# Imprecise continuous-time Markov chains

Thomas Krak [a,*], Jasper De Bock [b], Arno Siebes [a]

[a] *Utrecht University, Department of Information and Computing Sciences, Princetonplein 5, De Uithof, 3584 CC Utrecht, The Netherlands*
[b] *Ghent University, Department of Electronics and Information Systems, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium*

A B S T R A C T

Continuous-time Markov chains are mathematical models that are used to describe the state-evolution of dynamical systems under stochastic uncertainty, and have found widespread applications in various fields. In order to make these models computationally tractable, they rely on a number of assumptions that—as is well known—may not be realistic for the domain of application; in particular, the ability to provide exact numerical parameter assessments, and the applicability of time-homogeneity and the eponymous Markov property. In this work, we extend these models to *imprecise continuous-time Markov chains* (ICTMC's), which are a robust generalisation that relaxes these assumptions while remaining computationally tractable.

More technically, an ICTMC is a *set* of "precise" continuous-time finite-state stochastic processes, and rather than computing expected values of functions, we seek to compute *lower expectations*, which are tight lower bounds on the expectations that correspond to such a set of "precise" models. Note that, in contrast to e.g. Bayesian methods, all the elements of such a set are treated on equal grounds; we do not consider a distribution over this set. Together with the conjugate notion of *upper expectation*, the bounds that we provide can then be intuitively interpreted as providing best- and worst-case scenarios with respect to all the models in our set of stochastic processes.

The first part of this paper develops a formalism for describing continuous-time finite-state stochastic processes that does not require the aforementioned simplifying assumptions. Next, this formalism is used to characterise ICTMC's and to investigate their properties. The concept of lower expectation is then given an alternative operator-theoretic characterisation, by means of a *lower transition operator*, and the properties of this operator are investigated as well. Finally, we use this lower transition operator to derive tractable algorithms (with polynomial runtime complexity w.r.t. the maximum numerical error) for computing the lower expectation of functions that depend on the state at any finite number of time points.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Continuous-time Markov chains are mathematical models that can describe the behaviour of dynamical systems under stochastic uncertainty. In particular, they describe the stochastic evolution of such a system through a discrete state space and over a continuous time-dimension. This class of models has found widespread applications in various fields, includ-

ing queueing theory [1,2], mathematical finance [3–5], epidemiology [6–8], system reliability analysis [9–11], and many others [12].

In order to model a problem by means of such a continuous-time Markov chain, quite a lot of assumptions need to be satisfied. For example, it is common practice to assume that the user is able to specify an exact value for all the parameters of the model. A second important assumption is the Markov condition, which states that the future behaviour of the system only depends on its current state, and not on its history. Other examples are homogeneity, which assumes that the dynamics of the system are independent of time, and some technical differentiability assumptions. As a result of all these assumptions, continuous-time Markov chains can be described by means of simple analytic expressions.

However, we would argue that in many cases, these assumptions are not realistic and are grounded more in pragmatism than in informed consideration of the underlying system. In those cases, despite the fact that such issues are to be expected in any modelling task, we think that it is best to try and avoid these assumptions. Of course, since they are typically imposed to obtain a tractable model, relaxing these assumptions while maintaining a workable model is not straightforward. Nevertheless, as we will see, this can be achieved by means of *imprecise continuous-time Markov chains* [13,14] (ICTMC's). These ICMTC's are quite similar to continuous-time Markov chains: they model the same type of dynamical systems, and they therefore have the same fields of application. However, they do not impose the many simplifying assumptions that are traditionally adopted, and are therefore far more robust. Notably, as we will show in this work, these models allow us to relax these assumptions while remaining computationally tractable.

The following provides a motivating toy example. It is clearly too simple to be of any practical use, but it does allow us to illustrate the simplifying assumptions that are usually adopted, and to provide a basic idea of how we intend to relax them.

**Example 1.1.** Consider a person periodically becoming sick, and recovering after some time. If we want to model this behaviour using a continuous-time Markov chain with a binary state space {healthy, sick}, we need to specify a *rate parameter* for each of the two possible state-transitions: from healthy to sick, and from sick to healthy. Loosely speaking, such a rate parameter characterises how quickly the corresponding state-transition happens. Technically, it specifies the derivative of a transition probability. For example, for the transition from healthy to sick, the corresponding rate parameter is the derivative of $P(X_s = \text{sick} | X_t = \text{healthy})$ with respect to $s$, for $s = t$, where $P(X_s = \text{sick} | X_t = \text{healthy})$ is the probability of a person being sick at time $s$, given that he or she is healthy at time $t$. Together with the initial probabilities of a person being sick or healthy at time zero, these rate parameters uniquely characterise a continuous-time Markov chain, which can then be used to answer various probabilistic queries of interest. For instance, to compute the probability that a person will be sick in ten days, given that he or she is healthy today.

In this example, the defining assumptions of a continuous-time Markov chain impose rather strict conditions on our model. First of all: it is necessary to provide exact values—that is, point-estimates—for the initial probabilities and for the rate parameters. If these values are incorrect, this will affect the resulting conclusions. Secondly, in order to be able to define the rate parameters, the transition probabilities of the model need to be differentiable. Thirdly, the Markov assumption implies that for any time points $r < t < s$,

$$P(X_s = \text{sick} | X_t = \text{healthy}, X_r = \text{sick}) = P(X_s = \text{sick} | X_t = \text{healthy}),$$

that is, once we know the person's health state at time $t$, his or her probability of being sick at time $s$ does not depend on whether he or she has ever been sick before time $t$. If it is possible to develop immunity to the disease in question, then clearly, such an assumption is not realistic. Also, it implies that the rate parameters can only depend on the current state, and not on the previous ones. Fourthly, the rate parameters are assumed to remain constant over time, which, for example, excludes the possibility of modelling seasonal variations. This fourth condition can easily be removed by considering a continuous-time Markov chain that is not homogeneous. However, in that case, the rate parameters become time-dependent, which requires us to specify (or learn from data) even more point-estimates. Similarly, although a more complex model would be able to account for history-dependence, for example by adding more states to the model, this would vastly increase the computational complexity of working with the model.

In the imprecise continuous-time Markov chains that we consider, these four conditions are relaxed in the following way. First of all, instead of providing a point-estimate for the initial probabilities and the rate parameters, we allow ourselves to specify a set of values. For the purpose of this example, we can take these sets to be intervals. The other three conditions are then dropped completely, provided that they remain compatible with these intervals. For example, the rate parameters do not need to be constant, but can vary in time in an arbitrary way, as long as they remain within their interval. Similarly, the rate parameters are also allowed to depend on the history of the process, that is, the value of the previous states. In fact, the rate parameters—as defined above—do not even need to exist, since we do not require differentiability either. ◇

The idea of relaxing the defining assumptions of a continuous-time Markov chain is not new. Various variations on it have been developed over the years. For example, there are plenty of results to be found that deal with non-homogeneous continuous-time Markov chains [15–17]. Dropping the Markov condition is less common, but nevertheless definitely possible [18]. However, the approaches that drop this Markov assumption will typically replace it by some other, weaker assumption, instead of dropping it altogether.

A common property of all of these approaches is that they still require the parameters of the model to be specified exactly. Furthermore, since these models are typically more complex, the number of parameters that needs to be specified is a lot larger than before. Therefore, in practice, specifying such a generalised model is a lot more difficult than specifying a normal, homogeneous continuous-time Markov chain. In contrast, our approach does not introduce a large number of new parameters, but simply considers imprecise versions of the existing parameters. In particular, we provide constraints on the traditional parameters of a continuous-time Markov chain, by requiring them to belong to some specified set of candidate parameters. In this sense, our approach can be regarded as a type of sensitivity analysis on the parameters of a continuous-time Markov chain. However, instead of simply varying the parameters of a continuous-time Markov chain, as a more traditional sensitivity analysis would do, we also consider what happens when these parameters are allowed to be time- and history-dependent. Furthermore, a sensitivity analysis is typically interested in the effect of infinitesimal parameter variations, whereas we consider the effect of variations within some freely chosen set of candidate parameters.

Our approach should also not be confused with Bayesian methods [19]. Although these methods also consider parameter uncertainty, they model this uncertainty by means of a prior distribution, thereby introducing even more (hyper)parameters. Furthermore, when integrating out this prior, a Bayesian method ends up with a single 'averaged' stochastic process. In contrast, our approach considers set-valued parameter assessments, and does not provide a prior distribution over these sets. Every possible combination of parameter values gives rise to a different process, and we treat all of these processes on equal grounds, without averaging them out.

Among the many extensions of continuous-time Markov chains that are able to deal with complex parameter variations, the ones that resemble our approach the most are continuous-time Markov decision processes [20] and continuous-time controlled Markov chains [21]. Similar to what we do, these models vary the parameters of a continuous-time Markov chain, and allow these variations to depend on the history of the process. However, their variations are more restrictive, because the parameters are assumed to remain constant in between two transitions, whereas we allow them to vary in more arbitrary ways. The most important difference with our approach though, is that in these models, the parameter changes are not at all uncertain. In fact, the parameters can be chosen freely, and the goal is to control the evolution of the process in some optimal way, by tuning its parameters as the process evolves.

Having situated our topic within the related literature, let us now take a closer look at what we actually mean by an imprecise continuous-time Markov chain. In order to formalise this concept, we turn in this work to the field of *imprecise probability* [22–24]. The basic premise of this field is that, whenever it is impossible or unrealistic to specify a single probabilistic model, say $P$, it is better to instead consider a *set* of probabilistic models $\mathscr{P}$, and to then draw conclusions that are robust with respect to variations in this set. In the particular case of an imprecise continuous-time Markov chain, $\mathscr{P}$ will be a set of stochastic processes. Some of these processes are homogeneous continuous-time Markov chains. However, the majority of them are not. As explained in Example 1.1, we also consider other, more general stochastic processes, which are not required to be homogeneous, do not need to satisfy the Markov condition, and do not even need to be differentiable.

From a practical point of view, once a probabilistic model has been formulated and its parameters have been specified, one is typically interested in computing inferences, such as the probability of some event, or the expectation of some function. For example, as in Example 1.1, we might want to know the probability that a person will be sick in ten days, given that he or she is healthy today. Similarly, for expectations, one might for example like to know the expected utility of some financial strategy [5], the expected time until some component in a system breaks down [9] or the expected speed at which the clinical symptoms of a disease develop [6]. However, these inferences might depend crucially on the estimated values of the model parameters, and on the defining assumptions of the model. It is of interest, therefore, to study the robustness of such inferences with respect to parameter changes and relaxations of the defining assumptions of the model.

In our approach, we investigate this type of robustness as follows: we simply consider the probability or expectation of interest for each of the stochastic processes in $\mathscr{P}$, and then report the corresponding lower and upper bound. Intuitively, this can be regarded as providing best- and worst-case scenarios with respect to all the models in $\mathscr{P}$. Of course, since the stochastic processes in $\mathscr{P}$ are not required to satisfy the Markov condition, a naive optimisation method will be highly inefficient, and in most cases not even possible. However, as we will see, it is possible to develop other, more efficient methods for computing the lower and upper bounds that we are after. If the lower and upper bounds that we obtain are similar, we can conclude that the corresponding inference is robust with respect to the variations that are represented by $\mathscr{P}$. If the lower and upper bound are substantially different, then the inference is clearly sensitive to these variations, and policy may then have to be adapted accordingly.

Readers that are familiar with the literature on imprecise continuous-time Markov chains [13,14] should recognise the main ideas behind the approach that we have just described, but will also notice that our presentation differs from the one that is adopted in References [13] and [14]. Indeed, the seminal work by Škulj [13], which provided the first—and so far only—theoretical study of imprecise continuous-time Markov chains, characterised these models by means of the conditional lower and upper expectations of functions that depend on a single time point. In particular, this work defined such lower and upper expectations directly, through a generalisation of the well-known differential equation characterisation of "normal" continuous-time Markov chains. This approach allowed the author to focus on developing algorithms for solving this generalised differential equation, thereby yielding methods for the efficient computation of lower and upper expectations of functions that depend on a single time point. These results have since been successfully applied to conduct a robust analysis of failure-rates and repair-times in power-grid networks [14].

However, due to its direct characterisation of lower and upper expectations, this pioneering work left open a number of questions about which sets of stochastic processes these quantities correspond to; for instance, the question of whether or not these sets also include non-Markov processes. Furthermore, due to the focus on functions that depend on a single time point, computational methods for more general functions do not follow straightforwardly from this earlier work.

In contrast, we will in this present paper address these issues directly, by characterising imprecise continuous-time Markov chains explicitly as sets of stochastic processes. There are a number of advantages to working with such sets of processes. First of all, it removes any ambiguity as to the elements of such a set, which allows us to state exactly which assumptions the model robustifies against. Secondly, this approach allows us to prove some properties of such a set's corresponding lower and upper expectations which, in turn, allow us to derive tractable algorithms to compute lower and upper expectations of functions that depend on any finite number of time points. Finally, our approach allows us to derive algorithms that, for the special case of functions that depend on a single time point, improve upon the computational complexity of the algorithm in Reference [13].

In summary then, our aims with the present paper are threefold. First of all, to solidify the theoretical foundations of imprecise continuous-time Markov chains. Secondly, to extend and generalise existing methods for computing the corresponding lower expectations and, as a particular case, upper expectations and lower and upper probabilities. Thirdly, to provide analytical tools that can be used for future analysis of these models, and for the development of new algorithms.

Our main contributions can be summarised as follows.

1. We provide a unified framework for describing finite-state stochastic processes in continuous time, using the formalism of *full conditional probabilities*. This framework covers the full range of (non-)homogeneous, (non-)Markovian, and (non-)differentiable stochastic processes.
2. We use this framework to formalise *imprecise continuous-time Markov chains*: sets of stochastic processes that are in a specific sense consistent with user-specified set-valued assessments of the parameters of a continuous-time Markov chain. We conduct a thorough theoretical study of the properties of these sets of processes, and of the lower expectations that correspond to them.
3. We introduce a *lower transition operator* for imprecise continuous-time Markov chains, and show that this operator satisfies convenient algebraic properties such as homogeneity, differentiability, and Markovian-like factorisation—even if the underlying set of processes does not. Furthermore, and perhaps most importantly, we show that we can use this operator to *compute lower expectations* of functions that depend on the state $X_t$ at an arbitrary but finite number of time points.

To be upfront, we would like to conclude this introduction with a cautionary remark to practitioners: the main (computational) methods that we present do not enforce the homogeneity and—in some cases also—the Markovian independence assumptions of a traditional Markov chain, but allow these to be relaxed. Therefore, if one is convinced that the (true) system that is being modelled *does* satisfy these properties, then it is in general quite likely that the lower- and upper bounds that are reported by our methodology will be conservative. This is not to say that the methods that we present lead to vacuous or non-informative conclusions—see, e.g., Reference [25] for a successful application of our approach in telecommunication. However, tighter bounds—i.e. more informative conclusions—might then be obtainable by methodologies that do enforce these properties, provided of course that such methods are available and tractable. If one is uncertain, however, about whether these properties hold for one's system of interest, then the methods that we present are exactly applicable, and the bounds that we derive will be tight with respect to this uncertainty.

## 1.1. Finding your way around in this paper

Given the substantial length of this paper, we provide in this section some suggestions as to what readers with various interests might wish to focus on. In principle, however, the paper is written to be read in chronological order, and is organised as follows.

First, after we introduce our notation in Section 2 and discuss some basic mathematical concepts that will be used throughout, Section 3 discusses some crucial algebraic notions that will allow us to describe stochastic processes. Section 4 then goes on to formally introduce stochastic processes and provides some powerful tools for describing their dynamics.

Next, once we have all our mathematical machinery in place, we shift the focus to imprecise continuous-time Markov chains. We start in Section 5 by considering the special case of—precise—continuous-time Markov chains and then, in Section 6, we finally formalise the imprecise version that is the topic of this paper, and prove some powerful theoretical properties of this model.

The next three sections discuss computational methods: Section 7 introduces a lower transition operator for imprecise continuous-time Markov chains, and in Sections 8 and 9, we use this operator to compute lower expectations of functions that depend on the state at an arbitrary but finite number of time points.

The last two sections provide some additional context: Section 10 relates and compares our results to previous work on imprecise continuous-time Markov chains, and in Section 11, we conclude the paper and provide some ideas for future work.

**Table 1**
Glossary table of important notation and terminology.

| Symbol or term | Explanation | Page |
|---|---|---|
| $u$ | finite, ordered sequence of time points | 457 |
| $\mathscr{U}$ | set of all finite, ordered sequences of time points | 457 |
| $\mathscr{U}_{<t}$ | set of all finite, ordered sequences of time points before time $t$ | 457 |
| $\mathscr{U}_{[t,s]}$ | set of all finite, ordered sequences of time points that partition the time interval $[t,s]$ | 457 |
| $\mathscr{X}$ | finite state space of a stochastic process | 457 |
| $\mathscr{X}_u$ | joint state space at time points $u \in \mathscr{U}$ | 457 |
| $\mathscr{L}(\mathscr{X})$ | set of all functions $f : \mathscr{X} \to \mathbb{R}$ | 457 |
| $\mathscr{L}(\mathscr{X}_u)$ | set of all functions $f : \mathscr{X}_u \to \mathbb{R}$ | 457 |
| $\lVert \cdot \rVert$ | maximum ($L_\infty$) norm of a function, operator, or set of matrices | 457 |
| $\mathscr{T}$ | transition matrix system: a family of transition matrices satisfying certain properties | 459 |
| $\mathscr{T}_Q$ | exponential transition matrix system that corresponds to the transition rate matrix $Q$ | 460 |
| $\mathscr{T}_P$ | transition matrix system that corresponds to the stochastic process $P$ | 467 |
| $\mathscr{T}^{\mathbf{I}}$ | restriction of the transition matrix system $\mathscr{T}$ to the closed time interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$ | 460 |
| $\otimes$ | concatenation operator for two restricted transition matrix systems | 460 |
| well-behaved | loosely speaking, this means "with bounded rate-of-change"; see Definitions 3.4 and 4.4 | 460, 466 |
| $\mathbb{P}$ | set of all continuous-time stochastic processes | 465 |
| $\mathbb{P}^{\mathrm{W}}$ | set of all well-behaved continuous-time stochastic processes | 466 |
| $\mathbb{P}^{\mathrm{WM}}$ | set of all well-behaved continuous-time Markov chains | 469 |
| $\mathbb{P}^{\mathrm{WHM}}$ | set of all well-behaved, homogeneous continuous-time Markov chains | 470 |
| $\overline{\partial}_+ T_{t,x_u}^t,\ \overline{\partial}_- T_{t,x_u}^t,\ \overline{\partial} T_{t,x_u}^t$ | outer partial derivatives of the transition matrix of a stochastic process, for time $t$ and history $x_u$ | 468 |
| $\mathscr{R}$ | set of all transition rate matrices | 458 |
| $\mathscr{Q}$ | set of transition rate matrices | 459 |
| $\mathscr{M}$ | initial set of probability mass functions | 459 |
| $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}},\ \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}},\ \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ | three different types of imprecise continuous-time Markov chains; also see Definition 6.3 and $\mathbb{P}^{\mathrm{W}}$, $\mathbb{P}^{\mathrm{WM}}$ and $\mathbb{P}^{\mathrm{WHM}}$ | 472 |
| $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}},\ \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}},\ \underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ | lower expectation operators of the imprecise continuous-time Markov chains $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ | 474 |
| $\underline{Q}$ | lower transition rate operator | 478 |
| $\underline{L}_t^s$ | lower transition operator from time point $t$ to $s$, corresponding to some given $\underline{Q}$ | 481 |
| $\underline{\mathscr{T}}_{\underline{Q}}$ | family of lower transition operators $\underline{L}_t^s$ corresponding to some given $\underline{Q}$ | 481 |

Finally, the proofs of our results are gathered in an appendix, where they are organised by section and ordered by chronological appearance. This appendix also contains some additional lemmas that are needed only in the proofs, but which may nevertheless be of independent interest. The appendix of the extended online version of this paper [26] also provides some extra material, including proofs for some of the claims in our examples and some basic information about the gambling interpretation of coherent conditional probabilities.

For readers with different interests, then, we recommend to focus on the following parts of the paper. First, if one is interested in the full mathematical theory, we strongly encourage the reader to go through the paper in chronological order. However, if one has a more passing interest in the mathematical formalism, and is content with a more conceptual understanding of what we mean by an imprecise continuous-time Markov chain, then he or she may wish to start in either Section 5 or 6. Finally, readers who are mainly interested in how to compute lower expectations for ICTMC's might want to focus on Sections 8 and 9, referring back to Section 6.3—for details about the specific types of ICTMC's that we consider—and Section 7.2—for the requisite details about lower transition rate operators.

Of course, skipping parts of the paper may introduce some ambiguity about the meaning and definition of the many technical concepts that we refer to throughout. In order to alleviate this as much as possible, Table 1 provides a short glossary of the most important notation and terminology.

## 2. Preliminaries

We denote the reals as $\mathbb{R}$, the non-negative reals as $\mathbb{R}_{\geq 0}$, the positive reals as $\mathbb{R}_{>0}$ and the negative reals as $\mathbb{R}_{<0}$. For any $c \in \mathbb{R}$, $\mathbb{R}_{\geq c}$, $\mathbb{R}_{>c}$ and $\mathbb{R}_{<c}$ have a similar meaning. The natural numbers are denoted by $\mathbb{N}$, and we also define $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$. The rationals will be denoted by $\mathbb{Q}$.

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

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

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

### 2.1. Sequences of time points

We will make extensive use of finite sequences of time points. Such a sequence is of the form $u \coloneqq t_0, t_1, \ldots, t_n$, with $n \in \mathbb{N}_0$ and, for all $i \in \{0, \ldots, n\}$, $t_i \in \mathbb{R}_{\geq 0}$. These sequences are taken to be ordered, meaning that for all $i, j \in \{0, \ldots, n\}$ with

$i < j$, it holds that $t_i \leq t_j$. Let $\mathscr{U}$ denote the set of all such finite sequences that are *non-degenerate*, meaning that for all $u \in \mathscr{U}$ with $u = t_0, \ldots, t_n$, it holds that $t_i \neq t_j$ for all $i, j \in \{0, \ldots, n\}$ such that $i \neq j$. Note that this does not prohibit *empty* sequences. We therefore also define $\mathscr{U}_\emptyset := \mathscr{U} \setminus \{\emptyset\}$.

For any finite sequence $u$ of time points, let $\max u := \max\{t_i : i \in \{0, \ldots, n\}\}$. For any time point $t \in \mathbb{R}_{\geq 0}$, we then write $t > u$ if $t > \max u$, and similarly for other inequalities. If $u = \emptyset$, then $t > u$ is taken to be trivially true, regardless of the value of $t$. We use $\mathscr{U}_{<t}$ to denote the subset of $\mathscr{U}$ that consists of those sequences $u \in \mathscr{U}$ for which $u < t$, and, again, similarly for other inequalities.

Since a sequence $u \in \mathscr{U}$ is a subset of $\mathbb{R}_{\geq 0}$, we can use set-theoretic notation to operate on such sequences. The result of such operations is again taken to be ordered. For example, for any $u, v \in \mathscr{U}$, we use $u \cup v$ to denote the ordered union of $u$ and $v$. Similarly, for any $s \in \mathbb{R}_{\geq 0}$ and any $u \in \mathscr{U}_{<s}$ with $u = t_0, \ldots, t_n$, we use $u \cup \{s\}$ to denote the sequence $t_0, \ldots, t_n, s$.

As a special case, we consider finite sequences of time points that partition a given time interval $[t, s]$, with $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$. Such a sequence is taken to include the end-points of this interval. Thus, the sequence is of the form $t = t_0 < t_1 < \cdots < t_n = s$. We denote the set of all such sequences by $\mathscr{U}_{[t,s]}$. Since these sequences are non-degenerate, it follows that $\mathscr{U}_{[t,t]}$ consists of a single sequence $u = t$. For any $u \in \mathscr{U}_{[t,s]}$ with $u = t_0, \ldots, t_n$, we also define the sequential differences $\Delta_i := t_i - t_{i-1}$, for all $i \in \{1, \ldots, n\}$. We then use $\sigma(u) := \max\{\Delta_i : i \in \{1, \ldots, n\}\}$ to denote the maximum such difference.

## 2.2. States and functions

Throughout this work, we will consider some fixed finite *state space* $\mathscr{X}$. A generic element of this set is called a state and will be denoted by $x$. Without loss of generality, we can assume the states to be ordered, and we can then identify $\mathscr{X}$ with the set $\{1, \ldots, |\mathscr{X}|\}$, where $|\mathscr{X}|$ is the number of states in $\mathscr{X}$.

We use $\mathscr{L}(\mathscr{X})$ to denote the set of all real-valued functions on $\mathscr{X}$. Because $\mathscr{X}$ is finite, a function $f \in \mathscr{L}(\mathscr{X})$ can be interpreted as a vector in $\mathbb{R}^{|\mathscr{X}|}$. Hence, we will in the sequel use the terms 'function' and 'vector' interchangeably when referring to elements of $\mathscr{L}(\mathscr{X})$.

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

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

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

## 2.3. Norms and operators

For any $u \in \mathscr{U}_\emptyset$ and any $f \in \mathscr{L}(\mathscr{X}_u)$, let the norm $\|f\|$ be defined as

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

As a special case, we then have for any $f \in \mathscr{L}(\mathscr{X})$ that $\|f\| = \max\{|f(x)| : x \in \mathscr{X}\}$.

Linear maps from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ will play an important role in this work, and will be represented by matrices. Because the state space $\mathscr{X}$ is fixed throughout, we will always consider square, $|\mathscr{X}| \times |\mathscr{X}|$, real-valued matrices.

As such, we will for the sake of brevity simply refer to them as 'matrices'. If $A$ is such a matrix, we will index its elements as $A(x, y)$ for all $x, y \in \mathscr{X}$, where the indexing is understood to be row-major. Furthermore, $A(x, \cdot)$ will denote the $x$-th row of $A$, and $A(\cdot, y)$ will denote its $y$-th column. The symbol $I$ will be reserved throughout to refer to the $|\mathscr{X}| \times |\mathscr{X}|$ identity matrix.

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

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

$$\|A\| := \sup\{\|Af\| : f \in \mathscr{L}(\mathscr{X}), \|f\| = 1\}. \tag{1}$$

If $A$ is a matrix, it is easily verified that then

$$\|A\| = \max\left\{\sum_{y \in \mathscr{X}} |A(x, y)| : x \in \mathscr{X}\right\}. \tag{2}$$

Finally, for any set $\mathscr{A}$ of matrices, we define $\|\mathscr{A}\| := \sup\{\|A\| : A \in \mathscr{A}\}$.

These norms satisfy the following properties; Reference [27] provides a proof for the non-trivial ones.

**Proposition 2.1.** *For all $f, g \in \mathscr{L}(\mathscr{X})$, all $A, B$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ that are non-negatively homogeneous, all $\lambda \in \mathbb{R}$ and all $x \in \mathscr{X}$, we have that*

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

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

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

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

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

N6: $\|A\| \geq 0$

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

N8: $\|A + B\| \leq \|A\| + \|B\|$

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

N10: $\|AB\| \leq \|A\| \, \|B\|$

N11: $\|Af\| \leq \|A\| \, \|f\|$

## 3. Transition matrix systems

We provide in this section some definitions that will later be useful for characterising continuous-time Markov chains. Because we have not yet formally introduced the concept of a continuous-time Markov chain, for now, the definitions below can be taken to be purely algebraic constructs. Nevertheless, whenever possible, we will of course attempt to provide them with some intuition.

For the purposes of this section, it suffices to say that a continuous-time Markov chain is a process which at each time $t$ is in some state $x \in \mathscr{X}$. As time elapses, the process moves through the state space $\mathscr{X}$ in some stochastic fashion. We here define tools with which this stochastic behaviour can be conveniently described.

### 3.1. Transition matrices and transition rate matrices

A *transition matrix* $T$ is a matrix that is row-stochastic, meaning that, for each $x \in \mathscr{X}$, the row $T(x, \cdot)$ is a probability mass function on $\mathscr{X}$.

**Definition 3.1** *(Transition matrix).* A real-valued matrix $T$ is said to be a *transition matrix* if

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

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

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

**Proposition 3.1.** *For any two transition matrices $T_1$ and $T_2$, their composition $T_1 T_2$ is also a transition matrix.*

The interpretation in the context of Markov chains goes as follows. The elements $T(x, y)$ of a transition matrix $T$ describe the probability of the Markov chain ending up in state $y$ at the next time point, given that it is currently in state $x$. In other words, the row $T(x, \cdot)$ contains the state-transition probabilities, conditional on currently being in state $x$. We will make this connection more explicit when we formalise continuous-time stochastic processes in Section 4.

For now, we note that in a continuous-time setting, this notion of "next" time point is less obvious than in a discrete-time setting, because the state-transition probabilities are then continuously dependent on the evolution of time. To capture this aspect, the notion of *transition rate matrices* [28] is used.

**Definition 3.2** *(Transition rate matrix).* A real-valued matrix $Q$ is said to be a *transition rate matrix*, or sometimes simply *rate matrix*, if

R1: $\sum_{y \in \mathscr{X}} Q(x, y) = 0$ for all $x \in \mathscr{X}$;

R2: $Q(x, y) \geq 0$ for all $x, y \in \mathscr{X}$ such that $x \neq y$.

We use $\mathscr{R}$ to denote the set of all transition rate matrices.

The connection between transition matrices and rate matrices is perhaps best illustrated as follows. Suppose that at some time point $t$, we want to describe for any state $x$ the probability of ending up in state $y$ at some time $s \geq t$. Let $T_t^s$ denote the transition matrix that contains all these probabilities. Note first of all that it is reasonable to assume that, if time does not evolve, then the system should not change. That is, if we are in state $x$ at time $t$, then the probability of still being in state $x$ at time $s = t$, should be one. Hence, we should have $T_t^t = I$, with $I$ the identity matrix.

A rate matrix $Q$ is then used to describe the transition matrix $T_t^{t+\Delta}$ after a small period of time, $\Delta$, has elapsed. Specifically, the scaled matrix $\Delta Q$ serves as a linear approximation of the change from $T_t^t$ to $T_t^{t+\Delta}$. The following proposition states that, for small enough $\Delta$, this linear change still results in a transition matrix.

**Proposition 3.2.** *Consider any transition rate matrix $Q \in \mathscr{R}$, and any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta \|Q\| \leq 1$. Then the matrix $(I + \Delta Q)$ is a transition matrix.*

This also explains the terminology used; a rate matrix describes the "rate of change" of a (continuously) time-dependent transition matrix over a small period of time.

Of course, this notion can also be reversed; given a transition matrix $T_t^{t+\Delta}$, what is the change that it underwent compared to $T_t^t = I$? The following proposition states that such a change can always be described using a rate matrix.

**Proposition 3.3.** *Consider any transition matrix $T$, and any $\Delta \in \mathbb{R}_{>0}$. Then, the matrix $^1\!/\!_\Delta(T - I)$ is a transition rate matrix.*

Note that Proposition 3.3 essentially states that the finite-difference $^1\!/\!_\Delta(T_t^{t+\Delta} - T_t^t)$ is a rate matrix. Intuitively, if we now take the limit as this $\Delta$ goes to zero, this states that the derivative of a continuously time-dependent transition matrix is given by some rate matrix $Q \in \mathscr{R}$—assuming that this limit exists, of course. We will make this connection more explicit in Section 4.

We next introduce a function that is often seen in the context of continuous-time Markov chains: the *matrix exponential* [29] $e^{Q\Delta}$ of $Q\Delta$, with $Q$ a rate matrix and $\Delta \in \mathbb{R}_{\geq 0}$. There are various equivalent ways in which such a matrix exponential can be defined. We refer to [29] for some examples, and will consider some specific definitions later on in this work. For now, we restrict ourselves to stating the following well-known result.

**Proposition 3.4.** *[28, Theorem 2.1.2] Consider a rate matrix $Q \in \mathscr{R}$ and any $\Delta \in \mathbb{R}_{\geq 0}$. Then $e^{Q\Delta}$ is a transition matrix.*

We conclude this section with some comments about *sets* of rate matrices. First, note that the set of *all* rate matrices, $\mathscr{R}$, is closed under finite sums and multiplication with non-negative scalars. Consider now any set $\mathscr{Q} \subseteq \mathscr{R}$ of rate matrices. Then $\mathscr{Q}$ is said to be *non-empty* if $\mathscr{Q} \neq \emptyset$ and $\mathscr{Q}$ is said to be *bounded* if $\|\mathscr{Q}\| < +\infty$. The following proposition provides a simple alternative characterisation of boundedness.

**Proposition 3.5.** *A set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ is bounded if and only if*

$$\inf\{Q(x,x) \colon Q \in \mathscr{Q}\} > -\infty \quad \text{for all } x \in \mathscr{X}. \tag{3}$$

*3.2. Transition matrix systems that are well-behaved*

In the previous section, we used the notation $T_t^s$ to refer to a transition matrix that contains the probabilities of moving from a state at time $t$, to a state at time $s$. We now consider *families* of these transition matrices. Such a family $\mathscr{T}$ specifies a transition matrix $T_t^s$ for every $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$.

We already explained in the previous section that it is reasonable to assume that $T_t^t = I$. If the transition matrices of a family $\mathscr{T}$ satisfy this property, and if they furthermore satisfy the *semi-group* property—see Equation (4) below—we call this family a *transition matrix system*. We will use $\mathcal{T}$ to refer to the set of all transition matrix systems.

**Definition 3.3** (Transition matrix system). A *transition matrix system* $\mathscr{T}$ is a family of transition matrices $T_t^s$, defined for all $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, such that for all $t, r, s \in \mathbb{R}_{\geq 0}$ with $t \leq r \leq s$, it holds that

$$T_t^s = T_t^r T_r^s, \tag{4}$$

and for all $t \in \mathbb{R}_{\geq 0}$, $T_t^t = I$.

It will turn out that there is a strong connection between transition matrix systems and continuous-time Markov chains. We will return to this in Section 5.

In the previous section, we have seen that for any transition matrix $T$ and any $\Delta \in \mathbb{R}_{>0}$, the matrix $^1\!/\!_\Delta(T - I)$ is a rate matrix, and therefore, in particular, that the finite difference $^1\!/\!_\Delta(T_t^{t+\Delta} - I)$ is a rate matrix. We here note that this is also the case for the term $^1\!/\!_\Delta(T_{t-\Delta}^t - I)$ whenever $(t - \Delta) \geq 0$.

We now consider this property in the context of a transition matrix system $\mathscr{T}$. For all $t \in \mathbb{R}_{\geq 0}$ and all $\Delta \in \mathbb{R}_{>0}$, such a transition matrix system specifies a transition matrix $T_t^{t+\Delta}$ and—if $(t - \Delta) \geq 0$—a transition matrix $T_{t-\Delta}^t$. We now consider the behaviour of these matrices for various values of $\Delta$. In particular, we look what happens to these finite differences if we take $\Delta$ to be increasingly smaller.

For each $\Delta \in \mathbb{R}_{>0}$, due to the property that we have just recalled, there will be a rate matrix that corresponds to these finite differences. If the norm of these rate matrices never diverges to $+\infty$ as we take $\Delta$ to zero, we call the family $\mathscr{T}$ *well-behaved*.

**Definition 3.4** (*Well-behaved transition matrix system*). A transition matrix system $\mathscr{T}$ is called *well-behaved* if

$$(\forall t \in \mathbb{R}_{\geq 0}) \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| < +\infty \text{ and } (\forall t \in \mathbb{R}_{> 0}) \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta}^{t} - I \right\| < +\infty . \tag{5}$$

Observe that this notion of well-behavedness does not imply differentiability; the limit $\lim_{\Delta \to 0^+} 1/\Delta (T_t^{t+\Delta} - I)$ need not exist. Rather, it implies that the rate of change of the transition matrices in $\mathscr{T}$ is bounded at all times. In this sense, this notion of well-behavedness is similar to a kind of local Lipschitz continuity. The locality stems from the fact that, although the rate of change must be bounded for each $t$, Equation (5) does not impose that it must be uniformly bounded (at all $t$) by a single "Lipschitz constant". That said, we do not stress this connection any further, because we will shortly consider a more involved notion of well-behavedness for which this connection is less immediate; see Section 4.3.

We finally consider an important special type of transition matrix systems. We have seen in the previous section that for any $Q \in \mathscr{R}$ and any $\Delta \in \mathbb{R}_{\geq 0}$, the matrix exponential $e^{Q\Delta}$ is a transition matrix. We here consider for any $Q \in \mathscr{R}$ the family $\mathscr{T}_Q$ that is generated by such transition matrices.

**Definition 3.5.** For any rate matrix $Q \in \mathscr{R}$, we use $\mathscr{T}_Q$ to denote the family of transition matrices that is defined by

$$T_t^s = e^{Q(s-t)} \quad \text{for all } t, s \in \mathbb{R}_{\geq 0} \text{ such that } t \leq s.$$

We call this family $\mathscr{T}_Q$ the *exponential transition matrix system* corresponding to $Q$.

**Proposition 3.6.** *For any $Q \in \mathscr{R}$, $\mathscr{T}_Q$ is a well-behaved transition matrix system.*

This exponential transition matrix system corresponding to a $Q \in \mathscr{R}$ will turn out to play a large role in the context of continuous-time Markov chains. We return to this in Section 5.

### 3.3. Restricted transition matrix systems

We finally consider yet another construct that will be useful later: the restriction of a transition matrix system $\mathscr{T}$ to a closed interval $\mathbf{I}$ in the time-line $\mathbb{R}_{\geq 0}$.

By a closed interval $\mathbf{I}$, we here mean a non-empty closed subset $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$ that is connected, in the sense that for any $t, s \in \mathbf{I}$ such that $t \leq s$, and any $r \in [t, s]$, it holds that $r \in \mathbf{I}$. Note that for any $c \in \mathbb{R}_{\geq 0}$, $[c, +\infty)$ is such a closed interval.

For any transition matrix system $\mathscr{T}$ and any such closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, we use $\mathscr{T}^{\mathbf{I}}$ to denote the restriction of $\mathscr{T}$ to $\mathbf{I}$. Such a restriction is a family of transition matrices $T_t^s$ that is defined for all $t, s \in \mathbf{I}$ such that $t \leq s$. We call such a family $\mathscr{T}^{\mathbf{I}}$ a *restricted transition matrix system* on $\mathbf{I}$. The set of all restricted transition matrix systems on $\mathbf{I}$ is denoted by $\mathcal{T}^{\mathbf{I}}$.

