Expected Time Averages in Markovian Imprecise Jump Processes: A Graph-Theoretic Characterisation of Weak Ergodicity

Yema Paul YEMA.PAUL@ISAE-SUPAERO.FR
Space Advanced Concepts Laboratory, ISAE-SUPAERO, Toulouse, France

Alexander Erreygers ALEXANDER.ERREYGERS@UGENT.BE
Jasper De Bock JASPER.DEBOCK@UGENT.BE
Foundations Lab for imprecise probabilities, Ghent University, Belgium

Abstract

Markovian imprecise jump processes provide a way to express model uncertainty about Markovian jump processes. The dynamics are not governed by a unique rate matrix, but are instead partially specified by a set of such matrices. Since the dynamics are partially specified, the resulting expected time averages are no longer uniquely determined either, and one then resorts to tight lower and upper bounds on them. In this paper, we are interested in the existence of an asymptotic limit of these upper and lower bounds, as the time horizon becomes infinite. When those limits exist and are furthermore independent of the choice of the process's initial state, we say that the process is weakly ergodic. Our main contribution is a necessary and sufficient condition for a Markovian imprecise jump process to be weakly ergodic, expressed in terms of simple graph-theoretic conditions on its set of rate matrices.

Keywords: Markovian jump process, imprecise probabilities, expected time averages, weak ergodicity, upper rate operator.

1. Introduction

The Point-Wise (or Mean) Ergodic Theorem – see, for example, [16, Theorem 3.8.1] or [17, Corollary 4.4.10] – is one of the most important results for (time-homogeneous) Markovian jump processes.[1] This theorem establishes that, under suitable conditions on the rate matrix Q of the Markovian jump process, for any real-valued function f on the state space, the limit for large T of the (time) average of $f(X_t)$ over $[0,T]$ converges to $E_{\pi_\infty}(f)$ almost surely, where the 'limit/stationary distribution' π_∞ can be easily obtained as the unique solution to a system of linear equations – see [16, Sections 3.5 and 3.6] or [17, Theorem 4.4.8]. It is not difficult to see that the same conditions on the rate matrix Q imply what we call *weak ergodicity*, meaning that the limit of the expected time average exists and is independent of the initial state, or more formally, that for all initial states x and real-valued functions f of the state,

$$\lim_{t\to+\infty} E\left(\frac{1}{t}\int_0^t f(X_\tau)\,\mathrm{d}\tau \,\middle|\, X_0 = x\right) = E_{\pi_\infty}(f).$$

In this contribution, we investigate this weak ergodicity directly, which suffices for many applications. The novelty of this contribution lies in the fact that we do not consider a single Markovian jump process but a so-called Markovian imprecise jump process [19, 14, 9]. We will see in Section 2 that this is a set of jump processes characterised by a set of rate matrices, and that we are interested in weak ergodicity for the corresponding tight upper (and lower) bounds on the expected time averages. In Section 3, we explain how this is tied to a notion of weak ergodicity for the upper rate operator induced by the set of rate matrices. Section 4 introduces our main result: a convenient graph-theoretic condition on an upper rate operator that is necessary and sufficient for it to be weakly ergodic. We prove this main result in Sections 5 and 6, and conclude this contribution in Section 7.

2. Imprecise Jump Processes

A Markovian imprecise jump process – called an imprecise continuous-time Markov chain in [14] – generalises the notion of a (homogenous) Markovian jump process [16, 17] to the framework of imprecise probabilities, in particular to the setting of sets of (conditional) probability charges/measures. In this contribution we work with the theory as introduced by Krak et al. [14] and extended by Erreygers and De Bock [9]; see also [7, 13]. Unfortunately the page restriction does not permit us to explain all details of this theory, so we will only give an informal and intuitive explanation.

[1] We follow the – perhaps a bit non-standard – terminology used in [7, 9]; Norris [16, Chapters 2 and 3] speaks of 'continuous-time Markov chains' whereas Stroock [17, Chapter 4] and Iosifescu [12] stick to 'Markov processes (in continuous time)'.

2.1. Jump Processes

Suppose a subject – which we will refer to as 'You' – is interested in some system. The state X_t of this system takes values in some finite state space $\mathcal{X}$ and changes over continuous time $t \in \mathbb{R}_{\geq 0}$, and You are uncertain about the (future) values of the state. One way to model this uncertainty is through a *jump process* P [14, Section 4.2]; informally, this means that we fix the *initial probabilities*, so $P(X_0 = x)$ for all $x \in \mathcal{X}$, and the *transition probabilities*, so

$$P(X_{t_n+\Delta} = y \mid X_{t_1} = x_1, \ldots, X_{t_n} = x_n)$$

for all $\Delta \in \mathbb{R}_{>0}$, $t_1, \ldots, t_n \in \mathbb{R}_{\geq 0}$ such that $t_1 < \cdots < t_n$ and $x_1, \ldots, x_n, y \in \mathcal{X}$.

Such a jump process P is *Markovian* – or satisfies the Markov property – if for all $t \in \mathbb{R}_{\geq 0}$, $\Delta \in \mathbb{R}_{>0}$, $t_1, \ldots, t_n \in \mathbb{R}_{\geq 0}$ such that $t_1 < \cdots < t_n < t$ and $x_1, \ldots, x_n, x, y \in \mathcal{X}$,

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

and a Markovian jump process P is called *homogeneous* if furthermore

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

It is well-known – see [14, Theorem 5.2] or [12, Section 8.3] – that for any initial probability mass function $\pi\colon \mathcal{X} \to [0,1]$ and *rate matrix* $Q\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ – meaning that $\sum_{y\in\mathcal{X}} Q(x,y) = 0$ for all $x \in \mathcal{X}$ and $Q(x,y) \geq 0$ for all $x, y \in \mathcal{X}$ with $x \neq y$ – there is a unique homogeneous Markovian jump process $P_{\pi,Q}$ with initial probabilities $P_{\pi,Q}(X_0 = x) = \pi(x)$ and transition probabilities

$$P_{\pi,Q}(X_{t+\Delta} = y \mid X_t = x) = e^{\Delta Q}(x,y),$$

where $e^{\Delta Q}$ is the matrix exponential of ΔQ. Conversely, under a mild continuity condition [12, Eqn. (8.4)], a homogeneous Markovian jump process P is uniquely determined by its initial probability mass function

$$\pi_P\colon \mathcal{X} \to [0,1]\colon x \mapsto P(X_0 = x)$$

and its rate matrix $Q_P\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ defined by

$$Q_P(x,y) := \lim_{\Delta \searrow 0} \frac{P(X_\Delta = y \mid X_0 = x) - I(x,y)}{\Delta}$$

for all $x, y \in \mathcal{X}$, with I the identity matrix. That is, we then have that $P = P_{\pi_P, Q_P}$.

2.2. Towards Imprecision

That said, it is not always possible or desirable to model Your uncertainty with a single jump process. Here, we assume that it is instead possible to adequately model Your uncertainty through a set $\mathcal{P}$ of jump processes; one way to look at this is through the lens of sensitivity analysis: You believe that one of the jump processes $P \in \mathcal{P}$ is the 'correct' one, but do not know which one. Regardless of interpretation, we call such a set $\mathcal{P}$ of jump processes an *imprecise jump process*.

Krak et al. [14] present an elegant method to construct such an imprecise jump process. They draw inspiration from how a homogeneous Markovian jump process is uniquely characterised by its initial probability mass function and rate matrix, but they relax the requirement that these parameters should be known precisely: instead of one initial probability mass function π and rate matrix Q, they consider a set $\mathcal{M}$ of initial probability mass functions and a bounded[2] set $\mathcal{Q}$ of rate matrices; we keep these sets fixed throughout the remainder of this contribution. It then makes sense to consider the set

$$\mathcal{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}} := \{P_{\pi,Q}\colon \pi \in \mathcal{M}, Q \in \mathcal{Q}\}$$

of corresponding homogeneous Markovian jump processes. However, they argue that this is not the most convenient imprecise jump process, at least not in the case when $\mathcal{Q}$ is infinite and we seek to determine tight lower and upper bounds on the probabilities and expectations with respect to the jump processes in $\mathcal{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$. Quite remarkably, this does not appear to be a problem if one relaxes the requirement of homogeneity, and possibly even that of Markovianity. That is, if instead of the set $\mathcal{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ of homogeneous Markovian jump processes, one considers the set $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ of (not necessarily homogeneous) Markovian jump processes that are 'consistent' with $\mathcal{M}$ and $\mathcal{Q}$, or even the set $\mathcal{P}_{\mathcal{M},\mathcal{Q}}$ of (not necessarily Markovian) jump processes that are 'consistent' with $\mathcal{M}$ and $\mathcal{Q}$.

The notion of consistency with $\mathcal{M}$ is inspired by the definition of π_P: a jump process P is consistent with $\mathcal{M}$ if there is some $\pi \in \mathcal{M}$ such that $P(X_0 = x) = \pi(x)$ for all $x \in \mathcal{X}$. The notion of consistency with $\mathcal{Q}$ is more involved, but to some extent also inspired by the definition of Q_P: loosely speaking, a jump process P is consistent with $\mathcal{Q}$ if for all $n \in \mathbb{Z}_{\geq 0}$, $t_1, \ldots, t_n, t \in \mathbb{R}_{\geq 0}$ such that $t_1 < \cdots < t_n < t$, $x_1, \ldots, x_n \in \mathcal{X}$ and $\Delta \in \mathbb{R}_{>0}$, the corresponding matrix $Q_\Delta\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$, defined for all $x, y \in \mathcal{X}$ by

$$Q_\Delta(x,y) := \frac{P(X_{t+\Delta} = y \mid X_{t_{1:n}} = x_{1:n}, X_t = x) - I(x,y)}{\Delta},$$

comes arbitrarily close to – or is eventually contained in – $\mathcal{Q}$ as Δ approaches 0. In any case, these three imprecise processes are nested by construction: $\mathcal{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$.

