible. □

Consider again an irreducible transition rate matrix Q and a lumping map $\Lambda \colon \hat{\mathcal{X}} \to \mathcal{X}$. The associated set of lumped transition rate matrices

$$\hat{\mathcal{Q}} \coloneqq \left\{ \hat{Q}_\pi \colon \pi \in \mathcal{D}(\mathcal{X}) \right\} \subseteq \mathcal{R}(\hat{\mathcal{X}}) \tag{34}$$

plays a vital role in obtaining our imprecise CTMC. In the remainder of this section, we are only concerned with some of its nice technical properties. Our proof for one of these properties requires the following lemma.

Lemma 25. *If π_1 and π_2 are two positive distributions on $\mathcal{X}$, α is a real number in the open unit interval $(0,1)$ and Λ is a lumping map, then π_α in $\mathcal{L}(\mathcal{X})$, defined for all x in $\mathcal{X}$ as*

$$\pi_\alpha(x) \coloneqq \frac{\alpha \left(\sum_{z \in \Gamma(\Lambda(x))} \pi_2(z) \right) \pi_1(x) + (1-\alpha) \left(\sum_{z \in \Gamma(\Lambda(x))} \pi_1(z) \right) \pi_2(x)}{\sum_{\hat{y} \in \hat{\mathcal{X}}} \left(\sum_{y \in \Gamma(\hat{y})} \pi_1(y) \right) \left(\sum_{y \in \Gamma(\hat{y})} \pi_2(y) \right)} \tag{35}$$

is a positive distribution on $\mathcal{X}$.

Proof. To reduce the notational burden in the remainder, we define

$$c \coloneqq \sum_{\hat{y} \in \hat{\mathcal{X}}} \left(\sum_{y \in \Gamma(\hat{y})} \pi_1(y) \right) \left(\sum_{y \in \Gamma(\hat{y})} \pi_2(y) \right).$$

Note that c is clearly positive due to the fact that both π_1 and π_2 are positive distributions, and that therefore $\pi_\alpha(x)$ is well-defined and—because a convex mixture of positive real numbers is a positive real number—positive for all x in $\mathcal{X}$. Furthermore, we observe that

$$\sum_{x\in\mathcal{X}} \pi_\alpha(x) = \sum_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\Gamma(\hat{x})} \pi_\alpha(x)$$
$$= \sum_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\Gamma(\hat{x})} \frac{\alpha\Big(\sum_{z\in\Gamma(\Lambda(x))} \pi_2(z)\Big)\pi_1(x) + (1-\alpha)\Big(\sum_{z\in\Gamma(\Lambda(x))} \pi_1(z)\Big)\pi_2(x)}{c}$$
$$= \sum_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\Gamma(\hat{x})} \frac{\alpha\Big(\sum_{z\in\Gamma(\hat{x})} \pi_2(z)\Big)\pi_1(x) + (1-\alpha)\Big(\sum_{z\in\Gamma(\hat{x})} \pi_1(z)\Big)\pi_2(x)}{c}.$$

Some straightforward rearranging yields

$$\begin{aligned}\sum_{x\in\mathcal{X}} \pi_\alpha(x) = \frac{\alpha}{c}&\left(\sum_{\hat{x}\in\hat{\mathcal{X}}}\left(\sum_{z\in\Gamma(\hat{x})}\pi_2(z)\right)\sum_{x\in\Gamma(\hat{x})}\pi_1(x)\right)\\ &+\frac{1-\alpha}{c}\left(\sum_{\hat{x}\in\hat{\mathcal{X}}}\left(\sum_{z\in\Gamma(\hat{x})}\pi_1(z)\right)\sum_{x\in\Gamma(\hat{x})}\pi_2(x)\right)\\ &=\frac{1}{c}\sum_{\hat{x}\in\hat{\mathcal{X}}}\left(\sum_{x\in\Gamma(\hat{x})}\pi_1(x)\right)\left(\sum_{z\in\Gamma(\hat{x})}\pi_2(z)\right)=\frac{c}{c}=1.\end{aligned}$$

We now have that π_α is a positive real-valued function on $\mathcal{X}$ with $\sum_{x\in\mathcal{X}} \pi_\alpha(x) = 1$, hence π_α is indeed a positive distribution on $\mathcal{X}$. □

Lemma 26. *Let Q be an irreducible transition rate matrix and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map. The associated set $\hat{\mathcal{Q}}$ of lumped transition rate matrices: (i) is non-empty and bounded, (ii) is convex and (iii) has separately specified rows. Furthermore, every $\hat{Q}$ in $\hat{\mathcal{Q}}$ is irreducible, and $\|\hat{Q}\| \leq \|Q\|$.*

Proof. We start with proving (i). Note that it is immediate from (34) that $\hat{\mathcal{Q}}$ is non-empty as $\mathcal{D}(\mathcal{X})$ is non-empty. The boundedness of $\hat{\mathcal{Q}}$ follows from the last sentence of the stated, which we will prove last.