**Proposition 3.7.** *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $\mathscr{T}^{\mathbf{I}}$ be a family of transition matrices $T_t^s$ that is defined for all $t, s \in \mathbf{I}$ with $t \leq s$. Then $\mathscr{T}^{\mathbf{I}}$ is a restricted transition matrix system on $\mathbf{I}$ if and only if, for all $t, r, s \in \mathbf{I}$ with $t \leq r \leq s$, it holds that $T_t^s = T_t^r T_r^s$ and $T_t^t = I$.*

We call a restricted transition matrix system $\mathscr{T}^{\mathbf{I}}$ well-behaved if it is the restriction to $\mathbf{I}$ of a well-behaved transition matrix system.

**Proposition 3.8.** *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $\mathscr{T}^{\mathbf{I}}$ be a restricted transition matrix system on $\mathbf{I}$. Then $\mathscr{T}^{\mathbf{I}}$ is well-behaved if and only if*

$$(\forall t \in \mathbf{I}^+) \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| < +\infty \text{ and } (\forall t \in \mathbf{I}^-) \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta}^{t} - I \right\| < +\infty , \tag{6}$$

*where $\mathbf{I}^+ := \mathbf{I} \setminus \{\sup \mathbf{I}\}$ and $\mathbf{I}^- := \mathbf{I} \setminus \{\min \mathbf{I}\}$.*

Now, because these restricted transition matrix systems are only defined on some given closed interval, it will be useful to define a concatenation operator between two such systems.

**Definition 3.6** (*Concatenation operator*). For any two closed intervals $\mathbf{I}, \mathbf{J} \subseteq \mathbb{R}_{\geq 0}$ such that $\max \mathbf{I} = \min \mathbf{J}$, and any two restricted transition matrix systems $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$, the concatenation of $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$ is denoted by $\mathscr{T}^{\mathbf{I}} \otimes \mathscr{T}^{\mathbf{J}}$, and defined as the family of transition matrices $T_t^s$ that is given by

$$T_t^s := \begin{cases} {}^i T_t^s & \text{if } t, s \in \mathbf{I} \\ {}^j T_t^s & \text{if } t, s \in \mathbf{J} \\ {}^i T_t^r {}^j T_r^s & \text{if } t \in \mathbf{I} \text{ and } s \in \mathbf{J} \end{cases} \quad \text{for all } t, s \in \mathbf{I} \cup \mathbf{J} \text{ such that } t \leq s,$$

where $r := \max \mathbf{I} = \min \mathbf{J}$, and ${}^i T_t^s$ and ${}^j T_t^s$ denote the transition matrices corresponding to $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$, respectively.

**Proposition 3.9.** *Consider two closed intervals* $\mathbf{I}, \mathbf{J} \subseteq \mathbb{R}_{\geq 0}$ *such that* $\max \mathbf{I} = \min \mathbf{J}$, *and any two restricted transition matrix systems* $\mathscr{T}^{\mathbf{I}}$ *and* $\mathscr{T}^{\mathbf{J}}$. *Then their concatenation* $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}} \coloneqq \mathscr{T}^{\mathbf{I}} \otimes \mathscr{T}^{\mathbf{J}}$ *is a restricted transition matrix system on* $\mathbf{I} \cup \mathbf{J}$. *Furthermore, if both* $\mathscr{T}^{\mathbf{I}}$ *and* $\mathscr{T}^{\mathbf{J}}$ *are well behaved, then* $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}}$ *is also well-behaved.*

**Example 3.1.** Consider any two rate matrices $Q_1, Q_2 \in \mathscr{R}$ such that $Q_1 \neq Q_2$, and let $\mathscr{T}_{Q_1}$ and $\mathscr{T}_{Q_2}$ be their exponential transition matrix systems, which, as we know from Proposition 3.6, are well-behaved. Now choose any $r \in \mathbb{R}_{\geq 0}$ and define

$$\mathscr{T} \coloneqq \mathscr{T}_{Q_1}^{[0,r]} \otimes \mathscr{T}_{Q_2}^{[r,+\infty)}.$$

It then follows from Proposition 3.9 that $\mathscr{T}$ is a well-behaved transition matrix system. Furthermore, for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$, the transition matrix $T_t^s$ that corresponds to $\mathscr{T}$ is given by $T_t^s = T_t^r T_r^s = e^{Q_1(r-t)} e^{Q_2(s-r)}$. ◇

We also introduce a metric $d$ between restricted transition matrix systems that are defined on the same interval $\mathbf{I}$. For any two such restricted transition matrix systems $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{S}^{\mathbf{I}}$, we let

$$d(\mathscr{T}^{\mathbf{I}}, \mathscr{S}^{\mathbf{I}}) \coloneqq \sup \left\{ \left\| T_t^s - S_t^s \right\| : t, s \in \mathbf{I}, t \leq s \right\}, \tag{7}$$

where, for all $t, s \in \mathbf{I}$, it is understood that $T_t^s$ corresponds to $\mathscr{T}^{\mathbf{I}}$ and $S_t^s$ to $\mathscr{S}^{\mathbf{I}}$. This metric allows us to state the following result.

**Proposition 3.10.** *Consider any interval* $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$ *and let* $d$ *be the metric that is defined in Equation* (7). *The metric space* $(\mathcal{T}^{\mathbf{I}}, d)$ *is then complete.*

Note that this result includes as a special case that the set $\mathcal{T}$ of all (unrestricted) transition matrix systems is complete. The following example illustrates how this result can be used.

**Example 3.2.** Consider some positive constant $c \in \mathbb{R}_{>0}$ and let $\{Q_i\}_{i \in \mathbb{N}_0}$ be a sequence of rate matrices such that, for all $i \in \mathbb{N}$, $\|Q_i - Q_{i-1}\| \leq c$. We can then construct the following sequence. For $i = 0$, we let $\mathscr{T}_0 \coloneqq \mathscr{T}_{Q_0}$, and for all $i \in \mathbb{N}$, we let

$$\mathscr{T}_i \coloneqq \mathscr{T}_{Q_i}^{[0,\delta_i]} \otimes \mathscr{T}_{i-1}^{[\delta_i,+\infty)} \tag{8}$$

where, for all $i \in \mathbb{N}_0$, $\delta_i \coloneqq 2^{-i}$. The resulting sequence $\{\mathscr{T}_i\}_{i \in \mathbb{N}_0}$ is then clearly a subset of $\mathcal{T}$ and, because of Propositions 3.6 and 3.9, every transition matrix system in this sequence is well-behaved. Furthermore, as we prove in [26, Appendix I], $\{\mathscr{T}_i\}_{i \in \mathbb{N}_0}$ is a Cauchy sequence, which basically means that its elements become arbitrarily close to each other as the sequence progresses.

The reason why this is of interest to us is because in a complete metric space, every Cauchy sequence converges to a limit that belongs to the same space. Hence, since $\{\mathscr{T}_i\}_{i \in \mathbb{N}_0}$ is Cauchy, Proposition 3.10 allows us to infer that $\{\mathscr{T}_i\}_{i \in \mathbb{N}_0}$ converges to a limit $\mathscr{T} \coloneqq \lim_{i \to \infty} \mathscr{T}_i$ in $\mathcal{T}$. ◇

As this example illustrates, Proposition 3.10 allows us to (a) establish the existence of limits of sequences of (restricted) transition matrix systems and (b) prove that these limits are restricted transition matrix systems themselves. In order to make this concept of a limit of transition matrix systems less abstract, we now provide, for a particular case of the sequence in Example 3.2, closed-form expressions for some of the transition matrices that correspond to its limit.

**Example 3.3.** Let $Q_1, Q_2 \in \mathscr{R}$ be two commuting rate matrices. For example, let $Q_1 \in \mathscr{R}$ be an arbitrary rate matrix and let $Q_2 \coloneqq \alpha Q_1$, with $\alpha \in \mathbb{R}_{\geq 0}$.

Now let $\{Q_i\}_{i \in \mathbb{N}_0}$ be defined by $Q_i \coloneqq Q_1$ if $i$ is odd and $Q_i \coloneqq Q_2$ if $i$ is even, define $\delta_i \coloneqq 2^{-i}$ for all $i \in \mathbb{N}_0$, and consider the corresponding sequence of transition matrix systems $\{\mathscr{T}_i\}_{i \in \mathbb{N}}$ that was defined in Example 3.2. Since $\|Q_i - Q_{i-1}\| = \|Q_1 - Q_2\|$ for all $i \in \mathbb{N}$, the sequence $\{Q_i\}_{i \in \mathbb{N}_0}$ clearly satisfies the conditions in Example 3.2—just choose $c = \|Q_1 - Q_2\|$—and therefore, as we have seen, $\{\mathscr{T}_i\}_{i \in \mathbb{N}}$ converges to a limit $\mathscr{T} \coloneqq \lim_{i \to \infty} \mathscr{T}_i$ in $\mathcal{T}$.

As we prove in [26, Appendix I], it then holds that for any $t \in (0, 1]$, the transition matrix from 0 to $t$ that corresponds to the transition matrix system $\mathscr{T}$ is equal to

$$T_0^t = e^{Q_1 \varphi_1(t) + Q_2 \varphi_2(t)}, \tag{9}$$

with

$$\varphi_1(t) \coloneqq \begin{cases} t - \frac{2}{3}\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ odd} \\ \frac{2}{3}\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ even} \end{cases} \tag{10}$$

and

$$\varphi_2(t) := \begin{cases} 2/3\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ odd} \\ t - 2/3\delta_{i+1} & \text{if } \delta_{i+1} \leq t \leq \delta_i \text{ with } i \text{ even.} \end{cases} \tag{11}$$

Furthermore, it can be shown that the transition matrix system $\mathscr{T}$ is well-behaved—again, see [26, Appendix I] for a proof. $\diamond$

The transition matrix system $\mathscr{T}$ in our previous example was well-behaved, and was constructed as a limit of well-behaved transition matrix systems. Therefore, one might think that the former is implied by the latter. However, as our next example illustrates, this is not the case. A limit of well-behaved transition matrix systems need not be well-behaved itself.

**Example 3.4.** Consider any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$ and, for all $i \in \mathbb{N}_0$, define $Q_i := iQ$ and let $\mathscr{T}_i$ and $\delta_i$ be defined as in Example 3.2. Then since $\{Q_i\}_{i \in \mathbb{N}_0}$ satisfies the conditions of Example 3.2 with $c = 1$, the sequence $\{\mathscr{T}_i\}_{i \in \mathbb{N}_0}$ has a limit $\mathscr{T} := \lim_{i \to \infty} \mathscr{T}_i$ in $\mathcal{T}$.

However, despite the fact that we know from Example 3.2 that each of the transition matrix systems $\mathscr{T}_i$, $i \in \mathbb{N}_0$, is well-behaved, the limit $\mathscr{T}$ itself is not well-behaved; see [26, Appendix I] for a proof. $\diamond$

## 4. Continuous-time stochastic processes

We will in this section formalise the notion of a continuous-time stochastic process. However, we do not adopt the classical—Kolmogorovian, measure-theoretic—setting, but will instead be using the framework of full conditional probabilities. Our reasons for doing so are the following.

First of all, our results on imprecise continuous-time Markov chains will be concerned with events or functions that depend on an finite number of time points only. Therefore, we do not require the use of typical measure-theoretic concepts such as $\sigma$-algebras, $\sigma$-additivity, measurability, etcetera. Instead, we will impose only the bare minimum of assumptions that are required for our results. The extra structural and continuity assumptions that are typically imposed in a measure-theoretic setting are regarded as optional.

Secondly, our approach does not suffer from some of the traditional issues with probability zero. In standard settings, conditional probabilities are usually derived from unconditional probabilities through Bayes's rule, which makes them undefined whenever the conditioning event has probability zero. Instead, we will regard conditional probabilities as primitive concepts. As a result, we can do away with some of the usual 'almost surely' statements and replace them with statements that are certain.

Finally, and most importantly, in our 'imprecise' setting, we will be working with a set of stochastic processes rather than with a single stochastic process. In this context, we will often need to prove that such a set contains a stochastic process that meets certain specific requirements. Full conditional probabilities provide a convenient framework for constructing such existence proofs, through the notion of coherence.

We start in Section 4.1 by introducing full conditional probabilities and explaining their connection with coherence. In Section 4.2, we then use these concepts to formalise continuous-time stochastic processes. Section 4.3 describes a specific subclass of stochastic processes, which we call well-behaved, and on which we will largely focus throughout this work. Finally, Section 4.4 provides some tools with which we can describe the dynamics of stochastic processes.

### 4.1. Full and coherent conditional probabilities

Consider a variable $X$ that takes values $\omega$ in some non-empty—possibly infinite—*outcome space* $\Omega$. The actual value of $X$ is taken to be uncertain, in the sense that it is unknown. This uncertainty may arise because $X$ is the outcome of a random experiment that has yet to be conducted, but it can also simply be a consequence of our lack of information about $X$. We call any subset $E$ of $\Omega$ an event, we use $\mathscr{E}(\Omega)$ to denote the set of all such events, and we let $\mathscr{E}(\Omega)_\emptyset := \mathscr{E}(\Omega) \setminus \{\emptyset\}$ be the set of all non-empty events. A subject's uncertainty about the value of $X$ can then be described by means of a full conditional probability [30].

**Definition 4.1** (*Full conditional probability*). A full conditional probability $P$ is a real-valued map from $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ to $\mathbb{R}$ that satisfies the following axioms. For all $A, B \in \mathscr{E}(\Omega)$ and all $C, D \in \mathscr{E}(\Omega)_\emptyset$:

F1: $P(A|C) \geq 0$;
F2: $P(A|C) = 1$ if $C \subseteq A$;
F3: $P(A \cup B|C) = P(A|C) + P(B|C)$ if $A \cap B = \emptyset$;
F4: $P(A \cap D|C) = P(A|D \cap C)P(D|C)$ if $D \cap C \neq \emptyset$.

For any $A \in \mathscr{E}(\Omega)$ and $C \in \mathscr{E}(\Omega)_\emptyset$, we call $P(A|C)$ the probability of $A$ conditional on $C$. Also, for any $A \in \mathscr{E}(\Omega)$, we use the shorthand notation $P(A) := P(A|\Omega)$ and then call $P(A)$ the probability of $A$. The following additional properties can easily be shown to follow from F1–F3; see Appendix B for a proof. For all $A \in \mathscr{E}(\Omega)$ and all $C \in \mathscr{E}(\Omega)_\emptyset$:

F5: $0 \leq P(A|C) \leq 1$;
F6: $P(A|C) = P(A \cap C|C)$;
F7: $P(\emptyset|C) = 0$;
F8: $P(\Omega|C) = 1$.

Basically, F1–F4 are just the standard rules of probability. However, there are four rather subtle differences with the more traditional approach. The first difference is that a full conditional probability takes conditional probabilities as its basic entities: $P(A|C)$ is well-defined even if $P(C) = 0$. The second difference, which is related to the first, is that Bayes's rule—F4—is stated in a multiplicative form; it is not regarded as a definition of conditional probabilities, but rather as a property that connects conditional probabilities to unconditional ones. The third difference is that we consider all events, and do not restrict ourselves to some specific subset of events—such as a $\sigma$-algebra. The fourth difference, which is related to the third, is that we only require finite additivity—F3—and do not impose $\sigma$-additivity.

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

**Example 4.1.** Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ be the set of possible values for the throw of a—possibly unfair—die and let $\mathscr{C} := \{(E_o, \Omega), (E_e, \Omega)\} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$, where the events $E_o = \{1, 3, 5\}$ and $E_e = \{2, 4, 6\}$ correspond to an odd or even outcome of the die throw, respectively. The map $P \colon \mathscr{C} \to \mathbb{R}$ that is defined by

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

then satisfies F1–F4 on its domain. However, if we extend the domain by adding the trivial couple $(\Omega, \Omega)$, it becomes impossible to satisfy F1–F4, because F2 and F3 would then require that

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

which is clearly a contradiction.                                                                                               $\Diamond$

In order to avoid the situation in this example, that is, in order to guarantee the possibility of extending the domain of a conditional probability in a sensible way, we use the concept of coherence [31–35].

**Definition 4.2** (*Coherent conditional probability*). Let $P$ be a real-valued map from $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ to $\mathbb{R}$. Then $P$ is said to be a *coherent conditional probability* on $\mathscr{C}$ if, for all $n \in \mathbb{N}$ and every choice of $(A_i, C_i) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$,[1]

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

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

The interested reader is invited to take a look at [26, Appendix H], where we provide this abstract concept with a brief—and well-known—intuitive gambling interpretation; see References [33–35,37,31,36] for an extensive discussion. However, for our present purposes, this interpretation is not required. Instead, our motivation for introducing coherence stems from the following two results. First, if $\mathscr{C} = \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$, then coherence is equivalent to the axioms of probability, that is, properties F1–F4.

**Theorem 4.1.** *[33, Theorem 3] Let $P$ be a real-valued map from $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ to $\mathbb{R}$. Then $P$ is a coherent conditional probability if and only if it is a full conditional probability.*

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

---

[1] Many authors replace the maximum in this expression by a supremum, and also impose an additional inequality, where the maximum—supremum—is replaced by a minimum—infimum—and where the inequality is reversed [36,31,33]. This is completely equivalent to our definition. First of all, if the maximum is replaced by a supremum, then since $n$ is finite and because, for every $i \in \{1, \ldots, n\}$, $\mathbb{I}_{A_i}$ and $\mathbb{I}_{C_i}$ can only take two values—0 or 1—it follows that this supremum is taken over a finite set of real numbers, which implies that it is actually a maximum. Secondly, replacing the maximum by a minimum and reversing the inequality is equivalent to replacing the $\lambda_i$ in our expression by their negation, which is clearly allowed because the coefficients $\lambda_i$ can take any arbitrary real value.

**Theorem 4.2.** *[33, Theorem 4] Let $P$ be a coherent conditional probability on $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$. Then for any $\mathscr{C} \subseteq \mathscr{C}^* \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$, $P$ can be extended to a coherent conditional probability on $\mathscr{C}^*$.*

In particular, it is therefore always possible to extend a coherent conditional probability $P$ on $\mathscr{C}$, to a coherent conditional probability on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$. Due to Theorem 4.1, this extension is a full conditional probability. The following makes this explicit.

**Corollary 4.3.** *Let $P$ be a real-valued map from $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ to $\mathbb{R}$. Then $P$ is a coherent conditional probability if and only if it can be extended to a full conditional probability.*

Note, therefore, that if $P$ is a coherent conditional probability on $\mathscr{C}$, we can equivalently say that it is the restriction of a full conditional probability. Hence, any coherent conditional probability on $\mathscr{C}$ is guaranteed to satisfy properties F1–F4. However, as was essentially already illustrated in Example 4.1, and as our next example makes explicit, the converse is not true.

**Example 4.2.** Let $\Omega$, $\mathscr{C}$, $E_\mathrm{o}$, $E_\mathrm{e}$ and $P \colon \mathscr{C} \to \mathbb{R}$ be defined as in Example 4.1. Then as we have seen in that example, $P$ satisfies F1–F4 on its domain $\mathscr{C}$. However, $P$ is not a coherent conditional probability on $\mathscr{C}$, because if it was, then according to Corollary 4.3, $P$ could be extended to a full conditional probability. Since $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ includes $(\Omega, \Omega)$, the argument at the end of Example 4.1 implies that this is impossible. A similar conclusion can be reached by verifying Definition 4.2 directly; we leave this as an exercise.                                                                                           ◇

### 4.2. Stochastic processes as a special case

A (continuous-time) stochastic process is now simply a coherent conditional probability on a specific domain $\mathscr{C}^{\mathrm{SP}}$, or equivalently, the restriction of a full conditional probability to this domain $\mathscr{C}^{\mathrm{SP}}$—see Definition 4.3. However, before we get to this definition, let us first provide some intuition.

Basically, a continuous-time stochastic process describes the behaviour of a system as it moves through the—finite—state space $\mathscr{X}$ over a continuous time line $\mathbb{R}_{\geq 0}$. A single realisation of this movement is called a path or a trajectory. We are typically uncertain about the specific path that will be followed, and a stochastic process quantifies this uncertainty by means of a probabilistic model, which, in our case, will be a coherent conditional probability. These ideas are formalised as follows.

A *path* $\omega$ is a function from $\mathbb{R}_{\geq 0}$ to $\mathscr{X}$, and we denote with $\omega(t)$ the value of $\omega$ at time $t$. For any sequence of time points $u \in \mathscr{U}$ and any path $\omega$, we will write $\omega|_u$ to denote the restriction of $\omega$ to $u \subset \mathbb{R}_{\geq 0}$. Using this notation, we write for any $x_u \in \mathscr{X}_u$ that $\omega|_u = x_u$ if, for all $t \in u$, it holds that $\omega(t) = x_t$.

The outcome space $\Omega$ of a stochastic process is a set of paths. Three commonly considered choices are to let $\Omega$ be the set of all paths, the set of all right-continuous paths [28], or the set of all cadlag paths (right-continuous paths with left-sided limits) [38]. However, our results do not require such a specific choice. For the purposes of this paper, all that we require is that

$$(\forall u \in \mathscr{U}_\emptyset)(\forall x_u \in \mathscr{X}_u)(\exists \omega \in \Omega)\ \omega|_u = x_u. \tag{12}$$

Thus, $\Omega$ must be chosen in such a way that, for any non-empty finite sequence of time points $u \in \mathscr{U}_\emptyset$ and any state assignment $x_u \in \mathscr{X}_u$ on those time points, there is at least one path $\omega \in \Omega$ that agrees with $x_u$ on $u$. Essentially, this condition guarantees that $\Omega$ is "large enough to be interesting". It serves as a nice exercise to check that the three specific sets $\Omega$ in the beginning of this paragraph each satisfy this condition.

For any set of events $\mathscr{E} \subseteq \mathscr{E}(\Omega)$, we use $\langle \mathscr{E} \rangle$ to denote the algebra that is generated by them. That is, $\langle \mathscr{E} \rangle$ is the smallest subset of $\mathscr{E}(\Omega)$ that contains all elements of $\mathscr{E}$, and that is furthermore closed under complements in $\Omega$ and finite unions, and therefore also under finite intersections. Furthermore, for any $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$, we define the elementary event

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

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

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

be the set of elementary events whose time point is either preceded by or belongs to $u$, and we let $\mathscr{A}_u \coloneqq \langle \mathscr{E}_u \rangle$ be the algebra that is generated by this set of elementary events.

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

$$(X_u = x_u) \coloneqq \{\omega \in \Omega : \omega|_u = x_u\}$$

belongs to $\mathscr{E}(\Omega)_\emptyset$, because it follows from Equation (12) that this event is non-empty. Hence, for any $A \in \mathscr{A}_u$ and $x_u \in \mathscr{X}_u$, we find that $(A, X_u = x_u) \in \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$. Since this is true for every $u \in \mathscr{U}$, it follows that

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

is a subset of $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$. It is also worth noting that if $u = \emptyset$, then $\omega|_u = x_u$ is vacuously true, which implies that in that case, $(X_u = x_u) = \Omega$.

That being said, we can now finally formalise our definition of a (continuous-time) stochastic process.

**Definition 4.3** *(Stochastic process).* A *stochastic process* is a coherent conditional probability on $\mathscr{C}^{\mathrm{SP}}$. We denote the set of all stochastic processes by $\mathbb{P}$.

**Corollary 4.4.** *Let $P$ be a real-valued map from $\mathscr{C}^{\mathrm{SP}}$ to $\mathbb{R}$. Then $P$ is a stochastic process if and only if it is the restriction of a full conditional probability.*

There are two reason why we restrict ourselves to the domain $\mathscr{C}^{\mathrm{SP}}$. The most important reason is simply that all of the results in this paper can be expressed using only this domain, because they are all concerned with events or functions that depend on a finite number of time points. The second reason is that this restriction will allow us to state uniqueness results that do not extend to larger domains; see for example Corollary 5.3. However, it is important to realise that our restriction of the domain does not impose any real limitations, because, as we know from Theorem 4.2, the domain of a coherent conditional probability—and hence also a stochastic process—can always be extended (although not uniquely). In fact, as briefly touched upon in the conclusions of this paper, it is even possible to consider extensions that are $\sigma$-additive.

Finally, we would like to point out that it is also possible to use a different—yet equivalent—definition for stochastic processes. Indeed, due to Corollary 4.4, a stochastic process can also be defined as the restriction of a full conditional probability to $\mathscr{C}^{\mathrm{SP}}$. Of course, given this realisation, one may then start to wonder why we have gone through to the trouble of introducing coherence, because this alternative definition would not require any notion of coherence. The reason why we nevertheless need coherence, is because it allows us to establish the existence of stochastic processes that have certain properties, which will often be necessary in the proofs of our results.

For instance, by means of an example, suppose that we are given an arbitrary function $p \colon \mathbb{R}_{>0} \to [0, 1]$, and that we want to know if there is a stochastic process $P$ for which, for some $x, y \in \mathscr{X}$,

$$P(X_t = y \mid X_0 = x) = p(t) \ \text{ for all } t \in \mathbb{R}_{>0}. \tag{13}$$

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

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

**Example 4.3.** Let $\mathscr{X}$ be a state space that contains at least two states, fix two—possibly equal—states $x, y \in \mathscr{X}$, and consider any function $p \colon \mathbb{R}_{>0} \to [0, 1]$. The aim of this example is to prove that there is a stochastic process $P$ that satisfies Equation (13).

The crucial step of the proof is to consider a function $\tilde{P}$ that is defined by

$$\tilde{P}(X_t = y \mid X_0 = x) \coloneqq p(t) \ \text{ for all } (X_t = y, X_0 = x) \in \mathscr{C}, \tag{14}$$

and to prove that this function is a coherent conditional probability on

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

or equivalently, that it satisfies Definition 4.2.

So consider any $n \in \mathbb{N}$ and, for all $i \in \{1, \ldots, n\}$, some $(X_{t_i} = y, X_0 = x) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. According to Definition 4.2, we now have to show that

$$\max\left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_x(\omega(0)) \left( \tilde{P}(X_{t_i} = y \mid X_0 = x) - \mathbb{I}_y(\omega(t_i)) \right) \ \middle| \ \omega \in C_0 \right\} \geq 0, \tag{15}$$

with $C_0 \coloneqq \bigcup_{i=1}^{n} (X_0 = x) = (X_0 = x)$. While doing so, we can assume without loss of generality that $i \neq j$ implies $t_i \neq t_j$, because if $t_i = t_j$ for some $i \neq j$, then we can simply add the corresponding two summands in Equation (15).

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

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

Furthermore, let $N_{<0} \coloneqq \{i \in \{1, \ldots, n\} \colon \lambda_i < 0\}$ and $N_{\geq 0} \coloneqq \{i \in \{1, \ldots, n\} \colon \lambda_i \geq 0\}$. Since $n$ is finite, Equation (12) now guarantees that there is some $\omega \in \Omega$ such that $\omega|_u = x_u$. Evaluating the sum in Equation (15) using this $\omega$, we find that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_x(\omega(0)) \left( \tilde{P}(X_{t_i} = y | X_0 = x) - \mathbb{I}_y(\omega(t_i)) \right) = \sum_{i=1}^{n} \lambda_i \left( \tilde{P}(X_{t_i} = y | X_0 = x) - \mathbb{I}_y(\omega(t_i)) \right)$$

$$= \sum_{i \in N_{<0}} \lambda_i \left( \tilde{P}(X_{t_i} = y | X_0 = x) - 1 \right) + \sum_{i \in N_{\geq 0}} \lambda_i \tilde{P}(X_{t_i} = y | X_0 = x)$$

$$\geq \sum_{i \in N_{<0}} \lambda_i \left( \tilde{P}(X_{t_i} = y | X_0 = x) - 1 \right)$$

$$= \sum_{i \in N_{<0}} |\lambda_i| \left( 1 - \tilde{P}(X_{t_i} = y | X_0 = x) \right) \geq 0,$$

where the two inequalities follow from the fact that $\tilde{P}(X_{t_i} = y | X_0 = x) = p(t_i) \in [0, 1]$. Furthermore, because $\omega(0) = x$, we also have that $\omega \in (X_0 = x) = C_0$. Therefore, we find that Equation (15) indeed holds. Hence, we conclude that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$.

The rest of the proof is now straightforward. Since $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$, and because $\mathscr{C}$ is a subset of $\mathscr{C}^{\mathrm{SP}}$, it follows from Theorem 4.2 that $\tilde{P}$ can be extended to a coherent conditional probability $P$ on $\mathscr{C}^{\mathrm{SP}}$, or equivalently, to a stochastic process $P$. Since this stochastic process $P$ is an extension of $\tilde{P}$, Equation (13) is now an immediate consequence of Equation (14). ◇

### 4.3. Well-behaved stochastic processes

Stochastic processes, as they are defined in the previous section, can behave in rather extreme ways. For instance, Example 4.3 tells us that for any two states $x, y \in \mathscr{X}$, there exists a stochastic process $P$ such that $P(X_t = y | X_0 = x) = \mathbb{I}_{\mathbb{Q}_{>0}}(t)$, where $\mathbb{I}_{\mathbb{Q}_{>0}}$ is the indicator of the positive rational numbers.

In order to avoid this kind of extreme behaviour, we will require that the rate of change of a stochastic processes remains bounded. We formalise this requirement through the notion of *well-behavedness*.

**Definition 4.4** (*Well-behaved stochastic process*). A stochastic process $P \in \mathbb{P}$ is said to be *well-behaved* if, for any−possibly empty−time sequence $u \in \mathscr{U}$, any $x_u \in \mathscr{X}_u$, any $x, y \in \mathscr{X}$ and any $t \in \mathbb{R}_{\geq 0}$ such that $t > u$:

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_{t+\Delta} = y | X_t = x, X_u = x_u) - \mathbb{I}_x(y)| < +\infty \tag{16}$$

and, if $t \neq 0$,

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_t = y | X_{t-\Delta} = x, X_u = x_u) - \mathbb{I}_x(y)| < +\infty. \tag{17}$$

The set of all well-behaved stochastic processes is denoted by $\mathbb{P}^{\mathrm{W}}$.

This definition of well-behavedness is related to continuity and differentiability, but stronger than the former and weaker than the latter. Our next example provides some intuition on this.

**Example 4.4.** Let $\mathscr{X}$ be a state space that contains at least two states, fix two states $x, y \in \mathscr{X}$ such that $x \neq y$, and consider any function $p : \mathbb{R}_{>0} \to [0, 1]$. Then as we know from Example 4.3, there is a stochastic process $P$ such that

$$P(X_\Delta = y | X_0 = x) = p(\Delta) \quad \text{for all } \Delta \in \mathbb{R}_{>0}.$$

Furthermore, since $x \neq y$, it follows from F6 and F7 that $P(X_0 = y | X_0 = x) = 0$. We now consider two specific choices for $p$.

If we let $p(\Delta) := \sqrt{\Delta}$ for $\Delta \in (0, 1]$ and $p(\Delta) := 1$ for $\Delta \geq 1$, then $P(X_\Delta = y | X_0 = x)$ is continuous on $\mathbb{R}_{\geq 0}$ because $P(X_0 = y | X_0 = x) = 0$. However, we also find that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_\Delta = y | X_0 = x) - \mathbb{I}_x(y)| = \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \sqrt{\Delta} = +\infty$$

and therefore, it follows from Equation (16)−with $t = 0$ and $u = \emptyset$−that $P$ is not well-behaved.

On the other hand, if we let $p(\Delta) := \Delta |\sin(1/\Delta)|$ for $\Delta \in \mathbb{R}_{>0}$, we find that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} |P(X_\Delta = y | X_0 = x) - \mathbb{I}_x(y)| = \limsup_{\Delta \to 0^+} |\sin(1/\Delta)| = 1.$$

In this case−at least for $t = 0$ and $u = \emptyset$−$P$ does exhibit the behaviour that we associate with a well-behaved stochastic process. Furthermore, as we invite the reader to check, $P(X_\Delta = y | X_0 = x)$ is again continuous on $\mathbb{R}_{\geq 0}$. However, $P(X_\Delta = y | X_0 = x)$ is not differentiable in $\Delta = 0$, because $1/\Delta P(X_\Delta = y | X_0 = x) = |\sin(1/\Delta)|$ oscillates as $\Delta$ approaches zero. ◇

[Definition 4.4](#) is also very similar to the definition of a well-behaved transition matrix system, as introduced in Section [3.2](#). Furthermore, here too, one could say that well-behavedness is intuitively related to local Lipschitz continuity. However, note that this time, this locality not only depends on the time point $t$, but also on the given history $x_u$.

### 4.4. Process dynamics

We end this section by introducing some tools to describe the behaviour of stochastic processes. Rather than work with the individual probabilities, it will be convenient to jointly consider probabilities that are related by the same conditioning event. To this end, we will next introduce the notion of transition matrices corresponding to a given stochastic process.

**Definition 4.5** *(Corresponding transition matrix).* Consider any stochastic process $P \in \mathbb{P}$. Then, for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, the *corresponding transition matrix* $T_t^s$ is a matrix that is defined by

$$T_t^s(x_t, x_s) \coloneqq P(X_s = x_s \,|\, X_t = x_t) \quad \text{for all } x_s, x_t \in \mathscr{X}.$$

We denote this family of matrices by $\mathscr{T}_P$.

Because we will also want to work with conditioning events that contain more than a single time point, we furthermore introduce the following generalisation.

**Definition 4.6** *(History-dependent corresponding transition matrix).* Let $P \in \mathbb{P}$ be any stochastic process. Then, for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the corresponding *history-dependent* transition matrix $T_{t,\,x_u}^s$ is a matrix that is defined by

$$T_{t,\,x_u}^s(x_t, x_s) \coloneqq P(X_s = x_s | X_t = x_t, X_u = x_u) \quad \text{for all } x_s, x_t \in \mathscr{X}.$$

For notational convenience, we allow $u$ to be empty, in which case $T_{t,\,x_u}^s = T_t^s$.

The following proposition establishes some simple properties of these corresponding (history-dependent) transition matrices.

**Proposition 4.5.** *Let $P \in \mathbb{P}$ be a stochastic process. Then, for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the corresponding (history dependent) transition matrix $T_{t,\,x_u}^s$ is—as its name suggests—a transition matrix, and $T_{t,\,x_u}^t = I$. Furthermore, $P$ is well-behaved if and only if, for every—possibly empty—time sequence $u \in \mathscr{U}$, any $x_u \in \mathscr{X}_u$ and any $t \in \mathbb{R}_{\geq 0}$ such that $t > u$:*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t,\,x_u}^{t+\Delta} - I \right\| < +\infty \tag{18}$$

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

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta,\,x_u}^t - I \right\| < +\infty. \tag{19}$$

**Remark 4.1.** Note that for any $P \in \mathbb{P}$, the corresponding transition matrix $T_{t,\,x_u}^s$ is a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, that can therefore be applied to any $f \in \mathscr{L}(\mathscr{X})$. Furthermore, for any $x_t \in \mathscr{X}$, we have that

$$\left[ T_{t,x_u}^s f \right](x_t) = \sum_{x_s \in \mathscr{X}} f(x_s) P(X_s = x_s \,|\, X_t = x_t, X_u = x_u) = \mathbb{E}_P \left[ f(X_s) \,|\, X_t = x_t, X_u = x_u \right],$$

where the expectation is taken with respect to $P(X_s \,|\, X_t = x_t, X_u = x_u)$. This observation will be useful when we later focus on expectations with respect to stochastic processes; for functions $f \in \mathscr{L}(\mathscr{X})$, their corresponding transition matrices serve as an alternative representation for the expectation operator. $\Diamond$

Because a stochastic process is defined on a continuous time line, and because its corresponding transition matrices $T_{t,\,x_u}^s$ only describe the behaviour of this process on fixed points in time, we will furthermore require some tools to capture the dynamics of a stochastic process. That is, we will be interested in how their transition matrices change over time.

One seemingly obvious way to describe these dynamics is to use the derivatives of the transition matrices that correspond to stochastic processes. However, because we do not impose differentiability assumptions on these processes, such derivatives may not exist. We will therefore instead introduce *outer partial derivatives* below. It will be instructive, however, to first consider ordinary *directional partial derivatives*.

**Definition 4.7** *(Directional partial derivatives).* For any stochastic process $P \in \mathbb{P}$, any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the *right-sided partial derivative* of $T_{t,x_u}^t$ is defined by

$$\partial_+ T_{t,x_u}^t \coloneqq \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T_{t,x_u}^{t+\Delta} - T_{t,x_u}^t) = \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T_{t,x_u}^{t+\Delta} - I)$$

and, if $t \neq 0$, the *left-sided partial derivative* of $T_{t,x_u}^t$ is defined by

$$\partial_- T_{t,x_u}^t \coloneqq \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T_{t-\Delta,x_u}^t - T_{t,x_u}^t) = \lim_{\Delta \to 0^+} \frac{1}{\Delta}(T_{t-\Delta,x_u}^t - I).$$

If these partial derivatives exist, then because of Proposition 3.3, they are guaranteed to belong to the set of rate matrices $\mathscr{R}$. If they both exist and coincide, we write $\partial T_{t,x_u}^t$ to denote their common value. If $t = 0$, we let $\partial T_{t,x_u}^t \coloneqq \partial_+ T_{t,x_u}^t$.

   The following example establishes that these directional partial derivatives need not exist. In particular, they need not exist even for well-behaved processes.