[2] A set $\mathcal{A} \subseteq \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is bounded if $\sup\{\|A\|\colon A \in \mathcal{A}\} < +\infty$, where for any matrix $A\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$, $\|A\| = \sup\{\sum_{y\in\mathcal{X}} |A(x,y)|\colon x \in \mathcal{X}\}$.

2.3. The Upper Expectation of Interest

Suppose we have some imprecise jump process $\mathcal{P}$, and that we want to say something about the average of $f(X_\tau)$ over $[0,t]$, with f some given real-valued function on $\mathcal{X}$. Erreygers and De Bock [9] show that – under a mild technical condition on $\mathcal{P}$ – for any $P \in \mathcal{P}$ and $x \in \mathcal{X}$, the expectation

$$E_P\left(\frac{1}{t}\int_0^t f(X_\tau)\mathrm{d}\tau \,\middle|\, X_0 = x\right) \tag{1}$$

is well defined. Often, it suffices to know the best- and/or worst-case value of this expected average. That is, it suffices to determine the lower and upper bound on the set of values of the expectation in (1); we denote the upper bound by

$$\overline{E}_{\mathcal{P}}\left(\frac{1}{t}\int_0^t f(X_\tau)\,\mathrm{d}\tau \,\middle|\, X_0 = x\right) \coloneqq \sup_{P\in\mathcal{P}} E_P\left(\frac{1}{t}\int_0^t f(X_\tau)\,\mathrm{d}\tau \,\middle|\, X_0 = x\right),$$

and leave it to the reader to verify that

$$-\overline{E}_{\mathcal{P}}\left(\frac{1}{t}\int_0^t -f(X_\tau)\,\mathrm{d}\tau \,\middle|\, X_0 = x\right)$$

is the lower bound; hence, it suffices to only study the upper bounds. Consequently, we can generalise the notion of weak ergodicity for Markovian jump processes (as informally introduced in the Introduction) to one for imprecise jump processes as follows.

Definition 1 *An imprecise jump process $\mathcal{P}$ is said to be* weakly ergodic *if for all $f\colon \mathcal{X}\to\mathbb{R}$ and $x\in\mathcal{X}$,*

$$\lim_{t\to+\infty} \overline{E}_{\mathcal{P}}\left(\frac{1}{t}\int_0^t f(X_\tau)\,\mathrm{d}\tau \,\middle|\, X_0 = x\right)$$

exists and is the same for all $x \in \mathcal{X}$.

2.4. A Computational Scheme

Erreygers and De Bock [9] show that for any imprecise jump process $\mathcal{P}$ such that $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$ and under some mild condition on $\mathcal{Q}$, we can determine the upper expectation of the average of $f(X_\tau)$ over $[0,t]$ recursively, where in every step we only need to solve an optimisation problem with $\mathcal{Q}$. This result motivates the remainder of this contribution, so let us introduce the necessary notation and notions needed to repeat it.

We let $\mathcal{L} \coloneqq \mathbb{R}^{\mathcal{X}}$ be the set of real functions on $\mathcal{X}$, which we equip with the supremum (or maximum) norm; this makes $\mathcal{L}$ a Banach space, which will be important in Section 3 further on. Whenever it does not lead to confusion, we identify any real number $\mu \in \mathbb{R}$ with the constant function $\mu \in \mathcal{L}$ with range $\{\mu\}$. Besides the constant functions, we will also often need indicator functions: for any subset A of $\mathcal{X}$, the corresponding *indicator* $\mathbb{I}_A$ takes the value 1 on A and 0 elsewhere; for the sake of clarity, for all $x\in\mathcal{X}$ we will shorten $\mathbb{I}_{\{x\}}$ to $\mathbb{I}_x$.

Henceforth, we identify any matrix $M\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ with the linear map from $\mathcal{L}$ to $\mathcal{L}$, which we will also denote by M, that maps $f\in\mathcal{L}$ to $M[f]\in\mathcal{L}$ defined by

$$M[f](x) \coloneqq \sum_{y\in\mathcal{X}} M(x,y)f(y) \quad \text{for all } x\in\mathcal{X};$$

note that $M[\mathbb{I}_y](x) = M(x,y)$ for all $x,y\in\mathcal{X}$. The (pointwise) upper envelope of $\mathcal{Q}$, then, maps any $f\in\mathcal{L}$ to

$$\overline{Q}_{\mathcal{Q}}[f]\colon \mathcal{X}\to\mathbb{R}\colon x\mapsto \sup\{Q[f](x)\colon Q\in\mathcal{Q}\}.$$

Now that we have introduced $\overline{Q}_{\mathcal{Q}}$, we are almost ready to repeat (a slightly rephrased version of) Theorem 68 in [9]. We do need one more notion, though. The bounded set $\mathcal{Q}$ of rate matrices is said to have *separately specified rows* [14, Definition 7.3] if for any selection $(Q_x)_{x\in\mathcal{X}}\in\mathcal{Q}^{\mathcal{X}}$ of rate matrices in $\mathcal{Q}$, the rate matrix

$$Q\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}\colon (x,y)\mapsto Q_x(x,y),$$

whose x-th row is the x-th row of Q_x, belongs to $\mathcal{Q}$; in other words, $\mathcal{Q}$ has separately specified rows if and only if its 'rows' are specified independently of each other.

Theorem 2 *Consider an imprecise jump process $\mathcal{P}$ such that $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$. Fix any $f\in\mathcal{L}$ and $t\in\mathbb{R}_{>0}$. For all $n\in\mathbb{N}$, let $\Delta_n \coloneqq t/n$ and define the sequence $(\tilde{f}_{n,0},\dots,\tilde{f}_{n,n})$ by the initial condition $\tilde{f}_{n,n}\coloneqq 0$ and, for all $k\in\{0,\dots,n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} \coloneqq \Delta_n f + \tilde{f}_{n,k+1} + \Delta_n\overline{Q}_{\mathcal{Q}}[\tilde{f}_{n,k+1}]. \tag{2}$$

If $\mathcal{Q}$ has separately specified rows, then for all $x\in\mathcal{X}$,

$$\overline{E}_{\mathcal{P}}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \,\middle|\, X_0 = x\right) = \frac{1}{t}\lim_{n\to+\infty}\tilde{f}_{n,0}(x).$$

Proof [9, Theorem 66] – with, for all $n\in\mathbb{N}$ and $k\in\{0,\dots,n-1\}$, $\tilde{f}_{n,k} = \Delta_n f + f_{n,k+1}$ – implies, for all $x\in\mathcal{X}$,

$$\overline{E}_{\mathcal{P}}\left(\int_0^t f(X_s)\mathrm{d}s \,\middle|\, X_0 = x\right) = \lim_{n\to+\infty}(I+\Delta_n\overline{Q}_{\mathcal{Q}})[\tilde{f}_{n,0}](x).$$

Now it follows from (R5) further on that for all $n\in\mathbb{N}$,

$$\left\|(I+\Delta_n\overline{Q}_{\mathcal{Q}})[\tilde{f}_{n,0}] - \tilde{f}_{n,0}\right\| \leq \Delta_n\|\overline{Q}_{\mathcal{Q}}\|\|\tilde{f}_{n,0}\|,$$

with $\|\overline{Q}_{\mathcal{Q}}\|$ as defined in (3) further on. Moreover, it is not difficult to show (see Lemma 23 in the Supplementary Material) that $\|\tilde{f}_{n,0}\| \leq t\|f\|$ for all $n\in\mathbb{N}$ such that $t\|\overline{Q}_{\mathcal{Q}}\|\leq 2n$. From these two observations, we infer that

$$\lim_{n\to+\infty}\left\|(I+\Delta_n\overline{Q}_{\mathcal{Q}})[\tilde{f}_{n,0}] - \tilde{f}_{n,0}\right\| \leq \lim_{n\to+\infty}\frac{t^2}{n}\|\overline{Q}_{\mathcal{Q}}\|\|f\| = 0.$$

The equality in the statement follows because $\overline{E}_{\mathcal{P}}$ is positively homogeneous since every E_P is. ■

3. Weak Ergodicity for Upper Rate Operators

Krak et al. [14, Eqn. (38)] argue that the upper envelope $\overline{Q}_{\mathcal{Q}}$ is a map from $\mathcal{L}$ to $\mathcal{L}$ because $\mathcal{Q}$ is bounded, and their Proposition 7.5 establishes that this upper envelope $\overline{Q}_{\mathcal{Q}}$ is an upper rate operator in the sense of their Definition 7.2, which we repeat here.

Definition 3 *An* upper rate operator $\overline{Q}$ *is a map from* $\mathcal{L}$ *to* $\mathcal{L}$ *such that for all* $f, g \in \mathcal{L}$, $\mu \in \mathbb{R}$ *and* $x, y \in \mathcal{X}$,

R1. $\overline{Q}[f+g] \leq \overline{Q}[f] + \overline{Q}[g]$;

R2. $\overline{Q}[\mu f] = \mu \overline{Q}[f]$ *whenever* $\mu \geq 0$;

R3. $\overline{Q}[\mu] = 0$;

R4. $\overline{Q}[-\mathbb{I}_y](x) \leq 0$ *whenever* $x \neq y$.