We therefore move on to proving (ii). To that end, we fix two arbitrary elements of $\hat{\mathcal{Q}}$, denoted by $\hat{Q}_1$ and $\hat{Q}_2$. Note that because of the way $\hat{\mathcal{Q}}$ is constructed, there is a positive distribution π_1 (π_2) on $\mathcal{X}$ such that $\hat{Q}_{\pi_1} = \hat{Q}_1$ ($\hat{Q}_{\pi_2} = \hat{Q}_2$). Fix now an arbitrary α in the open unit interval $(0, 1)$, and let

$\hat{Q}_\alpha \coloneqq \alpha\hat{Q}_1 + (1-\alpha)\hat{Q}_2$. Then for all $\hat{x}$ and $\hat{y}$ in $\hat{\mathcal{X}}$,

$$\begin{aligned}
&\hat{Q}_\alpha(\hat{x},\hat{y}) = \alpha\hat{Q}_1(\hat{x},\hat{y}) + (1-\alpha)\hat{Q}_2(\hat{x},\hat{y}) \\
&= \sum_{x\in\Gamma(\hat{x})}\left(\alpha\left(\frac{\pi_1(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_1(z)}\right) + (1-\alpha)\left(\frac{\pi_2(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_2(z)}\right)\right)\sum_{y\in\Gamma(\hat{y})} Q(x,y) \\
&= \sum_{x\in\Gamma(\hat{x})}\frac{\alpha\left(\sum_{z\in\Gamma(\hat{x})}\pi_2(z)\right)\pi_1(x) + (1-\alpha)\left(\sum_{z\in\Gamma(\hat{x})}\pi_1(z)\right)\pi_2(x)}{\left(\sum_{z\in\Gamma(\hat{x})}\pi_1(z)\right)\left(\sum_{z\in\Gamma(\hat{x})}\pi_2(z)\right)}\sum_{y\in\Gamma(\hat{y})} Q(x,y).
\end{aligned}$$

Dividing both the numerator and the denominator of the fraction in the expression above by $\sum_{\hat{z}\in\hat{\mathcal{X}}}\left(\sum_{z\in\Gamma(\hat{z})}\pi_1(z)\right)\left(\sum_{z\in\Gamma(\hat{z})}\pi_2(z)\right)$ yields

$$\hat{Q}_\alpha(\hat{x},\hat{y}) = \sum_{x\in\Gamma(\hat{x})}\frac{\pi_\alpha(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_\alpha(z)}\sum_{y\in\Gamma(\hat{y})} Q(x,y),$$

where π_α is defined as in Lemma 25. Since we know from Lemma 25 that π_α is a positive distribution on $\mathcal{X}$, it follows from (34) that $\hat{Q}_\alpha$ is an element of $\hat{\mathcal{Q}}$. As $\hat{Q}_1$, $\hat{Q}_2$ and α were arbitrary, this proves that the set $\hat{\mathcal{Q}}$ is convex.

Next, we prove (iii). To that end, we fix an arbitrary $|\hat{\mathcal{X}}|$-tuple $(\hat{Q}_{\hat{x}} \colon \hat{x} \in \hat{\mathcal{X}})$ of which the entries—one for every state—are all elements of $\hat{\mathcal{Q}}$. We know from (34) that, for any $\hat{x}$ in $\hat{\mathcal{X}}$, there is a positive distribution $\pi_{\hat{x}}$ on $\mathcal{X}$ such that $\hat{Q}_{\pi_{\hat{x}}} = \hat{Q}_{\hat{x}}$. Following Definition 4, we now construct a matrix $\hat{Q}^\star$, defined by

$$\hat{Q}^\star(\hat{x},\hat{y}) \coloneqq \hat{Q}_{\hat{x}}(\hat{x},\hat{y}) \text{ for all } \hat{x},\hat{y} \in \mathcal{X}.$$

We need to prove that $\hat{Q}^\star$ is an element of $\hat{\mathcal{Q}}$. To verify this, we observe that for all $\hat{x}$ and $\hat{y}$ in $\hat{\mathcal{X}}$,

$$\hat{Q}^\star(\hat{x},\hat{y}) = \sum_{x\in\Gamma(\hat{x})}\frac{\pi_{\hat{x}}(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_{\hat{x}}(z)}\sum_{y\in\Gamma(\hat{y})} Q(x,y).$$