**Example 4.5.** Let $Q_1, Q_2 \in \mathscr{R}$ be two commuting rate matrices such that $Q_1 \neq Q_2$—for example, let $Q_1 \neq 0$ be an arbitrary rate matrix and let $Q_2 \coloneqq \alpha Q_1$, with $\alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}$—and consider a well-behaved stochastic process $P \in \mathbb{P}^W$ of which, for all $t \in (0, 1]$, the transition matrix $T_0^t$ is given by Equation (9) in Example 3.3.
   For now, we simply assume that this is possible. A formal proof for the existence of such a process requires some additional machinery, and we therefore postpone it to Example 5.1, where we construct a well-behaved continuous-time Markov chain that is compatible with Equation (9).
   The aim of the present example is to show that for any such process, the right-sided partial derivative $\partial_+ T_0^0$—which corresponds to choosing $t = 0$ and $u = \emptyset$ in Definition 4.7—does not exist. The reason for this is that—as we prove in [26, Appendix I]—for any $\lambda \in [1/3, 2/3]$, there is a sequence $\{\Delta_i\}_{i \in \mathbb{N}_0} \to 0^+$ such that

$$\lim_{i \to +\infty} \frac{1}{\Delta_i}(T_0^{\Delta_i} - I) = Q_\lambda \tag{20}$$

with $Q_\lambda \coloneqq \lambda Q_1 + (1 - \lambda)Q_2$. The reason why this indeed implies that $\partial_+ T_0^0$ does not exist, is because if it would exist, then Equation (20) would imply that $\partial_+ T_0^0 = Q_\lambda$ for all $\lambda \in [1/3, 2/3]$. The only way for this to be possible would be if $Q_1 = Q_2$, but that case was excluded in the beginning of this example.                                                                                          ◇

   Observe, therefore, that the problem is essentially that each of the two finite-difference expressions $1/\Delta(T_{t,x_u}^{t+\Delta} - I)$ and $1/\Delta(T_{t-\Delta,x_u}^t - I)$, parameterised in $\Delta$, can have multiple accumulation points as we take $\Delta$ to 0. Therefore, it will be more convenient to instead work with what we call *outer partial derivatives*. These can be seen as a kind of set-valued derivatives, containing all these accumulation points obtained as $\Delta \to 0^+$.

**Definition 4.8** *(Outer partial derivatives).* For any stochastic process $P \in \mathbb{P}$, any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the *right-sided outer partial derivative* of $T_{t,x_u}^t$ is defined by

$$\overline{\partial}_+ T_{t,x_u}^t \coloneqq \left\{ Q \in \mathscr{R} \colon \left( \exists \{\Delta_i\}_{i \in \mathbb{N}} \to 0^+ \colon \lim_{i \to +\infty} \frac{1}{\Delta_i}(T_{t,x_u}^{t+\Delta_i} - I) = Q \right) \right\} \tag{21}$$

and, if $t \neq 0$, the *left-sided outer partial derivative* of $T_{t,x_u}^t$ is defined by

$$\overline{\partial}_- T_{t,x_u}^t \coloneqq \left\{ Q \in \mathscr{R} \colon \left( \exists \{\Delta_i\}_{i \in \mathbb{N}} \to 0^+ \colon \lim_{i \to +\infty} \frac{1}{\Delta_i}(T_{t-\Delta_i,x_u}^t - I) = Q \right) \right\}. \tag{22}$$

Furthermore, the *outer partial derivative* of $T_{t,x_u}^t$ is defined as

$$\overline{\partial} T_{t,x_u}^t \coloneqq \overline{\partial}_+ T_{t,x_u}^t \cup \overline{\partial}_- T_{t,x_u}^t \text{ if } t > 0 \text{ and } \overline{\partial} T_{t,x_u}^t \coloneqq \overline{\partial}_+ T_{t,x_u}^t \text{ if } t = 0.$$

   For well-behaved processes $P \in \mathbb{P}^W$, as our next result shows, these outer partial derivatives are always non-empty, bounded and closed.

**Proposition 4.6.** *Consider any $P \in \mathbb{P}^W$. Then $\overline{\partial}_+ T_{t,x_u}^t$, $\overline{\partial}_- T_{t,x_u}^t$ and $\overline{\partial} T_{t,x_u}^t$ are non-empty, bounded and closed subsets of $\mathscr{R}$.*

   The following two examples provide this result with some intuition. Example 4.6 illustrates the validity of the result, while Example 4.7 shows that the requirement that $P$ must be well-behaved is essential for the result to be true.

**Example 4.6.** Consider again the well-behaved stochastic process $P \in \mathbb{P}^{\mathrm{W}}$ from Example 4.5 of which, for all $t \in (0, 1]$, the transition matrix $T_0^t$ is given by Equation (9). As we prove in [26, Appendix I], it holds for this particular process that

$$\overline{\partial}_+ T_0^0 = \left\{ Q_\lambda : \lambda \in [1/3, 2/3] \right\},$$

where, for every $\lambda \in [1/3, 2/3]$, $Q_\lambda \coloneqq \lambda Q_1 + (1 - \lambda) Q_2$ as in Example 4.5. ◇

**Example 4.7.** Fix any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$, let $\mathscr{T}$ be the transition matrix system of Example 3.4, and consider any stochastic process $P \in \mathbb{P}$ of which the corresponding family of transition matrices $\mathscr{T}_P$ is equal to $\mathscr{T}$.

For now, we simply assume that such a process exists. A formal proof again requires some additional machinery—as in Example 4.5—and we therefore postpone it to Example 5.1, where we construct a continuous-time Markov chain whose family of transition matrices $\mathscr{T}_P$ is equal to the transition matrix system $\mathscr{T}$.

As we prove in [26, Appendix I], for such a stochastic process $P$, the right-sided outer partial derivative $\overline{\partial}_+ T_0^0$ is then empty. ◇

We end this section with two additional properties of the outer partial derivatives of well-behaved stochastic processes. First, as we establish in our next result, they satisfy an $\varepsilon$–$\delta$ expression that is similar to the limit expression of a partial derivative.

**Proposition 4.7.** *Consider any well-behaved stochastic process $P \in \mathbb{P}^{\mathrm{W}}$. Then, for any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, any state assignment $x_u \in \mathscr{X}_u$, and any $\varepsilon > 0$, there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$:*

$$(\exists Q \in \overline{\partial}_+ T_{t, x_u}^t) \left\| \frac{1}{\Delta} (T_{t, x_u}^{t+\Delta} - I) - Q \right\| < \varepsilon \tag{23}$$

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

$$(\exists Q \in \overline{\partial}_- T_{t, x_u}^t) \left\| \frac{1}{\Delta} (T_{t-\Delta, x_u}^t - I) - Q \right\| < \varepsilon. \tag{24}$$

Secondly, these outer partial derivatives are a proper generalisation of directional partial derivatives. In particular, if the latter exist, their values correspond exactly to the single element of the former.

**Corollary 4.8.** *Consider any $P \in \mathbb{P}^{\mathrm{W}}$. Then $\overline{\partial}_+ T_{t, x_u}^t$ is a singleton if and only if $\partial_+ T_{t, x_u}^t$ exists and, in that case, $\overline{\partial}_+ T_{t, x_u}^t = \{\partial_+ T_{t, x_u}^t\}$. Analogous results hold for $\overline{\partial}_- T_{t, x_u}^t$ and $\partial_- T_{t, x_u}^t$, and for $\overline{\partial} T_{t, x_u}^t$ and $\partial T_{t, x_u}^t$.*

## 5. Continuous-time Markov chains

Having introduced continuous-time stochastic processes in Section 4, we will in this section focus on a specific class of such processes: *continuous-time Markov chains*.

**Definition 5.1** *(Markov property, Markov chain).* A stochastic process $P \in \mathbb{P}$ satisfies the *Markov property* if for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any time sequence $u \in \mathscr{U}_{<t}$, any $x_u \in \mathscr{X}_u$, and any states $x, y \in \mathscr{X}$:

$$P(X_s = y \mid X_t = x, X_u = x_u) = P(X_s = y \mid X_t = x).$$

A stochastic process that satisfies this property is called a *Markov chain*. We denote the set of all Markov chains by $\mathbb{P}^{\mathrm{M}}$ and use $\mathbb{P}^{\mathrm{WM}}$ to refer to the subset that only contains the well-behaved Markov chains.

We already know from Proposition 4.5 that the transition matrices of a stochastic process—and therefore also, in particular, of a Markov chain—satisfy some simple properties. For the specific case of a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$, the family of transition matrices $\mathscr{T}_P$ also satisfies an additional property. In particular, for any $t, r, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$, these transition matrices satisfy

$$T_t^s = T_t^r T_r^s. \tag{25}$$

In this context, this property is known as the *Chapman-Kolmogorov equation* or the semi-group property [39]. Indeed, this is the same semi-group property that we defined in Section 3 to hold for transition matrix systems $\mathscr{T}$. The following result should therefore not be surprising.

**Proposition 5.1.** *Consider a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ and let $\mathscr{T}_P$ be the corresponding family of transition matrices. Then $\mathscr{T}_P$ is a transition matrix system. Furthermore, $\mathscr{T}_P$ is well-behaved if and only if $P$ is well-behaved.*

At this point we know that every (well-behaved) Markov chain has a corresponding (well-behaved) transition matrix system. Our next result establishes that the converse is true as well: every (well-behaved) transition matrix system has a corresponding (well-behaved) Markov chain, and for a given initial distribution, this Markov chain is even unique.

**Theorem 5.2.** *Let $p$ be any probability mass function on $\mathscr{X}$ and let $\mathscr{T}$ be a transition matrix system. Then there is a unique Markov chain $P \in \mathbb{P}^M$ such that $\mathscr{T}_P = \mathscr{T}$ and, for all $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$. Furthermore, $P$ is well-behaved if and only if $\mathscr{T}$ is well-behaved.*

Hence, Markov chains—and well-behaved Markov chains in particular—are completely characterised by their transition matrices and their initial distribution. Our next example uses this result to formally establish the existence of the Markov chains that were used in Examples 4.5 and 4.7. Furthermore, it also illustrates that not every Markov chain is well-behaved.

**Example 5.1.** For any transition matrix system $\mathscr{T}$, it follows from Theorem 5.2—with $p$ chosen arbitrarily—that there exists a continuous-time Markov chain $P \in \mathbb{P}^M \subseteq \mathbb{P}$ such that $\mathscr{T}_P = \mathscr{T}$ and, furthermore, that $P$ is well-behaved if and only if $\mathscr{T}$ is well-behaved.

For example, for any rate matrix $Q \in \mathscr{R}$ such that $\|Q\| = 1$, if we let $\mathscr{T}$ be the transition matrix system of Example 3.4, we find—as already claimed in Example 4.7—that there is a continuous-time Markov chain $P \in \mathbb{P}^M \subseteq \mathbb{P}$ such that $\mathscr{T}_P = \mathscr{T}$. Furthermore, since we know from Example 3.4 that $\mathscr{T}$ is not well-behaved, it follows that $P$ is not well-behaved either.

As another example, for any two commuting rate matrices $Q_1, Q_2 \in \mathscr{R}$, if we let $\mathscr{T}$ be the well-behaved transition matrix system of Example 3.3, we find—as already claimed in Example 4.5—that there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{WM} \subseteq \mathbb{P}^M \subseteq \mathbb{P}$ such that, for all $t \in (0, 1]$, the transition matrix $T_0^t$ is given by Equation (9) in Example 3.3. ◇

As a final note, observe that not only does the Markov property simplify the conditional probabilities of a Markov chain, it also simplifies its dynamics. In particular, for any $t \in \mathbb{R}_{\geq 0}$, any $u \in \mathscr{U}_{<t}$, and any $x_u \in \mathscr{X}_u$, it holds that $\overline{\partial}_- T_{t, x_u}^t = \overline{\partial}_- T_t^t$, $\overline{\partial}_+ T_{t, x_u}^t = \overline{\partial}_+ T_t^t$ and $\overline{\partial} T_{t, x_u}^t = \overline{\partial} T_t^t$. We now focus on a number of special cases.

### 5.1. Homogeneous Markov chains

**Definition 5.2** *(Homogeneous Markov chain).* A Markov chain $P \in \mathbb{P}^M$ is called *time-homogeneous*, or simply *homogeneous*, if its transition matrices $T_t^s$ do not depend on the absolute value of $t$ and $s$, but only on the time-difference $s - t$:

$$T_t^s = T_0^{s-t} \quad \text{for all } t, s \in \mathbb{R}_{\geq 0} \text{ such that } t \leq s. \tag{26}$$

We denote the set of all homogeneous Markov chains by $\mathbb{P}^{HM}$ and use $\mathbb{P}^{WHM}$ to refer to the subset that consists of the well-behaved homogeneous Markov chains.

Recall now from Section 3 the exponential transition matrix system $\mathscr{T}_Q$ corresponding to some $Q \in \mathscr{R}$. As we have seen, the transition matrices of such a system were defined by $T_t^s = e^{Q(s-t)}$. This family $\mathscr{T}_Q$ therefore clearly satisfies Equation (26). Furthermore, by Proposition 3.6, $\mathscr{T}_Q$ is well-behaved. Hence, we have the following result.

**Corollary 5.3.** *Consider any rate matrix $Q \in \mathscr{R}$ and let $p$ be an arbitrary probability mass function on $\mathscr{X}$. Then there is a unique Markov chain $P \in \mathbb{P}^M$ such that $\mathscr{T}_P = \mathscr{T}_Q$ and, for all $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$. Furthermore, this unique Markov chain is well-behaved and homogeneous.*

Our next result strengthens this connection between well-behaved homogeneous Markov chains and exponential transition matrix systems.

**Theorem 5.4.** *For any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{WHM}$, there is a unique rate matrix $Q \in \mathscr{R}$ such that $\mathscr{T}_P = \mathscr{T}_Q$.*[2]

Hence, any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{WHM}$ is completely characterised by its initial distribution and a rate matrix $Q \in \mathscr{R}$. We will denote this rate matrix by $Q_P$.

The dynamic behaviour of well-behaved homogeneous Markov chains is furthermore particularly easy to describe, as shown by the next result.

---

[2] Although our proof for this result starts from scratch, this result is essentially well known. Our version of it should be regarded as a (re)formulation that is adapted to our terminology and notation and, in particular, to our use of coherent and/or full conditional probabilities.

**Proposition 5.5.** *Consider any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$ and let $Q_P \in \mathscr{R}$ be its corresponding rate matrix. Then $\partial T_t^t = \partial_+ T_t^t = \partial_- T_t^t = Q_P$ and $\overline{\partial} T_t^t = \overline{\partial}_+ T_t^t = \overline{\partial}_- T_t^t = \{Q_P\}$.*

### 5.2. Non-homogeneous Markov chains

In contrast to homogeneous Markov chains, a Markov chain for which Equation (26) does not hold is called—rather obviously—*non-homogeneous*. While we know from Theorem 5.4 that well-behaved homogeneous Markov chains can be characterised (up to an initial distribution) by a fixed rate matrix $Q \in \mathscr{R}$, this is not the case for well-behaved non-homogeneous Markov chains.

Instead, such systems are typically described by a function $Q_t$ that gives for each time point $t \in \mathbb{R}_{\geq 0}$ a rate matrix $Q_t \in \mathscr{R}$. For any such function $Q_t$, the existence and uniqueness of a corresponding non-homogeneous Markov chain then depend on the specific properties of $Q_t$. Rather than attempt to treat all these different cases here, we instead refer to some examples from the literature.

Typically, some kind of continuity of $Q_t$ in terms of $t$ is assumed. The specifics of these assumptions may then depend on the intended generality of the results, computational considerations, the domain of application, and so forth. For example, Reference [16] assumes that $Q_t$ is left-continuous and has bounded right-hand limits. As a stronger restriction, Reference [17] uses a collection $Q_1, \dots, Q_n$ of commuting rate matrices, and defines $Q_t$ as a weighted linear combination of these component rate matrices wherein the weights vary continuously with $t$. In Reference [15], a right-continuous and piecewise-constant $Q_t$ is used, meaning that $Q_t$ takes different values on various (half-open) intervals of $\mathbb{R}_{\geq 0}$, but fixed values within those intervals.

This idea of using a time-dependent rate matrix $Q_t$ has the advantage of being rather intuitive, but it is rather difficult to formalise. Essentially, the problem with this approach is that it does not allow us to distinguish between left and right derivatives. Intuitively, $Q_t$ is supposed to be 'the' derivative. However, this is impossible if $Q_t$ is discontinuous—for example in the piecewise constant case. Therefore, in our present work, instead of using a function $Q_t$, we will characterise non-homogeneous Markov chains by means of their transition matrix system and their initial distribution, making use of the results in Proposition 5.1 and Theorem 5.2.

One technique for constructing transition matrix systems that is particularly important for our work, and especially in our proofs, is to combine restrictions of exponential transition matrix systems to form a new transition matrix system that is, loosely speaking, piecewise constant. Example 3.1 provided a simple illustration of this technique. More generally, these transition matrix systems will be of the form

$$\mathscr{T}_{Q_0}^{[0,t_0]} \otimes \mathscr{T}_{Q_1}^{[t_0,t_1]} \otimes \cdots \otimes \mathscr{T}_{Q_n}^{[t_{n-1},t_n]} \otimes \mathscr{T}_{Q_{n+1}}^{[t_n,\infty)}. \tag{27}$$

For example, the transition matrix systems $\mathscr{T}_i$, $i \in \mathbb{N}_0$, that we defined in Equation (8) are all of this form. As we know from Propositions 3.6 and 3.9, transition matrix systems that are of this form are always well-behaved. The following result is therefore a trivial consequence of Theorem 5.2.

**Proposition 5.6.** *Let $p$ be an arbitrary probability mass function on $\mathscr{X}$, let $u = t_0, \dots, t_n$ be a finite sequence of time points in $\mathscr{U}_\emptyset$, and let $Q_0, \dots, Q_{n+1} \in \mathscr{Q}$ be a collection of rate matrices. Then there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}^{\mathrm{WM}}$ such that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$ and such that $\mathscr{T}_P$ is given by Equation (27).*

## 6. Imprecise continuous-time Markov chains

In Sections 4 and 5, we formalised stochastic processes and provided ways to characterise them. We now turn to the field of *imprecise probability* [24,22] to formalise the notion of an *imprecise continuous-time Markov chain*. Basically, rather than look at a single stochastic process $P \in \mathbb{P}$, we instead consider jointly some *set* of processes $\mathscr{P} \subseteq \mathbb{P}$. We start by looking at how such sets can be described.

### 6.1. Sets of consistent stochastic processes

Recall from Section 4.4 that for a given stochastic process $P \in \mathbb{P}$, its dynamics can be described by means of the outer partial derivatives $\overline{\partial} T_{t,x_u}^t$ of its transition matrices, which can depend both on the time $t \in \mathbb{R}_{\geq 0}$ and on the history $x_u \in \mathscr{X}_u$. Furthermore, we also found that—at least for well-behaved processes—these outer partial derivatives are non-empty bounded sets of rate matrices. If all the partial derivatives of a process belong to the same non-empty bounded set of rate matrices $\mathscr{Q}$, we call this process consistent with $\mathscr{Q}$.

**Definition 6.1** *(Consistency with a set of rate matrices).* Consider a non-empty bounded set of rate matrices $\mathscr{Q}$ and a stochastic process $P \in \mathbb{P}$. Then $P$ is said to be *consistent* with $\mathscr{Q}$ if

$$(\forall t \in \mathbb{R}_{\geq 0})(\forall u \in \mathscr{U}_{<t})(\forall x_u \in \mathscr{X}_u) : \overline{\partial} T_{t,x_u}^t \subseteq \mathscr{Q}.$$

If $P$ is consistent with $\mathscr{Q}$, we will write $P \sim \mathscr{Q}$.

Thus, when a process is consistent with a set of rate matrices $\mathscr{Q}$, we know that its dynamics can always be described using the rate matrices in that set. However, we do not know which of these rate matrices $Q \in \mathscr{Q}$ describe the dynamics at any given time $t \in \mathbb{R}_{\geq 0}$ or for any given history $x_u \in \mathscr{X}_u$. Furthermore, consistency of a process with a set of rate matrices $\mathscr{Q}$ does not tell us anything about the initial distribution of the process. Therefore, we also introduce the concept of consistency with a set of initial distributions $\mathscr{M}$.

**Definition 6.2** *(Consistency with a set of initial distributions).* Consider any non-empty bounded set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$ and any stochastic process $P \in \mathbb{P}$. We then say that $P$ is *consistent* with $\mathscr{M}$, and write $P \sim \mathscr{M}$, if $P(X_0) \in \mathscr{M}$.

In the remainder of this paper, we will focus on sets of processes that are jointly consistent with some given $\mathscr{Q}$ and $\mathscr{M}$. However, rather than look at the set of *all* processes consistent with some $\mathscr{Q}$ and $\mathscr{M}$, we will instead consider the consistent subset of some given set of processes $\mathscr{P} \subseteq \mathbb{P}$.

**Definition 6.3** *(Consistent subset of processes).* Consider a non-empty bounded set of rate matrices $\mathscr{Q}$ and a non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$ and a set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$. Then, the *subset of $\mathscr{P}$ consistent with $\mathscr{Q}$ and $\mathscr{M}$* is denoted by $\mathscr{P}_{\mathscr{Q},\mathscr{M}}$, and defined as

$$\mathscr{P}_{\mathscr{Q},\mathscr{M}} \coloneqq \{ P \in \mathscr{P} : P \sim \mathscr{Q}, \; P \sim \mathscr{M} \} \,.$$

When $\mathscr{M}$ is the set of *all* probability mass functions on $\mathscr{X}$, we will write $\mathscr{P}_{\mathscr{Q}}$ for the sake of brevity.

For some fixed $\mathscr{Q}$ and $\mathscr{M}$, different choices for $\mathscr{P}$ will result in different sets of consistent processes $\mathscr{P}_{\mathscr{Q},\mathscr{M}}$. Three specific choices of $\mathscr{P}$ will be particularly important in this paper because, as we will now show, they lead to three different types of imprecise continuous-time Markov chains.

### 6.2. Types of imprecise continuous-time Markov chains

In Sections 4 and 5, we introduced three sets $\mathbb{P}^{\mathrm{W}}$, $\mathbb{P}^{\mathrm{WM}}$ and $\mathbb{P}^{\mathrm{WHM}}$ of well-behaved stochastic processes with different qualitative properties. $\mathbb{P}^{\mathrm{W}}$ is the set of all well-behaved stochastic processes, $\mathbb{P}^{\mathrm{WM}}$ consists of the processes in $\mathbb{P}^{\mathrm{W}}$ that are continuous-time Markov chains, and $\mathbb{P}^{\mathrm{WHM}}$ is the set of all homogeneous Markov chains that are well-behaved, which is therefore a subset of $\mathbb{P}^{\mathrm{WM}}$. We now use Definition 6.3 to define three sets of consistent processes that also have these respective qualitative properties.

**Definition 6.4** *(Imprecise continuous-time Markov chain).* For any non-empty bounded set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we define the following three sets of stochastic processes that are jointly consistent with $\mathscr{Q}$ and $\mathscr{M}$:

- $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is the consistent set of all well-behaved stochastic processes;
- $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ is the consistent set of all well-behaved Markov chains;
- $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ is the consistent set of all well-behaved homogeneous Markov chains.

We call each of these three sets an *imprecise continuous-time Markov chain*, and abbreviate this as ICTMC.[3] Following Definition 6.3, we will write $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$ when we take $\mathscr{M}$ to be the set of *all* probability mass functions on $\mathscr{X}$, and similarly for $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q}}$.

Since the sets $\mathbb{P}^{\mathrm{WHM}}$, $\mathbb{P}^{\mathrm{WM}}$ and $\mathbb{P}^{\mathrm{W}}$ are nested, it should be clear that this also true for the corresponding types of ICTMC's.

**Proposition 6.1.** *Consider any bounded set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then,*

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

Observe furthermore that for $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$, the extra properties of their elements allow us to simplify the notion of consistency in Definition 6.1, which leads to the following alternative characterisations:

---

[3] For the set $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, one might wonder why we choose to call it an imprecise *Markov* chain, since it contains processes that do not satisfy the Markov property. As we will see in Section 8.1, this choice of terminology is motivated by that fact that the set $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ itself—rather than its elements—satisfies a so-called *imprecise Markov property*.

$$\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}} = \left\{ P \in \mathbb{P}^{\mathrm{WM}} : (\forall t \in \mathbb{R}_{\geq 0}) \quad \overline{\partial} T_t^t \subseteq \mathscr{Q}, \ P(X_0) \in \mathscr{M} \right\}, \tag{28}$$

and

$$\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}} = \left\{ P \in \mathbb{P}^{\mathrm{WHM}} : Q_P \in \mathscr{Q}, \ P(X_0) \in \mathscr{M} \right\}.$$

This first equality follows from the Markov property of the elements of $\mathbb{P}^{\mathrm{WM}}$, which ensures that $\overline{\partial} T_{t,x_u}^t = \overline{\partial} T_t^t$. The second equality follows from the Markov property and the homogeneity of the processes $P \in \mathbb{P}^{\mathrm{WHM}}$, which, by Proposition 5.5, ensures that $\overline{\partial} T_t^t = \{Q_P\}$ for all $t \in \mathbb{R}_{\geq 0}$.

The following example further illustrates the difference between the three types of ICTMC's that we consider.

**Example 6.1.** Let $Q_1$ and $Q_2$ be two different transition rate matrices and consider the set $\mathscr{Q} \coloneqq \{Q_1, Q_2\}$. Let furthermore $\mathscr{M} \coloneqq \{p\}$, with $p$ an arbitrary probability mass function on $\mathscr{X}$.

The set $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$ then contains exactly two stochastic processes $P_1$ and $P_2$, each of which is a well-behaved homogeneous Markov chain. They both have $p$ as their initial distribution, in the sense that $P_1(X_0) = P_2(X_0) = p$, but their transition rate matrices $Q_{P_i}$, $i \in \{1, 2\}$—whose existence is guaranteed by Theorem 5.4—are different: $Q_{P_1} = Q_1$, and $Q_{P_2} = Q_2$. The transition matrix systems of the Markov chains $P_1$ and $P_2$ are given by the exponential transition matrix systems $\mathscr{T}_{Q_1}$ and $\mathscr{T}_{Q_2}$, in that order, and the transition matrices of $P_1$ and $P_2$ are therefore given by ${}^1T_t^s = e^{Q_1(s-t)}$ and ${}^2T_t^s = e^{Q_2(s-t)}$, respectively.

Since $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$ is a subset of $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$, each of the two homogeneous Markov chains $P_1$ and $P_2$ belongs to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ as well. However, $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ also contains additional processes, which are not homogeneous and therefore do not belong to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$. For instance, for any $r > 0$, it follows from Proposition 6.2 further on that there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ that has $\mathscr{T}_{Q_1}^{[0,r]} \otimes \mathscr{T}_{Q_2}^{[r,+\infty)}$ as its transition matrix system. This Markov chain is clearly not homogeneous because $T_0^r = e^{Q_1 r}$ is different from $T_r^{2r} = e^{Q_2 r}$, and therefore it does not belong to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$.

Since $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ is a subset of $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$, each of the processes that we have considered so far belong to the set $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$ as well. However this latter set contains more complicated processes still. For instance, for any $u \in \mathscr{U}$ and any $x_u, y_u \in \mathscr{X}_u$ such that $x_u \neq y_u$, $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$ will for example contain a stochastic process $P$ such that, for all $t > u$, $\overline{\partial} T_{t,x_u}^t = \{Q_1\}$ and $\overline{\partial} T_{t,y_u}^t = \{Q_2\}$. For all $s > t > u$, the history-dependent transition matrices $T_{t,x_u}^s$ and $T_{t,y_u}^s$ of this process $P$ will be given by $T_{t,x_u}^s = e^{Q_1(s-t)}$ and $T_{t,y_u}^s = e^{Q_2(s-t)}$, which implies that this process $P$ does not satisfy the Markov property, and therefore, that it does not belong to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$.  $\diamond$

**Proposition 6.2.** *Consider any non-empty bounded set of rate matrices $\mathscr{Q}$ and let $\mathscr{M}$ be any non-empty set of probability mass functions on $\mathscr{X}$. Then for any $p \in \mathscr{M}$, any ordered finite sequence of time points $u = t_0, \ldots, t_n$ in $\mathscr{U}_\emptyset$ and any collection of rate matrices $Q_0, \ldots, Q_{n+1} \in \mathscr{Q}$, there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ such that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$ and such that $\mathscr{T}_P$ is given by Equation* (27).

We conclude this section with some notes about closure properties of the different types of ICTMC's that we consider, which are particularly useful for existence proofs. In particular, we focus on closure properties under recombination of known elements—colloquially, the "piecing together" of two or more processes to construct a new process that belongs to the same ICTMC.

The example above already suggested how to do this for $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$, by combining two well-behaved homogeneous Markov chains $P_1, P_2 \in \mathbb{P}^{\mathrm{WHM}}$ to form a new process $P \in \mathbb{P}^{\mathrm{WM}}$. Similarly, but more generally, for any two processes $P_1, P_2 \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$, we can combine their transition matrix systems $\mathscr{T}_{P_1}$ and $\mathscr{T}_{P_2}$ to construct a new transition matrix system $\mathscr{T} \coloneqq \mathscr{T}_{P_1}^{[0,r]} \otimes \mathscr{T}_{P_2}^{[r,+\infty)}$, with $r > 0$ chosen arbitrarily. Theorem 5.2 then guarantees that there exists a Markov chain $P$ such that $\mathscr{T}_P = \mathscr{T}$ and $P(X_0) = P_1(X_0)$. It is straightforward to verify that, because $P_1, P_2 \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$, also $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$; we leave this as an exercise for the reader.

Clearly, a similar procedure is impossible for $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$, because the combination of two processes would make the resultant one lose the homogeneity property, which is required to be an element of $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WHM}}$.

However, for our most general type of ICTMC, which is $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$, it turns out to be possible to recombine elements in an even more general, history-dependent way. That is, if $\mathscr{Q}$ is convex, then it is possible, for fixed time points $u \in \mathscr{U}$, to choose for every history $x_u \in \mathscr{X}_u$ a different process $P_{x_u} \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$, and to recombine these into a process $P$ that agrees with these $P_{x_u}$ conditional on the specific history $x_u$. Furthermore, the distribution on the time points $u$ can be chosen to agree with any element $P_\emptyset \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$. The following result guarantees that this new process $P$ will again belong to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$.

**Theorem 6.3.** *Consider a non-empty convex set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Fix a finite sequence of time points $u \in \mathscr{U}$. Choose any $P_\emptyset \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$ and, for all $x_u \in \mathscr{X}_u$, choose some $P_{x_u} \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$. Then there is a stochastic process $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}$ such that, for all $u_1, u_2 \subseteq u$ such that $u_1 < u_2$, all $x_u \in \mathscr{X}_u$ and all $A \in \mathscr{A}_u$:*

$$P(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) = P_\emptyset(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) \tag{29}$$

*and*

$$P(A|X_u = x_u) = P_{x_u}(A|X_u = x_u). \tag{30}$$

*6.3. Lower expectations for ICTMC's*

From a practical point of view, after having specified a (precise) stochastic process, one is typically interested in the expected value of some function of interest, or the probability of some event. Similarly, in this work, our main objects of consideration will be the *lower and upper expectations* and *lower and upper probabilities* that correspond to the ICTMC's that we introduced in the previous section.

**Definition 6.5** *(Lower expectation).* For any non-empty set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$, the *(conditional) lower expectation with respect to $\mathscr{P}$ is defined as*

$$\underline{\mathbb{E}}[\cdot \,|\, \cdot] \coloneqq \inf\{\mathbb{E}_P[\cdot \,|\, \cdot] : P \in \mathscr{P}\}, \tag{31}$$

*where $\mathbb{E}_P[\cdot \,|\, \cdot]$ denotes the (conditional) expectation taken with respect to $P$.*

In particular, for any non-empty bounded set of rate matrices $\mathscr{Q}$ and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we let

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}[\cdot \,|\, \cdot] \coloneqq \inf\left\{\mathbb{E}_P[\cdot \,|\, \cdot] : P \in \mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\right\}, \tag{32}$$

and similarly for $\underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\underline{\mathbb{E}}^{\mathsf{WHM}}_{\mathscr{Q},\mathscr{M}}$. If $\mathscr{M}$ is the set of *all* probability mass functions on $\mathscr{X}$, then as in Definition 6.3, we will write $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q}}$ instead of $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$, and similarly for $\underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q}}$ and $\underline{\mathbb{E}}^{\mathsf{WHM}}_{\mathscr{Q}}$.

Upper expectations can be defined analogously, simply by replacing the infimum by a supremum: $\overline{\mathbb{E}}[\cdot \,|\, \cdot] \coloneqq \sup\{\mathbb{E}_P[\cdot \,|\, \cdot] : P \in \mathscr{P}\}$. However, there is no need to study these upper expectations separately, because they are in one-to-one correspondence to lower expectations through the conjugacy relation $\overline{\mathbb{E}}[\cdot \,|\, \cdot] = -\underline{\mathbb{E}}[- \cdot \,|\, \cdot]$. Lower and upper probabilities also have analogous definitions. However, these too do not need to be studied separately, because they correspond to special cases of lower and upper expectations. For example, for any $u \in \mathscr{U}$, $x_u \in \mathscr{X}_u$ and $A \in \mathscr{A}_u$, we have that

$$\underline{P}(A|X_u = x_u) \coloneqq \inf\{P(A|X_u = x_u) : P \in \mathscr{P}\} = \inf\{\mathbb{E}_P[\mathbb{I}_A|X_u = x_u] : P \in \mathscr{P}\} = \underline{\mathbb{E}}[\mathbb{I}_A|X_u = x_u], \tag{33}$$

and similarly, we also have that $\overline{P}(A|X_u = x_u) = \overline{\mathbb{E}}[\mathbb{I}_A|X_u = x_u]$, which, due to conjugacy, implies that $\overline{P}(A|X_u = x_u) = -\underline{\mathbb{E}}[-\mathbb{I}_A|X_u = x_u]$. Hence, from a computational point of view, it clearly suffices to focus on lower expectations, which is what we will do in the remainder of this work.

In particular, one of the main aims of this paper is to provide methods to compute lower expectations for the different types of ICTMC's that we have introduced in Section 6.2; however, as we will see at the end of this section, doing this for the set $\mathbb{P}^{\mathsf{WHM}}_{\mathscr{Q},\mathscr{M}}$ is particularly difficult. We will therefore largely focus on performing these computations for the ICTMCs $\mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathsf{WM}}_{\mathscr{Q},\mathscr{M}}$. We start by giving some useful properties of the three types of lower expectations that we are interested in.

First of all, it can be shown that if $\mathscr{Q}$ is a non-empty, bounded, convex and closed set of rate matrices, then the conditional lower expectations $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q}}$, $\underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q}}$ and $\underline{\mathbb{E}}^{\mathsf{WHM}}_{\mathscr{Q}}$ are actually minima—rather than infima—because they are always reached by some element of their corresponding ICTMC. However, the proof of this claim is rather involved and requires some technical machinery that is outside of the scope of this paper; we intend to publish these and related results separately in future work.

Moving on, we know from Section 6.2 that the sets $\mathbb{P}^{\mathsf{WHM}}_{\mathscr{Q},\mathscr{M}}$, $\mathbb{P}^{\mathsf{WM}}_{\mathscr{Q},\mathscr{M}}$ and $\mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$ are nested subsets of each other. As an immediate consequence, their corresponding lower expectations provide (lower) bounds for each other.

**Proposition 6.4.** *Consider any non-empty bounded set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then,*

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}[\cdot \,|\, \cdot] \leq \underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q},\mathscr{M}}[\cdot \,|\, \cdot] \leq \underline{\mathbb{E}}^{\mathsf{WHM}}_{\mathscr{Q},\mathscr{M}}[\cdot \,|\, \cdot].$$

Generally speaking, the inequalities in this proposition can be—and often are—strict; for the first inequality, we will illustrate this further on in this section, in Fig. 1, whereas for the second inequality, this can be seen by comparing Examples 9.1 and 9.2 in Section 9.2.

For now, we first provide a useful property of the lower expectation $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$ that corresponds to $\mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$. To this end, consider a function $f \in \mathscr{L}(\mathscr{X}_{u \cup v \cup w})$ defined on the union $u \cup v \cup w$ of three finite sets of time points $u, v, w \in \mathscr{U}$ such that $u < v < w$. It then follows from the basic properties of expectations that for any stochastic process $P$, the corresponding expectation of $f$, conditional on $X_u$, decomposes as follows:

$$\mathbb{E}_P[f(X_u, X_v, X_w)|X_u] = \mathbb{E}_P\big[\mathbb{E}_P[f(X_u, X_v, X_w)|X_u, X_v]\,\big|\,X_u\big]. \tag{34}$$

This equality is well-known, and is called the *law of iterated expectation*. Rather remarkably, if $\mathscr{Q}$ is convex, then the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ satisfies a similar property. The proof for this so-called law of iterated *lower* expectation is based on Theorem 6.3.

**Theorem 6.5.** *Let $\mathscr{Q}$ be an arbitrary non-empty, bounded, and convex set of rate matrices, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for any $u, v, w \in \mathscr{U}$ such that $u < v < w$ and any $f \in \mathscr{L}(\mathscr{X}_{u \cup v \cup w})$:*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w) \,|\, X_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}\Big[\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w) \,|\, X_u, X_v] \,\Big|\, X_u \Big]. \tag{35}$$

The reason why this result is useful, is because it essentially allows us to compute lower expectations recursively. In particular, instead of computing a lower expectation for all time points simultaneously, we can focus on each of the time points separately, and eliminate them one by one. We will revisit this idea in Section 9, where we will use it to develop efficient algorithms.

For now, we would like to draw attention to the fact that Theorem 6.5 relies heavily on the fact that the individual elements of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ do not need to satisfy the Markov property, in the sense that they can have history-dependent probabilities. Therefore, unfortunately, the result in Theorem 6.5 does not extend to $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ or $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. Since Theorem 6.5 simplifies considerably the process of computing the lower expectation $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$, this suggests that the qualitative differences between the three types of ICTMC's that we consider make some of their corresponding lower expectations harder to compute than others. In fact, we will show in Section 9 that $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$ is in some sense the easiest set to do this for, exactly because of Theorem 6.5.

Furthermore, rather ironically, computing lower expectations for the set $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$—which, intuitively, is the simplest of our three types of ICTMC's—seems to be much harder than for the sets $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q},\mathscr{M}}$ or $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q},\mathscr{M}}$. The problem is, essentially, that while homogeneous Markov chains are easy to work with numerically, the set $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ does not provide enough "degrees of freedom" to easily solve the optimisation problem that is involved in computing $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$. The following serves as an illustration.

Suppose that we want to compute the lower expectation $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t) \,|\, X_0 = x_0]$ of some function $f \in \mathscr{L}(\mathscr{X})$ at time $t$, conditional on the information that the state $X_0$ at time 0 takes the value $x_0 \in \mathscr{X}_0$. It then follows from Definition 6.5, Remark 4.1 and Proposition 5.5 that

$$\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t) \,|\, X_0 = x_0] = \inf\Big\{\mathbb{E}_P[f(X_t) \,|\, X_0 = x_0] : P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}\Big\} = \inf\Big\{[e^{Q_P t} f](x_0) : P \in \mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}\Big\}$$
$$= \inf\Big\{[e^{Qt} f](x_0) : Q \in \mathscr{Q}\Big\}.$$

Therefore, computing $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t) \,|\, X_0 = x_0]$ is at its core a non-linear, constrained optimisation problem over the set $\mathscr{Q}$, where the non-linearity stems from the term $e^{Qt}$, and the specific form of the constraints depends on the choice of $\mathscr{Q}$. The following example illustrates the non-linearity of this type of optimisation problem in a simple case.