Its conjugate lower rate operator $\underline{Q}\colon \mathcal{L} \to \mathcal{L}$ *is defined by*

$$\underline{Q}[f] := -\overline{Q}[-f] \quad \text{for all } f \in \mathcal{L}.$$

Throughout this contribution, we will need the (operator) norm of an upper rate operator $\overline{Q}$, which De Bock [4, Eqn. (4)] defines as

$$\|\overline{Q}\| := \sup\{\|\overline{Q}[f]\|\colon f \in \mathcal{L}, \|f\| = 1\} < +\infty, \tag{3}$$

and can easily be computed with the help of Proposition 4 in [8]. Moreover, it follows immediately from this definition and (R2) that

R5. $\|\overline{Q}[f]\| \leq \|\overline{Q}\|\|f\|$ for all $f \in \mathcal{L}$.

Let us now revisit the recursion in (2), but this time for a general upper rate operator $\overline{Q}$. For all $f \in \mathcal{L}$, we let

$$\overline{Q}_f\colon \mathcal{L} \to \mathcal{L}\colon g \mapsto f + \overline{Q}[g].$$

This way, for all $f, g \in \mathcal{L}$ and $\Delta \in \mathbb{R}_{>0}$,

$$\Delta f + g + \Delta\overline{Q}[g] = g + \Delta\overline{Q}_f[g] = (I + \Delta\overline{Q}_f)[g].$$

With this in mind, we see that for all $n \in \mathbb{N}$, $\tilde{f}_{n,0}$ as defined in Theorem 2 can be thought of as an approximation through Euler's method with uniform step size $\Delta_n = t/n$ – see for example [15, Section 5.1] – of the solution of the initial value problem

$$\begin{cases} u'(t) = \overline{Q}_f[u(t)] \text{ for } t \in \mathbb{R}_{\geq 0} \\ u(0) = 0. \end{cases} \tag{4}$$

It should not come as a surprise, then, that this initial value problem has a solution; as we will presently see, this solution is in fact unique.

Fix some $f \in \mathcal{L}$, and consider the initial value problem

$$\begin{cases} u'(t) = \overline{Q}_f[u(t)] \text{ for } t \in \mathbb{R}_{\geq 0} \\ u(0) = g \end{cases} \tag{5}$$

in $\mathcal{L}$, where $g \in \mathcal{L}$ is an arbitrary initial condition. The Cauchy–Lipschitz Theorem – see for example [2, Theorem 7.3] or [15, Theorem 1.1] – gives existence and uniqueness of a solution in $C^1(\mathbb{R}_{\geq 0}, \mathcal{L})$ of this initial value problem, but this requires that $\overline{Q}_f$ is Lipschitz. This is clearly the case because $\overline{Q}$ is Lipschitz [4, R11] and, for all $g, h \in \mathcal{L}$,

$$\|\overline{Q}_f[g] - \overline{Q}_f[h]\| = \|\overline{Q}[g] - \overline{Q}[h]\|.$$

Hence, by the Cauchy–Lipschitz Theorem, the initial value problem in (5) has a unique solution, which we denote by $e^{\bullet\overline{Q}_f}[g]\colon \mathbb{R}_{\geq 0} \to \mathcal{L}\colon t \mapsto e^{t\overline{Q}_f}[g]$.

We can now rewrite Theorem 2 with $e^{t\overline{Q}_f}$ because, since Euler's Method is convergent [see 15, Section 5.2],

$$e^{t\overline{Q}_f}[g] = \lim_{n\to+\infty}\left(I + \frac{t}{n}\overline{Q}_f\right)^n[g] \tag{6}$$

for all $t \in \mathbb{R}_{\geq 0}$ and $g \in \mathcal{L}$.

Corollary 4 *Consider an imprecise jump process* $\mathcal{P}$ *such that* $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$, *and let* $\overline{Q} := \overline{Q}_{\mathcal{Q}}$. *Fix any* $f \in \mathcal{L}$ *and* $t \in \mathbb{R}_{>0}$. *If* $\mathcal{Q}$ *has separately specified rows, then for all* $x \in \mathcal{X}$,

$$\overline{E}_{\mathcal{P}}\left(\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s \,\middle|\, X_0 = x\right) = \frac{1}{t}e^{t\overline{Q}_f}[0](x).$$

The combination of this result and Definition 1 motivates our notion of weak ergodicity for upper rate operators.

Definition 5 *An upper rate operator* $\overline{Q}$ *is called* weakly ergodic *if for all* $f \in \mathcal{L}$, *the limit* $\lim_{t\to+\infty}\frac{1}{t}e^{t\overline{Q}_f}[0]$ *exists and is constant over* $\mathcal{X}$.

This way, we have turned the question whether $\mathcal{P}$ – with $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$ and $\mathcal{Q}$ separately specified – is weakly ergodic into the question whether $\overline{Q} := \overline{Q}_{\mathcal{Q}}$ is weakly ergodic, and this is progress because the latter is easier to investigate.

Before we continue, let us briefly investigate the special case $f = 0$. Since $\overline{Q}_0 = \overline{Q}$, it follows from (6) that for all $t \in \mathbb{R}_{\geq 0}$ and $g \in \mathcal{L}$,

$$e^{t\overline{Q}_0}[g] = \lim_{n\to+\infty}\left(I + \frac{t}{n}\overline{Q}\right)^n[g] = e^{t\overline{Q}}[g],$$

where $e^{t\overline{Q}}$ is the (non-linear) operator exponential of $t\overline{Q}$ as defined and investigated by [19, 4, 14].

4. A Necessary and Sufficient Condition

Our main contribution is an easy to check condition on $\overline{Q}$ that is necessary and sufficient for $\overline{Q}$ to be weakly ergodic. As we will see in Section 4.2, said condition can be elegantly stated by means of the notions of upper and lower reachability. We will work with these notions for upper rate operators *and* upper transition operators – see Section 4.3 further on – so we present these notions for an arbitrary operator; this is simply a unified treatment of the reachability relations in [11, 4, 18], see also [1, Sections 4.1 and 4.2].

4.1. Upper and Lower Reachability

Consider an operator $F\colon \mathcal{L} \to \mathcal{L}$. We let $\mathcal{G}(F) = (\mathcal{V}_F, \mathcal{E}_F)$ be the directed graph with vertices $\mathcal{V}_F := \mathcal{X}$ and directed edges

$$\mathcal{E}_F := \big\{(x,y) \in \mathcal{X}^2 \colon x \neq y, F[\mathbb{I}_y](x) > 0\big\}.$$

For any two states $x, y \in \mathcal{X}$, we say that y is *upper reachable* from x, and write $x \rightsquigarrow y$, if (i) $x = y$; or (ii) there is a directed path from x to y in $\mathcal{G}(F)$, so a sequence $x_0, x_1, \dots, x_n$ such that $x_0 = x$, $x_n = y$ and $(x_{k-1}, x_k) \in \mathcal{E}_F$ for all $k \in \{1, \dots, n\}$. Moreover, we say that x and y *communicate*, denoted by $x \leftrightsquigarrow y$, if $x \rightsquigarrow y$ and $y \rightsquigarrow x$. The equivalence relation $\bullet \leftrightsquigarrow \bullet$ partitions $\mathcal{X}$ into equivalence classes, called *communication classes*, which partition $\mathcal{X}$. Such a communication class $\mathcal{S}$ is *closed* if $x \not\rightsquigarrow y$ for all $x \in \mathcal{S}$ and $y \in \mathcal{S}^{\mathrm{c}}$. The upper reachability relation $\bullet \rightsquigarrow \bullet$ induces a partial order $\prec$ on the communication classes: $\mathcal{S}_1 \prec \mathcal{S}_2$ if and only if $x \rightsquigarrow y$ for one (and then all) $(x, y) \in \mathcal{S}_1 \times \mathcal{S}_2$. Now we say that F has a *top class* if there is a communication class that dominates all other communication classes in the partial order $\prec$; T'Joens and De Bock [18, Section 5] show that this is the case if and only if the set

$$\{x \in \mathcal{X} \colon (\forall y \in \mathcal{X})\ y \rightsquigarrow x\}$$

is non-empty, in which case this set is the top class. Furthermore, whenever it exists, the top class is the unique closed communication class.

We also need the notion of lower reachability. Let $K := |\mathcal{X}|$. For all non-empty $A \subseteq \mathcal{X}$ and $x \in \mathcal{X}$, we say that A is *lower reachable* from x if $x \in A_K$, where $(A_0, \dots, A_K)$ is the sequence defined by the initial condition $A_0 := A$ and, for all $k \in \{0, \dots, K-1\}$, by the recursive relation

$$A_{k+1} := A_k \cup \big\{x \in \mathcal{X} \setminus A_k \colon F[-\mathbb{I}_{A_k}](x) < 0\big\}.$$

Finally, a (non-empty) set $\mathcal{S} \subseteq \mathcal{X}$ is said to be *F-absorbing* if $\mathcal{S}$ is lower reachable from all $x \in \mathcal{X} \setminus \mathcal{S}$.

4.2. The Main Result

As previously mentioned, we are by no means the first to use the notions of upper and lower reachability. For an upper rate operator $\overline{Q}$, the corresponding notions of upper and lower reachability are precisely those introduced by De Bock [4, Defintions 7 and 8] – the equivalence holds because for all $A \subseteq \mathcal{X}$, $\underline{Q}[\mathbb{I}_A] > 0 \Leftrightarrow \overline{Q}[-\mathbb{I}_A] < 0$. They use these notions to establish a necessary and sufficient condition for ergodicity, defined as follows [4, Definition 6].

Definition 6 *An upper rate operator $\overline{Q}$ is called* ergodic *if for all $f \in \mathcal{L}$, $\lim_{t \to +\infty} e^{t\overline{Q}}[f]$ exists and is a constant function.*

Their Theorem 19, then, establishes that $\overline{Q}$ is ergodic if and only if $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing.

The notions of ergodicity and weak ergodicity of $\overline{Q}$ are quite different. It is therefore quite remarkable that, due to [4, Theorem 19] and the following result, which is the main contribution of our paper, they actually turn out to be equivalent.