We now divide both the numerator and the denominator of the fraction in the expression above by $\sum_{z\in\mathcal{X}}\pi_{\Lambda(x)}(x)$, to yield

$$\hat{Q}^\star(\hat{x},\hat{y}) = \sum_{x\in\Gamma(\hat{x})}\frac{\pi^\star(x)}{\sum_{z\in\Gamma(\hat{x})}\pi^\star(z)}\sum_{y\in\Gamma(\hat{y})} Q(x,y),$$

where $\pi^\star$ is the positive distribution—one can easily verify that this is indeed the case—on $\mathcal{X}$ defined by

$$\pi^\star(x) \coloneqq \frac{\pi_{\Lambda(x)}(x)}{\sum_{z\in\mathcal{X}}\pi_{\Lambda(z)}(z)} \quad \text{for all } x \in \mathcal{X}.$$

Because this final equality holds for all $\hat{x}$ and $\hat{y}$ in $\hat{\mathcal{X}}$, we find conclude that $\hat{Q}^\star$ is indeed an element of $\hat{\mathcal{Q}}$.

Next, we fix an arbitrary $\hat{Q}$ in $\hat{\mathcal{Q}}$. Let π be the positive distribution on $\mathcal{X}$ such that $\hat{Q}_\pi = \hat{Q}$. Then $\hat{Q}$ is irreducible by Lemma 24. Furthermore,

$$\begin{aligned}
\|\hat{Q}\| &= 2\max\Big\{|\hat{Q}(\hat{x},\hat{x})| \colon \hat{x} \in \hat{\mathcal{X}}\Big\} \\
&= 2\max\left\{\left|\sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} \sum_{y\in\Gamma(\hat{x})} Q(x,y)\right| \colon \hat{x} \in \hat{\mathcal{X}}\right\} \\
&\leq 2\max\left\{\sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} \left|\sum_{y\in\Gamma(\hat{x})} Q(x,y)\right| \colon \hat{x} \in \hat{\mathcal{X}}\right\} \\
&\leq 2\max\left\{\sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} |Q(x,x)| \colon \hat{x} \in \hat{\mathcal{X}}\right\} \\
&\leq 2\max\Big\{\max\{|Q(x,x)| \colon x \in \Gamma(\hat{x})\} \colon \hat{x} \in \hat{\mathcal{X}}\Big\} \\
&= 2\max\{|Q(x,x)| \colon x \in \mathcal{X}\} = \|Q\|,
\end{aligned}$$

where the first and last equality follow from (9), the second equality follows from (33), the first inequality follows from the triangle inequality, the second inequality follows from the properties of transition rate matrices—i.e., non-negative off-diagonal elements and rows that sum to zero— and where for the third inequality we use the fact that a convex combination of real numbers is always lower than the maximum of these real numbers. □

E.2 The Lower Transition (Rate) Operator Corresponding to the Set of Lumped Transition Rate Matrices

Since $\hat{\mathcal{Q}}$ is non-empty and bounded by Lemma 26, we know from Appendix C.2 that it has an associated lower transition rate operator $\underline{\hat{Q}} \colon \mathcal{L}(\hat{\mathcal{X}}) \to \mathcal{L}(\hat{\mathcal{X}})$, defined by (13):

$$[\underline{\hat{Q}}\hat{f}](\hat{x}) = \inf\{[\hat{Q}\hat{f}](\hat{x}) \colon \hat{Q} \in \hat{\mathcal{Q}}\} \quad \text{for all } \hat{x} \text{ in } \hat{\mathcal{X}} \text{ and all } \hat{f} \text{ in } \mathcal{L}(\hat{\mathcal{X}}). \tag{36}$$

Note that (7), the definition for $\underline{\hat{Q}}$ in the main text, differs from (36), its proper definition. These two definitions turn out to be equal in this case, as is stated in the following result.

Proposition 27. *If Q is an irreducible transition rate matrix and $\Lambda \colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then for all $\hat{x}$ in $\hat{\mathcal{X}}$ and $\hat{f}$ in $\mathcal{L}(\hat{\mathcal{X}})$,*

$$[\underline{\hat{Q}}\hat{f}](\hat{x}) = \min\left\{\sum_{\hat{y}\in\hat{\mathcal{X}}} \hat{f}(\hat{y}) \sum_{y\in\Gamma(\hat{y})} Q(x,y) \colon x \in \Gamma(\hat{x})\right\}.$$

Proof. Let

$$f_x \coloneqq \sum_{\hat{y}\in\hat{\mathcal{X}}} \hat{f}(\hat{y}) \sum_{y\in\Gamma(\hat{y})} Q(x,y) \quad \text{for all } x \in \Gamma(\hat{x}).$$

Then we need to prove that

$$[\underline{\hat{Q}}\hat{f}](\hat{x}) = \min\{f_x : x \in \Gamma(\hat{x})\}.$$

By combining (34) and (36), we find that

$$[\underline{\hat{Q}}\hat{f}](\hat{x}) = \inf\{[\hat{Q}\hat{f}](\hat{x}) : \hat{Q} \in \hat{\mathcal{Q}}\} = \inf\{[\hat{Q}_\pi\hat{f}](\hat{x}) : \pi \in \mathcal{D}(\mathcal{X})\}.$$