**Example 6.2.** Consider an ordered ternary state space $\mathscr{X} := \{a, b, c\}$, let $f \in \mathscr{L}(\mathscr{X})$ be defined as $f := [1 \ 0 \ 2]^\top$—in the sense that $f(a) := 1$, $f(b) := 0$ and $f(c) := 2$—and consider the set of transition rate matrices

$$\mathscr{Q} := \left\{ \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -0.01 & 0.01 \\ 0 & 0 & 0 \end{bmatrix} : \lambda \in [0.01, 0.5] \right\}.$$

Every process $P$ in the imprecise Markov chain $\mathbb{P}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}$ is then a homogeneous Markov chain of which the unique transition rate matrix $Q$ is completely determined by some $\lambda$ in $[0.01, 0.5]$. Furthermore, as we know from Remark 4.1 and Theorem 5.4, the conditional expectation $\mathbb{E}_P[f(X_t)|X_0 = a]$ that corresponds to this homogeneous Markov chain is equal to $[e^{Qt}f](a)$. Due to this equality, it is a matter of applying some basic—yet cumbersome—algebra to find that

$$\mathbb{E}_P[f(X_t)|X_0 = a] = \begin{cases} 2 + e^{-\lambda t} + 2\dfrac{e^{-\lambda t} - 100\lambda e^{-t/100}}{100\lambda - 1} & \text{if } \lambda \in (0.01, 0.5] \\[4mm] 2 - e^{-t/100} - \dfrac{1}{50} t e^{-t/100} & \text{if } \lambda = 0.01. \end{cases}$$

Obtaining the value of $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t) \,|\, X_0 = a]$ now corresponds to minimising this expression as $\lambda$ ranges over the interval $[0.01, 0.5]$. Fig. 1 illustrates that, even in this simple ternary case, this minimisation problem is already non-trivial, because—depending on the value of $t$—the minimum is not guaranteed to be obtained for one of the end points of $[0.01, 0.5]$, but may only be achieved by an internal point.

Nevertheless, in this simple case, $\underline{\mathbb{E}}^{\mathrm{WHM}}_{\mathscr{Q},\mathscr{M}}[f(X_t) \,|\, X_0 = a]$ can of course be closely approximated by taking a very fine discretisation of the interval $[0.01, 0.5]$, and performing an exhaustive search through this discretised parameter space to
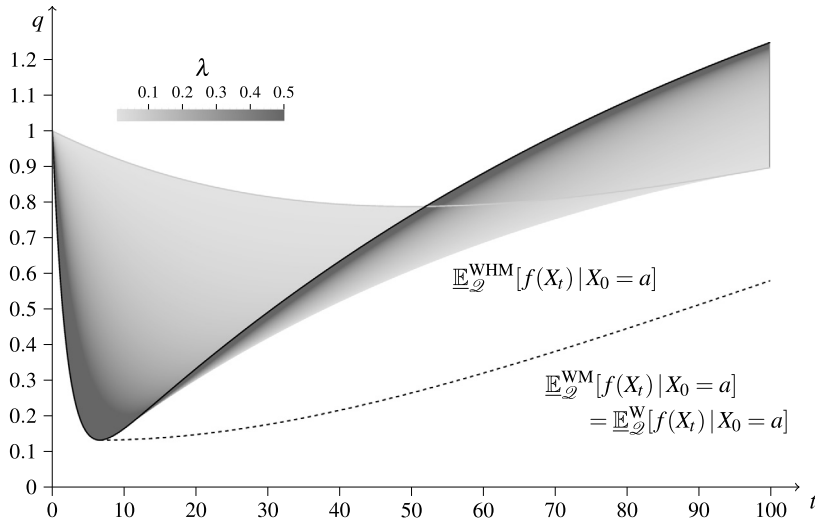
**Fig. 1.** Plot of the induced set of expected values $\mathbb{E}_P[f(X_t)|X_0 = a]$, for time points $t \in [0, 100]$, corresponding to all $P \in \mathbb{P}_{\mathscr{Q}}^{\text{WHM}}$ obtained as we vary $\lambda \in [0.01, 0.5]$. Observe that the lower expectation with respect to this set is initially reached by choosing $\lambda = 0.5$. This changes around the time point $t \approx 6.6$, after which the minimising value of $\lambda$ becomes a (changing) internal point of the interval $[0.01, 0.5]$. The dashed line corresponds to the lower expectation with respect to the sets $\mathbb{P}_{\mathscr{Q}}^{\text{WM}}$ and $\mathbb{P}_{\mathscr{Q}}^{\text{W}}$, which also include non-homogeneous Markov chains and, for $\mathbb{P}_{\mathscr{Q}}^{\text{W}}$, even more general processes. Note that the lower expectations with respect to $\mathbb{P}_{\mathscr{Q}}^{\text{WHM}}$, $\mathbb{P}_{\mathscr{Q}}^{\text{WM}}$ and $\mathbb{P}_{\mathscr{Q}}^{\text{W}}$ are all equal for $t < 6.6$. However, as time evolves further, the lower expectation with respect to $\mathbb{P}_{\mathscr{Q}}^{\text{WHM}}$ starts to diverge from the other two.

estimate the value of the lower expectation and the range of (precise) expectations that can be obtained within the set; this leads to the solution that is depicted in Fig. 1.

Fig. 1 also depicts $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{W}}[f(X_t)|X_0 = a]$ and $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{W}}[f(X_t)|X_0 = a]$, which happen to coincide. The reason why they coincide in this case, and the method by which we have computed them, will be explained in Section 8. For now, it suffices to notice that for large enough values of $t$, their common value differs from $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{WHM}}[f(X_t)|X_0 = a]$, thereby illustrating that the first inequality in Proposition 6.4 can indeed be strict. $\diamond$

While this example shows that for sufficiently "nice" sets $\mathscr{Q}$ and in low dimensions, a "discretise and exhaustive-search" method allows us to numerically solve the type of non-linear optimisation problems that are associated with $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{WHM}}$, the computational complexity of such an approach will in general quickly explode as the size of $\mathscr{X}$ increases.[4] Since we are not aware of other computational methods, this suggests that computing lower expectations for ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{WHM}}$ is typically very difficult, and it will therefore often be necessary to resort to approximation methods. For the particular case where $\mathscr{Q}$ corresponds to an interval matrix, such methods have been explored in, for example, References [40] and [41].

In the remainder of this paper, we will not consider ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{WHM}}$. Instead, we will focus on the lower expectations $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{WM}}$ and $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\text{W}}$ that correspond to ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{WM}}$ or $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{W}}$, and we will develop efficient methods for computing them. In order to do that, we start by introducing the notion of a lower transition operator.

## 7. Towards lower transition operators for ICTMC's

As explained in Remark 4.1, the transition matrix $T_t^s$ of a stochastic process $P$ serves as an alternative representation of the expectation operator $\mathbb{E}_P[f(X_s)|X_t]$, in the sense that

$$[T_t^s f](x_t) = \mathbb{E}_P[f(X_s)|X_t = x_t] \text{ for all } f \in \mathscr{L}(\mathscr{X}) \text{ and } x_t \in \mathscr{X}. \tag{36}$$

Furthermore, if $P$ is a well-behaved homogeneous Markov chain, then as explained in Section 5.1, $T_t^s$ is completely determined by a unique transition rate matrix $Q$, in the sense that $T_t^s = e^{Q(s-t)}$.

We will in this section introduce a generalisation of these transition matrices and transition rate matrices, called *lower transition operators* and *lower transition rate operators*, respectively. Furthermore, and most importantly, we will introduce the

---

[4] In the sense that in general, the dimensionality of the parameter-space grows quadratically in the number of states—since the number of elements of a rate matrix grows quadratically in this number, and we are considering sets of rate matrices—and hence the size of any discretisation of this parameter space (for a given precision) scales exponentially in $|\mathscr{X}|^2$. Clearly, performing an exhaustive search through such a discretised parameter-space will quickly become infeasible.

notion of a *corresponding lower transition operator*, which, much like in the precise homogeneous case, will be completely determined by some given lower transition rate operator.

Further on in this paper, we will then show that this type of lower transition operator serves as an alternative representation for the lower expectations of imprecise continuous-time Markov chains, thereby establishing an analogy with Equation (36). However, for now, in this section, we focus on introducing the relevant concepts, and on deriving this operator of interest. We end in Section 7.4 by showing that this operator satisfies a number of convenient properties.

## 7.1. Lower transition operators

The central concept that we will be interested in throughout this section is that of a lower transition operator $\underline{T}$.

**Definition 7.1** *(Lower transition operator).* A map $\underline{T}$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ is called a *lower transition operator* if, for all $f, g \in \mathscr{L}(\mathscr{X})$, all $\lambda \in \mathbb{R}_{\geq 0}$, and all $x \in \mathscr{X}$:

LT1: $[\underline{T} f](x) \geq \min\{f(y) : y \in \mathscr{X}\}$;
LT2: $[\underline{T}(f + g)](x) \geq [\underline{T} f](x) + [\underline{T} g](x)$;
LT3: $[\underline{T}(\lambda f)](x) = \lambda [\underline{T} f](x)$.

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

Such lower transition operators furthermore satisfy the following properties—see Reference [27] for a proof. For any lower transition operator $\underline{T}$, any $f, f_1, f_2 \in \mathscr{L}(\mathscr{X})$, any $\mu \in \mathbb{R}$ and any two non-negatively homogeneous operators $A, B$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$:

LT4: $\|\underline{T}\| \leq 1$;
LT5: $f_1 \geq f_2 \Rightarrow \underline{T} f_1 \geq \underline{T} f_2$;
LT6: $\underline{T}(f + \mu) = \underline{T}(f) + \mu$;
LT7: $\|\underline{T} A - \underline{T} B\| \leq \|A - B\|$.

**Proposition 7.1.** *For any two lower transition operators $\underline{T}, \underline{S} \in \underline{\mathbb{T}}$, their composition $\underline{T}\,\underline{S}$ is again a lower transition operator.*

The first thing to note is that any transition matrix $T$ will also satisfy properties LT1–LT3, and hence is also a lower transition operator. It is therefore clear that lower transition operators are a generalisation of transition matrices.

A first way to motivate this specific generalisation is to note the following. For any lower transition operator $\underline{T}$ and any $x \in \mathscr{X}$, consider the map $\underline{T}_x : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$, defined for all $f \in \mathscr{L}(\mathscr{X})$ as

$$\underline{T}_x f := [\underline{T} f](x). \tag{37}$$

Due to properties LT1–LT3, it then follows that $\underline{T}_x$ is a map from $\mathscr{L}(\mathscr{X})$ to $\mathbb{R}$ that is super-additive, non-negatively homogeneous, and bounded below by the minimum operator. By definition [22, Definition 2.3.3], that means that $\underline{T}_x$ is a *coherent lower prevision* on $\mathscr{L}(\mathscr{X})$. Note that the term "prevision" here is a synonym for "expectation"—be it with a different interpretation attached to it—so this essentially states that $\underline{T}_x$ is a "coherent lower expectation" on $\mathscr{L}(\mathscr{X})$.

For the reader that is unfamiliar with this notion of coherent lower previsions, this is perhaps best clarified as follows. Consider some arbitrary set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, and consider a map $\underline{\mathbb{E}} : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$, defined for all $f \in \mathscr{L}(\mathscr{X})$ as

$$\underline{\mathbb{E}} f := \inf\left\{ \sum_{x \in \mathscr{X}} p(x) f(x) : p \in \mathscr{M} \right\}.$$

We call this map $\underline{\mathbb{E}}$ the *lower envelope* of $\mathscr{M}$, and it should be clear that this map computes a lower expectation with respect to $\mathscr{M}$. Furthermore, note that for any $p \in \mathscr{M}$, the quantity $\sum_{x \in \mathscr{X}} p(x) f(x)$ is bounded from below by $\min\{f(y) : y \in \mathscr{X}\}$, and since this is true for any $p \in \mathscr{M}$, it follows that $\underline{\mathbb{E}} f$ is also bounded below by this minimum. Similarly, it is easily verified that $\underline{\mathbb{E}}$ is super-additive and non-negatively homogeneous, due to the properties of the inf operator. Hence, by the definition cited above, the lower expectation operator $\underline{\mathbb{E}}$ that corresponds to $\mathscr{M}$ is a coherent lower prevision on $\mathscr{L}(\mathscr{X})$.

Our consideration of the coherent lower prevision $\underline{T}_x$ above is essentially the same idea, but *without* considering explicitly any link to some set of probability mass functions $\mathscr{M}_x$. Suffice it to say that, for any coherent lower prevision $\underline{T}_x$, such a set $\mathscr{M}_x$ will always exist [42, Section 10.2]. It should furthermore be noted that the qualifier "coherent" here has connections to the notion of coherence that appeared in Section 4.1, and that it can also be given a direct gambling interpretation; we refer to [23,22] for further discussion on this.

In any case, for our present purposes, it suffices to realise that, because $\underline{T}_x$ is a coherent lower prevision for any $x \in \mathscr{X}$, the lower transition operator $\underline{T}$ can be seen as a vector of coherent lower previsions. Therefore, and because each of these coherent lower previsions $\underline{T}_x$ has some set $\mathscr{M}_x$ of probability mass functions of which $\underline{T}_x$ computes the lower envelope, we can combine these sets of probability mass functions to form a set of transition matrices

$$\mathbb{T}_{\underline{T}} = \{T \in \mathbb{T} \colon (\forall x \in \mathscr{X})\, T(x, \cdot) \in \mathscr{M}_x\}$$

of which $\underline{T}$ is the lower envelope, in the sense that

$$[\underline{T} f](x) = \inf\{[Tf](x) \colon T \in \mathbb{T}_{\underline{T}}\} \text{ for all } f \in \mathscr{L}(\mathscr{X}) \text{ and } x \in \mathscr{X}.$$

Due to the correspondence between transition matrices and conditional expectation operators—see Remark 4.1—it should therefore be clear that lower transition operators are an intuitive starting point to try and find alternative characterisations for the lower expectations that correspond to an ICTMC. What remains is to find the specific lower transition operator $\underline{T}$ whose set of dominating transition matrices $\mathbb{T}_{\underline{T}}$ corresponds to the set of transition matrices that is induced by a given ICTMC; the remainder of this section will provide the machinery required to do this.

We conclude with the following result about the set $\underline{\mathbb{T}}$ of all lower transition operators, which states that this set is a complete metric space with respect to our usual norm.

**Proposition 7.2.** *The metric space $(\underline{\mathbb{T}}, d)$ is complete with respect to the metric $d$ that is induced by our usual norm $\|\cdot\|$.*

*7.2. Lower transition rate operators*

We next focus on the generalisation of transition rate matrices $Q$ to lower transition rate operators $\underline{Q}$, as follows.

**Definition 7.2** *(Lower transition rate operator).* A map $\underline{Q}$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ is called a *lower transition rate operator* if, for all $f, g \in \mathscr{L}(\mathscr{X})$, all $\lambda \in \mathbb{R}_{\geq 0}$, all constant functions $\mu \in \mathscr{L}(\mathscr{X})$, and all $x \in \mathscr{X}$:

LR1: $\left[\underline{Q}\,\mu\right](x) = 0$;
LR2: $\left[\underline{Q}\,\mathbb{I}_y\right](x) \geq 0$ for all $y \in \mathscr{X}$ such that $x \neq y$;
LR3: $\left[\underline{Q}\,(f + g)\right](x) \geq \left[\underline{Q}\,f\right](x) + \left[\underline{Q}\,g\right](x)$;
LR4: $\left[\underline{Q}\,(\lambda f)\right](x) = \lambda\left[\underline{Q}\,f\right](x)$.

Such lower transition rate operators furthermore satisfy the following properties—see Reference [27] for a proof. For any lower transition rate operator $\underline{Q}$ and any two non-negatively homogeneous operators $A, B$ from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$:

LR5: $\left\|\underline{Q}\right\| \leq 2\max\left\{\left|[\underline{Q}\,\mathbb{I}_x](x)\right| \colon x \in \mathscr{X}\right\} < +\infty$;
LR6: $\left\|\underline{Q}\,A - \underline{Q}\,B\right\| \leq 2\left\|\underline{Q}\right\| \|A - B\|$.

Note that properties LR1 and LR2 essentially preserve properties R1 and R2 from Definition 3.2. The main difference lies in the fact that a rate matrix $Q$ is a linear map, whereas properties LR3 and LR4 merely require that a lower transition rate operator $\underline{Q}$ is super-additive and non-negatively homogeneous. Therefore, every rate matrix is clearly a lower transition rate operator, with the latter concept providing a generalisation of the former.

A first reason why this specific generalisation is of interest, is because it preserves the relation between transition matrices and rate matrices that was established in Propositions 3.2 and 3.3. Indeed, the following two results generalise these relations to our current setting.

**Proposition 7.3** *(Reference [27, Proposition 5]).* *Consider any lower transition rate operator $\underline{Q}$, and any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta \left\|\underline{Q}\right\| \leq 1$. Then the operator $(I + \Delta\underline{Q})$ is a lower transition operator.*

**Proposition 7.4** *(Reference [27, Proposition 6]).* *Consider any lower transition operator $\underline{T}$, and any $\Delta \in \mathbb{R}_{>0}$. Then the operator $1/\Delta(\underline{T} - I)$ is a lower transition rate operator.*

Now, in our discussion of lower transition operators in Section 7.1, we mentioned that lower transition operators can be interpreted as the lower envelope of a set of transition matrices. As we are about to show, there is a similar connection between lower transition rate operators and sets of rate matrices.

So consider any non-empty bounded set $\mathscr{Q} \subseteq \mathscr{R}$ of rate matrices. Then for any $f \in \mathscr{L}(\mathscr{X})$, if we let

$$[\underline{Q}\,f](x) := \inf\{[Q f](x) \colon Q \in \mathscr{Q}\} \text{ for all } x \in \mathscr{X}, \tag{38}$$

the resulting function $\underline{Q}\,f$ is again an element of $\mathscr{L}(\mathscr{X})$,[5] and therefore, $\underline{Q}$ is a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$. We call this operator $\underline{Q}$, as defined by Equation (38), the *lower envelope* of $\mathscr{Q}$. It is a matter of straightforward verification to see that $\underline{Q}$ is a lower transition rate operator.

---

[5] Since $\mathscr{Q}$ is bounded, N11 implies that, for all $Q \in \mathscr{Q}$, $\|Q f\| \leq \|Q\| \|f\| \leq \|\mathscr{Q}\| \|f\| < +\infty$. Therefore, and since $\mathscr{Q}$ is non-empty, the components of $\underline{Q}\,f$ are bounded below by $-\|\mathscr{Q}\| \|f\|$, which implies that $\underline{Q}\,f$ is a real-valued function on $\mathscr{X}$.

**Proposition 7.5.** *For any non-empty bounded set $\mathcal{Q} \subseteq \mathcal{R}$ of rate matrices, the corresponding operator $\underline{Q} : \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$, as defined by Equation* (38)*, is a lower transition rate operator.*

Inspired by this result, we will also refer to the lower envelope of $\mathcal{Q}$ as the *lower transition rate operator that corresponds to $\mathcal{Q}$*. However, this correspondence is not one-to-one. As the following example establishes, different non-empty bounded sets of rate matrices may have the same corresponding lower transition rate operator.

**Example 7.1.** For the sake of simplicity, we assume that the state space $\mathcal{X}$ has only two elements, which allows us to work with $2 \times 2$ matrices. Consider now two rate matrices

$$A := \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \quad \text{and} \quad B := \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix},$$

let $C := 1/2(A + B)$ be their convex mixture, which is clearly also a rate matrix, and use these matrices to define the sets $\mathcal{Q}_1 := \{A, B\}$ and $\mathcal{Q}_2 := \{A, B, C\}$. Then clearly, $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are two different, non-empty and bounded sets of rate matrices. Let $\underline{Q}_1$ and $\underline{Q}_2$ be the lower transition rate operators that correspond to $\mathcal{Q}_1$ and $\mathcal{Q}_2$, respectively. Then as we prove in [26, Appendix I], these lower transition rate operators are identical, i.e. $\underline{Q}_1 = \underline{Q}_2$.  ◇

Of course, this should not really be surprising—Example 7.1 essentially establishes that different sets can have the same infimum. However, it does lead to a natural question: what do these different sets have in common?

Therefore, we next consider some fixed lower transition rate operator $\underline{Q}$. All the non-empty bounded sets $\mathcal{Q}$ of rate matrices that have $\underline{Q}$ as their lower envelope then share a common property: they consist of rate matrices $Q$ that dominate $\underline{Q}$, in the sense that $Qf \geq \underline{Q}f$ for all $f \in \mathcal{L}(\mathcal{X})$. Therefore, each of these sets $\mathcal{Q}$ is contained in the following set of dominating rate matrices:

$$\mathcal{Q}_{\underline{Q}} := \left\{ Q \in \mathcal{R} : Qf \geq \underline{Q}f \text{ for all } f \in \mathcal{L}(\mathcal{X}) \right\}. \tag{39}$$

As our next result shows, this set $\mathcal{Q}_{\underline{Q}}$ is non-empty and bounded, and has $\underline{Q}$ as its lower envelope. Even stronger, the infimum in Equation (38) is reached—can be replaced by a minimum.

**Proposition 7.6.** *Consider a lower transition rate operator $\underline{Q}$ and let $\mathcal{Q}_{\underline{Q}}$ be the corresponding set of dominating rate matrices, as defined by Equation* (39)*. Then $\mathcal{Q}_{\underline{Q}}$ is non-empty and bounded and, for all $f \in \mathcal{L}(\mathcal{X})$, there is some $Q \in \mathcal{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$.*

Because of this result, and since—as discussed above—every non-empty bounded set of rate matrices that has $\underline{Q}$ as its lower envelope is a subset of $\mathcal{Q}_{\underline{Q}}$, it follows that $\mathcal{Q}_{\underline{Q}}$ is the largest non-empty bounded set of rate matrices that has $\underline{Q}$ as its lower envelope. Furthermore, as we show in Proposition 7.7 below, this set $\mathcal{Q}_{\underline{Q}}$ is also closed and convex, and has what we call *separately specified rows*.

Intuitively, we say that a set $\mathcal{Q}$ of rate matrices has separately specified rows if it is closed under taking arbitrary combinations of rows from its elements. More formally, if for every $x \in \mathcal{X}$ we let $\mathcal{Q}_x := \{Q(x, \cdot) : Q \in \mathcal{Q}\}$ denote the set of $x$-rows of the matrices in $\mathcal{Q}$, then we say that $\mathcal{Q}$ has separately specified rows if $\mathcal{Q}$ contains every matrix $Q$ that can be constructed by selecting, for all $x \in \mathcal{X}$, an arbitrary row $Q(x, \cdot)$ from $\mathcal{Q}_x$.

**Definition 7.3.** A set of rate matrices $\mathcal{Q} \subseteq \mathcal{R}$ has separately specified rows if

$$\mathcal{Q} = \{Q \in \mathcal{R} : (\forall x \in \mathcal{X}) \, Q(x, \cdot) \in \mathcal{Q}_x\},$$

where, for every $x \in \mathcal{X}$, $\mathcal{Q}_x := \{Q(x, \cdot) : Q \in \mathcal{Q}\}$ is some given set of rows from which the $x$-row $Q(x, \cdot)$ of the rate matrices $Q$ in $\mathcal{Q}$ is selected, independently of the other rows.[6]

**Proposition 7.7.** *Consider a lower transition rate operator $\underline{Q}$ and let $\mathcal{Q}_{\underline{Q}}$ be the corresponding set of dominating rate matrices, as defined by Equation* (39)*. Then $\mathcal{Q}_{\underline{Q}}$ is closed and convex, and has separately specified rows.*

These additional properties characterise $\mathcal{Q}_{\underline{Q}}$ completely, in the sense that no other set satisfies them.

**Proposition 7.8.** *Consider any non-empty, bounded, closed and convex set of rate matrices $\mathcal{Q} \subseteq \mathcal{R}$ with separately specified rows that has $\underline{Q}$ as its lower envelope. Then $\mathcal{Q} = \mathcal{Q}_{\underline{Q}}$.*

---

[6] This concept is related to the separately specified local models that are often used in credal networks—see, e.g., Reference [43]—and, for readers that are familiar with the literature on this latter subject, can be regarded as the continuous-time analogue of this notion.

**Example 7.2.** Let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be constructed as in Example 7.1, and let $\underline{Q} := \underline{Q}_1 = \underline{Q}_2$ be their common lower transition rate operator. As we are about to show, the corresponding set of dominating rate matrices $\mathcal{Q}_{\underline{Q}}$ is then equal to

$$\mathcal{Q}^* := \{Q \in \mathcal{R} : (\forall x \in \mathcal{X}) \, Q(x, \cdot) \in \mathcal{Q}_x\}, \tag{40}$$

where, for all $x \in \mathcal{X}$, $\mathcal{Q}_x$ is given by

$$\mathcal{Q}_x := \{\lambda A(x, \cdot) + (1 - \lambda) B(x, \cdot) : \lambda \in [0, 1]\}. \tag{41}$$

First of all, as we shown in [26, Appendix I], $\underline{Q}$ is also the lower transition operator corresponding to $\mathcal{Q}^*$. Furthermore, $\mathcal{Q}^*$ is clearly non-empty, bounded, closed, convex, has separately specified rows and, as we have just mentioned, has $\underline{Q}$ as its lower transition rate operator. Therefore, by Proposition 7.8, it follows that $\mathcal{Q}^* = \mathcal{Q}_{\underline{Q}}$.                                  $\diamond$

We conclude from all of this that non-empty bounded sets of rate matrices are more informative than lower transition rate operators, in the following sense. Different non-empty bounded sets of rate matrices $\mathcal{Q}$ may have the same lower transition rate operator $\underline{Q}$ and therefore, in general, knowledge of $\underline{Q}$ does not suffice to reconstruct $\mathcal{Q}$; we can only reconstruct an outer approximation $\mathcal{Q}_{\underline{Q}}$, which is guaranteed to include $\mathcal{Q}$. This changes if, besides non-empty and bounded, $\mathcal{Q}$ is also closed and convex and has separately specified rows. In that case, $\underline{Q}$ serves as an alternative representation for $\mathcal{Q}$ because, since $\mathcal{Q} = \mathcal{Q}_{\underline{Q}}$, we can use $\underline{Q}$ to reconstruct $\mathcal{Q}$. In other words: there is a one-to-one correspondence between lower transition rate operators and non-empty, bounded, closed and convex sets of rate matrices that have separately specified rows.

### 7.3. Corresponding lower transition operators

Proposition 7.3 already established that we can fairly easily construct a lower transition operator from a given lower transition rate operator $\underline{Q}$: if $\Delta \geq 0$ is sufficiently small, then $I + \Delta \underline{Q}$ will be a lower transition operator. In this section, we construct a somewhat more complicated lower transition operator from a given lower transition rate operator, and it is this specific lower transition operator on which we will focus for the remainder of this work. In particular, we will introduce the *lower transition operator corresponding to a given lower transition rate operator*.

To this end, we will assume here that we are given some arbitrary lower transition rate operator $\underline{Q}$, and any two time points $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$. For any $u \in \mathcal{U}_{[t,s]}$ such that $u = t_0, \ldots, t_n$, we then define the auxiliary operator

$$\Phi_u := \prod_{i=1}^{n}(I + \Delta_i \underline{Q}), \tag{42}$$

where, as in Section 2.1, for every $i \in \{1, \ldots, n\}$, $\Delta_i = t_i - t_{i-1}$ denotes the difference between two consecutive time points in $u$, and $\sigma(u) := \max\{\Delta_i : i \in \{1, \ldots, n\}\}$ is the maximum such difference. Clearly, if $\sigma(u)$ is small enough, Proposition 7.3 guarantees that each of the terms $I + \Delta_i \underline{Q}$ is a lower transition operator, and it then follows from Proposition 7.1 that $\Phi_u$—since it is a composition of lower transition operators—is also a lower transition operator. The so-called *lower transition operator corresponding to $\underline{Q}$* will be defined below as the limit of these lower transition operators $\Phi_u$, obtained as we take $u$ to be an increasingly finer partition of the interval $[t, s]$.

However, before we can do that, we first need to establish that this limit indeed exists. To this end, we start by providing a bound on the distance between two operators $\Phi_u$ and $\Phi_{u*}$.

**Proposition 7.9.** *Consider any $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, any $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\| \underline{Q} \right\| \leq 1$, and any $u, u^* \in \mathcal{U}_{[t,s]}$ such that $\sigma(u) \leq \delta$ and $\sigma(u^*) \leq \delta$. Let $C := s - t$. Then*

$$\|\Phi_u - \Phi_{u*}\| \leq 2\delta C \left\| \underline{Q} \right\|^2.$$

Note, therefore, that the distance $\|\Phi_u - \Phi_{u*}\|$ vanishes as we make $\sigma(u)$ and $\sigma(u^*)$ smaller and smaller. This allows us to state the following result.

**Corollary 7.10.** *For every sequence $\{u_i\}_{i \in \mathbb{N}}$ in $\mathcal{U}_{[t,s]}$ such that $\lim_{i \to \infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i \in \mathbb{N}}$ is Cauchy.*

Since we already know that, for partitions $u$ that are sufficiently fine, $\Phi_u$ is a lower transition operator, Proposition 7.2 now implies that this Cauchy sequence converges to a limit, and that this limit is again a lower transition operator.

**Corollary 7.11.** *For every sequence $\{u_i\}_{i \in \mathbb{N}}$ in $\mathcal{U}_{[t,s]}$ such that $\lim_{i \to \infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i \in \mathbb{N}}$ converges to a lower transition operator.*

Finally, as our next result establishes, this limit is unique, in the sense that it is independent of the choice of $\{u_i\}_{i \in \mathbb{N}}$.

**Theorem 7.12.** *For any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and any lower transition rate operator $\underline{Q}$, there is a unique lower transition operator $\underline{T} \in \underline{\mathbb{T}}$ such that*

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall u \in \mathscr{U}_{[t,s]} : \sigma(u) \leq \delta)\ \left\| \underline{T} - \Phi_u \right\| \leq \varepsilon. \tag{43}$$

Note that the $\varepsilon$–$\delta$ expression in Theorem 7.12 is a limit statement. Specifically, it is a limit of operators $\Phi_u$ corresponding to increasingly finer partitions $u$ of the interval $[t, s]$. In the sequel, whenever such a unique limit exists and equals some lower transition operator $\underline{T}$, we will denote it as

$$\lim_{\sigma(u) \to 0} \left\{ \Phi_u : u \in \mathscr{U}_{[t,s]} \right\} = \underline{T}. \tag{44}$$

Here, the notation is understood to indicate that the limit of these operators $\Phi_u$ is independent of the exact choice of $\{u_i\}_{i \in \mathbb{N}}$ in $\mathscr{U}_{[t,s]}$, so long as $\lim_{i \to \infty} \sigma(u_i) = 0$.

We are now ready to define the *lower transition operator corresponding to $\underline{Q}$*, which is the operator in which we will be interested for the remainder of this work.

**Definition 7.4** (*Corresponding lower transition operator*). Consider any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and let $\underline{Q}$ be an arbitrary lower transition rate operator. The *corresponding lower transition operator $\underline{L}_t^s$* is a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\overline{\mathscr{X}})$, defined by

$$\underline{L}_t^s := \lim_{\sigma(u) \to 0} \left\{ \Phi_u : u \in \mathscr{U}_{[t,s]} \right\},$$

where the limit is understood as in Equation (44).

### 7.4. Properties of corresponding lower transition operators

We will next establish that this operator $\underline{L}_t^s$ satisfies a number of convenient properties. In particular, we will focus on the family $\underline{\mathscr{T}}_{\underline{Q}}$ of lower transition operators corresponding to a given lower transition rate operator $\underline{Q}$.

**Definition 7.5** (*Lower transition operator system*). Let $\underline{Q}$ be an arbitrary lower transition rate operator. Then, the *lower transition operator system* corresponding to $\underline{Q}$ is the family $\underline{\mathscr{T}}_{\underline{Q}}$ of lower transition operators $\underline{L}_t^s$ corresponding to $\underline{Q}$, defined for all $t, s \in \mathbb{R}_{\geq 0}$, with $t \leq s$, as in Definition 7.4.

Our first result is that this family $\underline{\mathscr{T}}_{\underline{Q}}$ satisfies the same semi-group property that was found to hold for the transition matrix system $\mathscr{T}_P$ of a Markov chain $P \in \mathbb{P}^{\mathrm{M}}$ in Section 5.

**Proposition 7.13.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, r, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$, it holds that*

$$\underline{L}_t^s = \underline{L}_t^r \underline{L}_r^s.$$

*Furthermore, for all $t \in \mathbb{R}_{\geq 0}$, we have that $\underline{L}_t^t = I$.*

Our next result is that this family $\underline{\mathscr{T}}_{\underline{Q}}$ is time-homogeneous:

**Proposition 7.14.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, we have that $\underline{L}_t^s = \underline{L}_0^{s-t}$.*

Finally, we find that the derivatives of these lower transition operators always exist, and that they furthermore satisfy the following simple equalities.

**Proposition 7.15.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, it holds that[7]*

$$\frac{\partial}{\partial t} \underline{L}_t^s = -\underline{Q}\,\underline{L}_t^s \quad and \quad \frac{\partial}{\partial s} \underline{L}_t^s = \underline{Q}\,\underline{L}_t^s.$$

---

[7] If $0 = t < s$, the derivative with respect to $t$ is taken to be a right derivative. If $t = s$, the derivative with respect to $s$ is taken to be a right derivative and the derivative with respect to $t$ is taken to be a left derivative (or becomes meaningless if $t = 0$).

We would like to point out here that the derivatives in this result are not taken pointwise, but are taken with respect to the operator norm. For example, for $t = 0$ and $s > 0$, Proposition 7.15 does not state that

$$\frac{d}{ds} L_0^s f = \underline{Q}\, L_0^s f \quad \text{for all } f \in \mathscr{L}(\mathscr{X}), \tag{45}$$

but rather that

$$\lim_{\Delta \to 0} \left\| \frac{L_0^{s+\Delta} - L_0^s}{\Delta} - \underline{Q}\, L_0^s \right\| = 0. \tag{46}$$

Of these two statements, the latter is the strongest one, in the sense that it trivially implies the former. Hence, although from an intuitive point of view, the reader may wish to interpret the results in Proposition 7.15 as in Equation (45)—which would be correct—one should keep in mind that from a technical point of view, the result is in fact stronger, and is intended to be read as in Equation (46).

It is also worth noting that, as a consequence of Propositions 7.15 and 7.13, the operator $L_t^s$ satisfies the following differential equation:

$$\frac{\partial}{\partial s} L_t^s = \underline{Q}\, L_t^s\,, \qquad L_t^t = I\,.$$

Observe, therefore, the strong correspondence between the operator $L_t^s$ corresponding to some $\underline{Q}$, and the matrix exponential $e^{Q\,(s-t)}$ of a rate matrix $Q \in \mathscr{R}$. In particular, as is very well known [29, Equation 4.4], this matrix exponential is the unique solution of the differential equation

$$\frac{\partial}{\partial s} e^{Q\,(s-t)} = Q\, e^{Q\,(s-t)}\,, \qquad e^{Q\,(t-t)} = I\,.$$

Now, recall that any rate matrix $Q$ is also a lower transition rate operator. It then follows from the above that the lower transition operator $L_t^s$ that corresponds to this $\underline{Q} = Q$ is given by $L_t^s = e^{Q\,(s-t)}$. Hence, more generally, $L_t^s$ can be regarded as a generalised version of the matrix exponential of a transition rate matrix, and—with some slight abuse of terminology—can be considered to be the 'matrix exponential' of the lower transition rate operator $\underline{Q}$.

Another closely related observation is that, if $\underline{Q} = Q$ for some $Q \in \mathscr{R}$, then the family $\underline{\mathscr{T}}_{\underline{Q}}$ is equal to $\mathscr{T}_Q$, which is the exponential transition matrix system from Definition 3.5 that, by Corollary 5.3, is known to correspond to a well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\text{WHM}}$. Interestingly, then, the family $\underline{\mathscr{T}}_{\underline{Q}}$ maintains the convenient properties of differentiability, time-homogeneity, and "Markovian-like" factorisation, when instead of some rate matrix $Q$ we replace it by a lower transition rate operator $\underline{Q}$.

Mathematical niceties aside, we are of course not really interested in the trivial case where $\underline{Q} = Q$. Instead, we wish to use the lower transition operator $L_t^s$ to compute lower expectations for imprecise continuous-time Markov chains. We will show in the next section that this is indeed possible.

## 8. Connecting ICTMC's and lower transition operators

As we know from Section 7.1, lower transition operators are essentially just lower envelopes of transition matrices. Combined with the fact that transition matrices are a convenient tool for representing and computing expectations in a Markov chain, it seems intuitive to expect that, similarly, lower transition operators can be used to represent and compute lower expectations in an imprecise Markov chain. We will show in this section that this is indeed the case.

In particular, we establish in this section that for ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\text{WM}}$ or $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\text{W}}$, with $\mathscr{Q}$ having separately specified rows, we can use the lower transition operator $L_t^s$ to represent and compute conditional lower expectations of functions $f(X_s)$ that depend on the state $X_s$ at a single time point $s$ in the future. The treatment of more general functions is deferred to Section 9.

### 8.1. Lower transition operators as a representational tool

In order to establish a connection between the operator $L_t^s$ and the lower expectations that correspond to an ICTMC, it is important to realise that the latter is derived from a set $\mathscr{Q}$ of transition rate matrices—as in Definition 6.4—whereas the former is derived from a lower transition rate operator $\underline{Q}$—as in Definition 7.4. Therefore, we clearly need to start by creating a link between $\mathscr{Q}$ and $\underline{Q}$. Fortunately, we have already seen in Section 7.2 that there is a strong connection between sets of rate matrices $\mathscr{Q}$ and lower transition rate operators $\underline{Q}$. In particular, any set $\mathscr{Q}$ has a corresponding lower transition rate operator $\underline{Q}$, which computes the lower envelope with respect to $\mathscr{Q}$, as in Equation (38). It is exactly this connection between sets of transition rate matrices and lower transition rate operators that we will use here to establish a connection between the operator $L_t^s$ and the lower expectations that correspond to an ICTMC.