Theorem 7 *An upper rate operator $\overline{Q}$ is weakly ergodic if and only if $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing.*

From a practical point of view, it is good to know that the two conditions in this result can be easily checked; we refer to [4, Algorithms 1 and 2] for more details. We also point out the following two corollaries, which follow by combining Theorem 7 with Corollary 4 and Theorem 19 in [4], respectively.

Corollary 8 *Consider an imprecise jump process $\mathcal{P}$ such that $\mathcal{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathcal{P}_{\mathcal{M},\mathcal{Q}}$, and suppose that $\mathcal{Q}$ has separately specified rows. Then $\mathcal{P}$ is weakly ergodic if and only if $\overline{Q}_{\mathcal{Q}}$ has a top class that is $\overline{Q}_{\mathcal{Q}}$-absorbing.*

Corollary 9 *An upper rate operator $\overline{Q}$ is weakly ergodic if and only if it is ergodic.*

The remainder of this contribution is devoted to our proof of Theorem 7. We show the sufficiency (or the converse implication) in Section 5 and the necessity (or the forward implication) in Section 6. Throughout these proofs, we will repeatedly rely on the notion of weak ergodicity for upper transition operators, the discrete-time counterpart of upper rate operators. It is for this reason that we discuss this notion in the following section.

4.3. Upper Transition Operators

An upper transition operator is to an imprecise Markov chain [3, 11, 18] what an upper rate operator is to a Markovian imprecise jump process.

Definition 10 *An* upper transition operator $\overline{T}$ *is a map from $\mathcal{L}$ to $\mathcal{L}$ such that for all $f, g \in \mathcal{L}$ and $\lambda \in \mathbb{R}_{\geq 0}$,*

T1. $\overline{T}[f + g] \leq \overline{T}[f] + \overline{T}[g]$;

T2. $\overline{T}[\lambda f] = \lambda \overline{T}[f]$;

T3. $\overline{T}[f] \leq \max f$.

It is well known – see, for example, [3, Appendix], [11, Section 1] or [18, Section 4] – that an upper transition operator $\overline{T}$ has the following properties: for all $f, g \in \mathcal{L}$ and $\mu \in \mathbb{R}$,

T4. $\min f \leq \overline{T}[f] \leq \max f$;

T5. $\overline{T}[\mu + f] = \mu + \overline{T}[f]$;

T6. $\overline{T}[f] \leq \overline{T}[g]$ whenever $f \leq g$;

T7. $\|\overline{T}[f]\| \leq \|f\|$;

T8. $\|\overline{T}[f] - \overline{T}[g]\| \leq \|f - g\|$.

An upper transition operator $\overline{T}$ is called *ergodic* if for all $f \in \mathcal{L}$, $\lim_{n\to+\infty} \overline{T}^n[f]$ exists and is a constant function [11, Definition 2]; note the resemblance between this definition and Definition 6. The notion of weak ergodicity of upper rate operators (Definition 5) also has a counterpart for upper transition operators. To state it, T'Joens and De Bock [18, Section 4] introduce the following derived operator: for any upper transition operator $\overline{T}$ and any $f \in \mathcal{L}$,

$$\overline{T}_f : \mathcal{L} \to \mathcal{L} : g \mapsto \overline{T}_f[g] := f + \overline{T}[g].$$

Then an upper transition operator $\overline{T}$ is *weakly ergodic* if for all $f \in \mathcal{L}$, $\lim_{n\to+\infty} \frac{1}{n}\overline{T}_f^n[0]$ exists and is a constant function [18, Section 4]. Again, this notion is similar to our counterpart for upper rate operators (Definition 5); this should not come as a surprise since T'Joens and De Bock [18, Section 3] also motivate their notion of weak ergodicity through the limit for $n \to +\infty$ of the upper expectation (with respect to some imprecise Markov chain) of $\frac{1}{n}\sum_{k=1}^n f(X_k)$.

Crucially, T'Joens and De Bock [18, Theorem 14] give a necessary and sufficient condition on $\overline{T}$ for weak ergodicity, and their condition is one in two parts. The first part is that $\overline{T}$ has a top class, but the second part is a bit more involved. With the help of results in [11] or [1], we can translate this second part into one about $\overline{T}$-absorption. To this end, note that any upper transition operator $\overline{T}$ and any $f \in \mathcal{L}$, it follows from (T5) and (T6) that both $\overline{T}$ and $\overline{T}_f$ are what is known as a 'topical map' [18, (T1) and (T2)] or a 'monotone additively homogeneous map' [1, Section 2.1]; $\overline{T}$ is furthermore convex due to (T1) and (T2).

Theorem 11 *For any upper transition operator $\overline{T}$, the following statements are equivalent.*

(i) *$\overline{T}$ is weakly ergodic.*

(ii) *for all $f, h \in \mathcal{L}$, $\lim_{n\to+\infty} \frac{1}{n}\overline{T}_f^n[h]$ exists, is a constant function and does not depend on h.*

(iii) *$\overline{T}$ has a top class $\mathcal{S}$ that is 'absorbing', meaning that*

$$(\forall x \in \mathcal{X} \setminus \mathcal{S})(\exists n \in \mathbb{N})\ \overline{T}^n[\mathbb{I}_{\mathcal{S}^c}](x) < 1.$$

(iv) *$\overline{T}$ has a top class that is $\overline{T}$-absorbing.*

(v) *For all $f \in \mathcal{L}$, there is some pair $(\mu, g) \in \mathbb{R} \times \mathcal{L}$ such that $\overline{T}_f[g] = \mu + g$.*

Proof Since $\overline{T}_f$ is a 'topical map', it follows from Lemma 3.1 in [10] that (i) implies (ii); (i) and (ii) are therefore equivalent because (ii) trivially implies (i). The equivalence between (i) and (iii) is exactly Theorem 14 in [18]. Since $\overline{T}$ is a 'monotone additively homogeneous map', the equivalence between (ii) and (v) follows from Theorem 2.1 in [1]. That (iv) is equivalent to the other conditions, follows from [11] or [1]: (iv) is equivalent to (iii) due to [11, Proposition 6], and equivalent to (ii) due to [1, Corollary 4.4 and Theorem 2.1]. Neither of these may be immediately obvious though, so let us elaborate this a bit.

To see why [11, Proposition 6] implies that (iv) is equivalent to (iii), observe that for $A \subseteq \mathcal{X}$ and $x \in \mathcal{X} \setminus A$, since $-\mathbb{I}_A = \mathbb{I}_{A^c} - 1$,

$$\overline{T}[-\mathbb{I}_A](x) < 0 \overset{\text{(T5)}}{\Leftrightarrow} \overline{T}[\mathbb{I}_{A^c}](x) < 1 \overset{\text{(T4)}}{\Leftrightarrow} \overline{T}[\mathbb{I}_{A^c}](x) \neq 1.$$

This makes the definition of lower reachability in Section 4.1 equivalent to the condition in [11, Proposition 6], provided we take into account that the sets A_k there correspond to the sets A_k^c here. There's a caveat though, because [11, Proposition 6] requires that the top class is regular. Going through the proof of the result, however, it becomes clear that this condition is in fact never used.

To see why [1, Corollary 4.4 and Theorem 2.1] imply the equivalence between (iv) and (ii), observe that for $x, y \in \mathcal{X}$,

$$+\infty = \lim_{\alpha\to+\infty} \overline{T}[\alpha\mathbb{I}_y](x) \overset{\text{(T2)}}{=} \lim_{\alpha\to+\infty} \alpha\overline{T}[\mathbb{I}_y](x)$$

if and only if $\overline{T}[\mathbb{I}_y](x) > 0$. Hence, the 'unique final class' of $\mathcal{G}_\infty(\overline{T})$ (in their notation) is precisely the top class $\mathcal{S}$ of $\overline{T}$. Similarly, for all $A \subseteq \mathcal{X}$ and $x \in \mathcal{X} \setminus A$,

$$-\infty = \lim_{\alpha\to-\infty} \overline{T}[\alpha\mathbb{I}_A](x) \overset{\text{(T2)}}{=} \lim_{\alpha\to-\infty} -\alpha\overline{T}[-\mathbb{I}_A](x)$$

if and only if $\overline{T}[-\mathbb{I}_A](x) < 0$. With a bit more work, and taking into account the explanation in [1, Section 4.1] (their edges are reversed compared to ours), we can verify that $\operatorname{reach}(\mathcal{S}, \mathcal{H}_\infty^-(\overline{T}))$ (in their notation) is exactly the set of states $y \in \mathcal{X}$ such that $\mathcal{S}$ is lower reachable from y. So $\operatorname{reach}(\mathcal{S}, \mathcal{H}_\infty^-(\overline{T})) = \mathcal{X}$ (the second part of the condition in [1, Corollary 4.4]) if and only if $\mathcal{S}$ is $\overline{T}$-absorbing. ■

While we have seen in Corollary 9 that the notions of weak ergodicity and ergodicity are equivalent for upper rate operators, this is not the case for upper transition operators! T'Joens and De Bock [18, Example 1] give an elegant counterexample of an upper transition operator $\overline{T}$ that is weakly ergodic, but *not* ergodic.

That said, there are important connections between the notions of weak ergodicity for upper rate operators and upper transition operators. These connections stem from the fact that we can use an upper rate operator $\overline{Q}$ to construct an upper transition operator in the following way.