Explicitly writing out the matrix-vector product $[\hat{Q}_\pi\hat{f}](\hat{x})$ yields

$$\begin{aligned}
[\hat{Q}_\pi\hat{f}](\hat{x}) &= \sum_{\hat{y}\in\hat{\mathcal{X}}} \hat{f}(\hat{y})\hat{Q}_\pi(\hat{x},\hat{y}) = \sum_{\hat{y}\in\hat{\mathcal{X}}} \hat{f}(\hat{y}) \sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} \sum_{y\in\Gamma(\hat{y})} Q(x,y) \\
&= \sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} \sum_{\hat{y}\in\hat{\mathcal{X}}} \hat{f}(\hat{y}) \sum_{y\in\Gamma(\hat{y})} Q(x,y) \\
&= \sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} f_x.
\end{aligned}$$

Hence, we need to prove that

$$\inf\left\{\sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} f_x : \pi \in \mathcal{D}(\mathcal{X})\right\} = \min\{f_x : x \in \Gamma(\hat{x})\}.$$

Note that the right hand side is clearly a lower bound for

$$\inf\left\{\sum_{x\in\Gamma(\hat{x})} \frac{\pi(x)}{\sum_{z\in\Gamma(\hat{x})}\pi(z)} f_x : \pi \in \mathcal{D}(\mathcal{X})\right\}.$$

We now show that it is the tightest lower bound—i.e., the infimum—of this set. To that end, we construct a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ in $\mathcal{D}$ such that the induced sequence

$$\left\{\sum_{x\in\Gamma(\hat{x})} \frac{\pi_n(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_n(z)} f_x\right\}_{n\in\mathbb{N}}$$

converges to $\min\{f_x : x \in \Gamma(\hat{x})\}$. Let $x^\star$ be an element of $\Gamma(\hat{x})$ such that $f_{x^\star} = \min\{f_x : x \in \Gamma(\hat{x})\}$, $c \coloneqq \frac{1}{|\mathcal{X}|}$ and $m \coloneqq |\Gamma(\hat{x})|$. For all n in $\mathbb{N}$, we define the positive distribution π_n on $\mathcal{X}$ by

$$\pi_n(x) \coloneqq \begin{cases} c & \text{if } x \notin \Gamma(\hat{x}) \\ cm - \dfrac{c}{n}(m-1) & \text{if } x = x^\star \\ \dfrac{c}{n} & \text{otherwise} \end{cases} \qquad \text{for all } x \text{ in } \mathcal{X}.$$

Then clearly,

$$\begin{aligned}\lim_{n\to+\infty}\sum_{x\in\Gamma(\hat{x})}\frac{\pi_n(x)}{\sum_{z\in\Gamma(\hat{x})}\pi_n(z)}f_x\\ &= \lim_{n\to+\infty}\left(\left(1-\frac{m-1}{mn}\right)f_{x^\star}+\sum_{x\in\Gamma(\hat{x})\colon x\neq x^\star}\frac{1}{mn}f_x\right)=f_{x^\star}\\ &= \min\{f_x\colon x\in\Gamma(\hat{x})\}.\end{aligned}$$

□

The following result states that $\underline{\hat{Q}}$ is irreducible, which is to be expected as it is the lower envelope of a set of irreducible transition rate matrices.

Corollary 28. *If Q is an irreducible transition rate matrix and $\Lambda \coloneqq \mathcal{X}\to\hat{\mathcal{X}}$ a lumping map, then $\underline{\hat{Q}}$ is* irreducible.

Proof. Recall from Lemma 26 that any $\hat{Q}$ in $\hat{\mathcal{Q}}$ is irreducible. Fix now any arbitrary $\hat{Q}^\star$ in $\hat{\mathcal{Q}}$. Then for any distinct $\hat{x}$ and $\hat{y}$ in $\hat{\mathcal{X}}$, there is a sequence $\hat{y}=\hat{x}_1,\dots,\hat{x}_n=\hat{x}$ in $\hat{\mathcal{X}}$ such that $\hat{Q}^\star(\hat{x}_{i-1},\hat{x}_i)>0$ for all i in $\{1,\dots,n\}$. By (36), it then clearly holds for any i in $\{1,\dots,n\}$ that

$$\begin{aligned}-[\underline{\hat{Q}}(-\mathbb{1}_{\hat{x}_i})](\hat{x}_{i-1}) &= -\inf\left\{-\hat{Q}(\hat{x}_{i-1},\hat{x}_i)\colon \hat{Q}\in\hat{\mathcal{Q}}\right\}\\ &= \sup\left\{\hat{Q}(\hat{x}_{i-1},\hat{x}_i)\colon \hat{Q}\in\hat{\mathcal{Q}}\right\}\geq \hat{Q}^\star(\hat{x}_{i-1},\hat{x}_i)>0.\end{aligned}$$