To start with, as the following result shows, for any lower transition rate operator $\underline{Q}$, the corresponding lower transition operator $L_t^s$ provides a lower bound on the conditional expectations $\mathbb{E}_P[f(X_s) \mid X_t = x_t, X_u = x_u]$ of any well-behaved stochastic process $P \in \mathbb{P}_{\mathscr{Q}}^{\text{W}}$ that is consistent with a set of rate matrices $\mathscr{Q}$ that has $\underline{Q}$ as its the lower envelope.

**Proposition 8.1.** *Consider a non-empty bounded set of rate matrices $\mathscr{Q}$ whose corresponding lower transition rate operator is $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then, for any $P \in \mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any $u \in \mathscr{U}_{<t}$, any $x_t \in \mathscr{X}$ and $x_u \in \mathscr{X}_u$, and any $f \in \mathscr{L}(\mathscr{X})$:*

$$\mathbb{E}_P[f(X_s) \mid X_t = x_t, X_u = x_u] \geq [L_t^s f](x_t).$$

Notice that this result is stated for stochastic processes $P$ in $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, whose initial distributions $P(X_0)$ are not required to belong to some given set of initial distributions $\mathscr{M}$. However, of course, since $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$ is a clearly a subset of $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$, the same result also holds for any choice of such $\mathscr{M}$.

Our next result establishes that the bound in Proposition 8.1 is tight if $\mathscr{Q}$ has separately specified rows. Specifically, we show that $L_t^s f$ can then be approximated to arbitrary precision by carefully choosing a Markov process $P$ from the set $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$.

**Proposition 8.2.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, and all $\varepsilon \in \mathbb{R}_{>0}$, there is a well-behaved Markov chain $P \in \mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$ such that*

$$\left| \mathbb{E}_P[f(X_s) \mid X_t = x_t] - [L_t^s f](x_t) \right| < \varepsilon \text{ for all } x_t \in \mathscr{X}.$$

Together, Propositions 8.1 and 8.2 establish a strong connection between the operator $L_t^s$ and the lower expectations that correspond to $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$ or $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$. In particular, for $\mathscr{Q}$ with separately specified rows, and for functions $f(X_s)$ that depend on the state $X_s$ at a single time point $s$ in the future, these three objects end up being identical.

**Corollary 8.3.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$ and $x_t \in \mathscr{X}$, and all $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t, X_u = x_u] = \left[L_t^s f\right](x_t).$$

Hence, we find that there is indeed a correspondence between the operator $L_t^s$ and the lower expectations that correspond to ICTMC's.

This result also helps to clarify why we choose to call $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$ an imprecise Markov chain, despite the fact that it contains processes that do not satisfy the Markov property. In order to see that, observe that Corollary 8.3 holds for *all* histories $x_u \in \mathscr{X}_u$ and *any* sequence of time points $u \in \mathscr{U}_{<t}$. Therefore, and because the definition of $L_t^s$ does not depend on this choice of $u$ and $x_u$, it follows that for $\mathscr{Q}$ with separately specified rows:

$$\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t] \tag{47}$$

and

$$\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_s) \mid X_t = x_t]. \tag{48}$$

In other words, the conditional lower expectations $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$ and $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$ satisfy an *imprecise Markov property*: conditional on the state at time $t$, the lower expectation of a function $f(X_s)$ at a future time point $s$ is functionally independent of the states at time points $u$ that precede $t$. For the lower expectation $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$, this is of course to be expected: since $\underline{\mathbb{E}}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$ is the lower envelope of the set of Markov processes $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}, \mathscr{M}}$, it is not surprising that this lower envelope itself satisfies a Markov property as well. In fact, for this reason, Equation (47) is clearly also true if $\mathscr{Q}$ does not have separately specified rows. The most important message here though is that $\underline{\mathbb{E}}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$ also satisfies such an imprecise Markov property. In this case, this result is far from trivial, because the individual processes in $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}, \mathscr{M}}$ are not required to—and usually do not—satisfy a Markov property. It remains an open question at this point whether Equation (48) also holds if $\mathscr{Q}$ does not have separately specified rows.

That being said, Corollary 8.3 also establishes that the correspondence between $L_t^s$ and ICTMC's is not one-to-one. For starters, $L_t^s$ computes the lower expectation for two different sets of processes: $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}}$. Furthermore, we know from Section 7.2 that different sets $\mathscr{Q}_1$ and $\mathscr{Q}_2$ may have the same corresponding lower transition rate operator $\underline{Q}$. Hence, whenever this is the case, $L_t^s$ will—assuming the conditions in Corollary 8.3 are met by both $\mathscr{Q}_1$ and $\mathscr{Q}_2$—compute the lower expectation with respect to the sets of stochastic processes $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}_1}$, $\mathbb{P}^{\mathrm{WM}}_{\mathscr{Q}_2}$, $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_1}$ and $\mathbb{P}^{\mathrm{W}}_{\mathscr{Q}_2}$.

A particularly interesting special case corresponds to the situation where $\mathscr{Q}_1$ is a non-empty bounded set of rate matrices $\mathscr{Q}$ that has separately specified rows, and $\mathscr{Q}_2$ is its closed convex hull, which, because of Proposition 7.8, is equal to $\mathscr{Q}_{\underline{Q}}$, where $\underline{Q}$ is the lower transition rate operator that corresponds to $\mathscr{Q}$. The two sets of transition rate matrices $\mathscr{Q}$ and $\mathscr{Q}_{\underline{Q}}$ then clearly (i) have the same lower corresponding lower transition rate operator $\underline{Q}$ and (ii) satisfy the conditions in

Corollary 8.3. Therefore, it follows from the preceding argument that the resulting lower expectations are identical and, in particular, that

$$\underline{\mathbb{E}}_{\mathcal{Q}}^{\mathrm{WM}}[f(X_s) \mid X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}_{\mathcal{Q}_{\underline{Q}}}^{\mathrm{W}}[f(X_s) \mid X_t = x_t, X_u = x_u] = [L_t^s f](x_t),$$

which in turn immediately implies that for any set of stochastic processes $\mathscr{P}$ such that $\mathbb{P}_{\mathcal{Q}}^{\mathrm{WM}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathcal{Q}_{\underline{Q}}}^{\mathrm{W}}$:

$$\underline{\mathbb{E}}[f(X_s) \mid X_t = x_t, X_u = x_u] = [L_t^s f](x_t), \tag{49}$$

where $\underline{\mathbb{E}}$ is the lower expectation with respect to $\mathscr{P}$, as defined in Equation (31).

A common feature of these sets of stochastic processes $\mathscr{P}$, is that each of their elements $P$ is well-behaved and consistent with $\mathcal{Q}_{\underline{Q}}$. An obvious question, then, is whether this feature is necessary in order for Equation (49) to hold. The following result establishes that this is indeed the case.

**Theorem 8.4.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, with $\mathcal{Q}_{\underline{Q}}$ its set of dominating rate matrices, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then the largest set of stochastic processes $\mathscr{P}$ for which the corresponding conditional lower expectation operator $\underline{\mathbb{E}}[\cdot \mid \cdot]$—as defined in Equation (31)—satisfies*

$$\underline{\mathbb{E}}[f(X_s) \mid X_t = x_t, X_u = x_u] = [L_t^s f](x_t)$$

*for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}$ and $x_u \in \mathscr{X}_u$, and every $f \in \mathscr{L}(\mathscr{X})$, is the set $\mathbb{P}_{\mathcal{Q}_{\underline{Q}}}^{\mathrm{W}}$.*

We regard this result as a vindication for our choice to focus on *well-behaved* stochastic processes—instead of more restricted ones, such as, say, continuous or differentiable stochastic processes. Since our aim here is to use $L_t^s$ as a representational and computational tool for lower expectations, it follows from this result that in order to be able to do this, it is indeed necessary to impose this minimal property of well-behavedness.

### 8.2. Lower transition operators as a computational tool

An important consequence of the fact that the lower expectations $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}^{\mathrm{WM}}$ and $\underline{\mathbb{E}}_{\mathcal{Q},\mathcal{M}}^{\mathrm{W}}$ can be conveniently represented by the operator $L_t^s$—at least for functions of the state at a single time point in the future—is that we can focus our computational efforts on evaluating this operator $L_t^s$, thereby abstracting away all the technicalities of dealing with lower expectations with respect to sets of stochastic process. Of course, in practice, we are only interested in a finite precision approximation of $L_t^s f$, and we will therefore focus on computing $L_t^s f$ within some guaranteed $\varepsilon$-bound.

The construction of the operator $L_t^s$ in Section 7 already suggests how we can do this; namely, by using a finite-precision approximation of $L_t^s$ using the auxiliary operator $\Phi_u \coloneqq \prod_{i=1}^n (I + \Delta_i \underline{Q})$. Recall from Section 7 that the approximation of $L_t^s$ by $\Phi_u$ becomes better as we take $u \in \mathscr{U}_{[t,s]}$ to be an increasingly finer partition of the interval $[t, s]$. The following result tells us exactly how fine this partition needs to be for a specific function $f \in \mathscr{L}(\mathscr{X})$, in order to guarantee an $\varepsilon$-error bound on $L_t^s f$.

**Proposition 8.5.** *Let $\underline{Q}$ be a lower transition rate operator, choose any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, and let $L_t^s$ be the lower transition operator corresponding to $\underline{Q}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon \in \mathbb{R}_{>0}$, if we choose any $n \in \mathbb{N}$ such that*

$$n \geq \max\left\{ (s - t) \left\| \underline{Q} \right\|, \frac{1}{2\varepsilon} (s - t)^2 \left\| \underline{Q} \right\|^2 \|f\|_{\mathrm{v}} \right\},$$

*with $\|f\|_{\mathrm{v}} \coloneqq \max f - \min f$, we are guaranteed that*

$$\left\| L_t^s f - \prod_{i=1}^n (I + \Delta \underline{Q}) f \right\| \leq \varepsilon,$$

*with $\Delta \coloneqq {}^{(s-t)}\!/_n$.*

Simply put, this result tells us that if we can compute $\underline{Q} g$ for all $g \in \mathscr{L}(\mathscr{X})$, then we can also approximate the quantity $L_t^s f$ to arbitrary precision, for any given $f \in \mathscr{L}(\mathscr{X})$. Therefore, and because of our results in Section 8.1, if $\underline{Q}$ is the lower transition rate operator that corresponds to a given set $\mathcal{Q}$ that is non-empty, bounded and has separately specified rows, the non-linear optimisation problem of computing $\underline{\mathbb{E}}_{\mathcal{Q}}^{\mathrm{WM}}[f(X_s) \mid X_t = x_t, X_u = x_u]$ or $\underline{\mathbb{E}}_{\mathcal{Q}}^{\mathrm{W}}[f(X_s) \mid X_t = x_t, X_u = x_u]$ reduces to the problem of computing $\underline{Q} g_i$ for $n$ different functions $g_i \in \mathscr{L}(\mathscr{X})$, with $n \in \mathbb{N}$ as in Proposition 8.5.

In the remainder of this section, we will numerically illustrate this method of computing the conditional lower expectation of a given function $f \in \mathscr{L}(\mathscr{X})$. In order to make this illustration less abstract, we consider the context of the simple disease model that was put forward in Example 1.1 in Section 1. To this end, we first construct a parameter set $\mathcal{Q}$ that we will use in the examples to come.

**Example 8.1.** Consider again the binary-state disease model from Example 1.1, modelling a person periodically becoming sick and recovering after some time. The state space here is of the form $\mathscr{X} = \{\texttt{healthy}, \texttt{sick}\}$, and we wish to specify numerical values for the rate at which the two possible transitions occur; that is, the transitions from $\texttt{healthy}$ to $\texttt{sick}$, and from $\texttt{sick}$ to $\texttt{healthy}$.

If we were using a precise homogeneous Markov chain $P \in \mathbb{P}^{\text{WHM}}$, it would suffice to select a single rate matrix $Q \in \mathscr{R}$, which, in this binary case, would be of the form

$$Q = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \text{ for some } a, b \in \mathbb{R}_{\geq 0}.$$

The parameter $a$ here is the rate at which a healthy person becomes sick. Technically, this means that if a person is healthy at time $t$, the probability that he or she will be sick at time $t + \Delta$, for small $\Delta$, is very close to $\Delta a$. More intuitively, if we take the time unit to be one week, it means that he or she will, on average, become sick after $1/a$ weeks. The parameter $b$ is the rate at which a sick person becomes healthy again, and has a similar interpretation. In other words, if we would for example say that $a = 1/52$ and $b = 1$, then this would mean that—on average—we expect that a healthy person will stay healthy for about one year—52 weeks—and that a sick person will remain sick for one week.

In practice, assessing the exact values of $a$ and $b$ may be difficult, especially if we want them to be reliable. In those cases, instead of modelling our problem by means of a precise continuous-time Markov chain, we can consider an ICTMC. We then no longer need to assess a single transition rate matrix $Q$, but only need to specify a set of rate matrices $\mathscr{Q}$ to which we think $Q$ might belong. Suppose for example that we feel confident in saying that the average time for a healthy person to become sick lies somewhere in between four months and a year, and that the average time for a sick person to recover is situated somewhere in between half a week and two weeks. The corresponding set of transition rate matrices is then given by

$$\mathscr{Q} \coloneqq \left\{ \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} : a \in \left[ \frac{1}{52}, \frac{3}{52} \right], b \in \left[ \frac{1}{2}, 2 \right] \right\}, \tag{50}$$

with the time unit again being one week. Note that $\mathscr{Q}$ here is clearly non-empty, bounded, closed, convex, and has separately specified rows. As such, this set will satisfy all the preconditions necessary to be used in the examples that follow. ◇

As mentioned above, for a given set $\mathscr{Q}$, Proposition 8.5 allows us to reduce the problem of computing conditional lower expectations to the problem of evaluating the lower transition operator $\underline{Q}$ of $\mathscr{Q}$, as defined in Equation (38). In theory, the latter could still be a difficult problem, since $\mathscr{Q}$ can be very complicated—we only require that it is non-empty and has separately specified rows. However, in practice, $\mathscr{Q}$ will typically be described by means of linear constraints, and computing $\underline{Q} g$ is then a standard constrained linear optimisation problem, which is easily solved by means of linear programming techniques. We will henceforth assume that the optimisation problem in Equation (38) is solvable. However, as we have just explained, the complexity of this problem will depend on the exact form of $\mathscr{Q}$.

For the set $\mathscr{Q}$ of Example 8.1, as defined by Equation (50), the constraints that we are dealing with are simple box-constraints. Furthermore, because the state space $\mathscr{X}$ contains only two states in this specific example model, computing $\underline{Q} g$ is particularly easy, regardless of the choice of $g \in \mathscr{L}(\mathscr{X})$. This is made explicit in the following example, where we also compute the quantity $\|\underline{Q}\|$ that is required for the numerical computations that follow.

**Example 8.2.** Consider the set $\mathscr{Q}$ of Example 8.1, as defined by Equation (50), let $\underline{Q}$ be the corresponding lower transition rate operator, and consider any $g \in \mathscr{L}(\mathscr{X})$. Furthermore, for the sake of notational convenience, let us abbreviate $\texttt{healthy}$ as h and $\texttt{sick}$ as s. It then follows from Equation (38) that

$$[\underline{Q} g](\text{h}) \coloneqq \inf\{[Q g](\text{h}) : Q \in \mathscr{Q}\} = \inf \left\{ \begin{bmatrix} -a & a \end{bmatrix} \begin{bmatrix} g(\text{h}) \\ g(\text{s}) \end{bmatrix} : a \in \left[ \frac{1}{52}, \frac{3}{52} \right] \right\}$$

$$= \inf \left\{ a\big(g(\text{s}) - g(\text{h})\big) : a \in \left[ \frac{1}{52}, \frac{3}{52} \right] \right\},$$

and therefore, that

$$[\underline{Q} g](\text{h}) = \begin{cases} 1/52\big(g(\text{s}) - g(\text{h})\big) & \text{if } g(\text{s}) \geq g(\text{h}) \\ 3/52\big(g(\text{s}) - g(\text{h})\big) & \text{if } g(\text{s}) \leq g(\text{h}). \end{cases} \tag{51}$$

In a completely analogous way, we also find that

$$[\underline{Q} g](\text{s}) = \begin{cases} 1/2\big(g(\text{h}) - g(\text{s})\big) & \text{if } g(\text{s}) \leq g(\text{h}) \\ 2\big(g(\text{h}) - g(\text{s})\big) & \text{if } g(\text{s}) \geq g(\text{h}). \end{cases} \tag{52}$$

Using these observations, we can now compute $\|\underline{Q}\|$. If $g \in \mathscr{L}(\mathscr{X})$ is such that $\|g\| = 1$, then clearly, it follows from Equation (51) that the highest possible value for $\left|[\underline{Q}\,g](\text{h})\right|$ is $3/52$—for $g(\text{s}) = -1$ and $g(\text{h}) = 1$—and similarly, using Equation (52), we find that the highest possible value for $\left|[\underline{Q}\,g](\text{s})\right|$ is $4$—for $g(\text{s}) = 1$ and $g(\text{h}) = -1$. Since

$$\|\underline{Q}\| = \sup\left\{\|\underline{Q}\,g\| : g \in \mathscr{L}(\mathscr{X}),\ \|g\| = 1\right\}$$
$$= \sup\left\{\max\left\{\left|[\underline{Q}\,g](x)\right| : x \in \mathscr{X}\right\} : g \in \mathscr{L}(\mathscr{X}),\ \|g\| = 1\right\},$$

this implies that $\|\underline{Q}\| = 4$.                                                                                                    ◇

We now have everything in place to demonstrate the use of Proposition 8.5 for numerically computing conditional lower expectations.

**Example 8.3.** Consider again the set of rate matrices $\mathscr{Q}$ from Example 8.1, which, as we have seen there, expresses that the rate $a$ at which a healthy person becomes sick belongs to the interval $[1/52, 3/52]$, and the rate $b$ at which a sick person becomes healthy belongs to the interval $[1/2, 2]$.

Depending on the extra assumptions that we impose, we can now consider three different ICTMC's. If we believe that the rates $a$ and $b$ remain constant over time, and interpret the intervals $[1/52, 3/52]$ and $[1/2, 2]$ as bounds on these unknown but constant rates, then we can model our problem by means of the ICTMC $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{WHM}}$. If, on the other hand, we think that the rates $a$ and $b$ might vary within these bounds as time progresses, then we should use the ICTMC's $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{WM}}$ or $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\text{W}}$ instead, where the former assumes that the time-dependent rates $a$ and $b$ cannot be influenced by the value of the past states of the process, whereas the latter does not make such an assumption. The choice of $\mathscr{M}$ does not matter, because it will not effect any of the computations in this example; we therefore drop it from our notation.

Suppose now that we are interested in the lower probability that a person is sick one week from now, given that he or she is currently also sick. That is, we want to compute the conditional lower probability $\underline{P}(X_1 = \text{sick} \mid X_0 = \text{sick})$, which—using the abbreviation of Example 8.2—we will denote by $\underline{P}(X_1 = \text{s} \mid X_0 = \text{s})$. Then as we know from Equation (33), this is equivalent to computing $\underline{\mathbb{E}}[\mathbb{I}_\text{s}(X_1) \mid X_0 = \text{s}]$.

The result of this computation will, generally speaking, depend on our choice of ICTMC. In this example, we consider $\mathbb{P}_{\mathscr{Q}}^{\text{WM}}$ or $\mathbb{P}_{\mathscr{Q}}^{\text{W}}$. The choice between these two does not make any difference, because it follows from Corollary 8.3 that[8]

$$\underline{\mathbb{E}}_{\mathscr{Q}}^{\text{W}}[\mathbb{I}_\text{s}(X_1) \mid X_0 = \text{s}] = \underline{\mathbb{E}}_{\mathscr{Q}}^{\text{WM}}[\mathbb{I}_\text{s}(X_1) \mid X_0 = \text{s}] = [L_0^1 \mathbb{I}_\text{s}](\text{s}).$$

Hence, our problem has now been reduced to the task of computing $[L_0^1 \mathbb{I}_\text{s}](\text{s})$.

Proposition 8.5 tells us that we can approximate this quantity by using a fine enough partition of the time interval $[0, 1]$. We will be using a maximum error for this computation of $\varepsilon := 10^{-3}$. Furthermore, in this case, the length of the time interval is equal to $(s - t) = 1$, the difference between the maximum and minimum of $\mathbb{I}_\text{s}$ is $\|\mathbb{I}_\text{s}\|_\text{v} = 1$, and due to our choice of $\mathscr{Q}$, we have $\|\underline{Q}\| = 4$, as in Example 8.2. Hence, if we subdivide the time interval into $n$ steps, with

$$n = \max\left\{(s - t)\|\underline{Q}\|,\ \frac{1}{2\varepsilon}(s - t)^2\|\underline{Q}\|^2\|f\|_\text{v}\right\} = \max\{4, 8000\} = 8000,$$

resulting in a step size of $\Delta = (s-t)/n = 1.25 \times 10^{-4}$, then Proposition 8.5 guarantees that we can approximate $[L_0^1 \mathbb{I}_\text{s}]$—with a maximum error of $\varepsilon$—by

$$\prod_{i=1}^{n}(I + \Delta\,\underline{Q})\mathbb{I}_\text{s} = \prod_{i=1}^{8000}(I + \Delta\,\underline{Q})\mathbb{I}_\text{s} = \left(\prod_{i=1}^{7999}(I + \Delta\,\underline{Q})\right)\left((I + \Delta\,\underline{Q})\mathbb{I}_\text{s}\right).$$

We start by computing the right-most factor on the right-hand side of this equation. Using Equations (51) and (52) in Example 8.2, we find that

$$[\underline{Q}\,\mathbb{I}_\text{s}](x) = \begin{cases} 1/52 & \text{if } x = \text{h} \\ -2 & \text{if } x = \text{s} \end{cases} \quad \text{for all } x \in \mathscr{X},$$

and therefore, it follows that

$$g_1(x) := [(I + \Delta\,\underline{Q})\mathbb{I}_\text{s}](x) = [\mathbb{I}_\text{s}](x) + \Delta[\underline{Q}\,\mathbb{I}_\text{s}](x) = \begin{cases} 2.4038 \times 10^{-5} & \text{if } x = \text{h} \\ 0.9975 & \text{if } x = \text{s} \end{cases} \quad \text{for all } x \in \mathscr{X}.$$

Having completed this first step, we now have that

---

[8]   Actually, in this particular—binary—case, it can be shown that $\mathbb{P}_{\mathscr{Q}}^{\text{WHM}}$ would lead to the exact same result. However, in general, as we have seen in Example 6.2, this will not be the case.

$$\prod_{i=1}^{n}(I+\Delta\underline{Q})\mathbb{I}_{\mathrm{s}}=\prod_{i=1}^{7999}(I+\Delta\underline{Q})g_1=\left(\prod_{i=1}^{7998}(I+\Delta\underline{Q})\right)\left((I+\Delta\underline{Q})g_1\right).$$

At this point, we simply keep on repeating this process, computing $g_2 \coloneqq (I+\Delta\underline{Q})g_1$, then $g_3 \coloneqq (I+\Delta\underline{Q})g_2$, and so on. In this way, after $j$ steps, we find that

$$\prod_{i=1}^{n}(I+\Delta\underline{Q})\mathbb{I}_{\mathrm{s}}=\prod_{i=1}^{8000-j}(I+\Delta\underline{Q})g_j$$

and after completing all $n$ steps, we eventually find that

$$\prod_{i=1}^{n}(I+\Delta\underline{Q})\mathbb{I}_{\mathrm{s}}=g_n=\begin{cases}0.0083 & \text{if } x=\mathrm{h}\\0.1410 & \text{if } x=\mathrm{s}\end{cases}\quad\text{for all } x\in\mathscr{X},$$

from which we conclude that

$$\underline{P}_{\mathscr{Q}}^{\mathrm{W}}(X_1=\mathrm{s}\,|\,X_0=\mathrm{s})=\underline{P}_{\mathscr{Q}}^{\mathrm{WM}}(X_1=\mathrm{s}\,|\,X_0=\mathrm{s})=\left[L_0^1\mathbb{I}_{\mathrm{s}}\right](\mathrm{s})=g_n(\mathrm{s})\pm\varepsilon=0.141\pm0.001,$$

where the third equality[9] follows from the error bound that is guaranteed by Proposition 8.5.                ◇

This iterative procedure for computing $L_t^s f$, as illustrated in the above example, is outlined in Algorithm 1. The algorithm first finds the number of steps required to reach the given precision $\varepsilon$ (Line 2), and computes from this the corresponding step size $\Delta$ (Line 3). Starting with the function $g_0 \coloneqq f$ (Line 4), the algorithm iteratively computes the function $g_i \coloneqq (I+\Delta\underline{Q})g_{(i-1)}$ (Line 6). After repeating this for $n$ steps (Line 5), the returned function $g_n$ (Line 8) corresponds to $L_t^s f \pm \varepsilon$, due to Proposition 8.5. The algorithm takes for granted that $\|\underline{Q}\|$ is known and/or can be derived from $\underline{Q}$; if this is not the case, then the algorithm should be adapted accordingly, by replacing the norm $\|\underline{Q}\|$ with an upper bound, such as the one that is provided in LR5.

---

**Algorithm 1** Numerically compute $L_t^s f$ for any $f \in \mathscr{L}(\mathscr{X})$.

**Input:** A lower transition rate operator $\underline{Q}$, two time points $t,s\in\mathbb{R}_{\geq0}$ such that $t\leq s$, a function $f\in\mathscr{L}(\mathscr{X})$ and a maximum numerical error $\varepsilon\in\mathbb{R}_{>0}$.

**Output:** A function $L_t^s f \pm \varepsilon$ in $\mathscr{L}(\mathscr{X})$.

1: **function** COMPUTELF($\underline{Q}, t, s, f, \varepsilon$)

2:     $n \leftarrow \left\lceil \max\left\{\left((s-t)^2\|\underline{Q}\|^2\|f\|_{\mathrm{v}}\right)/2\varepsilon,\ (s-t)\|\underline{Q}\|\right\}\right\rceil$

3:     $\Delta \leftarrow (s-t)/n$

4:     $g_0 \leftarrow f$

5:     **for** $i \in \{1,\ldots,n\}$ **do**

6:         $g_i \leftarrow g_{(i-1)} + \Delta\underline{Q}\,g_{(i-1)}$

7:     **end for**

8:     **return** $g_n$

9: **end function**

---

## 9. A general framework for computing lower expectations

Having shown in the previous Section 8.1 that the operator $L_t^s$ allows us to compute conditional lower expectations for functions $f \in \mathscr{L}(\mathscr{X})$ that are defined on the state at a single future time point, we will now turn our attention to functions defined on multiple time points.

We will start by considering conditional lower expectations, where the conditioning is done with respect to the states $x_u$ at a finite sequence of time points $u$. In this context, we distinguish between two different classes of functions. We begin in Section 9.1 by considering functions $f \in \mathscr{L}(\mathscr{X}_{u\cup s})$ and lower expectations of the form $\underline{\mathbb{E}}[f(X_u, X_s)\,|\,X_u = x_u]$, with $s$ a single future time point. Thus, although the function $f$ depends on multiple time points, all but one of these time points coincide with the time points on which we are conditioning. This is then generalised in Section 9.2 to functions $f \in \mathscr{L}(\mathscr{X}_{u\cup v})$ and lower expectations $\underline{\mathbb{E}}[f(X_u, X_v)\,|\,X_u = x_u]$, where $v$ is now a finite sequence of future time points. We end in Section 9.3 by showing how to compute unconditional lower expectations of the form $\underline{\mathbb{E}}[f(X_u)]$, where $u$ is again a finite sequence of time points.

---

[9] Formally, we should write this equality as $[L_0^1\mathbb{I}_{\mathrm{s}}](\mathrm{s}) \in [g_n(\mathrm{s})-\varepsilon,\ g_n(\mathrm{s})+\varepsilon]$ or, completely equivalently, as $g_n(\mathrm{s}) \in \left[[L_0^1\mathbb{I}_{\mathrm{s}}](\mathrm{s})-\varepsilon,\ [L_0^1\mathbb{I}_{\mathrm{s}}](\mathrm{s})+\varepsilon\right]$. However, we find the shorthand $[L_0^1\mathbb{I}_{\mathrm{s}}](\mathrm{s}) = g_n(\mathrm{s}) \pm \varepsilon$ equally clear. Furthermore, we will later consider similar statements for functions, which we can write much more succinctly with this minor abuse of notation—with the understanding that the above inclusion should then be taken point-wise.

*9.1. Multi-variable functions on a single future time point*

Consider a non-empty sequence of time points $u = t_0, \ldots, t_n \in \mathcal{U}_\emptyset$ and a single future time point $s > u$. Let $f \in \mathcal{L}(\mathscr{X}_{u \cup s})$ be a function that depends on the states at these time points. Then for any stochastic process $P$ and any history $x_u \in \mathscr{X}_u$, the corresponding conditional expectation of such a function is well known to satisfy the following simple property:

$$\mathbb{E}_P[f(X_u, X_s) \mid X_u = x_u] = \mathbb{E}_P[f(x_u, X_s) \mid X_u = x_u].$$

Therefore, and because a lower expectation is an infimum over expectations, we find that also

$$\underline{\mathbb{E}}[f(X_u, X_s) \mid X_u = x_u] = \underline{\mathbb{E}}[f(x_u, X_s) \mid X_u = x_u]. \tag{53}$$

While writing down these equations, we have implicitly introduced a notational convention that should be intuitively clear: we interpret the function $f(x_u, X_s)$ as the restriction of $f(X_u, X_s)$ to the state $\mathscr{X}_s$, for the fixed state assignment $x_u$. A more formal—but completely equivalent—way of doing this, is to identify $f(x_u, X_s)$ with a function $f_{x_u} \in \mathcal{L}(\mathscr{X}_s) = \mathcal{L}(\mathscr{X})$, defined by $f_{x_u}(x_s) \coloneqq f(x_u, x_s)$ for all $x_s \in \mathscr{X}$. With this notation, the equation above then turns into

$$\underline{\mathbb{E}}[f(X_u, X_s) \mid X_u = x_u] = \underline{\mathbb{E}}[f_{x_u}(X_s) \mid X_u = x_u],$$

and in this way, it becomes clear that the problem of computing $\underline{\mathbb{E}}[f(X_u, X_s) \mid X_u = x_u]$ is completely equivalent to the problem that we have considered in the previous section, which is to compute lower expectations of functions that depend on the state at a single future time point. In particular, under the conditions of Corollary 8.3, we find that

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u = x_u] = \underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u = x_u] = \left[L^s_{t_n} f_{x_u}\right](x_{t_n}).$$

However, note that the quantity $[L^s_{t_n} f_{x_u}](x_{t_n})$ still depends on the *entire* state assignment $x_u \in \mathscr{X}_u$, and not just on the state $x_{t_n}$.

In order to unify our notation, we therefore stipulate the following convention. For any $s \in \mathbb{R}_{>0}$ and any $u \in \mathcal{U}_{<s}$ with $u \neq \emptyset$ such that $u = t_0, \ldots, t_n$, we allow $L^s_{t_n}$ to be applied to any $f \in \mathcal{L}(\mathscr{X}_{u \cup s})$, by applying it to the restriction of $f$ to the *latest* time point at which it is defined—the time point $s$, in this case. Because this restriction depends on the state assignment $x_u$ at the other time points, the result is a function $[L^s_{t_n} f] \in \mathcal{L}(\mathscr{X}_u)$. In short, we stipulate for any $f \in \mathcal{L}(\mathscr{X}_{u \cup s})$ that

$$[L^s_{t_n} f](x_u) \coloneqq [L^s_{t_n} f_{x_u}](x_{t_n}) \coloneqq [L^s_{t_n} f(x_u, X_s)](x_{t_n}) \quad \text{for all } x_u \in \mathscr{X}_u. \tag{54}$$

Using this notational convention, the following corollary formalises the fact that our previous results also apply to functions in $\mathcal{L}(\mathscr{X}_{u \cup s})$.

**Corollary 9.1.** *Let $\mathscr{M}$ be an arbitrary non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be an arbitrary non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $s \in \mathbb{R}_{>0}$, all $u \in \mathcal{U}_{<s}$ such that $u \neq \emptyset$ and all $f \in \mathcal{L}(\mathscr{X}_{u \cup s})$:*

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u] = \underline{\mathbb{E}}^{\mathsf{WM}}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u] = \left[L^s_{t_n} f\right](X_u).$$

Now, since $L^s_{t_n} f$ must be read as a separate application of the operator $L^s_{t_n}$ to every $f_{x_u}$, for $x_u \in \mathscr{X}_u$, it follows that we can compute $L^s_{t_n} f$ by applying Algorithm 1 multiple times. The method for this computation is outlined in Algorithm 2, which can be read as follows.

The algorithm starts by allocating space for a new function $g \in \mathcal{L}(\mathscr{X}_u)$ (Line 2), which will be the result of the algorithm. Then, for each state assignment $x_u \in \mathscr{X}_u$ (Line 3), it computes the value of $g(x_u)$ (Lines 4–6), as follows. First, we take the restriction of $f \in \mathcal{L}(\mathscr{X}_{u \cup s})$ to the state $\mathscr{X}_s$, for a specific state assignment $x_u$ (Line 4). Since the resulting function $f_{x_u} \coloneqq f(x_u, X_s)$ belongs to $\mathcal{L}(\mathscr{X})$, we can then approximate $L^s_{t_n} f_{x_u}$ using Algorithm 1 (Line 5), where $t_n$ is the last time point in $u$. The resulting function $\hat{f}_{x_u} = L^s_{t_n} f_{x_u} \pm \varepsilon$ is then a function in $\mathcal{L}(\mathscr{X}_{t_n})$, which, as we know from Equation (54), can be evaluated in $x_{t_n}$—the restriction of $x_u$ to the time point $t_n$—to obtain the value $g(x_u) = [L^s_{t_n} f](x_u) \pm \varepsilon$ of $g$ in $x_u$ (Line 6). Therefore, up to a maximum error of $\varepsilon$, the returned function $g$ (Line 8) will be equal to $L^s_{t_n} f$.

*9.2. Multi-variable functions on multiple future time points*

We next consider functions $f \in \mathcal{L}(\mathscr{X}_{u \cup v})$, where $u = t_0, \ldots, t_n$ is a sequence of time points in a process's history, on which we intend to condition, and $v = s_0, \ldots, s_m$ is a sequence of time points in a process's future; hence, we have that $s_0 > t_n$. For functions of this kind, it—surprisingly—turns out that computing the conditional lower expectation with respect to $\mathbb{P}^{\mathsf{W}}_{\mathscr{Q}, \mathscr{M}}$ is easier than for $\mathbb{P}^{\mathsf{WM}}_{\mathscr{Q}, \mathscr{M}}$. We start in this section by showing how to do this for the former set, that is, how to compute the conditional lower expectation

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_v) \mid X_u], \tag{55}$$

---

**Algorithm 2** Numerically compute $L_{t_n}^s f$ for any $f \in \mathscr{L}(\mathscr{X}_{u \cup s})$.

---

**Input:** A lower transition rate operator $\underline{Q}$, a time point $s \in \mathbb{R}_{\geq 0}$, a sequence of time points $u \in \mathscr{U}_{<s}$ with $u = t_0, \ldots, t_n$, a function $f \in \mathscr{L}(\mathscr{X}_{u \cup s})$, and a maximum numerical error $\varepsilon \in \mathbb{R}_{>0}$.

**Output:** A function $L_{t_n}^s f \pm \varepsilon$ in $\mathscr{L}(\mathscr{X}_u)$.

1: **function** COMPUTELUF($\underline{Q}, u, s, f, \varepsilon$)
2:     $g \leftarrow g \in \mathscr{L}(\mathscr{X}_u)$                                    ▷ Allocate space for $g \in \mathscr{L}(\mathscr{X}_u)$
3:     **for** $x_u \in \mathscr{X}_u$ **do**                                         ▷ Compute $g(x_u)$ for all $x_u \in \mathscr{X}_u$
4:         $f_{x_u} \leftarrow f(x_u, X_s)$
5:         $\hat{f}_{x_u} \leftarrow$ COMPUTELF($\underline{Q}, t_n, s, f_{x_u}, \varepsilon$)                    ▷ Run Algorithm 1
6:         $g(x_u) \leftarrow \hat{f}_{x_u}(x_{t_n})$
7:     **end for**
8:     **return** $g$
9: **end function**

---

and then go on to provide a counter example to illustrate that, unfortunately, this method does not work for the latter set, that is, for lower expectations with respect to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$.

As we have seen in the previous sections, we can use the operator $L_t^s$ to compute lower expectations of functions that depend on the state at a single future time point, provided that $\mathscr{Q}$ is non-empty, bounded, and has separately specified rows. We here show that if $\mathscr{Q}$ is additionally convex, then we can do the same for functions that depend on multiple future time points. The reason why this is the case is Theorem 6.5. In particular, because of that theorem, if $\mathscr{Q}$ is convex, we can decompose the lower expectation in Equation (55) as follows:

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}\big[\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u, X_{v \setminus \{s_m\}}] \,\big|\, X_u\big], \tag{56}$$

where $s_m$ is the last time point in $v = s_1, \ldots, s_m$. The essential feature of this decomposition is that the inner lower expectation on the right-hand side of the equality is conditioned on all the time points in $u$ and $v \setminus \{s_m\}$, or equivalently, that this lower expectation is taken with respect to a single future time point $s_m$. Therefore, it follows from the discussion in Section 9.1, and from Corollary 9.1 in particular, that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u, X_{v \setminus \{s_m\}}] = \big[L_{s_{m-1}}^{s_m} f\big](X_u, X_{v \setminus \{s_m\}}),$$

which, by substituting the inner lower expectation in Equation (56), implies that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}\big[\big[L_{s_{m-1}}^{s_m} f\big](X_u, X_{v \setminus \{s_m\}}) \,\big|\, X_u\big].$$

In this way, our original problem, which was to compute $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u]$, has been reduced to a completely analogous—but smaller-sized—problem; the only difference is that $f$ has been replaced by a new function, which no longer depends on $X_u$ and $X_v$, but only on $X_u$ and $X_{v \setminus s_m}$.