Lemma 12 *Consider an upper rate operator $\overline{Q}$ and some $\Delta \in \mathbb{R}_{\geq 0}$. Then $I + \Delta\overline{Q}$ is an upper transition operator if and only if $\Delta\|\overline{Q}\| \leq 2$.*

Proof Since $\|\overline{Q}[f]\| = \|\underline{Q}[-f]\|$, it is clear that $\|\overline{Q}\|$ as defined in (3) is equal to $\|\underline{Q}\|$ as used in [8]. Now Proposition 3 in [8] says that $(I + \Delta\underline{Q})$ is a *lower transition operator* (in the sense of their Definition 1) if and only if $\Delta\|\overline{Q}\| = \Delta\|\underline{Q}\| \leq 2$. Our statement follows because $\overline{T}$ is an upper transition operator if and only if the conjugate operator $\underline{T}\colon \mathcal{L} \to \mathcal{L}\colon f \mapsto -\overline{T}[-f]$ is a lower transition operator. ■

Quite remarkably, the condition on $\overline{Q}$ in Theorem 7 is necessary and sufficient to guarantee that the constructed upper transition operator $(I + \Delta\overline{Q})$ is weakly ergodic.

Theorem 13 *Consider an upper rate operator $\overline{Q}$ and some $\Delta \in \mathbb{R}_{>0}$ such that $\Delta\|\overline{Q}\| \leq 2$. Then $I + \Delta\overline{Q}$ is weakly ergodic if and only if $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing.*

Due to Theorem 11, we can divide our proof in the following two lemmas.

Lemma 14 *Consider an upper rate operator $\overline{Q}$ and some $\Delta \in \mathbb{R}_{>0}$.Then the notions of upper reachability for $\overline{Q}$ and $I+\Delta\overline{Q}$ are equivalent. Consequently, the communication classes of $\overline{Q}$ and $I + \Delta\overline{Q}$ are equivalent; in particular, $\overline{Q}$ has a top class if and only if $I + \Delta\overline{Q}$ has one, and these are necessarily equal if they exist.*

Proof Observe that for all $x, y \in \mathcal{X}$ such that $x \neq y$,

$$0 < \overline{Q}[\mathbb{I}_y](x) \Leftrightarrow 0 < \mathbb{I}_y(x)+\Delta\overline{Q}[\mathbb{I}_y](x) = (I+\Delta\overline{Q})[\mathbb{I}_y](x).$$

It follows immediately from this equality that the notions of upper reachability for $\overline{Q}$ and $I + \Delta\overline{Q}$ are equivalent, which implies the statement. ■

Lemma 15 *Consider an upper rate operator $\overline{Q}$ and some $\Delta \in \mathbb{R}_{>0}$. A (non-empty) subset A of $\mathcal{X}$ is $\overline{Q}$-absorbing if and only if it is $(I + \Delta\overline{Q})$-absorbing.*

Proof Observe that for all $B \subseteq \mathcal{X}$ and $x \in \mathcal{X} \setminus B$,

$$\Delta\overline{Q}[-\mathbb{I}_B](x) = -\mathbb{I}_B(x)+\Delta\overline{Q}[-\mathbb{I}_B](x) = (I+\Delta\overline{Q})[-\mathbb{I}_B](x).$$

It follows immediately from this equality that the notions of lower reachability for $\overline{Q}$ and $I + \Delta\overline{Q}$ are equivalent, and this implies the statement. ■

Proof of Theorem 13 Recall from Lemma 12 that $I+\Delta\overline{Q}$ is an upper transition operator. Due to Theorem 11, the result therefore follows immediately from Lemmas 14 and 15. ■

At first sight, one might be tempted to think that Theorem 13 provides a direct path to Theorem 7. Recall from Definition 5 and (6) that $\overline{Q}$ is weakly ergodic if and only if for all $f \in \mathcal{L}$,

$$\lim_{t\to+\infty} \frac{1}{t} e^{t\overline{Q}_f}[0] = \lim_{t\to+\infty} \lim_{n\to+\infty} \frac{n}{t}\frac{1}{n}\Big(I + \frac{t}{n}\overline{Q}_f\Big)^n[0] \qquad (7)$$

exists and is constant. Now if $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing, then for all $t \in \mathbb{R}_{>0}$ and $n \in \mathbb{N}$ such that $t\|\overline{Q}\| \leq 2n$, Theorem 13 implies that for all $f \in \mathcal{L}$,

$$\lim_{n\to+\infty} \frac{1}{n}\Big(I + \frac{t}{n}\overline{Q}_f\Big)^n[0]$$

exists and is constant. Unfortunately, this does not immediately help us to prove (7). Theorem 13 would help us if we could establish that

$$\lim_{t\to+\infty} \frac{1}{t} e^{t\overline{Q}_f}[0] = \lim_{\Delta\searrow 0} \lim_{n\to+\infty} \frac{1}{n\Delta}(I + \Delta\overline{Q}_f)^n[0],$$

but this is no easy feat. For this reason, we choose to pursue an avenue that builds on Theorem 13, but also on Theorem 11 and some intermediary results in [18].

5. Proof of the Sufficiency

The goal of this section is to prove the sufficiency in Theorem 7, so the following result.

Proposition 16 *Consider an upper rate operator $\overline{Q}$. If $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing, then $\overline{Q}$ is weakly ergodic.*

Suppose $\overline{Q}$ satisfies the condition in Proposition 16, fix some $\Delta \in \mathbb{R}_{>0}$ such that $\Delta\|\overline{Q}\| \leq 2$ and let $\overline{T} := I + \Delta\overline{Q}$. Then we know from Theorem 13 that $\overline{T} = I + \Delta\overline{Q}$ is weakly ergodic. Hence, for all $f \in \mathcal{L}$, it follows from Theorem 11 (with Δf here as f there) that $\overline{T}_{\Delta f} = I + \Delta\overline{Q}_f$ has an additive eigenvector, in the sense that there is some couple $(\alpha, g) \in \mathbb{R} \times \mathcal{L}$ such that $(I + \Delta\overline{Q}_f)[g] = \alpha + g$. With $\mu := \alpha/\Delta$, this implies that $\overline{Q}_f[g] = \mu$. Let us use this information to prove Proposition 16.

As an intermediary step, we set out to prove that under our conditions on $\overline{Q}$, $e^{t\overline{Q}_f}$ has an additive eigenvector for all $t \in \mathbb{R}_{\geq 0}$ and $f \in \mathcal{L}$. In our proof, we use the following property of the upper rate operator $\overline{Q}$ [4, R6]:

R6. $\overline{Q}[f + \mu] = \overline{Q}[f]$ for all $f \in \mathcal{L}$ and $\mu \in \mathbb{R}$.

Lemma 17 *Consider an upper rate operator $\overline{Q}$, and suppose $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing. Then for all $f \in \mathcal{L}$, there is a couple $(\mu, g) \in \mathbb{R} \times \mathcal{L}$ such that $e^{t\overline{Q}_f}[g] = t\mu + g$ for all $t \in \mathbb{R}_{\geq 0}$.*

Proof Fix any $f \in \mathcal{L}$. Then as explained in the main text, it follows from our assumptions that there is a couple $(\mu, g) \in \mathbb{R} \times \mathcal{L}$ such that $\overline{Q}_f[g] = \mu$.

Since the statement is trivial for $t = 0$ – because $e^{0\overline{Q}_f} = I$ – we fix some $t \in \mathbb{R}_{>0}$. Then for all $n \in \mathbb{N}$, it follows from the preceding and $(n-1)$ applications of (R6) that

$$\left(I + \frac{t}{n}\overline{Q}_f\right)^n [g] = \left(I + \frac{t}{n}\overline{Q}_f\right)^{n-1}\left(g + \frac{t}{n}\mu\right) = \cdots = t\mu + g.$$

Due to (6), this implies that

$$e^{t\overline{Q}_f}[g] = \lim_{n\to+\infty}\left(I + \frac{t}{n}\overline{Q}_f\right)^n [g] = t\mu + g,$$

as required. ■

We are almost there, but not quite. Indeed, from Lemma 17 we learn that under the condition on $\overline{Q}$ in Proposition 16, for all $f \in \mathcal{L}$ there is some couple $(\mu, g) \in \mathbb{R} \times \mathcal{L}$ such that

$$\lim_{t\to+\infty}\frac{1}{t}e^{t\overline{Q}_f}[g] = \lim_{t\to+\infty}\frac{t\mu + g}{t} = \mu.$$

This is almost weak ergodicity, be it not that we need this property for $g = 0$. Fortunately, it is easy to prove that this is implied.

Lemma 18 *Consider an upper rate operator $\overline{Q}$ and some $f \in \mathcal{L}$. If the limit $\lim_{t\to\infty}\frac{1}{t}e^{t\overline{Q}_f}[g]$ exists for some $g \in \mathcal{L}$, then*

$$\lim_{t\to\infty}\frac{1}{t}e^{t\overline{Q}_f}[h] = \lim_{t\to\infty}\frac{1}{t}e^{t\overline{Q}_f}[g] \quad \textit{for all } h \in \mathcal{L}.$$

Proof Fix any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta\|\overline{Q}\| \leq 2$, and recall from Lemma 12 that $I + \Delta\overline{Q}$ is an upper transition operator. Then for all $h_1, h_2 \in \mathcal{L}$,

$$\begin{aligned}&\left\|(I + \Delta\overline{Q}_f)[h_1] - (I + \Delta\overline{Q}_f)[h_2]\right\| \\ &\quad = \left\|(I + \Delta\overline{Q})[h_1] - (I + \Delta\overline{Q})[h_2]\right\| \leq \|h_1 - h_2\|,\end{aligned}$$

where the inequality follows from (T8).