Consequently, $\hat{y}\rightarrowtail\hat{x}$ for any arbitrary $\hat{x}$ and $\hat{y}$, which proves the stated. □

E.3 Laying Down the Last Pieces of the Puzzle

For any positive and irreducible CTMC P with initial distribution π_0 and any lumping map $\Lambda\colon\mathcal{X}\to\hat{\mathcal{X}}$, we define the *lumped initial distribution*

$$\hat{\pi}_0\colon\hat{\mathcal{X}}\to\mathbb{R}\colon\hat{x}\mapsto\hat{\pi}_0(\hat{x})\coloneqq\sum_{x\in\Gamma(\hat{x})}\pi_0(x). \tag{37}$$

It can be immediately verified that $\hat{\pi}_0$ is a positive distribution on $\hat{\mathcal{X}}$. Furthermore, using (21), Lemma 17, (3) and (37) yields that

$$\begin{aligned}\hat{P}(\hat{X}_0=\hat{x}) &= \hat{P}(\hat{X}_0=\hat{x}\mid\hat{X}_\emptyset=\hat{x}_\emptyset)\\ &= \frac{\sum_{x\in\Gamma(\hat{x})}P(X_0=x\mid X_\emptyset=x_\emptyset)P(X_\emptyset=x_\emptyset)}{P(X_\emptyset=x_\emptyset)}\\ &= \sum_{x\in\Gamma(\hat{x})}P(X_0=x)=\sum_{x\in\Gamma(\hat{x})}\pi_0(x)=\hat{\pi}_0(\hat{x}).\end{aligned}$$

Hence, if we let $\hat{\mathcal{M}}\coloneqq\{\hat{\pi}_0\}$, we see that the lumped stochastic process $\hat{P}$ is *consistent* with $\hat{\mathcal{M}}$, see Appendix C.1. The following intermediary result, which follows immediately from Proposition 18 and (34), states that it is also consistent with $\hat{\mathcal{Q}}$; again, see Appendix C.1.

Corollary 29. *Let P be a positive and irreducible CTMC and let $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ be a lumping map. Fix any t in $\mathbb{R}_{\geq 0}$, any u in $\mathcal{U}_{<t}$ and any $\hat{x}_u$ in $\hat{\mathcal{X}}_u$. Then there is a unique element $\hat{Q}_{(u,\hat{x}_u,t)}$ of $\hat{\mathcal{Q}}$ such that, for all $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$,*

$$\lim_{\Delta \to 0^+} \frac{\hat{P}(\hat{X}_{t+\Delta} = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} = \hat{Q}_{(u,\hat{x}_u,t)}(\hat{x}, \hat{y}) \tag{38}$$

and, if $t \neq 0$,

$$\lim_{\Delta \to 0^+} \frac{\hat{P}(\hat{X}_t = \hat{y} \mid \hat{X}_u = \hat{x}_u, \hat{X}_{t-\Delta} = \hat{x}) - \mathbb{1}_{\hat{x}}(\hat{y})}{\Delta} = \hat{Q}_{(u,\hat{x}_u,t)}(\hat{x}, \hat{y}). \tag{39}$$

Proof. Follows immediately from Proposition 18, (33) and (34). □

We now combine several of our intermediary results concerning the lumped stochastic process $\hat{P}$ to finally end up with the result we need to prove the results in Sect. 4.

Corollary 30. *If P is a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then the associated lumped stochastic process $\hat{P}$ is contained in $\mathbb{P}_{\pi_0,Q,\Lambda} \coloneqq \mathbb{P}^{\mathrm{W}}_{\hat{\mathcal{Q}},\hat{\mathcal{M}}}$.*

Proof. Recall that $\hat{P}$ is well-behaved by Corollary 23. Furthermore, as we have just seen in this section, $\hat{P}$ is consistent with $\hat{\mathcal{Q}}$ and $\hat{\mathcal{M}}$. The stated now follows because, by definition, $\mathbb{P}_{\pi_0,Q,\Lambda}$ contains all well-behaved stochastic processes that are consistent with $\hat{\mathcal{Q}}$ and $\hat{\mathcal{M}}$. □

F Proofs of the Results in Sect. 4

In the main text, we limited ourselves to determining bounds on marginal and limit expectations of functions f in $\mathcal{L}(\mathcal{X})$ that are lumpable with respect to Λ, mainly due to length constraints. Since this length constraint is not present in this extended pre-print, we here drop this restriction.