We can now simply apply this step over and over again, by repeatedly removing the last remaining time point $s_{m-1}, s_{m-2}, \ldots, s_0$ and, at each step, replacing the inner lower expectation with the operator $L_{s_{m-2}}^{s_{m-1}}, L_{s_{m-3}}^{s_{m-2}}, \ldots, L_{t_n}^{s_0}$. Because $v$ is finite, this process eventually stops, and we then obtain the following result.

**Corollary 9.2.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty, bounded and convex set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for any $u = t_0, \ldots, t_n$ and $v = s_0, \ldots, s_m$ in $\mathscr{U}_\emptyset$ such that $u < v$, and any $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathrm{W}}[f(X_u, X_v) \,|\, X_u] = L_{t_n}^{s_0} L_{s_0}^{s_1} \cdots L_{s_{m-1}}^{s_m} f. \tag{57}$$

The following example numerically illustrates the use of Corollary 9.2 for computing the conditional lower expectation of a function that depends on the state at multiple time points.

**Example 9.1.** Consider again the set of rate matrices $\mathscr{Q}$ from Example 8.1, and suppose that we are interested in the lower probability $\underline{P}(X_1 = X_2 \,|\, X_0)$. That is, given that we start in some initial state $X_0$, we want to know the lower probability that $X_1$ and $X_2$ will have identical values. As we know from Equation (33), this lower probability is given by

$$\underline{P}(X_1 = X_2 \,|\, X_0) = \underline{\mathbb{E}}[\mathbb{I}_{X_1 = X_2}(X_1, X_2) \,|\, X_0],$$

with $\mathbb{I}_{X_1 = X_2}$ the indicator of the event $X_1 = X_2$, as defined by

$$\mathbb{I}_{X_1 = X_2}(x_1, x_2) := \begin{cases} 1 & \text{if } x_1 = x_2 \\ 0 & \text{if } x_1 \neq x_2 \end{cases} \quad \text{for all } x_1, x_2 \in \mathscr{X}.$$

As explained in Example 8.3, the value of such a lower expectation can depend on the particular type of ICTMC that we choose to use, and in fact, in this particular case, as we will see in Example 9.2, this choice does indeed make a difference.

For now, in this example, we use the ICTMC $\mathbb{P}_{\mathscr{Q}}^{\mathrm{W}}$. The aim is therefore to compute the conditional lower expectation $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{W}}[\mathbb{I}_{X_1=X_2}(X_1, X_2) \,|\, X_0]$. Technically speaking, we cannot do this by means of Corollary 9.2, because $\mathbb{I}_{X_1=X_2}$ does not formally depend on $X_0$. However, in practice, this is of course not a problem, because we can trivially regard $\mathbb{I}_{X_1=X_2}$ as a function of $X_0$, $X_1$ and $X_2$ whose value remains constant on $X_0$. In order to formalise this, we introduce the function $f(X_0, X_1, X_2)$, defined by

$$f(x_0, x_1, x_2) := \mathbb{I}_{X_1=X_2}(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = x_2 \\ 0 & \text{if } x_1 \neq x_2 \end{cases} \quad \text{for all } x_0, x_1, x_2 \in \mathscr{X}. \tag{58}$$

Since $f(X_0, X_1, X_2)$ is clearly identical to $\mathbb{I}_{X_1=X_2}(X_1, X_2)$, our problem now consists of computing $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{W}}[f(X_0, X_1, X_2) \,|\, X_0]$, which, because of Corollary 9.2, is given by

$$\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{W}}[f(X_0, X_1, X_2) \,|\, X_0] = \left[ \underline{L}_0^1 \underline{L}_1^2 f \right](X_0). \tag{59}$$

In order to compute the left hand side of this equality, we will resolve the composition of operators $\underline{L}_0^1 \underline{L}_1^2$ by starting from the latest time point, and working back to the earliest time point. Hence, we start by looking at $\underline{L}_1^2 f$.

The function $f$ depends on multiple—in this case three—time points, and we will therefore use Algorithm 2 to compute $\underline{L}_1^2 f$, which consists in applying Algorithm 1 four times. In particular, for every $x_0 \in \mathscr{X} = \{\mathrm{h}, \mathrm{s}\}$ and $x_1 \in \mathscr{X} = \{\mathrm{h}, \mathrm{s}\}$, we use Algorithm 1 to compute

$$g(x_0, x_1) := [\underline{L}_1^2 f](x_0, x_1) := [\underline{L}_1^2 f(x_0, x_1, X_2)](x_1) = \begin{cases} \left[ \underline{L}_1^2 \mathbb{I}_{\mathrm{h}} \right](\mathrm{h}) = 0.956 \pm \varepsilon & \text{if } x_1 = \mathrm{h} \\ \left[ \underline{L}_1^2 \mathbb{I}_{\mathrm{s}} \right](\mathrm{s}) = 0.141 \pm \varepsilon & \text{if } x_1 = \mathrm{s} \end{cases}$$

up to some desired maximum error $\varepsilon$, which we have here chosen to be $\varepsilon = 10^{-3}$. Note that the value of $\left[ \underline{L}_1^2 \mathbb{I}_{\mathrm{s}} \right](\mathrm{s})$ is identical to the value that we had obtained for $\left[ \underline{L}_0^1 \mathbb{I}_{\mathrm{s}} \right](\mathrm{s})$ in Example 8.3; this is to be expected, as it follows from the time-homogeneity property in Proposition 7.14.

Next, we apply the operator $\underline{L}_0^1$ to the function $\underline{L}_1^2 f$, or equivalently, to the function $g$. The approach is completely similar. Again, as in Algorithm 2, we simply apply Algorithm 1 multiple times. In particular, for every $x_0 \in \mathscr{X} = \{\mathrm{h}, \mathrm{s}\}$, we use Algorithm 1 to compute

$$[\underline{L}_0^1 \underline{L}_1^2 f](x_0) = [\underline{L}_0^1 g](x_0) := [\underline{L}_0^1 g(x_0, X_1)](x_0) = \begin{cases} 0.920 \pm 2\varepsilon & \text{if } x_0 = \mathrm{h} \\ 0.453 \pm 2\varepsilon & \text{if } x_0 = \mathrm{s}, \end{cases} \tag{60}$$

where the maximum error is now $2\varepsilon = 2 \cdot 10^{-3}$, which is the sum of our previous error bound on $\underline{L}_1^2 f$, which was $\varepsilon$, and the maximum extra error that may have arisen while computing Equation (60); the additivity of the two error bounds is a result of LT5 and LT6 and the fact that $\underline{L}_t^s$ is a lower transition operator.

Finally, by combining Equations (59) and (60), we conclude that

$$\underline{P}_{\mathscr{Q}}^{\mathrm{W}}(X_1 = X_2 \,|\, X_0 = \mathrm{h}) = 0.920 \pm 0.002 \quad \text{and} \quad \underline{P}_{\mathscr{Q}}^{\mathrm{W}}(X_1 = X_2 \,|\, X_0 = \mathrm{s}) = 0.453 \pm 0.002.$$

$$\diamond$$

A general method for computing the right hand side of Equation (57) is outlined in Algorithm 3. This algorithm starts by constructing a buffer $w := v = s_0, \ldots, s_m$ of the time points for which the operators $\underline{L}_{t_n}^{s_0}, \underline{L}_{s_1}^{s_2}, \ldots, \underline{L}_{s_{m-1}}^{s_m}$ have not yet been resolved, which initially corresponds to just $v$ (Line 2). We rename the function $f$ to $g_m$ for indexing purposes (Line 3). The algorithm then iteratively resolves the composition of operators in a backward fashion, working from the last time point in $v$ back to the first time point in $v$ (Line 4). One step here amounts to removing the current latest remaining time point $s_i$ from the buffer $w$ (Line 5), and then computing $\underline{L}_{s_{i-1}}^{s_i} g_i$ using Algorithm 2. The outcome of this computation is a function $g_{(i-1)} \in \mathscr{L}(\mathscr{X}_{u \cup w})$. After all time points $s_m, s_{m-1}, \ldots, s_0$ have been resolved, it follows from Proposition 8.5 that the returned function $g_{(-1)}$ (Line 8) is equal to $\underline{L}_{t_n}^{s_0} \underline{L}_{s_0}^{s_1} \cdots \underline{L}_{s_{m-1}}^{s_m} f \pm \varepsilon$.

Note that the error bound of $\pm \varepsilon$ holds because we compute each step using a maximum error of $\varepsilon/m{+}1$ (Line 6). That is, the finite-precision approximation error scales linearly in the number of time points that have to be computed. The reason why the accumulated error bounds can simply be added, is because $\underline{L}_{s_{i-1}}^{s_i}$ is a lower transition operator and therefore satisfies LT5 and LT6. In particular, using these two properties, it is easily verified that for any $\mu > 0$, $\underline{L}_{s_{i-1}}^{s_i}(g_i \pm \mu) = \underline{L}_{s_{i-1}}^{s_i} g_i \pm \mu$, where, in this case, we choose $\mu = \varepsilon/m{+}1$.

As we already claimed in the beginning of this section, we cannot use Corollary 9.2—nor, therefore, Algorithm 3—to compute lower expectations of functions $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$ with respect to a set of Markov chains $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$. In the remainder of this section, we provide this claim with some intuition and illustrate it by means of a numerical counterexample.

The core of the problem is that for a function $f \in \mathscr{L}(\mathscr{X}_{u \cup s})$, the conditional lower expectation $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}[f(X_u, X_s) \,|\, X_u]$ is not necessarily obtained by a single Markov process. That is, for every $\varepsilon > 0$ and every $x_u \in \mathscr{X}_u$, it follows from Equation (31) that there is a Markov chain $P_{x_u} \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ such that

---

**Algorithm 3** Numerically compute $L_{t_n}^{s_0} L_{s_0}^{s_1} \cdots L_{s_{m-1}}^{s_m} f$ for any $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$.

**Input:** A lower transition rate operator $\underline{Q}$, two sequences of time points $u = t_0, \ldots, t_n$ and $v = s_0, \ldots, s_n$ in $\mathscr{U}_\emptyset$ such that $u < v$, a function $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$, and a maximum numerical error $\varepsilon \in \mathbb{R}_{>0}$.

**Output:** A function $L_{t_n}^{s_0} L_{s_0}^{s_1} \cdots L_{s_{m-1}}^{s_m} f \pm \varepsilon$ in $\mathscr{L}(\mathscr{X}_u)$.

1: **function** COMPUTELUVF($\underline{Q}, u, v, f, \varepsilon$)
2:     $w \leftarrow v$
3:     $g_m \leftarrow f$
4:     **for** $i \in \{m, m-1, \ldots, 0\}$ **do**          ▷ Iterate backward over $v = s_0, \ldots, s_m$
5:         $w \leftarrow w \setminus s_i$          ▷ Remove last point of $w$
6:         $g_{(i-1)} \leftarrow$ COMPUTELUF($\underline{Q}, u \cup w, s_i, g_i, \varepsilon/m+1$)          ▷ Run Algorithm 2
7:     **end for**
8:     **return** $g_{(-1)}$
9: **end function**

---

$$\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}[f(X_u, X_s)|X_u = x_u] \leq \mathbb{E}_{P_{x_u}}[f(X_u, X_s)|X_u = x_u] \leq \underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}[f(X_u, X_s)|X_u = x_u] + \varepsilon,$$

but there may *not* be a single Markov chain $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ that does this simultaneously for *all* histories $x_u \in \mathscr{X}_u$.

For example, if we were to assume that

$$\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}[f(X_u, X_t, X_s) \mid X_u] = \underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}\big[\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}[f(X_u, X_t, X_s) \mid X_u, X_t] \mid X_u\big],$$

thereby mimicking Equation (56) for $v = \{t, s\}$, we would essentially be assuming that the individual Markov chains $P_{x_u, x_t} \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ for which the conditional lower expectations

$$\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}[f(X_u, X_t, X_s) \mid X_u = x_u, X_t = x_t]$$

are obtained—or rather, are approximated from above up to some arbitrary $\varepsilon$—can be combined into a single new Markov chain $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ such that

$$P(X_s | X_u = x_u, X_t = x_t) = P_{x_u, x_t}(X_s | X_u = x_u, X_t = x_t) = P_{x_u, x_t}(X_s | X_t = x_t) \tag{61}$$

for all $x_u \in \mathscr{X}_u$ and $x_t \in \mathscr{X}$. However, in most cases, this will not be possible. The issue is that for a stochastic process $P$ that satisfies Equation (61), the Markov property will—generally speaking—fail, because the right hand side of the expression still depends on $x_u$.

All of this changes if we replace $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ with $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{W}}$. In that case, we are no longer forced to obey the Markov property, and it then does become possible to combine different stochastic processes in a history-dependent way, as in Equation (61)—using Theorem 6.3. Essentially, this is the main reason why we were able to decompose the lower expectations $\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{W}}$—as we did in Theorem 6.5 and Equation (56)—and ultimately, why we could use the operator $L_t^s$ to compute them.[10]

In the example below, we reconsider the problem of Example 9.1, but for a different ICTMC: we now use $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ instead of $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{W}}$. The result ends up being different, and thereby illustrates—by means of a counterexample—that Algorithm 3 is not in general applicable when working with sets of Markov chains such as $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$.

**Example 9.2.** Consider again the set of rate matrices $\mathscr{Q}$ from Example 8.1 and the problem of computing $\underline{P}(X_1 = X_2 | X_0)$. However, this time, we will compute this quantity with respect to the ICTMC $\mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}}$ instead of $\mathbb{P}_{\mathscr{Q}}^{\mathrm{W}}$. Using an argument that is analogous to that in Example 8.1, we find that this problem is equivalent to computing $\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}[f(X_0, X_1, X_2) | X_0]$, with $f$ as in Equation (58).

The issue, however, is that we do not have a general method for computing such lower expectations for sets of Markov chains. Fortunately, in the very specific case of a binary state space $\mathscr{X}$—as we use here in this example—and for functions $f$ that depend on a small number of states $X_t$—only three in this case—it is possible to numerically solve the optimisation problem given by the right-hand side of

$$\underline{\mathbb{E}}_{\mathscr{Q}}^{\mathrm{WM}}[f(X_0, X_1, X_2) | X_0] = \inf\left\{ \mathbb{E}_P[f(X_1, X_2) | X_0] : P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{WM}} \right\}$$

by combining brute-force optimisation techniques with some clever tricks. We will not report on this method here, because it only works in very specific cases such as the one we consider here, and does not scale well to larger problems. For our present purposes, it suffices to know that this method yields the following results:

---

[10] For readers that are closely familiar with the theory of credal networks [44], it might be helpful to compare the difference between $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ and $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{W}}$ with the difference between credal networks under epistemic irrelevance [45,46] and credal networks under strong independence [47]. There too, dropping the restrictions that are imposed by the Markov property—as is done for credal networks under epistemic irrelevance—allows for lower expectations to be decomposed and, as such, for efficient recursive computational methods to be developed.

$$\underline{P}_{\mathscr{Q}}^{\mathsf{WM}}(X_1 = X_2 \,|\, X_0 = \mathrm{h}) \approx 0.939 \quad \text{and} \quad \underline{P}_{\mathscr{Q}}^{\mathsf{WM}}(X_1 = X_2 \,|\, X_0 = \mathrm{s}) \approx 0.467.$$

Comparing to the results for $\mathbb{P}_{\mathscr{Q}}^{\mathsf{W}}$ in Example 9.1, we see that the obtained lower probabilities are indeed different for these two sets of processes. Furthermore, as guaranteed by Proposition 6.4, the lower probabilities with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathsf{W}}$—which, as we know from Section 6.3, is just a special type of lower expectation—provides a lower bound on the lower probabilities with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathsf{WM}}$. ◇

We end with some remarks about the tractability of the algorithms that we have presented. In particular, observe that Algorithm 2 and 3 have a runtime complexity that is exponential in the number of time points. For example, for the computation of the lower expectation of a function $f \in \mathscr{L}(\mathscr{X}_{u \cup s})$, Algorithm 2 requires a separate execution of Algorithm 1 for every $x_u \in \mathscr{X}_u$. If we write $u = t_0, \ldots, t_n$ and let $|\mathscr{X}|$ denote the number of states in $\mathscr{X}$, then, clearly, this requires $|\mathscr{X}|^{n+1}$ executions of Algorithm 1. Similarly, for the computation of the lower expectation of a function $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$, with $v = s_0, \ldots, s_m$, Algorithm 3 requires a number of executions of Algorithm 1 that is of the order $|\mathscr{X}|^{n+m+2}$. Of course, this should not really be surprising—even the simple enumeration of the different values of $f$ takes this many steps, so we cannot in general expect the computation of its lower expectation to be of a lower complexity. What is important though—but beyond the scope of this paper—is that if $f$ has some specific structure that reduces the complexity of its specification, then the complexity of our algorithms will scale down accordingly; we discuss these and other avenues for future research in our conclusions.

## 9.3. Unconditional lower expectations

Having shown in the previous sections how to compute conditional lower expectations, we now shift our attention to unconditional ones. In particular, we consider a function $f \in \mathscr{L}(\mathscr{X}_u)$, with $u \in \mathscr{U}_\emptyset$, and are interested in computing $\underline{\mathbb{E}}[f(X_u)]$.

As we will see, the difference between unconditional and conditional lower expectations is mainly related to the role of the set of initial distributions $\mathscr{M}$. In fact, as an astute reader may have noticed, all our previous results about computing conditional lower expectations with the operator $L_t^s$ did not depend on the particular choice of this set $\mathscr{M}$. In contrast, and rather unsurprisingly, unconditional lower expectations do depend on this choice.

In particular, the choice of $\mathscr{M}$ will influence our computations through the corresponding lower expectation operator $\underline{\mathbb{E}}_{\mathscr{M}}$, defined by

$$\underline{\mathbb{E}}_{\mathscr{M}}[f] \coloneqq \inf\left\{ \sum_{x \in \mathscr{X}} p(x)f(x) \colon p \in \mathscr{M} \right\} \quad \text{for all } f \in \mathscr{L}(\mathscr{X}). \tag{62}$$

In principle, evaluating this operator can be difficult, because the only restriction we impose on the set $\mathscr{M}$ is that it should be non-empty. However, in practice, $\mathscr{M}$ will typically have a rather simple structure, and computing $\underline{\mathbb{E}}_{\mathscr{M}}[f]$ will then be straightforward. For example, if $\mathscr{M}$ consists of a finite number of probability mass functions—or is equal to their convex hull—then evaluating $\underline{\mathbb{E}}_{\mathscr{M}}[f]$ is just a matter of minimising over that finite set of probability mass functions. Another typical situation is when $\mathscr{M}$ is specified by means of linear constraints. In that case, $\underline{\mathbb{E}}_{\mathscr{M}}[f]$ can be computed by means of linear programming techniques. We will proceed under the assumption that the optimisation problem in Equation (62) is solvable.

Regardless of the type of ICTMC that we consider, $\underline{\mathbb{E}}_{\mathscr{M}}$ allows us to compute the lower expectation of functions that depend on the initial state $X_0$.

**Proposition 9.3.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$ and let $\mathscr{Q}$ be a non-empty bounded set of rate matrices. Then for all $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[f].$$

For functions that depend on some non-initial state $X_s$, with $s > 0$, the operator $\underline{\mathbb{E}}_{\mathscr{M}}$ still allows us to compute their lower expectation, by combining it with Algorithm 1. However, this only works for ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ or $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$.

**Proposition 9.4.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$ and let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$. Then for any $s \in \mathbb{R}_{\geq 0}$ and any $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_s)] = \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f].$$

For ICTMC's that are of the type $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$, this result generalises to functions that depend on multiple time points. Computing the lower expectation of such a function can then be done by combining $\underline{\mathbb{E}}_{\mathscr{M}}$ with Algorithm 3.

**Proposition 9.5.** *Let $\mathcal{M}$ be a non-empty set of probability mass functions on $\mathcal{X}$ and let $\mathcal{Q}$ be a non-empty, bounded, convex set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$. Then for any $u = t_0, \ldots, t_n$ in $\mathcal{U}_\emptyset$ such that $t_0 = 0$, and any $f \in \mathcal{L}(\mathcal{X}_u)$:*

$$\underline{\mathbb{E}}^W_{\mathcal{Q}, \mathcal{M}}[f(X_u)] = \underline{\mathbb{E}}_{\mathcal{M}}[\underline{L}^{t_1}_{t_0} \underline{L}^{t_2}_{t_1} \cdots \underline{L}^{t_n}_{t_{n-1}} f].$$

In this result, the requirement $t_0 = 0$ is a purely formal one, and imposes no actual restrictions. Whenever $0 \notin u$, it suffices to first replace $f(X_u)$ by its trivial extension $f^*(X_0, X_u)$ to $\mathcal{X}_{0 \cup u}$, defined by $f^*(x_0, x_u) := f(x_u)$ for all $x_0 \in \mathcal{X}$ and $x_u \in \mathcal{X}_u$. Since $f^*(X_0, X_u)$ is then identical to $f(X_u)$, this addition of $X_0$ to the domain is merely a notational trick, and will clearly not influence the result of the computations.

## 10. Relation to previous work

To the best of our knowledge, the concept of an imprecise continuous-time Markov chain was first introduced in the literature by the seminal work of Škulj [13]. There, $\underline{Q}$ was defined as the lower envelope of a non-empty, bounded, closed, and convex set of rate matrices $\mathcal{Q}$ with separately specified rows, and the concept of an imprecise continuous-time Markov chain was then introduced in two different ways.

In the main body of his paper, Škulj characterises this concept in the following way. For any $f \in \mathcal{L}(\mathcal{X})$, he considers a time-dependent function $f_s$, which—using our notation—is defined as $f_s := T^s_0 f$, and then requires that $f_0 = f$,

$$\underline{Q} f_s \leq D^+ f_s := \liminf_{\Delta \to 0^+} \frac{f_{s+\Delta} - f_s}{\Delta} \quad \text{and} \quad \limsup_{\Delta \to 0^+} \frac{f_{s+\Delta} - f_s}{\Delta} =: D_+ f_s \leq \overline{Q} f_s, \tag{63}$$

where $\overline{Q}$ is the upper envelope of $\mathcal{Q}$, or equivalently, the conjugate upper transition rate operator of $\underline{Q}$, defined by $\overline{Q} g := -\underline{Q}(-g)$ for all $g \in \mathcal{L}(\mathcal{X})$, and where $D^+ f_s$ and $D_- f_s$ are called Dini derivatives. In this way, he generalises the usual differential equation characterisation of a precise homogeneous Markov chain, where the Dini derivatives would be replaced by derivatives, where the operator $\underline{Q}$ would be replaced by a transition rate matrix $Q$, and where the inequalities would become equalities.

Since this generalisation considers inequalities, the solution $f_s$ is not uniquely determined, and the problem is therefore to find lower and upper bounds for $f_s$. Škulj solves this problem, by showing that $\underline{f}_s \leq f_s \leq \overline{f}_s$, where $\underline{f}_s$ and $\overline{f}_s$ are the unique solutions of the differential equations

$$\frac{d\underline{f}_s}{ds} = \underline{Q} \, \underline{f}_s \quad \text{and} \quad \frac{d\overline{f}_s}{ds} = \overline{Q} \, \overline{f}_s, \tag{64}$$

with $\underline{f}_0 = \overline{f}_0 = f$. In practice, it suffices to restrict attention to either one of these two bounds, because they are related by conjugacy, in the sense that $\underline{f}_0 = -\overline{f}_0$ implies that $\underline{f}_s = -\overline{f}_s$.

Using this characterisation, Škulj focuses on computing $\overline{f}_s$. Since, for all $x \in \mathcal{X}$, $\overline{f}_s(x)$ is a tight upper bound on $f_s(x) := [T^s_0 f](x) = \mathbb{E}[f(X_s)|X_0 = x]$, this can be interpreted as computing a conditional upper expectation. In particular, for all $x \in \mathcal{X}$, we can interpret $\overline{f}_s(x)$ as

$$\overline{f}_s(x) = \overline{\mathbb{E}}[f(X_s)|X_0 = x] = -\underline{\mathbb{E}}[-f(X_s)|X_0 = x] = -\underline{[-f]}_s(x),$$

where the final two equalities follow from conjugacy.

The above characterisation leaves some ambiguity about the set of stochastic processes $\mathcal{P}$ with respect to which these lower and upper expectations are taken. However, as mentioned above, Škulj also provides another definition, which in his paper precedes the rigorous characterisation in terms of lower- and upper bounds that we have summarised above. There, he defines an imprecise continuous-time Markov chain as a random process $X \colon \mathbb{R} \to \mathcal{X}$ whose transition rate matrix is an unspecified function $Q_s \in \mathcal{Q}$ [13, Definition 1]. However, he does not provide a definition for $Q_s$—as being a derivative or some other representation of the infinitesimal behaviour of the process' transition probabilities—nor for the notion of a random process, and in that sense, this definition is rather informal.

On the one hand, this other definition of an imprecise continuous-time Markov chain clearly suggests that $\mathcal{P}$ is the set of all random processes whose time-dependent transition rate matrix $Q_s$ takes values in $\mathcal{Q}$. Since $Q_s$ is typically defined as a derivative, this suggests that $T^s_0$ should be differentiable in some way—or, at the very least, that $Q_s$ should be integrable. Furthermore, the fact that $Q_s$ is indexed only by the time $s$ suggests that $Q_s$ cannot be history-dependent—thereby implying that the corresponding random process satisfies the Markov property. On the other hand, Škulj's main definition suggests a different set of processes $\mathcal{P}$, because Equation (63) does not impose a Markov condition, and because the use of the Dini derivatives suggests that $T^s_0$ is not required to be differentiable. In summary then, Škulj clearly intends for $\mathcal{P}$ to be a set of stochastic processes of which the infinitesimal differences are compatible with $\mathcal{Q}$, but leaves open the question of whether or not these processes are Markov chains, and whether or not their transition operators $T^s_0$ should be differentiable.

Our approach, in contrast, removes this ambiguity, by making it formally explicit which set of stochastic processes $\mathcal{P}$ is being considered. Interestingly, but perhaps unsurprisingly, our results end up being very closely related to the work of

Škulj. In particular, by comparing Equations (45) and (64), we see that $\underline{f}_s = L_0^s f$, and therefore, it follows from Corollary 8.3 that

$$\underline{f}_s(x) = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)|X_0 = x] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_s)|X_0 = x],$$

thereby providing the function $\underline{f}_s$ with a clear interpretation in terms of sets of stochastic processes. Due to this connection, many of the results in Section 7 can be interpreted as properties of $\underline{f}_s$. For example, Definition 7.4 provides an alternative characterisation for $\underline{f}_s$, and Proposition 7.15 shows that the defining differential equation for $\underline{f}_s$—Equation (64)—can be stated uniformly, that is, with respect to the operator norm instead of pointwise.

Besides the differences that are related to the role of the set of processes $\mathscr{P}$, an other difference between our work and that of Škulj is the type of functions that are considered. Due to its focus on $\overline{f}_s$ (and, by conjugacy, $\underline{f}_s$), the work of Škulj applies to functions $f$ that depend on the state $X_s$ at a single time point $s$. In contrast, our work also considers functions that depend on the state at any finite number of time points. In this more general context, as we have seen, it actually becomes crucial to be aware of the exact set of stochastic processes with respect to which the lower expectation is taken, because this choice influences the value of the lower expectation, as well as whether or not it can be computed efficiently.

A final difference between our work and that of Škulj are the computational methods that are presented. In both cases, these methods are concerned with approximating $\underline{f}_s = L_0^s f$ up to some desired finite precision $\varepsilon$. However, the particular approximation methods differ. Škulj considers three distinct methods: a uniform grid discretisation, an adaptive grid discretisation, and a combination of both. We consider only one method, which corresponds to Algorithm 1.

Škulj's uniform grid discretisation method works similarly to our Algorithm 1, in that it considers a uniform partition $u = t_0, t_1, \ldots, t_n$ of the time interval $[0, s]$ and, at each time point $t_i$, requires solving a local optimisation problem. In our case, this local computation consists in evaluating $(I + \Delta \underline{Q})g = g + \Delta \underline{Q} g$ for some $g \in \mathscr{L}(\mathscr{X})$, with $\Delta = s/n$. In the uniform grid discretisation method of Škulj, this local computation consists in evaluating $e^{\Delta Q} g$, where $Q \in \mathscr{Q}$ needs to be selected in such a way that $Q g = \underline{Q} g$.

It turns out that among these two methods, our Algorithm 1 is much more efficient, for the following two reasons. First, it does not require the computation of a matrix exponential, which, for large state spaces, can be computationally demanding. Secondly, the number of steps $n$ that we require is much smaller. For instance, for the numerical example presented at the end of [13, Section 4.1], Škulj finds that his partition requires more than $n = 6000$ steps. In contrast, Algorithm 1 requires only $n = 50$ steps for that particular example. Similarly, in our Example 8.3 we found that we required $n = 8000$ steps. Applying the method from [13], by comparison, requires that $n \approx 2.3 \times 10^9$. These observations confirm empirical findings in Reference [14], where the discretisation method of Algorithm 1 was applied in a practical problem. The theoretical guarantee that is provided by Corollary 8.5 was not available at the time, but the authors of Reference [14] did observe that the number of steps required to obtain empirical convergence was much lower than Škulj's theoretical upper bound in Reference [13].

In order to avoid having to use a too fine grid, as his uniform discretisation method tends to require, Škulj also considers another computational method, which he calls adaptive grid discretisation. In this method, the size of the discretisation steps is not forced to be constant, but is instead varied in order to try and reduce the total number of steps required. This second method is computationally more efficient, but has the disadvantage that it can only be applied to certain cases. Therefore, Škulj also considers a third method, which consists in combining his first two methods in such a way as to profit from their respective advantages.

Since the limit in Definition 7.4 does not require the partition $u$ to be uniform, our computational methods could, in principle, consider non-uniform 'adaptive grid' discretisations as well, thereby mimicking Škulj's second and third method. However, since we did not investigate the—theoretical and empirical—efficiency of such an approach, we are unable to compare it with the methods of Škulj. As discussed in our conclusions below, we consider this to be a promising avenue for future research.

## 11. Conclusions & future work

In this work, we formalised the concept of *imprecise continuous-time Markov chains*; a generalisation of continuous-time Markov chains that robustifies this model class with respect to often-made simplifying assumptions. In particular, it allows one to relax the requirement of having to specify exact values—point-estimates—for the numerical parameters of a continuous-time Markov chain, and furthermore relaxes the assumptions of time-homogeneity and Markovian probabilistic independence. Since these assumptions are often made pragmatically to ensure a tractable model class, it should not be obvious that this generalisation results in a practically workable model class. Nevertheless, we were still able to derive polynomial runtime-complexity algorithms for computing quantities of interest.

In the remainder of this section, given the length of this paper, we start by providing a brief summary of our most important technical contributions. We then finally close by discussing some future lines of research that we consider to be promising.

The starting point of this work was to formally define *continuous-time stochastic processes* within the framework of full conditional probabilities, using the notion of coherent conditional probabilities. After restricting ourselves to *well-behaved* stochastic processes, we then introduced a way to characterise the dynamics of such stochastic processes by means of their *outer partial derivatives*—sets of *rate matrices*, which we have shown to be non-empty, bounded and closed, and which

describe the infinitesimal state-transition rates of these processes. In particular, these outer partial derivatives allow us to describe this behaviour without imposing differentiability assumptions on the conditional state-transition probabilities. Next, by imposing the Markov property on such stochastic processes, we obtained continuous-time Markov chains as a special case. Finally, by additionally assuming the time-homogeneity of the state-transition probabilities, we have re-derived, within the framework of full conditional probabilities, the well known homogeneous continuous-time Markov chains, which are those stochastic processes whose matrix of conditional state-transition probabilities $T_t^s$ is given by the matrix exponential $e^{Q(s-t)}$, for some rate matrix $Q$.

Using these definitions, we have defined *imprecise continuous-time Markov chains*, which we abbreviate as ICTMC's, as sets of well-behaved stochastic processes whose dynamics are *consistent* with a given set of rate matrices $\mathscr{Q}$ and a given set of initial distributions $\mathscr{M}$. In particular, an ICTMC is a set of stochastic processes whose outer partial derivatives are contained within $\mathscr{Q}$ and whose initial distribution belongs to $\mathscr{M}$. We distinguished between three different ICTMC's, some of which impose additional conditions. The ICTMC $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ imposes no additional conditions. The ICTMC $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$ restricts attention to Markov processes, and the ICTMC $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}$ only considers Markov processes that are time-homogeneous. We have investigated the closure-properties of these three different ICTMC's and, in particular, have studied their closure under specific types of recombinations of their elements.

The *lower expectation* of an ICTMC was defined as the infimum of the expectations that correspond to the different stochastic processes in the ICTMC, that is, the highest lower bound on these expectations. Upper expectations and lower and upper probabilities were defined similarly, but since they correspond to special cases, we focused on lower expectations. The difficulty of computing such lower expectations was shown to vary between the sets $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$, $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$, and $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}$. For example, we have seen that the law of iterated lower expectation applies only to the set $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$.

In order to make the computation of lower expectations tractable, we introduced the notion of a lower transition rate operator $\underline{Q}$, and used it to define the corresponding *lower transition operator* $\underline{L}_t^s$. First, we showed that $\underline{Q}$ can be regarded as the lower envelope of a non-empty bounded set of rate matrices $\mathscr{Q}$, and that there is a one-to-one correspondence between lower transition rate operators $\underline{Q}$ and sets of rate matrices $\mathscr{Q}$ that are non-empty, bounded, closed and convex, and have *separately specified rows*. Next, we showed that the operator $\underline{L}_t^s$ satisfies various convenient algebraic properties, including time-homogeneity, differentiability and the semi-group property, and that it can be regarded as a generalisation of the well-known matrix exponential. Finally, we established that the operator $\underline{L}_t^s$ corresponds to a uniform—rather than pointwise—solution of the differential equation (45), which was previously proposed as a definition for the conditional lower expectation $\underline{\mathbb{E}}[f(X_s)|X_t]$ of an ICTMC in Reference [13]—thereby bypassing the role of sets of stochastic processes.

For the ICTMC's $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ and $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$, we have shown that the corresponding lower expectations satisfy a property that is analogous to the definition in Reference [13]. In particular, if the lower transition rate operator $\underline{Q}$ is the lower envelope of $\mathscr{Q}$, and if $\mathscr{Q}$ has separately specified rows, then $\underline{\mathbb{E}}[f(X_s)|X_t, X_u]$ is equal to $[\underline{L}_t^s f](X_t)$, where the lower expectation $\underline{\mathbb{E}}[\cdot|\cdot]$ can be taken with respect to either $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ or $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$. Moreover, we have seen that the largest set of stochastic processes for which this is the case, is the set of all well-behaved stochastic processes that are consistent with $\mathscr{Q}_{\underline{Q}}$, where $\mathscr{Q}_{\underline{Q}}$ is the unique largest set of rate matrices that dominates $\underline{Q}$.

Using this connection between the operator $\underline{L}_t^s$ and the lower expectations of $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ and $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$, we then went on to develop algorithms for the numerical computation of these lower expectations, for functions that depend on the state at a finite number of time points. For $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$, we showed that this is—in general—only possible if all but one of these time points is situated in the past. For $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$, such a restriction was not necessary, and we presented algorithms that are able to deal with functions that depend on the state at multiple future time points, provided that $\mathscr{Q}$ is convex. Our algorithms apply to conditional as well as unconditional lower expectations, and are able to compute these quantities within a guaranteed $\varepsilon$-error bound.

In summary, then, we have provided in this work a generalisation of continuous-time Markov chains to imprecise continuous-time Markov chains, which robustifies these models with respect to both numerical parameter assessments and the simplifying assumptions of time-homogeneity, Markovian probabilistic independence and differentiability of the state-transition probabilities. Notably, our lower transition operator provides a convenient analytical tool to represent such a model's lower expectations, and these quantities can be tractably computed using the algorithms that we presented. Therefore, this work can be seen as providing the required rigorous foundations for working with ICTMC's, both theoretically and practically. As such, we expect that it will provide plenty of avenues for future research.

On a technical level, a first important line of future research would be to extend our results to functions that are allowed to depend on the state $X_t$ at all time points $t \in \mathbb{R}_{\geq 0}$—rather than only a finite number of time points. Being able to compute the lower and upper expectation of such functions would allow for the study of some practically important aspects that our current framework cannot tackle yet, such as the time until the system reaches some desired state, the average time that is spent in a state, etcetera. Our framework is well-suited to be extended to such a setting, because the coherence of the stochastic processes that we consider ensures that their domain can be extended to include events that depend on the state at all time points. Furthermore—should this be desired—the results in Reference [36] allow for such an extension to be established in a $\sigma$-additive way.

A second technical line of future research would be to drop the assumption that $\mathscr{X}$ should be finite, and to extend our framework to the case where $\mathscr{X}$ is countably—or perhaps even uncountably—infinite.

Thirdly, still on the technical level, it would be interesting to investigate whether the requirement that $\mathscr{Q}$ should be convex—which we currently need in order to guarantee that Algorithm 3 yields a correct result—is really necessary. In fact,

we conjecture that it is not. Basically, the reason why Algorithm 3 currently requires $\mathscr{Q}$ to be convex, is because it relies on Theorem 6.5, the proof of which in turn relies on Theorem 6.3. While we believe that the convexity of $\mathscr{Q}$ is necessary in order for Theorem 6.3 to hold, we think that this is not the case for Theorem 6.5, and we think that it should be possible to provide the latter with an alternative proof that does not require $\mathscr{Q}$ to be convex.

Yet another interesting line of future research would be to generalise the models that we consider by allowing the set of rate matrices $\mathscr{Q}$ to be time-dependent, that is, to consider a set-valued function $\mathscr{Q}_t$. As one anonymous reviewer pointed out, this could be used for example to model the fact that over time, we become less and less certain about what rate matrix to use. The subtlety will then lie in choosing the kind of continuity assumptions that should be imposed on such a function $\mathscr{Q}_t$. If it is piecewise-constant, then we expect that our methods should translate fairly naturally to such a generalised model. For more complex time-variations, however, technical and computational difficulties may need to be overcome.