For all $t \in \mathbb{R}_{>0}$ and $n \in \mathbb{N}$ such that $t\|\overline{Q}\| \leq 2n$, we infer from repeated application of the preceding that

$$\left\|\left(I + \frac{t}{n}\overline{Q}_f\right)^n [g] - \left(I + \frac{t}{n}\overline{Q}_f\right)^n [h]\right\| \leq \|g - h\|.$$

It is now easy to use this observation and (6) to prove that for all $t \in \mathbb{R}_{>0}$,

$$\|e^{t\overline{Q}_f}[g] - e^{t\overline{Q}_f}[h]\| \leq \|g - h\|.$$

Hence, we see that

$$\begin{aligned}&\lim_{t\to+\infty}\left\|\frac{1}{t}e^{t\overline{Q}_f}[g] - \frac{1}{t}e^{t\overline{Q}_f}[h]\right\| \\ &\quad = \lim_{t\to+\infty}\frac{1}{t}\left\|e^{t\overline{Q}_f}[g] - e^{t\overline{Q}_f}[h]\right\| \\ &\quad \leq \lim_{t\to+\infty}\frac{1}{t}\|g - h\| = 0,\end{aligned}$$

as required. ■

Proof of Proposition 16 Follows immediately from Lemmas 17 and 18. ■

6. Proof of the Necessity

This section is devoted to the proof of the necessity in Theorem 7, which is precisely the following result.

Proposition 19 *If $\overline{Q}$ is weakly ergodic, then $\overline{Q}$ has a top class that is $\overline{Q}$-absorbing.*

Our proof uses similar arguments as those of T'Joens and De Bock [18]. Their arguments are for the discrete time setting, but we can nevertheless use them because of the following quintessential result; since our proof is two pages long, we have relegated it to the Supplementary Material.

Theorem 20 *Consider an upper rate operator $\overline{Q}$, and fix some $f \in \mathcal{L}$ and $\Delta \in \mathbb{R}_{>0}$. Then for all $h \in \mathcal{L}$,*

$$\lim_{n\to+\infty}\frac{1}{n\Delta}\left\|e^{n\Delta\overline{Q}_f}[h] - \left(I + \frac{\Delta}{n^2}\overline{Q}_f\right)^{n^3}[h]\right\| = 0.$$

Other than that, our proof of Proposition 19 also relies on two lemmas. The first is Lemma 36 in [18], which we repeat here for convenience.

Lemma 21 *Consider an upper transition operator $\overline{T}$. Then for any closed communication class $\mathcal{S}$ and any $h, g \in \mathcal{L}$ such that $h(x) = g(x)$ for all $x \in \mathcal{S}$,*

$$\overline{T}_h^n[0](x) = \overline{T}_g^n[0](x) \quad \textit{for all } n \in \mathbb{N}, x \in \mathcal{S}.$$

The second is a 'rephrasing' – consequence is perhaps more accurate – of Lemma 44 in [18].

Lemma 22 *Consider an upper rate operator $\overline{Q}$ and some $\Delta \in \mathbb{R}_{>0}$ such that $\Delta\|\overline{Q}\| \leq 2$. Suppose $\overline{Q}$ has a top class $\mathcal{S}$ that is not $\overline{Q}$-absorbing. Then there is a non-empty subset A of $\mathcal{S}^c$ such that $\mathbb{I}_A \leq (I + \Delta\overline{Q})[\mathbb{I}_A]$.*

Proof It follows from the assumptions in the statement and Lemmas 14 and 15 that $(I + \Delta\overline{Q})$ also has $\mathcal{S}$ as top class and that $\mathcal{S}$ is not $(I + \Delta\overline{Q})$-absorbing. Since we know from Lemma 12 that $(I + \Delta\overline{Q})$ is an upper transition operator, the statement now follows immediately from Theorem 11 ((iii)⇔(iv)) and Lemma 44 in [18]. ■

Proof of Proposition 19 Fix some $\varDelta \in \mathbb{R}_{>0}$ such that $\varDelta\|\overline{Q}\| \leq 2$. Then for all $n \in \mathbb{N}$, with $\delta_n := \varDelta/n^2$, it follows from Lemma 12 that $I + \delta_n\overline{Q}$ is an upper transition operator. We will use this implicitly throughout the proof.

Since $\overline{Q}$ is weakly ergodic, we know that for all $f \in \mathcal{L}$, there is some $\alpha_f \in \mathbb{R}$ such that

$$\lim_{t\to+\infty} \frac{1}{t} e^{t\overline{Q}_f}[0](x) = \alpha_f \text{ for all } x \in \mathcal{X}.$$

Due to Theorem 20, this implies that for all $f \in \mathcal{L}$,

$$\lim_{n\to+\infty} \frac{1}{n\varDelta}\left(I + \frac{\varDelta}{n^2}\overline{Q}_f\right)^{n^3}[0](x) = \alpha_f \text{ for all } x \in \mathcal{X}. \tag{8}$$

Consider now any closed communication class $C \subseteq \mathcal{X}$ of $\overline{Q}$ and any $\alpha \in \mathbb{R}$ and $f \in \mathcal{L}$ such that $f(x) = \alpha$ for all $x \in C$. Then for all $n \in \mathbb{N}$, due to Lemma 14, C is also a closed communication class for $I + \delta_n\overline{Q}$, and it therefore follows from Lemma 21 (with $h = \delta_n f$ and $g = \delta_n\alpha$) that, for all $k \in \mathbb{N}$ and $x \in C$,

$$\begin{aligned}(I + \delta_n\overline{Q}_f)^k[0](x) &= (I + \delta_n\overline{Q})^k_{\delta_n f}[0](x)\\ &= (I + \delta_n\overline{Q})^k_{\delta_n\alpha}[0](x) = k\delta_n\alpha,\end{aligned}$$

where for the last equality we used (R3) k times. In particular, for $k = n^3$, we find that

$$\frac{1}{n\varDelta}\left(I + \frac{\varDelta}{n^2}\overline{Q}_f\right)^{n^3}[0](x) = \frac{1}{n\varDelta}n^3\frac{\varDelta}{n^2}\alpha = \alpha$$

for all $n \in \mathbb{N}$ and $x \in C$. Using (8), it therefore follows that

$$\alpha_f = \alpha \text{ for all } x \in C. \tag{9}$$

Our proof will now be in two parts. In the first part, we prove that $\overline{Q}$ has a top class, and we will do so by means of a proof by contradiction – essentially using the same strategy as the one in [18, Proof of Proposition 12]. So assume *ex absurdo* that $\overline{Q}$ has no top class. Then by a classic graph-theoretic argument, there are (at least) two (disjoint) closed communication classes $\mathcal{S}_1$ and $\mathcal{S}_2$ – see for example [18, Corollary 26]. Fix some $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\alpha_1 \neq \alpha_2$, and let $f := \alpha_1\mathbb{I}_{\mathcal{S}_1} + \alpha_2\mathbb{I}_{\mathcal{S}_2}$. For $\ell \in \{1,2\}$, it then follows from (9) (with $C = \mathcal{S}_\ell$ and $\alpha = \alpha_\ell$) that $\alpha_f = \alpha_\ell$, contradicting the fact that $\alpha_1 \neq \alpha_2$.

In the second part of this proof, we establish that the top class $\mathcal{S}$ of $\overline{Q}$ is $\overline{Q}$-absorbing. Our proof will again be one by contradiction, similar in spirit to the proof of Proposition 13 in [18]. Assume *ex absurdo* that the top class $\mathcal{S}$ is not $\overline{Q}$-absorbing. Then for all $n \in \mathbb{N}$, we know from Lemma 22 that there is some non-empty subset A_{δ_n} of $\mathcal{S}^c$ such that $\mathbb{I}_{A_{\delta_n}} \leq (I + \delta_n\overline{Q})[\mathbb{I}_{A_{\delta_n}}]$.

Since $\mathcal{S}^c \subseteq \mathcal{X}$ is finite, there is some non-empty subset A of $\mathcal{S}^c$ that occurs infinitely often in the sequence $(A_{\delta_n})_{n\in\mathbb{N}}$. Let $(n_k)_{k\in\mathbb{N}}$ be the increasing sequence of indices $n \in \mathbb{N}$ such that $A_{\delta_n} = A$, and observe that by construction,

$$\mathbb{I}_A \leq (I + \delta_{n_k}\overline{Q})[\mathbb{I}_A] \quad \text{for all } k \in \mathbb{N}. \tag{10}$$

Fix some $k \in \mathbb{N}$. We set out to prove that

$$\ell\delta_{n_k}\mathbb{I}_A \leq (I + \delta_{n_k}\overline{Q}_{\mathbb{I}_A})^\ell[0] \quad \text{for all } \ell \in \mathbb{N}. \tag{11}$$

Due to (R3),

$$(I + \delta_{n_k}\overline{Q}_{\mathbb{I}_A})[0] = \delta_{n_k}\mathbb{I}_A + (I + \delta_{n_k}\overline{Q})[0] = \delta_{n_k}\mathbb{I}_A,$$

so (11) clearly holds for $\ell = 1$. For the inductive step, we assume that (11) holds for some $\ell \in \mathbb{N}$, and set out to show that it also holds for $\ell+1$. Observe that, by definition of $\overline{Q}_{\mathbb{I}_A}$,

$$(I + \delta_{n_k}\overline{Q}_{\mathbb{I}_A})^{\ell+1}[0] = \delta_{n_k}\mathbb{I}_A + (I + \delta_{n_k}\overline{Q})(I + \delta_{n_k}\overline{Q}_{\mathbb{I}_A})^\ell[0].$$

It follows from this, the induction hypothesis, (T6), (T2) and (10) that

$$\begin{aligned}(I + \delta_{n_k}\overline{Q}_{\mathbb{I}_A})^{\ell+1}[0] &\geq \delta_{n_k}\mathbb{I}_A + (I + \delta_{n_k}\overline{Q})[\ell\delta_{n_k}\mathbb{I}_A]\\ &= \delta_{n_k}\mathbb{I}_A + \ell\delta_{n_k}(I + \delta_{n_k}\overline{Q})[\mathbb{I}_A]\\ &\geq (\ell+1)\delta_{n_k}\mathbb{I}_A,\end{aligned}$$

as required. Applying (11) for $\ell = n_k^3$, we find that

$$\frac{1}{n_k\varDelta}\left(I + \frac{\varDelta}{n_k^2}\overline{Q}_{\mathbb{I}_A}\right)^{n_k^3}[0](x) \geq 1 \quad \text{for all } x \in A. \tag{12}$$

Since (12) is true for any $k \in \mathbb{N}$, and furthermore $A \neq \emptyset$ and $\lim_{k\to+\infty} n_k = +\infty$, it follows follows from (8) that $\alpha_{\mathbb{I}_A} \geq 1$.