Let $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ be a lumping map. Then the reduction to $\hat{\mathcal{X}}$ of a non-lumpable f in $\mathcal{L}(\mathcal{X})$ is not unequivocally defined. Two restrictions that will turn out to be useful in our setting are $\hat{f}_L$ and $\hat{f}_U$ in $\mathcal{L}(\hat{\mathcal{X}})$, defined for all $\hat{x}$ in $\hat{\mathcal{X}}$ as

$$\hat{f}_L(\hat{x}) \coloneqq \min\{f(x)\colon x \in \Gamma(\hat{x})\} \quad \text{and} \quad \hat{f}_U(\hat{x}) \coloneqq \max\{f(x)\colon x \in \Gamma(\hat{x})\}.$$

Note that if f is lumpable with respect to Λ, then $\hat{f}_L = \hat{f} = \hat{f}_U$. Moreover, we have the following two properties.

Lemma 31. *If $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ is a lumping map, then for all x in $\mathcal{X}$ and f in $\mathcal{L}(\mathcal{X})$,*

$$\hat{f}_L(\Lambda(x)) \leq f(x) \leq \hat{f}_U(\Lambda(x)).$$

Proof. Follows immediately from the definition of $\hat{f}_L$ and $\hat{f}_U$. □

Lemma 32. *Let P be a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map. Then for all f in $\mathcal{L}(\mathcal{X})$ and t in $\mathbb{R}_{\geq 0}$,*

$$\hat{E}(\hat{f}_L(\hat{X}_t)) \leq E(f(X_t)) \leq \hat{E}(\hat{f}_U(\hat{X}_t)).$$

Proof. We start by proving the lower bound. Note that by Lemma 31 and the monotonicity of E,

$$E(\hat{f}_L(\Lambda(X_t))) \leq E(f(X_t)).$$

Some straightforward manipulations yield

$$\begin{aligned} E(\hat{f}_L(\Lambda(X_t))) &= \sum_{x \in \mathcal{X}} \hat{f}_L(\Lambda(x)) P(X_t = x) = \sum_{\hat{x} \in \hat{\mathcal{X}}} \hat{f}_L(\hat{x}) \sum_{x \in \Gamma(\hat{x})} P(X_t = x) \\ &= \sum_{\hat{x} \in \hat{\mathcal{X}}} \hat{f}_L(\hat{x}) P\left(\bigcup_{x \in \Gamma(\hat{x})} (X_t = x) \right) = \sum_{\hat{x} \in \hat{\mathcal{X}}} \hat{f}_L(\hat{x}) \hat{P}(\hat{X}_t = \hat{x}) \\ &= \hat{E}(\hat{f}_L(\hat{X}_t)), \end{aligned}$$

where the fourth equality follows from (19) and (21). Combining this equality with the previously obtained inequality immediately yields the lower bound of the stated.

To prove the upper bound, we apply the lower bound on the function $g \coloneqq -f$. Note that $\hat{g}_L = -\hat{f}_U$, and that

$$\hat{E}(-\hat{f}_U(\hat{X}_t)) = \hat{E}(\hat{g}_L(\hat{X}_t)) \leq E(g(X_t)) = E(-f(X_t))$$

clearly implies that $E(f(X_t)) \leq \hat{E}(\hat{f}_U(\hat{X}_t))$. □

The following result is slightly more general than Theorem 1. Recall from Sect. 3.3 that we use $\hat{\underline{T}}_t$ to denote the lower transition operator over t associated with $\underline{\hat{Q}}$ according to (8).

Proposition 33. *If P is a positive and irreducible CTMC and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then for all f in $\mathcal{L}(\mathcal{X})$ and t in $\mathbb{R}_{\geq 0}$,*

$$\hat{\pi}_0 \hat{\underline{T}}_t \hat{f}_L \leq E(f(X_t)) = \pi_0 T_t f \leq -\hat{\pi}_0 \hat{\underline{T}}_t (-\hat{f}_U).$$

Proof. We start by proving the lower bound. By Lemma (32),

$$\hat{E}(\hat{f}_L(\hat{X}_t)) \leq E(f(X_t)).$$

It follows from Corollary 30 that

$$\underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}}, \hat{\mathcal{M}}}(\hat{f}_L(\hat{X}_t)) \leq \hat{E}(\hat{f}_L(\hat{X}_t)).$$

Moreover, from Proposition 12 and Proposition 11 (in that order)—which we may both use due to Lemma 26—it follows that

$$\underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}}, \hat{\mathcal{M}}}(\hat{f}_L(\hat{X}_t)) = \underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}}, \hat{\mathcal{M}}}(\underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}}, \hat{\mathcal{M}}}(\hat{f}_L(\hat{X}_t) \mid \hat{X}_0)) = \underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}}, \hat{\mathcal{M}}}([\hat{\underline{T}}_t \hat{f}_L](\hat{X}_0)).$$