As far as the computational aspects of our work are concerned, there is also still quite a lot of room for future work. In particular, we believe that it should be possible to improve upon the efficiency of our methods for computing $L_t^s$. On the one hand, we did not yet study the effect of allowing for non-uniform discretisations—as Škulj does in his adaptive discretisation method [13]. On the other hand, our approach—as well as that of Škulj—does not yet exploit the fact that $L_t^s = L_0^{s-t}$ is guaranteed to converge to a limit as $s - t$ approaches infinity [27]; for large values of $s - t$, this feature should surely make it easier to compute $L_t^s$, whereas our—and Škulj's—methods become less efficient as $s - t$ increases. Some initial results in this direction can already be found in Reference [48].

Finally, on an algorithmic level, a practically important future line of research would be to develop specialised versions of our algorithms, which are tailored to deal with specific functions. The reason why we consider this to be important is because, for functions that depend on the state at a finite number of time points, the runtime-complexity of our algorithms is currently exponential in this number of time points. This is not problematic—and is in fact to be expected—because for general functions, even simply specifying the function has such a complexity. However, in many practical applications, there will typically be some underlying structure of the functions that we are interested in. For instance, for time points $v = s_0, \dots, s_m$, there may be functions $g_i \in \mathscr{L}(\mathscr{X}_{s_i})$, $i \in \{0, \dots, m\}$, such that $f \in \mathscr{L}(\mathscr{X}_v)$ is of the form $f(x_v) = \sum_{i=0}^m g_i(x_{s_i})$ or $f(x_v) = \prod_{i=0}^m g_i(x_{s_i})$. In the discrete time case, that is, for imprecise discrete time Markov chains, it has been shown that for these specific types of functions, inference algorithms such as Algorithm 3 simplify considerably, to the extent that their runtime complexity becomes linear in the number of time points [45,49]. We believe that it is possible—and even relatively straightforward—to extend these techniques to the continuous time framework of this paper. Furthermore, since similar techniques also lie at the heart of existing algorithms for imprecise discrete-time *hidden* Markov chains [50,51], we believe that this line of research could also serve as a first step towards the development of imprecise continuous-time *hidden* Markov chains—where the states are observed indirectly through noisy outputs—and in particular, towards the development of efficient algorithms for problems such as filtering and smoothing. In fact, some first steps in this direction have already been taken in Reference [48].

## Acknowledgements

## Appendix A. Proofs and lemmas for the results in Section 3

**Proposition 3.1.** *For any two transition matrices $T_1$ and $T_2$, their composition $T_1 T_2$ is also a transition matrix.*

**Proof.** Simply check each of the properties.  □

**Proposition 3.2.** *Consider any transition rate matrix $Q \in \mathscr{R}$, and any $\Delta \in \mathbb{R}_{\geq 0}$ such that $\Delta \| Q \| \leq 1$. Then the matrix $(I + \Delta Q)$ is a transition matrix.*

**Proof.** T1 follows from R1: for all $x \in \mathscr{X}$, R1 implies that

$$\sum_{y \in \mathscr{X}} [I + \Delta Q](x, y) = \sum_{y \in \mathscr{X}} I(x, y) + \Delta \sum_{y \in \mathscr{X}} Q(x, y) = 1 + \Delta 0 = 1.$$

T2 follows from R2 and because $0 \leq \Delta \| Q \| \leq 1$: for all $x, y \in \mathscr{X}$ such that $x \neq y$, $0 \leq \Delta \| Q \| \leq 1$ implies that $[I + \Delta Q](x, x) = 1 + \Delta Q(x, x) \geq 1 - \Delta \| Q \| \geq 0$, and R2 and $\Delta \geq 0$ imply that $[I + \Delta Q](x, y) = \Delta Q(x, y) \geq 0$.  □

**Proposition 3.3.** *Consider any transition matrix $T$, and any $\Delta \in \mathbb{R}_{>0}$. Then, the matrix $1/\Delta(T - I)$ is a transition rate matrix.*

**Proof.** This proof is analogous to that of Proposition 3.2; simply verify each of the properties in Definition 3.2. □

**Proposition 3.5.** *A set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ is bounded if and only if*

$$\inf\{Q(x, x) \colon Q \in \mathscr{Q}\} > -\infty \quad \text{for all } x \in \mathscr{X}. \tag{3}$$

**Proof.** We start by proving that Equation (3) implies $\|\mathscr{Q}\| < +\infty$. To this end, assume that Equation (3) holds. Since $\mathscr{X}$ is finite, it then follows that

$$\|\mathscr{Q}\| = \sup\{\|Q\| \colon Q \in \mathscr{Q}\} = \sup \left\{ \max \left\{ \sum_{y \in \mathscr{X}} |Q(x, y)| : x \in \mathscr{X} \right\} : Q \in \mathscr{Q} \right\}$$

$$= \max \left\{ \sup \left\{ \sum_{y \in \mathscr{X}} |Q(x, y)| : Q \in \mathscr{Q} \right\} : x \in \mathscr{X} \right\}.$$

Hence, there is some $x' \in \mathscr{X}$ such that

$$\|\mathscr{Q}\| = \sup_{Q \in \mathscr{Q}} \sum_{y \in \mathscr{X}} \left| Q(x', y) \right| = \sup_{Q \in \mathscr{Q}} \left( 2 \left| Q(x', x') \right| \right) = 2 \sup_{Q \in \mathscr{Q}} \left| Q(x', x') \right|$$

$$= -2 \inf_{Q \in \mathscr{Q}} \left( - \left| Q(x', x') \right| \right) = -2 \inf_{Q \in \mathscr{Q}} Q(x', x') < +\infty,$$

where the second and last equality follows from Definition 3.2, and the final inequality follows from Equation (3).

We next show that $\|\mathscr{Q}\| < +\infty$ implies Equation (3). To this end, consider any set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ such that $\|\mathscr{Q}\| < +\infty$, and assume *ex absurdo* that Equation (3) is not true. Then clearly, there is some $x' \in \mathscr{X}$ and $Q' \in \mathscr{Q}$ such that $Q'(x, x) < -\|\mathscr{Q}\|$. However, this implies that $\|\mathscr{Q}\| \geq \|Q'\| \geq \left| Q'(x, x) \right| > \|\mathscr{Q}\|$, which is a contradiction. Hence, it follows that Equation (3) must be true. □

**Lemma A.1.** *[28, Theorem 2.1.1] For any $Q \in \mathscr{R}$, we have that*

$$\frac{d}{d\Delta} e^{Q\Delta} \big|_{\Delta=0} := \lim_{\Delta \to 0} \frac{1}{\Delta} (e^{Q\Delta} - I) = Q.$$

**Proposition 3.6.** *For any $Q \in \mathscr{R}$, $\mathscr{T}_Q$ is a well-behaved transition matrix system.*

**Proof.** We start by showing that $\mathscr{T}_Q$ is a transition matrix system. Because of Proposition 3.4, $\mathscr{T}_Q$ is clearly a family of transition matrices. Consider now any $t, r, s \in \mathbb{R}_{\geq 0}$ such that $t \leq r \leq s$. It then follows from the definition of $\mathscr{T}_Q$ and [28, Theorem 2.1.1] that $T_t^s = T_t^r T_r^s$, and $T_t^t = I$. Because the $t, r, s$ are arbitrary, it follows from Definition 3.3 that $\mathscr{T}_Q$ is a transition matrix system.

To prove that $\mathscr{T}_Q$ is well-behaved, note that for any $t \in \mathbb{R}_{\geq 0}$, because of Definition 3.5,

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| = \limsup_{\Delta \to 0^+} \left\| \frac{1}{\Delta} (T_t^{t+\Delta} - I) - Q + Q \right\|$$

$$\leq \limsup_{\Delta \to 0^+} \left\| \frac{1}{\Delta} (T_t^{t+\Delta} - I) - Q \right\| + \|Q\| = \|Q\| < \infty,$$

where the first inequality follows from Proposition 2.1, the second equality follows from Lemma A.1 and the final inequality follows from the fact that $Q$ is real-valued. Because this holds for any $t \in \mathbb{R}_{\geq 0}$, the first condition in Equation (5) is satisfied. A similar argument shows that also the second condition is satisfied for all $t \in \mathbb{R}_{>0}$, and hence $\mathscr{T}_Q$ is well-behaved. □

**Proposition 3.7.** *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $\mathscr{T}^{\mathbf{I}}$ be a family of transition matrices $T_t^s$ that is defined for all $t, s \in \mathbf{I}$ with $t \leq s$. Then $\mathscr{T}^{\mathbf{I}}$ is a restricted transition matrix system on $\mathbf{I}$ if and only if, for all $t, r, s \in \mathbf{I}$ with $t \leq r \leq s$, it holds that $T_t^s = T_t^r T_r^s$ and $T_t^t = I$.*

**Proof.** If $\mathscr{T}^{\mathbf{I}}$ is a restricted transition matrix system, then, by definition, it is the restriction to $\mathbf{I}$ of some transition matrix system $\mathscr{T}$. Therefore, the 'only if' part of this result follows trivially from Definition 3.3.

For the 'if' part, we need to prove that for any family of transition matrices $\mathscr{T}^{\mathbf{I}}$ such that, for all $t, r, s \in \mathbf{I}$ with $t \leq r \leq s$, it holds that $T_t^s = T_t^r T_r^s$ and $T_t^t = I$, there is a transition matrix system $\mathscr{T}$ that coincides with $\mathscr{T}^{\mathbf{I}}$ on $\mathbf{I}$. In order to prove

this, it suffices to show that the unique family of transition matrices $\mathscr{T}$ that coincides with $\mathscr{T}^{\mathbf{I}}$ on $\mathbf{I}$ and that is otherwise defined by

$$
T_t^s := \begin{cases} I & \text{if } s < \min \mathbf{I} \\ T_{\min \mathbf{I}}^s & \text{if } t < \min \mathbf{I} \text{ and } s \in \mathbf{I} \\ T_{\min \mathbf{I}}^{\sup \mathbf{I}} & \text{if } t < \min \mathbf{I} \text{ and } \sup \mathbf{I} < s \\ T_t^{\sup \mathbf{I}} & \text{if } t \in \mathbf{I} \text{ and } \sup \mathbf{I} < s \\ I & \text{if } \sup \mathbf{I} < t \end{cases} \quad \text{for } t, s \in \mathbb{R}_{\geq 0} \text{ with } t \leq s \text{ and } [t, s] \nsubseteq \mathbf{I}, \tag{A.1}
$$

is a transition matrix system. This is a matter of straightforward verification. ☐

**Proposition 3.8.** *Consider any closed interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$, and let $\mathscr{T}^{\mathbf{I}}$ be a restricted transition matrix system on $\mathbf{I}$. Then $\mathscr{T}^{\mathbf{I}}$ is well-behaved if and only if*

$$
(\forall t \in \mathbf{I}^+) \ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_t^{t+\Delta} - I \right\| < +\infty \ \text{ and } \ (\forall t \in \mathbf{I}^-) \ \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T_{t-\Delta}^t - I \right\| < +\infty, \tag{6}
$$

*where $\mathbf{I}^+ := \mathbf{I} \setminus \{\sup \mathbf{I}\}$ and $\mathbf{I}^- := \mathbf{I} \setminus \{\min \mathbf{I}\}$.*

**Proof.** If $\mathscr{T}^{\mathbf{I}}$ is well-behaved, then, by definition, it is the restriction to $\mathbf{I}$ of a well-behaved transition matrix system $\mathscr{T}$. Therefore, the 'only if' part of this result follows trivially from Definition 3.4.

For the 'if' part, we need to show that for any restricted transition matrix system $\mathscr{T}^{\mathbf{I}}$ on $\mathbf{I}$ that satisfies Equation (6), there is a well-behaved transition matrix system $\mathscr{T}$ that coincides with $\mathscr{T}^{\mathbf{I}}$ on $\mathbf{I}$. Let $\mathscr{T}$ be constructed as in the proof of Proposition 3.7. Then, as explained in that proof, $\mathscr{T}$ is a transition matrix system that coincides with $\mathscr{T}^{\mathbf{I}}$ on $\mathbf{I}$. Therefore, it suffices to prove that $\mathscr{T}$ is well-behaved. We start by proving the first part of Equation (5). So consider any $t \in \mathbb{R}_{\geq 0}$. If $t \in \mathbf{I}^+$, then the desired inequality follows from Equation (6). If $t \notin \mathbf{I}^+$, then either $t < \min \mathbf{I}$ or $t \geq \sup \mathbf{I}$, and therefore, for sufficiently small $\Delta > 0$, it follows from Equation (A.1) that $T_t^{t+\Delta} = I$, thereby making the desired inequality trivially true. The second part of Equation (5) can be proved similarly. ☐

**Proposition 3.9.** *Consider two closed intervals $\mathbf{I}, \mathbf{J} \subseteq \mathbb{R}_{\geq 0}$ such that $\max \mathbf{I} = \min \mathbf{J}$, and any two restricted transition matrix systems $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$. Then their concatenation $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}} := \mathscr{T}^{\mathbf{I}} \otimes \mathscr{T}^{\mathbf{J}}$ is a restricted transition matrix system on $\mathbf{I} \cup \mathbf{J}$. Furthermore, if both $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$ are well behaved, then $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}}$ is also well-behaved.*

**Proof.** For all $t, s \in \mathbf{I} \cup \mathbf{J}$ such that $t \leq s$, it follows from Proposition 3.1 that the matrix $T_t^s$ that corresponds to $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}}$ is a transition matrix. Furthermore, for all $t \in \mathbf{I} \cup \mathbf{J}$, we have that either $t \in \mathbf{I}$ or $t \in \mathbf{J}$. In either case, we have that $T_t^t = I$, because either $T_t^t = {}^iT_t^t = I$ or $T_t^t = {}^jT_t^t = I$, with ${}^iT_t^t$ and ${}^jT_t^t$ corresponding to $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$, respectively. Next, we show that for all $t, q, s \in \mathbf{I} \cup \mathbf{J}$ with $t \leq q \leq s$, it holds that

$$
T_t^s = T_t^q T_q^s.
$$

If both $t, s \in \mathbf{I}$ or if both $t, s \in \mathbf{J}$, this clearly holds. Therefore, we may assume that $t \in \mathbf{I}$ and $s \in \mathbf{J}$. Suppose furthermore that $q \in \mathbf{I}$. Then, from the definition of the concatenation operator $\otimes$, we have that $T_q^s = T_q^r T_r^s$, with $r = \max \mathbf{I} = \min \mathbf{J}$. Because $t, q, r \in \mathbf{I}$, we know that $T_t^q T_q^r = T_t^r$, and hence, by the definition of the concatenation operator,

$$
T_t^q T_q^s = T_t^q T_q^r T_r^s = T_t^r T_r^s = T_t^s.
$$

An exactly analogous argument proves the case for $q \in \mathbf{J}$. Therefore, it follows from Proposition 3.7 that $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}}$ is a restricted transition matrix system.

It remains to prove that if $\mathscr{T}^{\mathbf{I}}$ and $\mathscr{T}^{\mathbf{J}}$ are both well-behaved, then $\mathscr{T}^{\mathbf{I} \cup \mathbf{J}}$ is also well-behaved. Due to Proposition 3.8, it suffices to prove Equation (6). We only prove the left part of this equation, that is, we prove that

$$
\left(\forall t \in (\mathbf{I} \cup \mathbf{J})^+\right) \ \limsup_{\Delta \to 0^+} \left\| \frac{1}{\Delta} \left( T_t^{t+\Delta} - I \right) \right\| < +\infty.
$$

The proof for the right part of Equation (6) is completely analogous. So consider any $t \in (\mathbf{I} \cup \mathbf{J})^+$. Since $\sup \mathbf{I} = \max \mathbf{I} = \min \mathbf{J}$, it follows that

$$
(\mathbf{I} \cup \mathbf{J})^+ := (\mathbf{I} \cup \mathbf{J}) \setminus \{\sup(\mathbf{I} \cup \mathbf{J})\} = (\mathbf{I} \cup \mathbf{J}) \setminus \{\sup \mathbf{J}\} = (\mathbf{I} \setminus \sup \mathbf{I}) \cup (\mathbf{J} \setminus \sup \mathbf{J}) = \mathbf{I}^+ \cup \mathbf{J}^+.
$$

Therefore, without loss of generality, we may assume that $t \in \mathbf{I}^+$. The desired result now follows by applying Proposition 3.8 to the well-behaved restricted transition matrix system $\mathscr{T}^{\mathbf{I}}$. ☐

The following lemma states a very useful norm inequality that we will use repeatedly in the proofs in this work. This result states that the distance between two composed transition matrices $T_1 T_2 \cdots T_n$ and $S_1 S_2 \cdots S_n$ is bounded from above by the sum of the distances $\|T_i - S_i\|$ of their component transition matrices.

**Lemma A.2.** *Let $T_1, \ldots, T_n$ and $S_1, \ldots, S_n$ be two finite sequences of transition matrices. Then*

$$\left\| \prod_{i=1}^n T_i - \prod_{i=1}^n S_i \right\| \leq \sum_{i=1}^n \|T_i - S_i\|.$$

**Proof.** This is a special case of Lemma E.4, which states a more general version. We have included this version separately because Lemma E.4 uses concepts that, on a chronological reading of this paper, would be undefined at this point. □

**Proposition 3.10.** *Consider any interval $\mathbf{I} \subseteq \mathbb{R}_{\geq 0}$ and let $d$ be the metric that is defined in Equation (7). The metric space $(\mathcal{T}^{\mathbf{1}}, d)$ is then complete.*

**Proof.** Consider any sequence $\{\mathscr{T}_i^{\mathbf{I}}\}_{i \in \mathbb{N}}$ in $\mathcal{T}^{\mathbf{1}}$ that is Cauchy. We will prove that this sequence converges to a limit that belongs to $\mathcal{T}^{\mathbf{1}}$. For all $i \in \mathbb{N}$ and any $t, s \in \mathbf{I}$ such that $t \leq s$, we will use ${}^i T_t^s$ to denote the transition matrix that corresponds to $\mathscr{T}_i^{\mathbf{I}}$.

Since $\{\mathscr{T}_i^{\mathbf{I}}\}_{i \in \mathbb{N}}$ is Cauchy, it follows from Equation (7) that

$$(\forall \varepsilon \in \mathbb{R}_{>0})\,(\exists n_\varepsilon \in \mathbb{N})\,(\forall k, \ell > n_\varepsilon)\,(\forall t, s \in \mathbf{I} \colon t \leq s)\ \left\| {}^k T_t^s - {}^\ell T_t^s \right\| < \varepsilon. \tag{A.2}$$

Clearly, for any $t, s \in \mathbf{I}$ such that $t \leq s$, this implies that the sequence $\{{}^i T_t^s\}_{i \in \mathbb{N}}$ is Cauchy. Since the set of all transition matrices—of the same finite dimension—is trivially complete, this implies that the sequence $\{{}^i T_t^s\}_{i \in \mathbb{N}}$ has a limit $T_t^s$, and that this limit is a transition matrix. We use $\mathscr{T}^{\mathbf{I}}$ to denote the family of transition matrices that consists of these limits.

Fix any $t, r, s \in \mathbf{I}$ such that $t \leq r \leq s$. Then for any $i \in \mathbb{N}$, because $\mathscr{T}_i^{\mathbf{I}}$ is a restricted transition matrix system, we know that ${}^i T_t^t = I$ and ${}^i T_t^s = {}^i T_t^r \, {}^i T_r^s$, which implies that $\|T_t^t - I\| = \|T_t^t - {}^i T_t^t\|$ and, due to Lemma A.2, that

$$\left\| T_t^s - T_t^r T_r^s \right\| \leq \left\| T_t^s - {}^i T_t^s \right\| + \left\| {}^i T_t^r \, {}^i T_r^s - T_t^r T_r^s \right\| \leq \left\| T_t^s - {}^i T_t^s \right\| + \left\| {}^i T_t^r - T_t^r \right\| + \left\| {}^i T_r^s - T_r^s \right\|.$$

Since we also know that $\lim_{i \to +\infty} {}^i T_t^t = T_t^t$, $\lim_{i \to +\infty} {}^i T_t^s = T_t^s$, $\lim_{i \to +\infty} {}^i T_t^r = T_t^r$ and $\lim_{i \to +\infty} {}^i T_r^s = T_r^s$, this implies that $\|T_t^t - I\| = 0$ and $\|T_t^s - T_t^r T_r^s\| = 0$, or equivalently, that $T_t^t = I$ and $T_t^s = T_t^r T_r^s$. Since this is true for any $t, r, s \in I$ such that $t \leq r \leq s$, and because we already know that the family $\mathscr{T}^{\mathbf{I}}$ consists of transition matrices, it follows from Proposition 3.7 that $\mathscr{T}^{\mathbf{I}}$ is a restricted transition matrix system. In the remainder of this proof, we will show that $\mathscr{T}^{\mathbf{I}} = \lim_{i \to \infty} \mathscr{T}_i^{\mathbf{I}}$.

Fix any $\varepsilon > 0$ and consider the corresponding $n_\varepsilon \in \mathbb{N}$ whose existence is guaranteed by Equation (A.2). Fix any $k > n_\varepsilon$. For any $t, s \in \mathbf{I}$ such that $t \leq s$, it then follows from Equation (A.2) that, for all $\ell > n_\varepsilon$:

$$\left\| {}^k T_t^s - T_t^s \right\| \leq \left\| {}^k T_t^s - {}^\ell T_t^s \right\| + \left\| {}^\ell T_t^s - T_t^s \right\| < \varepsilon + \left\| {}^\ell T_t^s - T_t^s \right\|.$$

Since $\lim_{\ell \to +\infty} {}^\ell T_t^s = T_t^s$, this implies that $\left\| {}^k T_t^s - T_t^s \right\| \leq \varepsilon$. Since this is true for all $t, s \in \mathbf{I}$ such that $t \leq s$, it follows from Equation (7) that $d(\mathscr{T}_k^{\mathbf{I}}, \mathscr{T}^{\mathbf{I}}) \leq \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we conclude that

$$(\forall \varepsilon \in \mathbb{R}_{>0})\,(\exists n_\varepsilon \in \mathbb{N})\,(\forall k > n_\varepsilon)\ d(\mathscr{T}_k^{\mathbf{I}}, \mathscr{T}^{\mathbf{I}}) \leq \varepsilon,$$

which implies that $\mathscr{T}^{\mathbf{I}} = \lim_{i \to \infty} \mathscr{T}_i^{\mathbf{I}}$. □

**Lemma A.3.** *Consider any $Q \in \mathscr{R}$ and any $\Delta \geq 0$. Then,*

$$\left\| e^{Q \Delta} - (I + \Delta Q) \right\| \leq \Delta^2 \, \|Q\|^2.$$

**Proof.** This is a special case of Lemma E.9, which states a more general version. We have included this version separately because Lemma E.9 uses concepts that, on a chronological reading of this paper, would be undefined at this point. □

**Lemma A.4.** *Consider any $Q \in \mathscr{R}$ and any $\Delta \geq 0$. Then,*

$$\left\| e^{Q \Delta} - I \right\| \leq \Delta \, \|Q\|.$$

**Proof.** This is a special case of Lemma E.10, which states a more general version. We have included this version separately because Lemma E.10 uses concepts that, on a chronological reading of this paper, would be undefined at this point. □

**Lemma A.5.** *If $Q_1$, $Q_2 \in \mathscr{R}$ and $\Delta \in \mathbb{R}_{\geq 0}$, then $\left\| e^{Q_1 \Delta} - e^{Q_2 \Delta} \right\| \leq \Delta \left\| Q_1 - Q_2 \right\|$.*

**Proof.** Consider any $n \in \mathbb{N}$. It then follows from Lemma A.2 that

$$\left\| e^{Q_1 \Delta} - e^{Q_2 \Delta} \right\| = \left\| \prod_{k=1}^{n} e^{Q_1 \frac{\Delta}{n}} - \prod_{k=1}^{n} e^{Q_2 \frac{\Delta}{n}} \right\| \leq \sum_{k=1}^{n} \left\| e^{Q_1 \frac{\Delta}{n}} - e^{Q_2 \frac{\Delta}{n}} \right\| = n \left\| e^{Q_1 \frac{\Delta}{n}} - e^{Q_2 \frac{\Delta}{n}} \right\|,$$

which, since we know from Lemma A.3 that

$$\left\| e^{Q_1 \frac{\Delta}{n}} - e^{Q_2 \frac{\Delta}{n}} \right\| \leq \left\| e^{Q_1 \frac{\Delta}{n}} - (I + \frac{\Delta}{n} Q_1) \right\| + \left\| \frac{\Delta}{n} (Q_1 - Q_2) \right\| + \left\| (I + \frac{\Delta}{n} Q_2) - e^{Q_2 \frac{\Delta}{n}} \right\|$$

$$\leq \frac{\Delta^2}{n^2} \left\| Q_1 \right\|^2 + \frac{\Delta}{n} \left\| Q_1 - Q_2 \right\| + \frac{\Delta^2}{n^2} \left\| Q_2 \right\|^2,$$

implies that

$$\left\| e^{Q_1 \Delta} - e^{Q_2 \Delta} \right\| \leq \frac{\Delta^2}{n} \left\| Q_1 \right\|^2 + \Delta \left\| Q_1 - Q_2 \right\| + \frac{\Delta^2}{n} \left\| Q_2 \right\|^2.$$

The result now follows by taking the limit for $n$ going to infinity. $\quad\square$

## Appendix B. Proofs for the results in Section 4

**Proof of F5–F8.** Consider any $A \in \mathscr{E}(\Omega)$ and $C \in \mathscr{E}(\Omega)_\emptyset$. It then follows from F2 and F3 that

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

and

$$P(A \cap C|C) = P(C|C) - P(C \setminus A|C) = 1 - P(C \setminus A|C). \tag{B.2}$$

F5 follows from Equation (B.1) and F1. F6 follows from Equations (B.1) and (B.2). F7 follows from F3, by letting $B := \emptyset$. F8 follows trivially from F2. $\quad\square$

**Corollary 4.3.** *Let $P$ be a real-valued map from $\mathscr{C} \subseteq \mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$ to $\mathbb{R}$. Then $P$ is a coherent conditional probability if and only if it can be extended to a full conditional probability.*

**Proof.** First assume that $P$ can be extended to a full conditional probability $P^*$. Theorem 4.1 then implies that $P^*$ is a coherent conditional probability, and therefore, since $P$ is the restriction of $P^*$ to $\mathscr{C}$, it clearly follows from Definition 4.2 that $P$ is a coherent conditional probability.

Conversely, if $P$ is a coherent conditional probability on $\mathscr{C}$, it follows from Theorem 4.2 that $P$ can be extended to a coherent conditional probability $P^*$ on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_\emptyset$, which, because of Theorem 4.1, is a full conditional probability. $\quad\square$

**Corollary 4.4.** *Let $P$ be a real-valued map from $\mathscr{C}^{\mathrm{SP}}$ to $\mathbb{R}$. Then $P$ is a stochastic process if and only if it is the restriction of a full conditional probability.*

**Proof.** Trivial consequence of Corollary 4.3. $\quad\square$

**Proposition 4.5.** *Let $P \in \mathbb{P}$ be a stochastic process. Then, for any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any sequence of time points $u \in \mathscr{U}_{<t}$, and any state assignment $x_u \in \mathscr{X}_u$, the corresponding (history dependent) transition matrix $T^s_{t, x_u}$ is—as its name suggests—a transition matrix, and $T^t_{t, x_u} = I$. Furthermore, $P$ is well-behaved if and only if, for every—possibly empty—time sequence $u \in \mathscr{U}$, any $x_u \in \mathscr{X}_u$ and any $t \in \mathbb{R}_{\geq 0}$ such that $t > u$:*

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T^{t+\Delta}_{t, x_u} - I \right\| < +\infty \tag{18}$$

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

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T^t_{t-\Delta, x_u} - I \right\| < +\infty. \tag{19}$$

**Proof.** The first part of the statement follows trivially from Corollary 4.4 and Definitions 3.1 and 4.1. The second part is an immediate consequence of Definition 4.4 and Equation (2). $\quad\square$

**Proposition 4.6.** *Consider any $P \in \mathbb{P}^{\mathsf{W}}$. Then $\overline{\partial}_+ T^t_{t, x_u}$, $\overline{\partial}_- T^t_{t, x_u}$ and $\overline{\partial} T^t_{t, x_u}$ are non-empty, bounded and closed subsets of $\mathscr{R}$.*

**Proof.** We only give the proof for $\overline{\partial}_+ T^t_{t, x_u}$. The proof for $\overline{\partial}_- T^t_{t, x_u}$ is completely analogous. The proof for $\overline{\partial} T^t_{t, x_u}$ then follows trivially because a union of two bounded, non-empty and closed sets is always bounded, non-empty and closed itself.

We start by establishing the boundedness of $\overline{\partial}_+ T^t_{t, x_u}$. Since $P$ is well-behaved, it follows from Proposition 4.5 that there is some $B > 0$ and $\delta > 0$ such that

$$(\forall 0 < \Delta < \delta) \; \left\| \frac{1}{\Delta} (T^{t+\Delta}_{t, x_u} - I) \right\| = \frac{1}{\Delta} \left\| (T^{t+\Delta}_{t, x_u} - I) \right\| \leq B. \tag{B.3}$$

Consider now any $Q \in \overline{\partial}_+ T^t_{t, x_u}$. Because of Equation (21), $Q$ is the limit of a sequence of matrices $\{Q_k\}_{k \in \mathbb{N}}$, defined by

$$Q_k := \frac{1}{\Delta_k} (T^{t+\Delta_k}_{t, x_u} - I) \quad \text{for all } k \in \mathbb{N}. \tag{B.4}$$

Because of Equation (B.3), the norms $\|Q_k\|$ of these matrices are eventually (for large enough $k$) bounded above by $B$. Hence, it follows that $\|Q\| \leq B$. Since this is true for any $Q \in \overline{\partial}_+ T^t_{t, x_u}$, we find that $\overline{\partial}_+ T^t_{t, x_u}$ is bounded.

In order to prove that $\overline{\partial}_+ T^t_{t, x_u}$ is non-empty, we consider any sequence $\{\Delta_k\}_{k \in \mathbb{N}} \to 0^+$. The corresponding sequence of matrices $\{Q_k\}_{k \in \mathbb{N}}$, as defined by Equation (B.4), is then bounded because $P$ is well-behaved—see Proposition 4.5—and therefore, it follows from the Bolzano–Weierstrass theorem that it has a convergent subsequence $\{Q_{k_i}\}_{i \in \mathbb{N}}$ of which we denote the limit by $Q^*$. Hence, we have found a sequence $\{\Delta_{k_i}\}_{i \in \mathbb{N}} \to 0^+$ such that $\{Q_{k_i}\}_{i \in \mathbb{N}} \to Q^*$. Since we know from Lemma 3.3 that each of the matrices in $\{Q_{k_i}\}_{i \in \mathbb{N}}$ is a rate matrix, the limit $Q^*$ is also a rate matrix, which therefore clearly belongs to $\overline{\partial}_+ T^t_{t, x_u}$.

We end by showing that $\overline{\partial}_+ T^t_{t, x_u}$ is closed, or equivalently, that for any converging sequence $\{Q^*_k\}_{k \in \mathbb{N}}$, of rate matrices in $\overline{\partial}_+ T^t_{t, x_u}$, the limit point $Q^* := \lim_{k \to +\infty} Q^*_k$ is again an element of $\overline{\partial}_+ T^t_{t, x_u}$. The argument goes as follows. First, since each of the rate matrices $Q^*_k$ belongs to the bounded set $\overline{\partial}_+ T^t_{t, x_u}$, their limit $Q^*$ is a (real-valued) rate matrix. Next, for any $k \in \mathbb{N}$, since $Q^*_k \in \overline{\partial}_+ T^t_{t, x_u}$, it follows from Equation (21) that there is some $0 < \Delta_k < 1/k$ such that $\|Q_k - Q^*_k\| \leq 1/k$, with $Q_k$ defined as in Equation (B.4). Consider now the sequences $\{Q_k\}_{k \in \mathbb{N}}$ and $\{\Delta_k\}_{k \in \mathbb{N}}$. Then on the one hand, we find that

$$0 \leq \limsup_{k \to +\infty} \left\| Q^* - Q_k \right\| \leq \limsup_{k \to +\infty} \left\| Q^* - Q^*_k \right\| + \limsup_{k \to +\infty} \left\| Q^*_k - Q_k \right\|$$

$$= \limsup_{k \to +\infty} \left\| Q^* - Q^*_k \right\| + \lim_{k \to +\infty} 1/k = \limsup_{k \to +\infty} \left\| Q^* - Q^*_k \right\| = 0,$$

which implies that sequence $\{Q_k\}_{k \in \mathbb{N}}$ converges to $Q^*$. On the other hand, we have that $\lim_{k \to +\infty} \Delta_k = 0$. Hence, because of Definition 4.8, it follows that $Q^* \in \overline{\partial}_+ T^t_{t, x_u}$. $\quad \square$

**Proposition 4.7.** *Consider any well-behaved stochastic process $P \in \mathbb{P}^{\mathsf{W}}$. Then, for any $t \in \mathbb{R}_{\geq 0}$, any sequence of time points $u \in \mathscr{U}_{<t}$, any state assignment $x_u \in \mathscr{X}_u$, and any $\varepsilon > 0$, there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$:*

$$(\exists Q \in \overline{\partial}_+ T^t_{t, x_u}) \left\| \frac{1}{\Delta} (T^{t+\Delta}_{t, x_u} - I) - Q \right\| < \varepsilon \tag{23}$$

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

$$(\exists Q \in \overline{\partial}_- T^t_{t, x_u}) \left\| \frac{1}{\Delta} (T^t_{t-\Delta, x_u} - I) - Q \right\| < \varepsilon. \tag{24}$$

**Proof.** Fix any $\varepsilon > 0$. Assume *ex absurdo* that

$$(\forall \delta > 0)(\exists 0 < \Delta < \delta)(\forall Q \in \overline{\partial}_+ T^t_{t, x_u}) \left\| \frac{1}{\Delta} (T^{t+\Delta}_{t, x_u} - I) - Q \right\| \geq \varepsilon.$$

Clearly, this implies the existence of a sequence $\{\Delta_k\}_{k \in \mathbb{N}} \to 0^+$ such that

$$\|Q_k - Q\| \geq \varepsilon \quad \text{for all } k \in \mathbb{N} \text{ and all } Q \in \overline{\partial}_+ T^t_{t, x_u}, \tag{B.5}$$

with $Q_k$ defined as in Equation (B.4). As we know from the proof of Proposition 4.6, the sequence $\{Q_k\}_{k \in \mathbb{N}}$ has a convergent subsequence $\{Q_{k_i}\}_{i \in \mathbb{N}}$ of which the limit $Q^*$ belongs to $\overline{\partial}_+ T^t_{t, x_u}$. On the one hand, since $\lim_{i \to +\infty} Q_{k_i} = Q^*$, we now have that $\lim_{i \to +\infty} \left\| Q_{k_i} - Q^* \right\| = 0$. On the other hand, since $Q^* \in \overline{\partial}_+ T^t_{t, x_u}$, it follows from Equation (B.5) that $\lim_{i \to +\infty} \left\| Q_{k_i} - Q^* \right\| \geq \varepsilon > 0$. From this contradiction, it follows that there must be some $\delta_1 > 0$ such that Equation (23) holds for all $0 < \Delta < \delta_1$. Similarly, using a completely analogous argument, we infer that if $t \neq 0$, there must be some $\delta_2 > 0$ such that Equation (24) holds for all $0 < \Delta < \delta_2$. Now let $\delta := \min\{\delta_1, \delta_2\}$ if $t \neq 0$ and let $\delta := \delta_1$ if $t = 0$. $\quad \square$

**Corollary 4.8.** *Consider any* $P \in \mathbb{P}^W$. *Then* $\overline{\partial}_+ T^t_{t, x_u}$ *is a singleton if and only if* $\partial_+ T^t_{t, x_u}$ *exists and, in that case,* $\overline{\partial}_+ T^t_{t, x_u} = \{\partial_+ T^t_{t, x_u}\}$. *Analogous results hold for* $\overline{\partial}_- T^t_{t, x_u}$ *and* $\partial_- T^t_{t, x_u}$, *and for* $\overline{\partial} T^t_{t, x_u}$ *and* $\partial T^t_{t, x_u}$.

**Proof.** This follows trivially from Proposition 4.7. □

## Appendix C. Proofs and lemmas for the results in Section 5

**Proposition 5.1.** *Consider a Markov chain* $P \in \mathbb{P}^M$ *and let* $\mathscr{T}_P$ *be the corresponding family of transition matrices. Then* $\mathscr{T}_P$ *is a transition matrix system. Furthermore,* $\mathscr{T}_P$ *is well-behaved if and only if* $P$ *is well-behaved.*

**Proof.** Consider any Markov chain $P \in \mathbb{P}^M$, with $\mathscr{T}_P$ its corresponding family of transition matrices. Then, because $P$ is a stochastic process, it follows from Proposition 4.5 that $T^t_t = I$ for all $t \in \mathbb{R}_{\geq 0}$.

Consider now any $t, r, s \in \mathbb{R}_{\geq 0}$ with $t \leq r \leq s$. We need to show that $T^s_t = T^r_t T^s_r$. If $t = r$, we have that $T^s_t = T^r_t T^s_r = I T^s_r = T^s_t$, and hence the result follows trivially. Similarly, the claim is trivial for $r = s$. Hence, it remains to show that the claim holds for $t < r < s$. It follows from Definition 5.1 that for all $x_t, x_r, x_s \in \mathscr{X}$,

$$P(X_s = x_s \mid X_r = x_r, X_t = x_t) = P(X_s = x_s \mid X_r = x_r).$$

Furthermore, because $P$ is a stochastic process, it follows from Corollary 4.4 that $P$ satisfies F4 and F3 on its domain. From F4, we infer that

$$P(X_s = x_s, X_r = x_r \mid X_t = x_t) = P(X_s = x_s \mid X_r = x_r, X_t = x_t) P(X_r = x_r \mid X_t = x_t)$$
$$= P(X_s = x_s \mid X_r = x_r) P(X_r = x_r \mid X_t = x_t),$$

where the second equality used the Markov property. From F3, we infer that

$$P(X_s = x_s \mid X_t = x_t) = \sum_{x_r \in \mathscr{X}} P(X_s = x_s, X_r = x_r \mid X_t = x_t).$$

From the definition of $\mathscr{T}_P$ in Definition 4.5, it now follows that, for any $x_t, x_s \in \mathscr{X}$,

$$T^s_t(x_t, x_s) = P(X_s = x_s \mid X_t = x_t) = \sum_{x_r \in \mathscr{X}} P(X_s = x_s, X_r = x_r \mid X_t = x_t)$$
$$= \sum_{x_r \in \mathscr{X}} P(X_s = x_s \mid X_r = x_r) P(X_r = x_r \mid X_t = x_t) = \sum_{x_r \in \mathscr{X}} T^r_t(x_t, x_r) T^s_r(x_r, x_s),$$

and hence, by the rules of matrix multiplication, we find that $T^s_t = T^r_t T^s_r$. Therefore, and because the $t, r, s \in \mathbb{R}_{\geq 0}$ were arbitrary, $\mathscr{T}_P$ is a transition matrix system.