On the other hand, since $A \subseteq \mathcal{S}^c$, we know that $\mathbb{I}_A(x) = 0$ for all $x \in \mathcal{S}$. Since the top class $\mathcal{S}$ is a closed communication class, it therefore follows from (9) (with $f = \mathbb{I}_A$, $\alpha = 0$ and $C = \mathcal{S}$) that $\alpha_{\mathbb{I}_A} = 0$, contradicting the fact that $\alpha_{\mathbb{I}_A} \geq 1$. ■

7. Conclusion

We introduced the notion of weak ergodicity for an imprecise jump process $\mathcal{P}$, and showed that in the case of Markovian imprecise jump processes defined by a bounded set $\mathcal{Q}$ of rate matrices with separately specified rows, this can be translated into the question whether the induced upper transition operator $\overline{Q}_{\mathcal{Q}}$ is weakly ergodic. Finally, we showed that an upper transition operator is weakly ergodic if and only if it is ergodic, or equivalently, if and only if it has a top class that is top class absorbing. This is similar to results in the discrete time setting, but with the crucial difference that in that setting weak ergodicity is in fact weaker than ergodicity.

In future work, we would like to see if we can extend to the imprecise setting the Point-Wise Ergodic Theorem that we started this paper with, mimicking similar such generalisations in a discrete time setting [5, 6].

Acknowledgments

Alexander Erreygers and Jasper De Bock are supported by the Research Foundation – Flanders (FWO) (project number 3G028919).

References

[1] Marianne Akian, Stéphane Gaubert, and Antoine Hochart. A game theory approach to the existence and uniqueness of nonlinear Perron–Frobenius eigenvectors. *Discrete and Continuous Dynamical Systems*, 40(1):207–231, 2020. doi: 10.3934/dcds.2020009.

[2] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011. doi: 10.1007/978-0-387-70914-7.

[3] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23(4):597—635, 2009. doi: 10.1017/S0269964809990039.

[4] Jasper De Bock. The limit behaviour of imprecise continuous-time Markov chains. *Journal of Nonlinear Science*, 27:159–196, 2017. doi: 10.1007/s00332-016-9328-3.

[5] Jasper De Bock and Natan T'Joens. Average behaviour of imprecise Markov chains: A single pointwise ergodic theorem for six different models. In *Proceedings of the Twelfth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'21)*, volume 147 of *Proceedings of Machine Learning Research*, pages 90–99. PMLR, 2021.

[6] G. de Cooman, J. De Bock, and S. Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76: 18–46, 2016. ISSN 0888-613X.

[7] Alexander Erreygers. *Markovian Imprecise Jump Processes: Foundations, Algorithms and Applications*. PhD thesis, Ghent University, 2021. Available at https://users.ugent.be/~aerreyge/.

[8] Alexander Erreygers and Jasper De Bock. Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'17)*, volume 62 of *Proceedings of Machine Learning Research*, pages 145–156. PMLR, 2017.

[9] Alexander Erreygers and Jasper De Bock. Markovian imprecise jump processes: Extension to measurable variables, convergence theorems and algorithms. *International Journal of Approximate Reasoning*, 147: 78–124, 2022. doi: 10.1016/j.ijar.2022.05.006.

[10] J. Gunawardena. From max-plus algebra to nonexpansive mappings: a nonlinear theory for discrete event systems. *Theoretical Computer Science*, 293(1): 141–167, 2003. ISSN 0304-3975.

[11] Filip Hermans and Gert de Cooman. Characterisation of ergodic upper transition operators. *International Journal of Approximate Reasoning*, 53(4):573–583, 2012. doi: 10.1016/j.ijar.2011.12.008.

[12] Marius Iosifescu. *Finite Markov Processes and their Applications*. Wiley, 1980.

[13] Thomas Krak. *Continuous-Time Imprecise-Markov Chains: Theory and Algorithms*. PhD thesis, Ghent University, 2021.

[14] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, pages 452–528, 2017. doi: 10.1016/j.ijar.2017.06.012.

[15] John D. Lambert. *Numerical Methods for Ordinary Differential Systems*. Wiley, 1991.

[16] James R. Norris. *Markov Chains*. Cambridge University Press, 1997. doi: 10.1017/CBO9780511810633.

[17] Daniel W. Stroock. *An Introduction to Markov Processes*. Springer, 2005. doi: 10.1007/b138428.

[18] Natan T'Joens and Jasper De Bock. Average behaviour in discrete-time imprecise Markov chains: A study of weak ergodicity. *International Journal of Approximate Reasoning*, 132:181–205, 2021. doi: 10.1016/j.ijar.2021.03.001.

[19] Damjan Škulj. 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.

Supplementary Material

Our proof for Theorem 20 builds on the following intermediary lemmata. In order not to unnecessarily repeat ourselves in this section, we fix some upper rate operator $\overline{Q}$ for the remainder. Furthermore, we let

$$D := \{\delta \in \mathbb{R}_{>0} : \delta\|\overline{Q}\| \leq 2\}$$

and for all $\delta \in D$, let $\overline{T}(\delta) := I + \delta\overline{Q}$; due to Lemma 12, $\overline{T}(\delta)$ is an upper transition operator whenever $\delta \in D$, and henceforth we will use this fact implicitly.

Lemma 23 *For all $\delta \in D$, $f, h \in \mathcal{L}$ and $n \in \mathbb{N}$,*

$$\|(I + \delta\overline{Q}_f)^n[h]\| \leq n\delta\|f\| + \|h\|. \tag{13}$$

Proof Let us prove the result by induction. For the base case $n = 1$, it follows from the definition of $\overline{Q}_f$ and (T7) that

$$\|(I + \delta\overline{Q}_f)[h]\| \leq \|\delta f\| + \|\overline{T}(\delta)[h]\| \leq \delta\|f\| + \|h\|,$$

as required. For the inductive step, we assume that (13) holds for $n = k$ with $k \in \mathbb{N}$, and set out to show that it then also holds for $n = k + 1$. From the definition of $\overline{Q}_f$, (T7) and the induction hypothesis, it follows immediately that

$$\begin{aligned}\|(I + \delta\overline{Q}_f)^{k+1}[h]\| &\leq \delta\|f\| + \|\overline{T}(\delta)(I + \delta\overline{Q}_f)^k[h]\| \\ &\leq \delta\|f\| + \|(I + \delta\overline{Q}_f)^k[h]\| \\ &\leq \delta(k + 1)\|f\| + \|h\|,\end{aligned}$$

as required. ■

The second intermediary lemma builds on Lemma 23.

Lemma 24 *Fix some $\delta \in D$ and $f, h \in \mathcal{L}$. Then for all $n \in \mathbb{N}$,*

$$\|(I + \delta\overline{Q}_f)^n[h] - h\| \leq n\delta c_1 + n^2\delta^2 c_2, \tag{14}$$

with $c_1 := \|f\| + \|\overline{Q}\|\|h\|$ and $c_2 := \|\overline{Q}\|\|f\|$.

Proof We again give a proof by induction. For the base case $n = 1$, note that

$$\begin{aligned}\|(I + \delta\overline{Q}_f)[h] - h\| &= \|\delta f + h + \delta\overline{Q}[h] - h\| \\ &\leq \delta\|f\| + \delta\|\overline{Q}\|\|h\| = \delta c_1,\end{aligned}$$

which implies the inequality in the statement for $n = 1$.

For the inductive step, we assume that (14) holds for $n = k$ with $k \in \mathbb{N}$, and set out to verify that it holds for $n = k + 1$ as well. Observe that

$$\begin{aligned}&(I + \delta\overline{Q}_f)^{k+1}[h] - h \\ &= \delta f + (I + \delta\overline{Q}_f)^k[h] - h + \delta\overline{Q}(I + \delta\overline{Q}_f)^k[h].\end{aligned}$$

Recall from (R5) that

$$\|\delta\overline{Q}(I + \delta\overline{Q}_f)^k[h]\| \leq \delta\|\overline{Q}\|\|(I + \delta\overline{Q}_f)^k[h]\|.$$

We infer from these two observations, the induction hypothesis and Lemma 23 that

$$\begin{aligned}&\|(I + \delta\overline{Q}_f)^{k+1}[h] - h\| \\ &\quad\leq \delta\|f\| + (k\delta c_1 + k^2\delta^2 c_2) + \delta\|\overline{Q}\|(k\delta\|f\| + \|h\|) \\ &\quad= (k + 1)\delta c_1 + k^2\delta^2 c_2 + k\delta^2 c_2.\end{aligned}$$

Since $k^2 + k \leq (k + 1)^2$, we infer from this that

$$\|(I + \delta\overline{Q}_f)^{k+1}[h] - h\| \leq (k + 1)\delta c_1 + (k + 1)^2\delta^2 c_2,$$

which is the inequality we were after. ■

Our next step is to use Lemma 24 to prove a 'generalisation' of Lemma E.5 in [14]. In this result, we need the fact that $\overline{Q}$ is Lipschitz:

R7. $\|\overline{Q}[f] - \overline{Q}[g]\| \leq \|\overline{Q}\|\|f - g\|$ for all $f, g \in \mathcal{L}$;

this is trivial if $\|\overline{Q}\| = 0$ and follows from Lemma 12 (with $\Delta = 2/\|\overline{Q}\|$) and (T8) (for $I + \Delta\overline{Q}$) otherwise, see also [4, R11] or [7, LR8].