Since $\hat{\mathcal{M}}$ is a singleton, it follows from [11, Proposition 9.3] that this can be rewritten as

$$\underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}},\hat{\mathcal{M}}}(\hat{f}_L(\hat{X}_t)) = \sum_{\hat{x}\in\hat{\mathcal{X}}} \hat{\pi}_0(\hat{x})[\hat{\underline{T}}_t \hat{f}_L](\hat{x}) = \hat{\pi}_0 \hat{\underline{T}}_t \hat{f}_L.$$

Finally, combining all that we have found so far yields the lower bound of the stated:

$$\hat{\pi}_0 \hat{\underline{T}}_t \hat{f}_L = \underline{E}^{\mathrm{M}}_{\hat{\mathcal{Q}},\hat{\mathcal{M}}}(\hat{f}_L(\hat{X}_t)) \leq \hat{E}(\hat{f}_L(\hat{X}_t)) = E(\hat{f}_L(\Lambda(X_t))) \leq E(f(X_t)).$$

To prove the upper bound, we simply apply the lower bound to $g \coloneqq -f$. This yields

$$\hat{\pi}_0 \hat{\underline{T}}_t(-\hat{f}_U) = \hat{\pi}_0 \hat{\underline{T}}_t \hat{g}_L \leq E(g(X_t)) = E(-f(X_t)),$$

which clearly implies the upper bound of the stated. □

Proof of Theorem 1. Since f is lumpable with respect to Λ, we know that $\hat{f}_L = \hat{f} = \hat{f}_U$. Therefore, the stated follows immediately from Proposition 33. □

Proposition 34. *If P is a CTMC with irreducible transition rate matrix Q and $\Lambda\colon \mathcal{X} \to \hat{\mathcal{X}}$ a lumping map, then for all f in $\mathcal{L}(\mathcal{X})$, δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$ and n in $\mathbb{N}_0$,*

$$\min(I + \delta\underline{\hat{Q}})^n \hat{f}_L \leq E_\infty(f) \leq -\min(I + \delta\underline{\hat{Q}})^n(-\hat{f}_U).$$

Furthermore, for fixed δ, the lower and upper bounds in this expression become monotonously tighter with increasing n, and each converges to a (possibly different) constant as n approaches $+\infty$.

Proof. As Q is irreducible, we know from Sect. 2.2 that there is a unique positive distribution, denoted by π_∞, such that $\pi_\infty Q = 0$. Hence, for all y in $\mathcal{X}$,

$$\sum_{x\in\mathcal{X}} \pi_\infty(x) Q(x,y) = 0.$$

Consider now $\hat{Q}_\infty \coloneqq \hat{Q}_{\pi_\infty}$. Then by Lemma 26, $\hat{Q}_\infty$ is irreducible. Let $\hat{\pi}'_\infty$ be the unique positive distribution on $\hat{\mathcal{X}}$ that satisfies the equilibrium condition $\hat{\pi}'_\infty \hat{Q}_\infty = 0$. We now claim that $\hat{\pi}'_\infty = \hat{\pi}_\infty$, where $\hat{\pi}_\infty$ is the positive distribution on $\hat{\mathcal{X}}$ defined by

$$\hat{\pi}_\infty(\hat{x}) = \sum_{x\in\Gamma(\hat{x})} \pi_\infty(x) \quad \text{for all } \hat{x} \text{ in } \hat{\mathcal{X}}.$$

To verify this claim, we fix any $\hat{y}$ in $\hat{\mathcal{X}}$ and see that

$$\begin{aligned}\sum_{\hat{x}\in\hat{\mathcal{X}}} \hat{\pi}_\infty(\hat{x})\hat{Q}_\infty(\hat{x},\hat{y}) &= \sum_{\hat{x}\in\hat{\mathcal{X}}} \hat{\pi}_\infty(\hat{x}) \sum_{x\in\Gamma(\hat{x})} \frac{\pi_\infty(x)}{\sum_{z\in\Gamma(\hat{z})}\pi_\infty(z)} \sum_{y\in\Gamma(\hat{y})} Q(x,y) \\ &= \sum_{\hat{x}\in\hat{\mathcal{X}}} \left(\sum_{x\in\Gamma(\hat{x})} \pi_\infty(x)\right) \sum_{x\in\Gamma(\hat{x})} \frac{\pi_\infty(x)}{\sum_{z\in\Gamma(\hat{z})}\pi_\infty(z)} \sum_{y\in\Gamma(\hat{y})} Q(x,y) \\ &= \sum_{\hat{x}\in\hat{\mathcal{X}}} \sum_{x\in\Gamma(\hat{x})} \pi_\infty(x) \sum_{y\in\Gamma(\hat{y})} Q(x,y) \\ &= \sum_{y\in\Gamma(\hat{y})} \sum_{x\in\mathcal{X}} \pi_\infty(x)Q(x,y) = 0.\end{aligned}$$

As $\hat{y}$ was arbitrary, we find that $\hat{\pi}_\infty$ satisfies the equilibrium condition $\hat{\pi}_\infty\hat{Q}_\infty = 0$. Since $\hat{\pi}'_\infty$ is the unique positive distribution that satisfies this equilibrium condition, we conclude that $\hat{\pi}_\infty = \hat{\pi}'_\infty$.