Lemma 25 *Fix some $\delta \in D$ and $f, h \in \mathcal{L}$. Then for all $n \in \mathbb{N}$,*

$$\|(I + \delta\overline{Q}_f)^n[h] - (I + n\delta\overline{Q}_f)[h]\| \leq n^2\delta^2 c_3 + n^3\delta^3 c_4,$$

with $c_3 := \|\overline{Q}\|\|f\| + \|\overline{Q}\|^2\|h\|$ and $c_4 := \|\overline{Q}\|^2\|f\|$.

Proof Our proof will be one by induction. The base case $n = 1$ is trivially satisfied. For the inductive step, we assume that the inequality in the statement holds for $n = k$ with $k \in \mathbb{N}$. To prove that the inequality in the statement holds for $n = k + 1$, we observe that

$$\begin{aligned}&(I + \delta\overline{Q}_f)^{k+1}[h] - (I + (k + 1)\delta\overline{Q}_f)[h] \\ &\quad= \delta f + (I + \delta\overline{Q}_f)^k[h] - (I + k\delta\overline{Q}_f)[h] \\ &\qquad\quad - \delta f - \delta\overline{Q}[h] + \delta\overline{Q}(I + \delta\overline{Q}_f)^k[h].\end{aligned}$$

It follows from this, the induction hypothesis, (R7) and Lemma 24 that

$$\begin{aligned}&\|(I + \delta\overline{Q}_f)^{k+1}[h] - (I + (k + 1)\delta\overline{Q}_f)[h]\| \\ &\quad\leq (k^2\delta^2 c_3 + k^3\delta^3 c_4) \\ &\qquad\quad + \delta\|\overline{Q}\|\|(I + \delta\overline{Q}_f)^k[h] - h\| \\ &\quad\leq (k^2\delta^2 c_3 + k^3\delta^3 c_4) + \delta\|\overline{Q}\|(k\delta c_1 + k^2\delta^2 c_2) \\ &\quad= (k^2 + k)\delta^2 c_3 + (k^3 + k^2)\delta^3 c_4\end{aligned}$$

$$\leq (k+1)^2\delta^2 c_3 + (k+1)^3\delta^3 c_4,$$

which is the inequality we were after. ∎

As a final intermediary step, we generalise Lemma 25; this result is to Lemma 25 what Lemma E.6 is to Lemma E.5 in [14].

Lemma 26 *Fix some $\delta \in D$, $f, h \in \mathcal{L}$ and $k \in \mathbb{N}$. Then for all $n \in \mathbb{N}$,*

$$\left\| \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{nk}[h] - (I + \delta\overline{Q}_f)^n[h] \right\| \leq n\delta^2 c_3 + n^2\delta^3 c_4,$$

with c_3 and c_4 as in Lemma 25.

Proof Let us prove the result by induction. For the base case $n = 1$, we apply Lemma 25 (with $\delta/k \in D$ here as δ there and k here as n there) to find that

$$\|(I + \frac{\delta}{k}\overline{Q}_f)^k[h] - (I + k\frac{\delta}{k}\overline{Q}_f)[h]\| \leq k^2\left(\frac{\delta}{k}\right)^2 c_3 + k^3\left(\frac{\delta}{k}\right)^3 c_4 = \delta^2 c_3 + \delta^3 c_4.$$

For the inductive step, we assume that the inequality in the statement holds for $n = \ell$ with $\ell \in \mathbb{N}$, and set out to establish the inequality in the statement for $n = \ell + 1$. Observe that

$$\begin{aligned}
&\left(I + \frac{\delta}{k}\overline{Q}_f\right)^{(\ell+1)k}[h] - (I + \delta\overline{Q}_f)^{\ell+1}[h] \\
&= \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{k}\left(I + \frac{\delta}{k}\overline{Q}_f\right)^{\ell k}[h] \\
&\quad - \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{k}(I + \delta\overline{Q}_f)^{\ell}[h] \\
&\quad + \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{k}(I + \delta\overline{Q}_f)^{\ell}[h] \\
&\quad - (I + \delta\overline{Q}_f)(I + \delta\overline{Q}_f)^{\ell}[h].
\end{aligned}$$

Let us denote the norm of the first two terms on the right hand side by $\eta_{1:2}$ and that of the last two terms by $\eta_{3:4}$, such that

$$\left\| \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{(\ell+1)k}[h] - (I + \delta\overline{Q}_f)^{\ell+1}[h] \right\| \leq \eta_{1:2} + \eta_{3:4}.$$

Since $\overline{T}(\delta/k)$ satisfies (T8) because $\delta/k \in D$, the same is true for $\overline{T}(\delta/k)_{\delta f/k}$ – we leave this for the reader to check – and therefore also for $\overline{T}(\delta/k)^k_{\delta f/k} = \left(I + \frac{\delta}{k}\overline{Q}_f\right)^k$; consequently,

$$\begin{aligned}
\eta_{1:2} &\leq \left\| \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{\ell k}[h] - (I + \delta\overline{Q}_f)^{\ell}[h] \right\| \\
&\leq \ell\delta^2 c_3 + \ell^2\delta^3 c_4,
\end{aligned}$$

where the second inequality is exactly the induction hypothesis. Moreover, it follows from Lemma 25 (with $(I + \delta\overline{Q}_f)^\ell[h]$ here as h there, k here as n there and $\delta/k \in D$ here as δ there) and Lemma 23 (with ℓ here as n there) that

$$\begin{aligned}
\eta_{3:4} &\leq \delta^2\left(\|\overline{Q}\|\|f\| + \|\overline{Q}\|^2\|(I + \delta\overline{Q}_f)^\ell[h]\|\right) + \delta^3 c_4 \\
&\leq \delta^2\left(\|\overline{Q}\|\|f\| + \|\overline{Q}\|^2\ell\delta\|f\| + \|\overline{Q}\|^2\|h\|\right) + \delta^3 c_4 \\
&= \delta^2 c_3 + \ell\delta^3 c_4 + \delta^3 c_4.
\end{aligned}$$

Combining our observations, we find that

$$\begin{aligned}
&\left\| \left(I + \frac{\delta}{k}\overline{Q}_f\right)^{(\ell+1)k}[h] - (I + \delta\overline{Q}_f)^{\ell+1}[h] \right\| \\
&\leq \ell\delta^2 c_3 + \ell^2\delta^3 c_4 + \delta^2 c_3 + \ell\delta^3 c_4 + \delta^3 c_4 \\
&= (\ell+1)\delta^2 c_3 + (\ell^2 + \ell + 1)\delta^3 c_4 \\
&\leq (\ell+1)\delta^2 c_3 + (\ell+1)^2\delta^3 c_4,
\end{aligned}$$

which is the inequality we were after. ∎

Proving Theorem 7 is now simply a matter of combining (6) and Lemma 26.

Proof of Theorem 20 Fix some $n \in \mathbb{N}$. Then for all $k \in \mathbb{N}$

$$\begin{aligned}
&e^{n\varDelta\overline{Q}_f}[h] - \left(I + \frac{\varDelta}{n^2}\overline{Q}_f\right)^{n^3}[h] \\
&= e^{n\varDelta\overline{Q}_f}[h] - \left(I + \frac{n\varDelta}{kn^3}\overline{Q}_f\right)^{kn^3}[h] \\
&\quad + \left(I + \frac{n\varDelta}{kn^3}\overline{Q}_f\right)^{kn^3}[h] - \left(I + \frac{\varDelta}{n^2}\overline{Q}_f\right)^{n^3}[h]
\end{aligned}$$

From (6) with $t = \varDelta n$, we know that

$$\begin{aligned}
e^{n\varDelta\overline{Q}_f}[h] &= \lim_{k\to+\infty}\left(I + \frac{n\varDelta}{k}\overline{Q}_f\right)^{k}[h] \\
&= \lim_{k\to+\infty}\left(I + \frac{n\varDelta}{kn^3}\overline{Q}_f\right)^{kn^3}[h].
\end{aligned}$$

Furthermore, if $\varDelta\|\overline{Q}\| \leq 2n^2$, it follows from Lemma 26 (with $\delta = \varDelta/n^2$ and n^3 here as n there) that for all $k \in \mathbb{N}$,

$$\begin{aligned}
&\left\| \left(I + \frac{n\varDelta}{kn^3}\overline{Q}_f\right)^{kn^3}[h] - \left(I + \frac{\varDelta}{n^2}\overline{Q}_f\right)^{n^3}[h] \right\| \\
&\leq n^3\left(\frac{\varDelta}{n^2}\right)^2 c_3 + n^6\left(\frac{\varDelta}{n^2}\right)^3 c_4 \\
&= \frac{1}{n}\varDelta^2 c_3 + \varDelta^3 c_4.
\end{aligned}$$

Combining the preceding and taking the limit for $k \to +\infty$ gives that, for all $n \in \mathbb{N}$ such that $\Delta\|\overline{Q}\| \leq 2n^2$,

$$\frac{1}{n\Delta}\left\| e^{n\Delta\overline{Q}_f}[h] - \left(I + \frac{\Delta}{n^2}\overline{Q}_f\right)^{n^3}[h]\right\| \leq \frac{1}{n^2}\Delta c_3 + \frac{1}{n}\Delta^2 c_4.$$

The right hand side of this inequality vanishes as $n \to +\infty$, which implies the statement. ∎