Fix now any f in $\mathcal{L}(\mathcal{X})$, δ in $\mathbb{R}_{>0}$ such that $\delta\|Q\| < 2$ and n in $\mathbb{N}_0$. Note that if $n = 0$, the stated trivially holds. Hence, we now consider the case $n > 0$, starting with the lower bound. Recall from Lemma 26 that $\|\hat{Q}_\infty\| \leq \|Q\|$, such that $\delta\|\hat{Q}_\infty\| < 2$. Hence, from Lemma 31 it follows that

$$\hat{\pi}_\infty\hat{f}_L \leq E_\infty(f) = \pi_\infty f.$$

As $\hat{Q}_\infty$ is irreducible with stationary distribution $\hat{\pi}_\infty$, it follows from Lemma 10 that

$$\min(I + \delta\hat{Q}_\infty)^n\hat{f}_L \leq \hat{\pi}_\infty\hat{f}_L.$$

Since $\hat{Q}_\infty$ is an element of $\hat{\mathcal{Q}}$ by construction, it follows from (36) that $\underline{\hat{Q}}\hat{g} \leq \hat{Q}_\infty\hat{g}$ for any $\hat{g}$ in $\mathcal{L}(\hat{\mathcal{X}})$, which implies that

$$(I + \delta\underline{\hat{Q}})\hat{g} \leq (I + \delta\hat{Q}_\infty)\hat{g}. \tag{40}$$

Clearly, (40) implies that $(I + \delta\underline{\hat{Q}})f_L \leq (I + \delta\hat{Q}_\infty)f_L$. In case $n = 1$, this is sufficient to prove the lower bound. In case $n > 1$, we need some more properties. Since $\delta\|\hat{Q}_\infty\| < 2$, it follows from Lemma 9 that $(I + \delta\hat{Q}_\infty)$ is a transition matrix. Consequently, repeated application of (40) and the monotonicity of this transition matrix yields

$$\begin{aligned}(I + \delta\hat{Q}_\infty)^n f_L = (I + \delta\hat{Q}_\infty)^{n-1}(I + \delta\hat{Q}_\infty)f_L &\geq (I + \delta\hat{Q}_\infty)^{n-1}(I + \delta\underline{\hat{Q}})\hat{f}_L \\ &\geq \cdots \geq (I + \delta\hat{Q}_\infty)(I + \delta\underline{\hat{Q}})^{n-1}\hat{f}_L \geq (I + \delta\underline{\hat{Q}})^n\hat{f}_L.\end{aligned}$$

Combining all intermediate results, we find that

$$\min(I + \delta\underline{\hat{Q}})^n\hat{f}_L \leq \min(I + \delta\hat{Q}_\infty)^n\hat{f}_L \leq \hat{\pi}_\infty\hat{f}_L \leq \pi_\infty f = E_\infty(f),$$

which proves the lower bound of the stated.

The upper bound follows from applying the lower bound to $g \coloneqq -f$. As $\hat{g}_L = -\hat{f}_U$, we find that

$$\min(I + \delta \underline{\hat{Q}})^n(-\hat{f}_U) = \min(I + \delta \underline{\hat{Q}})^n(\hat{g}_L) \leq E_\infty(g) = E_\infty(-f).$$

The upper bound now follows immediately from this inequality:

$$E_\infty(f) = -E_\infty(-f) \leq -\min(I + \delta \underline{\hat{Q}})^n(-\hat{f}_U).$$

We end this proof by verifying the statement concerning the monotonous convergence of the lower bound. Observe that by Lemma's 13 and 26,

$$\|\underline{\hat{Q}}\| = \sup\left\{ \|\hat{Q}\| \colon \hat{Q} \in \hat{\mathcal{Q}} \right\} \leq \|Q\|.$$

Hence, since $\delta\|Q\| < 2$, we find that $\delta\|\underline{\hat{Q}}\| < 2$. The monotonous convergence now follows immediately from Corollaries 14 and 28. □

Proof of Theorem 2. We know from (9) that $\|Q\| = 2\max\{|Q(x,x)| \colon x \in \mathcal{X}\}$. Furthermore, since f is lumpable with respect to $\mathit{\Lambda}$, we know that $\hat{f}_L = \hat{f} = \hat{f}_U$. Therefore, the stated follows immediately from Proposition 34. □