The fact that $\mathscr{T}_P$ is well-behaved if and only if $P$ is well-behaved follows immediately from Definition 3.4 and Proposition 4.5 because the Markov property implies that $T^{t+\Delta}_{t, x_u} = T^{t+\Delta}_t$ and $T^t_{t-\Delta, x_u} = T^t_{t-\Delta}$. □

**Theorem 5.2.** *Let* $p$ *be any probability mass function on* $\mathscr{X}$ *and let* $\mathscr{T}$ *be a transition matrix system. Then there is a unique Markov chain* $P \in \mathbb{P}^M$ *such that* $\mathscr{T}_P = \mathscr{T}$ *and, for all* $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$. *Furthermore,* $P$ *is well-behaved if and only if* $\mathscr{T}$ *is well-behaved.*

**Proof.** Let

$$\mathscr{C} := \{(X_s = y, X_u = x_u) \in \mathscr{C}^{SP} : u \in \mathscr{U}_\emptyset, \ s > u, \ x_u \in \mathscr{X}_u, \ y \in \mathscr{X}\} \cup \{(X_0 = y, X_\emptyset = x_\emptyset) \in \mathscr{C}^{SP} : y \in \mathscr{X}\}$$

and consider a real-valued function $\tilde{P}$ on $\mathscr{C}$ that is defined by

$$\tilde{P}(X_s = y \mid X_u = x_u) := \begin{cases} p(y) & \text{if } u = \emptyset \\ T^s_{\max u}(x_{\max u}, y) & \text{otherwise} \end{cases} \quad \text{for all } (X_s = y, X_u = x_u) \in \mathscr{C}. \tag{C.1}$$

We first prove that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$. So consider any $n \in \mathbb{N}$ and, for all $i \in \{1, \ldots, n\}$, choose $(A_i, C_i) = (X_{s_i} = y_i, X_{u_i} = x_{u_i}) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max \left\{ \sum_{i=1}^n \lambda_i \mathbb{I}_{C_i}(\omega) \left( \tilde{P}(A_i \mid C_i) - \mathbb{I}_{A_i}(\omega) \right) \ \middle| \ \omega \in C_0 \right\} \geq 0, \tag{C.2}$$

with $C_0 := \cup_{i=1}^n C_i$. Since every sequence $u_i$ is finite, there is some finite set $w = \{w_0, w_1, \ldots, w_m\} \subset \mathbb{R}_{\geq 0}$ of time points, with $m \in \mathbb{N}$, such that $0 = w_0 < w_1 < \cdots < w_m$ and, for all $i \in \{1, \ldots, n\}$, $u_i \subseteq w$ and $s_i \in w$. Let $P_w$ be the restriction

of $\tilde{P}$ to $\mathscr{C}_w$, with $\mathscr{C}_w$ defined as in Lemma C.1. Then since $P_w$ clearly satisfies the conditions of Lemma C.1, it follows that $P_w$ is a coherent conditional probability. Because of Theorem 4.2, this implies that $P_w$ can be extended to a coherent conditional probability $\tilde{P}_w$ on $\mathscr{E}(\Omega) \times \mathscr{E}(\Omega)_{\emptyset}$, which, because of Theorem 4.1, is also a full conditional probability. Since $\tilde{P}_w$ is a coherent conditional probability, it now follows from Definition 4.2 that

$$\max \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega) \big( \tilde{P}_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega) \big) \,\middle|\, \omega \in C_0 \right\} \geq 0. \tag{C.3}$$

By comparing Equations (C.2) and (C.3), we see that in order to prove that $\tilde{P}$ is coherent, it suffices to show that, for all $i \in \{1, \dots, n\}$, $\tilde{P}_w(A_i|C_i) = \tilde{P}(A_i|C_i)$. So fix any $i \in \{1, \dots, n\}$. If $u_i = \emptyset$, then $s_i = 0 = w_0$ and therefore $(A_i, C_i) \in \mathscr{C}_w$, which implies that $\tilde{P}_w(A_i|C_i) = P_w(A_i|C_i) = \tilde{P}(A_i|C_i)$. If $u_i \neq \emptyset$, then since $u_i \subseteq w$, $s_i \in w$ and $s_i > u_i$, it follows from Lemma C.2 that $\tilde{P}_w(A_i|C_i) = \tilde{P}(A_i|C_i)$. Hence, $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$.

Therefore, due to Theorem 4.2, and because $\mathscr{C} \subseteq \mathscr{C}^{\mathrm{SP}}$, $\tilde{P}$ can be extended to a coherent conditional probability $P$ on $\mathscr{C}^{\mathrm{SP}}$, which, according to Definition 4.3, is a stochastic process. Due to Equation (C.1), this implies that $P$ is a Markov chain such that $\mathscr{T}_P = \mathscr{T}$ and, for all $y \in \mathscr{X}$, $P(X_0 = y) = p(y)$. Lemma C.3 implies that this Markov chain is unique and, since $\mathscr{T}_P = \mathscr{T}$, Proposition 5.1 implies that $\mathscr{T}$ is well-behaved if and only if $P$ is well-behaved. $\quad\square$

**Lemma C.1.** *Let $w = \{w_0, w_1, \dots, w_m\} \subset \mathbb{R}_{\geq 0}$ be a finite set of time points, with $m \in \mathbb{N}_0$, such that $w_0 < w_1 < \cdots < w_m$. Let $P_w$ be a real-valued function on*

$$\mathscr{C}_w \coloneqq \big\{ (X_{w_j} = y, X_u = x_u) \colon j \in \{0, \dots, m\},\ u = \{w_0, \dots, w_{j-1}\},\ y \in \mathscr{X},\ x_u \in \mathscr{X}_u \big\}$$

*such that, for any $j \in \{0, \dots, m\}$, $u = \{w_0, \dots, w_{j-1}\}$ and $x_u \in \mathscr{X}_u$, $P_w(X_{w_j} = x | X_u = x_u)$, as a function of $x \in \mathscr{X}$, is a probability mass function on $\mathscr{X}$. Then $P_w$ is a coherent conditional probability.*

**Proof.** We provide a proof by induction. Assume that this statement is true for any $m' < m$—this is trivially the case for $m = 0$. We will show that this implies that it is also true for $m$.

Consider any $n \in \mathbb{N}$ and, for all $i \in \{1, \dots, n\}$, choose $(A_i, C_i) \in \mathscr{C}_w$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max \left\{ \sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega) \big( P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega) \big) \,\middle|\, \omega \in C_0 \right\} \geq 0, \tag{C.4}$$

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

For any $i \in \{1, \dots, n\}$, since $(A_i, C_i) \in \mathscr{C}_w$, there is some $j_i \in \{0, \dots, m\}$ and, for all $\ell \in \{0, \dots, j_i\}$, some $z_{\ell,i} \in \mathscr{X}$ such that

$$A_i = (X_{w_{j_i}} = z_{j_i,i}) \quad \text{and} \quad C_i = (X_{w_0} = z_{0,i}, \dots, X_{w_{j_i-1}} = z_{j_i-1,i}).$$

Let $S = \{i \in \{1, \dots, n\} \colon j_i < m\}$. If $S \neq \emptyset$, then by the induction hypothesis, we know that

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

with $C_0^* \coloneqq \cup_{i \in S} C_i$. It follows that there is some $\omega^* \in C_0^* \subseteq C_0$ such that

$$\sum_{i \in S} \lambda_i \mathbb{I}_{C_i}(\omega^*) \big( P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega^*) \big) \geq 0. \tag{C.5}$$

If $S = \emptyset$, then let $\omega^*$ be any element of $C_0$. Equation (C.5) is then trivially satisfied. Hence, in all cases, we have found some $\omega^* \in C_0$ that satisfies Equation (C.5).

Let $C^* \coloneqq \cap_{1 \leq \ell < m}(X_{w_\ell} = \omega^*(w_\ell))$ and $S^* \coloneqq \{i \in \{1, \dots, n\} \colon C_i = C^*\}$. Then by the assumptions of this lemma, there is some probability mass function $p$ on $\mathscr{X}$ such that, for all $x \in \mathscr{X}$, $P_w(X_{w_m} = x | C^*) = p(x)$. For all $x \in \mathscr{X}$, let $\lambda_x \coloneqq \sum_{\{i \in S^* \colon z_{m,i} = x\}} \lambda_i$. Since $p$ is a probability mass function, it then follows that

$$\sum_{i \in S^*} \lambda_i P_w(A_i|C^*) = \sum_{x \in \mathscr{X}} \lambda_x p(x) \geq \sum_{x \in \mathscr{X}} \Big( \min_{y \in \mathscr{X}} \lambda_y \Big) p(x) = \Big( \min_{y \in \mathscr{X}} \lambda_y \Big) \sum_{x \in \mathscr{X}} p(x) = \min_{y \in \mathscr{X}} \lambda_y.$$

Now let $y^*$ be any element of $\mathscr{X}$ such that $\min_{y \in \mathscr{X}} \lambda_y = \lambda_{y^*}$ (since $\mathscr{X}$ is finite, this is always possible). Let $\omega^{**}$ be any path in $\Omega$ such that $\omega^{**} \in C^*$ and $\omega^{**}(w_m) = y^*$; Equation (12) guarantees that this $\omega^{**} \in \Omega$ exists. Then

$$\sum_{i \in S^*} \lambda_i \big( P_w(A_i|C^*) - \mathbb{I}_{A_i}(\omega^{**}) \big) \geq \min_{y \in \mathscr{X}} \lambda_y - \sum_{i \in S^*} \lambda_i \mathbb{I}_{A_i}(\omega^{**}) = \lambda_{y^*} - \lambda_{y^*} = 0,$$

where the first equality holds because, for every $i \in S^*$, $A_i = (X_{w_m} = z_{m,i})$.

Let $S^{**} \coloneqq \{1, \ldots, n\} \setminus (S \cup S^*)$. Since $\omega^{**} \in C^*$, we find that $\mathbb{I}_{C_i}(\omega^{**}) = \mathbb{I}_{C_i}(\omega^*)$ and $\mathbb{I}_{A_i}(\omega^{**}) = \mathbb{I}_{A_i}(\omega^*)$ for all $i \in S$, that $\mathbb{I}_{C_i}(\omega^{**}) = 1$ for all $i \in S^*$, and that $\mathbb{I}_{C_i}(\omega^{**}) = 0$ for all $i \in S^{**}$. Hence, it follows from Equation (C.5) that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega^{**})\big(P_w(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{**})\big) \geq \sum_{i \in S^*} \lambda_i \big(P_w(A_i|C^*) - \mathbb{I}_{A_i}(\omega^{**})\big).$$

By combining this inequality with the previous one, we find that in order to show that Equation (C.4) holds, it suffices to prove that $\omega^{**} \in C_0$.

In order to prove this, it suffices to notice that the question of whether or not a path $\omega \in \Omega$ belongs to $C_0$, only depends on the values $\omega(t)$ of $\omega$ at time points $t \in \{w_0, \ldots, w_{m-1}\}$. Indeed, since we infer from $\omega^{**} \in C^*$ that the value of $\omega^*$ and $\omega^{**}$ at these time points is the same, and because $\omega^* \in C_0$, this implies that $\omega^{**} \in C_0$. $\quad\square$

**Lemma C.2.** *Let* $w = \{w_0, w_1, \ldots, w_m\} \subset \mathbb{R}_{\geq 0}$ *be a finite set of time points, with* $m \in \mathbb{N}_0$, *such that* $w_0 < w_1 < \cdots < w_m$. *Let* $\mathscr{T}$ *be a transition matrix system and let* $\tilde{P}_w$ *be any full conditional probability such that for all* $j \in \{1, \ldots, m\}$ *and* $x_{w_\ell} \in \mathscr{X}, \ell \in \{0, \ldots, j\}$:

$$\tilde{P}_w(X_{w_j} = x_{w_j} | X_{w_0} = x_{w_0}, \ldots, X_{w_{j-1}} = x_{w_{j-1}}) = T_{w_{j-1}}^{w_j}(x_{w_{j-1}}, x_{w_j}).$$

*Then for any* $s \in w$ *and* $u \subseteq w$ *such that* $s > u$ *and* $u \neq \emptyset$, *any* $y \in \mathscr{X}$ *and any* $x_u \in \mathscr{X}_u$, *we have that*

$$\tilde{P}_w(X_s = y | X_u = x_u) = T_{\max u}^s(x_{\max u}, y).$$

**Proof.** Since $\emptyset \neq u \subseteq w$, $s \in w$ and $s > u$, it follows that there is some $j \in \{1, \ldots, m\}$ such that $s = w_j$ and $u \subseteq \{w_0, \ldots, w_{j-1}\}$.

We provide a proof by induction. If $s = w_1$, then $u = \{w_0\}$, and therefore, the result follows trivially from the assumptions in this lemma. Assume now that the result is true for $s = w_j$, with $1 \leq j < m$. We will prove that this implies that it is also true for $s = w_{j+1}$. We consider two cases: $\max u = w_j$ and $\max u < w_j$.

If $\max u = w_j$, then with $v \coloneqq \{w_0, \ldots, w_j\} \setminus u$:

$$\begin{aligned}
\tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u) &= \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y, X_v = z_v | X_u = x_u) \\
&= \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u, X_v = z_v) \tilde{P}_w(X_v = z_v | X_u = x_u) \\
&= \sum_{z_v \in \mathscr{X}_v} T_{\max u}^{w_{j+1}}(x_{\max u}, y) \tilde{P}_w(X_v = z_v | X_u = x_u) \\
&= T_{\max u}^{w_{j+1}}(x_{\max u}, y) \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_v = z_v | X_u = x_u) = T_{\max u}^{w_{j+1}}(x_{\max u}, y),
\end{aligned}$$

where the first equality follows from F3, the second equality follows from F4, the third equality follows from the assumptions in this lemma and the fact that $\max u = w_j$, and the last equality follows from F3 and F8.

If $\max u < w_j$, then with $v \coloneqq \{w_0, \ldots, w_{j-1}\} \setminus u$:

$$\begin{aligned}
\tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u) &= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y, X_{w_j} = z_{w_j}, X_v = z_v | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_{w_{j+1}} = y | X_u = x_u, X_{w_j} = z_{w_j}, X_v = z_v) \\
&\qquad\qquad \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \tilde{P}_w(X_{w_j} = z_{w_j} | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} T_{w_j}^{w_{j+1}}(z_{w_j}, y) \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \tilde{P}_w(X_{w_j} = z_{w_j} | X_u = x_u) \\
&= \sum_{z_{w_j} \in \mathscr{X}} \sum_{z_v \in \mathscr{X}_v} T_{w_j}^{w_{j+1}}(z_{w_j}, y) \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) \\
&= \sum_{z_{w_j} \in \mathscr{X}} T_{w_j}^{w_{j+1}}(z_{w_j}, y) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) \sum_{z_v \in \mathscr{X}_v} \tilde{P}_w(X_v = z_v | X_u = x_u, X_{w_j} = z_{w_j}) \\
&= \sum_{z_{w_j} \in \mathscr{X}} T_{w_j}^{w_{j+1}}(z_{w_j}, y) T_{\max u}^{w_j}(x_{\max u}, z_{w_j}) = T_{\max u}^{w_{j+1}}(x_{\max u}, y),
\end{aligned}$$

where the first equality follows from F3, the second equality follows from F4, the third equality follows from the assumptions in this lemma, the fourth equality follows from the induction hypothesis, the sixth equality follows from F3 and F8, and the last equality follows from Equation (4).  □

**Lemma C.3.** *Consider two Markov chains $P_1, P_2 \in \mathbb{P}^M$ such that $\mathcal{T}_{P_1} = \mathcal{T}_{P_2}$ and, for all $y \in \mathcal{X}$, $P_1(X_0 = y) = P_2(X_0 = y)$. Then $P_1 = P_2$.*

**Proof.** Let $\mathcal{T} := \mathcal{T}_{P_1} = \mathcal{T}_{P_2}$ be the common transition matrix system of $P_1$ and $P_2$ and let $p$ be their common initial probability mass function, as defined by $p(y) := P_1(X_0 = y) = P_2(X_0 = y)$ for all $y \in \mathcal{X}$. Let $\tilde{P}$ be a real-valued function on $\mathscr{C}$, with $\mathscr{C}$ and $\tilde{P}$ defined as in the proof of Theorem 5.2. It then follows from Definition 5.1 that the restriction of $P_1$ and $P_2$ to $\mathscr{C}$ is equal to $\tilde{P}$. Furthermore, for any $s > 0$, $y \in \mathcal{X}$ and $j \in \{1, 2\}$, we find that

$$P_j(X_s = y) = \sum_{x \in \mathcal{X}} P_j(X_s = y, X_0 = x) = \sum_{x \in \mathcal{X}} P_j(X_s = y | X_0 = x) P_j(X_0 = x) = \sum_{x \in \mathcal{X}} \tilde{P}(X_s = y | X_0 = x) \tilde{P}(X_0 = x).$$

Hence, the restrictions of $P_1$ and $P_2$ to

$$\mathscr{C}^* := \mathscr{C} \cup \{(X_s = y, X_\emptyset = x_\emptyset) : s \in \mathbb{R}_{>0}, \ y \in \mathcal{X}\}$$

$$= \{(X_s = y, X_u = x_u) : u \in \mathcal{U}, \ s \in \mathbb{R}_{\geq 0}, \ s > u, \ x_u \in \mathcal{X}_u, \ y \in \mathcal{X}\}$$

are identical. We denote this common restriction by $\tilde{P}^*$.

Consider now any $(A, X_u = x_u) \in \mathscr{C}^{SP}$. Then since $A \in \mathscr{A}_u$, there is some finite set of time points $v = \{v_1, v_2, \ldots, v_n\} \subseteq \mathbb{R}_{\geq 0}$, with $n \in \mathbb{N}$, such that $\max u < v_1 < v_2 < \cdots < v_n$, and some set $S \subseteq \mathcal{X}_{u \cup v}$ such that $A = \cup_{z_{u \cup v} \in S} (X_{u \cup v} = z_{u \cup v})$. Let $S_v := \{z_v \in \mathcal{X}_v : (x_u, z_v) \in S\}$. For any $j \in \{1, 2\}$, we then find that

$$
\begin{aligned}
P_j(A | X_u = x_u) &= \sum_{z_{u \cup v} \in S} P_j(X_{u \cup v} = z_{u \cup v} | X_u = x_u) \\
&= \sum_{z_v \in S_v} P_j(X_{v_1} = z_{v_1}, X_{v_2} = z_{v_2}, \ldots, X_{v_n} = z_{v_n} | X_u = x_u) \\
&= \sum_{z_v \in S_v} \prod_{i=1}^{n} P_j(X_{v_i} = z_{v_i} | X_u = x_u, X_{v_1} = z_{v_1}, \ldots, X_{v_{i-1}} = z_{v_{i-1}}) \\
&= \sum_{z_v \in S_v} \prod_{i=1}^{n} \tilde{P}^*(X_{v_i} = z_{v_i} | X_u = x_u, X_{v_1} = z_{v_1}, \ldots, X_{v_{i-1}} = z_{v_{i-1}}),
\end{aligned}
$$

which implies that $P_1(A | X_u = x_u) = P_2(A | X_u = x_u)$. Since this is the case for any $(A, X_u = x_u) \in \mathscr{C}^{SP}$, it follows that $P_1 = P_2$.  □

**Corollary 5.3.** *Consider any rate matrix $Q \in \mathcal{R}$ and let $p$ be an arbitrary probability mass function on $\mathcal{X}$. Then there is a unique Markov chain $P \in \mathbb{P}^M$ such that $\mathcal{T}_P = \mathcal{T}_Q$ and, for all $y \in \mathcal{X}$, $P(X_0 = y) = p(y)$. Furthermore, this unique Markov chain is well-behaved and homogeneous.*

**Proof.** Since we know from Proposition 3.6 that $\mathcal{T}_Q$ is a well-behaved transition matrix system, it follows from Theorem 5.2 that there is a unique Markov chain $P \in \mathbb{P}^M$ such that $\mathcal{T}_P = \mathcal{T}_Q$ and, for all $y \in \mathcal{X}$, $P(X_0 = y) = p(y)$, and that this Markov chain is furthermore well-behaved. Since it—trivially—follows from Definition 3.5 that $\mathcal{T}_Q$ satisfies Equation (26), Definition 5.2 implies that $P$ is homogeneous.  □

**Theorem 5.4.** *For any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{WHM}$, there is a unique rate matrix $Q \in \mathcal{R}$ such that $\mathcal{T}_P = \mathcal{T}_Q$.*[11]

**Proof.** Because of Proposition 4.6, we know that $\overline{\partial}_+ T_0^0$ is a non-empty bounded set of rate matrices, which implies that there is some real $B > 0$ such that $\|Q'\| \leq B$ for all $Q' \in \overline{\partial}_+ T_0^0$. Let $Q$ be any element of $\overline{\partial}_+ T_0^0$.

Fix any $c \geq 0$, $\varepsilon > 0$ and $\delta > 0$. It then follows from Proposition 4.7 and N9 that there is some $\delta^* > 0$ such that

$$(\forall 0 < \Delta^* < \delta^*) \, (\exists Q^* \in \overline{\partial}_+ T_0^0) \, \left\| T_0^{\Delta^*} - (I + \Delta^* Q^*) \right\| < \Delta^* \varepsilon. \tag{C.6}$$

---

[11] Although our proof for this result starts from scratch, this result is essentially well known. Our version of it should be regarded as a (re)formulation that is adapted to our terminology and notation and, in particular, to our use of coherent and/or full conditional probabilities.

Furthermore, because of Equation (21) and N9, there is some $0 < \Delta < \min\{\delta, \delta^*\}$ such that

$$\left\| T_0^\Delta - (I + \Delta Q) \right\| < \Delta \varepsilon. \tag{C.7}$$

If we now define $n := \lfloor c/\Delta \rfloor$ and $d := c - n\Delta$, then $n\Delta \le c < (n+1)\Delta$ and therefore also $0 \le d < \Delta$. Because of Proposition 5.1, Equation (4) and Definition 5.2, we know that

$$T_0^c = \left( \prod_{j=1}^{n} T_{(j-1)\Delta}^{j\Delta} \right) T_{n\Delta}^c = \left( T_0^\Delta \right)^n T_0^d$$

and therefore, it follows from Lemma A.2 that

$$\left\| e^{Qc} - T_0^c \right\| = \left\| \left( T_0^\Delta \right)^n T_0^d - \left( e^{Q\Delta} \right)^n e^{Qd} \right\| \le n \left\| T_0^\Delta - e^{Q\Delta} \right\| + \left\| T_0^d - e^{Qd} \right\|. \tag{C.8}$$

From Equation (C.7) and Lemma A.3, we infer that

$$\left\| T_0^\Delta - e^{Q\Delta} \right\| \le \left\| T_0^\Delta - (I + \Delta Q) \right\| + \left\| (I + \Delta Q) - e^{Q\Delta} \right\| \le \Delta \varepsilon + \Delta^2 \left\| Q \right\|^2. \tag{C.9}$$

Since $d < \Delta < \delta^*$, we infer from Equation (C.6) that there is some $Q^* \in \overline{\partial}_+ T_0^0$ such that $\left\| T_0^d - (I + dQ^*) \right\| < d\varepsilon$. Hence, also using Lemma A.3, we find that

$$\left\| T_0^d - e^{Qd} \right\| \le \left\| T_0^d - (I + dQ^*) \right\| + \left\| (I + dQ^*) - (I + dQ) \right\| + \left\| (I + dQ) - e^{Qd} \right\|$$
$$\le d\varepsilon + d \left\| Q^* - Q \right\| + d^2 \left\| Q \right\|^2 \le d\varepsilon + d \left\| Q^* \right\| + d \left\| Q \right\| + d^2 \left\| Q \right\|^2. \tag{C.10}$$

By combining Equations (C.8), (C.9) and (C.10), it follows that

$$\left\| e^{Qc} - T_0^c \right\| \le n\Delta \varepsilon + n\Delta^2 \left\| Q \right\|^2 + d\varepsilon + d \left\| Q^* \right\| + d \left\| Q \right\| + d^2 \left\| Q \right\|^2.$$

Taking into account that $\| Q \| \le B$, $\| Q^* \| \le B$, $n\Delta \le c$ and $d < \Delta < \delta$, this implies that

$$\left\| e^{Qc} - T_0^c \right\| \le c\varepsilon + c\delta B^2 + \delta\varepsilon + 2\delta B + \delta^2 B^2.$$

Since this is true for any $\varepsilon > 0$ and $\delta > 0$, it follows that $\left\| e^{Qc} - T_0^c \right\| \le 0$, which implies that $T_0^c = e^{Qc}$. Since this is true for all $c \ge 0$, it follows from Definition 5.2 that

$$T_t^s = T_0^{s-t} = e^{Q(s-t)} \quad \text{for all } 0 \le t \le s, \tag{C.11}$$

or equivalently, that $\mathscr{T}_P = \mathscr{T}_Q$.

Finally, we prove that $Q$ is unique. Assume *ex absurdo* that this is not the case, or equivalently, that there are rate matrices $Q_1$ and $Q_2$, with $Q_1 \ne Q_2$, such that $\mathscr{T}_P = \mathscr{T}_{Q_1}$ and $\mathscr{T}_P = \mathscr{T}_{Q_2}$. For all $\Delta > 0$, we then have that $T_0^\Delta = e^{Q_1\Delta} = e^{Q_2\Delta}$, and therefore, it follows from Lemma A.1 that $\partial_+ T_0^0 = Q_1$ and $\partial_+ T_0^0 = Q_2$, which implies that $Q_1 = Q_2$. From this contradiction, it follows that $Q$ is indeed unique. $\quad\square$

**Proposition 5.5.** *Consider any well-behaved homogeneous Markov chain $P \in \mathbb{P}^{\mathrm{WHM}}$ and let $Q_P \in \mathscr{R}$ be its corresponding rate matrix. Then $\partial T_t^t = \partial_+ T_t^t = \partial_- T_t^t = Q_P$ and $\overline{\partial} T_t^t = \overline{\partial}_+ T_t^t = \overline{\partial}_- T_t^t = \{Q_P\}$.*

**Proof.** The result about the partial derivatives is an immediate consequence of Lemma A.1 and the fact that $T_t^{t+\Delta} = T_{t-\Delta}^t = e^{Q\Delta}$. The result about the outer partial derivatives then follows from Corollary 4.8. $\quad\square$

## Appendix D. Proofs for the results in Section 6

**Proposition 6.1.** *Consider any bounded set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then,*

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

**Proof.** This is immediate from Definitions 6.3 and 6.4 and the fact that $\mathbb{P}^{\mathrm{WHM}} \subseteq \mathbb{P}^{\mathrm{WM}} \subseteq \mathbb{P}^{\mathrm{W}}$. $\quad\square$

**Proposition 6.2.** *Consider any non-empty bounded set of rate matrices $\mathscr{Q}$ and let $\mathscr{M}$ be any non-empty set of probability mass functions on $\mathscr{X}$. Then for any $p \in \mathscr{M}$, any ordered finite sequence of time points $u = t_0, \ldots, t_n$ in $\mathscr{U}_\emptyset$ and any collection of rate matrices $Q_0, \ldots, Q_{n+1} \in \mathscr{Q}$, there is a well-behaved continuous-time Markov chain $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathrm{WM}}$ such that $P(X_0 = y) = p(y)$ for all $y \in \mathscr{X}$ and such that $\mathscr{T}_P$ is given by Equation (27).*

**Proof.** Proposition 5.6 implies the existence of a process $P \in \mathbb{P}^{WM}$ such that $P(X_0 = y)$ for all $y \in \mathscr{X}$ and such that $\mathscr{T}_P$ satisfies Equation (27). It remains to show that $P \in \mathbb{P}^{WM}_{\mathscr{Q}, \mathscr{M}}$. Because of Equation (28), and since we already know that $p \in \mathscr{M}$, this means that we have to show that $\overline{\partial} T_t^t \subseteq \mathscr{Q}$ for all $t \in \mathbb{R}_{\geq 0}$. To this end, consider any $t \in \mathbb{R}_{\geq 0}$.

We consider several cases. If $t < t_0$, then $\overline{\partial} T_t^t$ corresponds to $\mathscr{T}_{Q_0}^{[0, t_0]}$, and it then follows from Proposition 5.5 that $\overline{\partial} T_t^t = \{Q_0\} \subseteq \mathscr{Q}$. If $t > t_n$, then $\overline{\partial} T_t^t$ corresponds to $\mathscr{T}_{Q_{n+1}}^{[t_n, \infty)}$, in which case $\overline{\partial} T_t^t = \{Q_{n+1}\} \subseteq \mathscr{Q}$. Similarly, if there is some $i \in \{1, \ldots, n\}$ such that $t \in (t_{i-1}, t_i)$, then $\overline{\partial} T_t^t$ corresponds to $\mathscr{T}_{Q_i}^{[t_{i-1}, t_i]}$, and therefore $\overline{\partial} T_t^t = \{Q_i\} \subseteq \mathscr{Q}$. The only remaining case is when $t = t_i \neq 0$ for some $i \in \{0, \ldots, n\}$. In this case, we have that $\overline{\partial}_+ T_t^t = \{Q_{i+1}\}$ and, if $t \neq 0$, that $\overline{\partial}_- T_t^t = \{Q_i\}$, and therefore, it follows from Definition 4.7 that $\overline{\partial} T_t^t \subseteq \mathscr{Q}$.  □

**Theorem 6.3.** *Consider a non-empty convex set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Fix a finite sequence of time points $u \in \mathscr{U}$. Choose any $P_\emptyset \in \mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}$ and, for all $x_u \in \mathscr{X}_u$, choose some $P_{x_u} \in \mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}$. Then there is a stochastic process $P \in \mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}$ such that, for all $u_1, u_2 \subseteq u$ such that $u_1 < u_2$, all $x_u \in \mathscr{X}_u$ and all $A \in \mathscr{A}_u$:*

$$P(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) = P_\emptyset(X_{u_2} = x_{u_2} | X_{u_1} = x_{u_1}) \tag{29}$$

*and*

$$P(A | X_u = x_u) = P_{x_u}(A | X_u = x_u). \tag{30}$$

**Proof.** This proof is rather lengthy, and consists of two parts. First, we will show that there exists a stochastic process $P$ that satisfies Equations (29) and (30), by constructing it as the extension of a coherent conditional probability on a set of events $\mathscr{C} \subset \mathscr{C}^{SP}$. Next, we will finish the proof by showing that $P \in \mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}$, as desired.

Let $\mathscr{C} := \mathscr{C}_\emptyset \cup (\bigcup_{x_u \in \mathscr{X}_u} \mathscr{C}_{x_u})$, with

$$\mathscr{C}_\emptyset := \left\{ (A, X_v = x_v) \in \mathscr{C}^{SP} : v \in \mathscr{U}_{< \max u} \text{ and } A \in \langle \{(X_t = x) : x \in \mathscr{X}, t \in [0, \max u]\} \rangle \right\} \tag{D.1}$$

and, for all $x_u \in \mathscr{X}_u$,

$$\mathscr{C}_{x_u} := \{ (A, X_v = x_v) \in \mathscr{C}^{SP} : u \subseteq v \in \mathscr{U}, x_{v \setminus u} \in \mathscr{X}_{v \setminus u}, A \in \mathscr{A}_{u \cup (v \setminus [0, \max u])} \}. \tag{D.2}$$

Consider a real-valued function $\tilde{P}$ on $\mathscr{C}$ that is defined, for all $(A, X_v = x_v) \in \mathscr{C}$, by

$$\tilde{P}(A | X_v = x_v) := \begin{cases} P_\emptyset(A | X_v = x_v) & \text{if } (A, X_v = x_v) \in \mathscr{C}_\emptyset \\ P_{x_u}(A | X_{u \cup (v \setminus [0, \max u])} = x_{u \cup (v \setminus [0, \max u])}) & \text{if } (A, X_v = x_v) \in \mathscr{C}_{x_u}. \end{cases} \tag{D.3}$$

We first prove that $\tilde{P}$ is a coherent conditional probability on $\mathscr{C}$. So consider any $n \in \mathbb{N}$ and, for all $i \in \{1, \ldots, n\}$, choose $(A_i, C_i) \in \mathscr{C}$ and $\lambda_i \in \mathbb{R}$. We need to show that

$$\max \left\{ \sum_{i=1}^n \lambda_i \mathbb{I}_{C_i}(\omega) \left( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega) \right) \, \middle| \, \omega \in C_0 \right\} \geq 0, \tag{D.4}$$

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

Let $S^* := \{i \in \{1, \ldots, n\} : (A_i, C_i) \in \mathscr{C}_\emptyset\}$. We first consider the case $S^* \neq \emptyset$. Then since $P_\emptyset$ is a stochastic process, it follows from Equation (D.3) and Definitions 4.3 and 4.2 that

$$\max \left\{ \sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega) \left( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega) \right) \, \middle| \, \omega \in C^* \right\} \geq 0,$$

with $C^* := \cup_{i \in S^*} C_i$. Therefore, there is some $\omega^* \in C^*$ such that

$$\sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega^*) \left( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega^*) \right) \geq 0. \tag{D.5}$$

If $S^* = \emptyset$, we let $\omega^*$ be any element of $C_0$ (this is always possible, because $C_0 \neq \emptyset$). Clearly, this path $\omega^*$ will then also satisfy Equation (D.5)—because the left-hand side is a sum over an empty set and therefore zero.

Now let $x_u^* \in \mathscr{X}_u$ be defined by $x_u^* := \omega^*|_u$. Then for all $i \in \{1, \ldots, n\}$ such that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$, we know from Equation (D.2) that there are $u \subseteq v_i \in \mathscr{U}$ and $x_{v_i \setminus u} \in \mathscr{X}_{v_i \setminus u}$ such that

$$C_i = (X_u = x_u^*) \cap (X_{v_i \setminus u} = x_{v_i \setminus u}) = C_i^* \cap C_i^{**}, \tag{D.6}$$

with $C_i^* := (X_{(v_i \setminus u) \cap [0, \max u]} = x_{(v_i \setminus u) \cap [0, \max u]})$, and

$$C_i^{**} := (X_u = x_u^*) \cap (X_{v_i \setminus [0, \max u]} = x_{v_i \setminus [0, \max u]}). \tag{D.7}$$

Using this notation, we define

$$S^{**} \coloneqq \{i \in \{1, \ldots, n\} \colon (A_i, C_i) \in \mathscr{C}_{x_u^*} \text{ and } \mathbb{I}_{C_i^*}(\omega^*) = 1\}. \tag{D.8}$$

We first consider the case $S^{**} \neq \emptyset$. Since $P_{x_u^*}$ is a stochastic process, it then follows from Definitions 4.3 and 4.2 that

$$\max \left\{ \sum_{i \in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega) \big( P_{x_u^*}(A_i | C_i^{**}) - \mathbb{I}_{A_i}(\omega) \big) \,\middle|\, \omega \in C^{**} \right\} \geq 0,$$

with $C^{**} \coloneqq \cup_{i \in S^{**}} C_i^{**}$. Because of Equation (D.3), this implies that

$$\max \left\{ \sum_{i \in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega) \big( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega) \big) \,\middle|\, \omega \in C^{**} \right\} \geq 0,$$

which allows us to infer that there is some $\omega^{**} \in C^{**}$ such that

$$\sum_{i \in S^{**}} \lambda_i \mathbb{I}_{C_i^{**}}(\omega^{**}) \big( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega^{**}) \big) \geq 0. \tag{D.9}$$

Furthermore, since $\omega^{**} \in C^{**}$, Equation (D.7) implies that

$$\omega^{**}|_u = x_u^* = \omega^*|_u. \tag{D.10}$$

If $S^{**} = \emptyset$, we let $\omega^{**} = \omega^*$. Clearly, also in this case, $\omega^{**}$ satisfies Equations (D.9) and (D.10).

For any $i \in \{1, \ldots, n\}$, because $(A_i, C_i) \in \mathscr{C}$, there exists some finite sequence of time points $w_{C_i} \in \mathscr{U}$ such that $C_i$ only depends on the time points in $w_{C_i}$. Furthermore, it follows from Equations (D.1) and (D.2) that $A_i$ is an element of some algebra $\mathscr{A}$ that is generated by a set of events that only depend on a finite number of time points. Therefore, there is also some finite sequence of time points $w_{A_i} \in \mathscr{U}$, such that $A_i$ only depends on the time points in $w_{A_i}$. If we now let $w_i \coloneqq w_{A_i} \cup w_{C_i}$, then $(A_i, C_i)$ only depends on the (finite) sequence of time points $w_i$.

Because this holds for any $i \in \{1, \ldots, n\}$, this implies the existence of some finite sequence $w \in \mathscr{U}$ such that $w_i \subseteq w$ for all $i \in \{1, \ldots, n\}$.

Now let $\omega^{***} \in \Omega$ be any path such that, for all $s \in w$,

$$\omega^{***}(s) \coloneqq \begin{cases} \omega^*(s) & \text{if } s < \max u \\ \omega^{**}(s) & \text{if } s \geq \max u \end{cases}$$

Equation (12) guarantees that this $\omega^{***} \in \Omega$ exists. Furthermore, because of Equation (D.10), we know that, for all $s \in w$,

$$\omega^{***}(s) = \omega^*(s) \quad \text{if } s \in [0, \max u] \tag{D.11}$$

and

$$\omega^{***}(s) = \omega^{**}(s) \quad \text{if } s \in u \cup [\max u, +\infty) \tag{D.12}$$

and therefore, it follows from Equation (D.6) that

$$\omega^{***} \in C_i \Leftrightarrow (\omega^{***} \in C_i^* \text{ and } \omega^{***} \in C_i^{**}) \Leftrightarrow (\omega^* \in C_i^* \text{ and } \omega^{**} \in C_i^{**}) \tag{D.13}$$

for all $i \in \{1, \ldots, n\}$ such that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$.

Next, for any $i \in S^*$, we infer from Equation (D.1) that the value of $\mathbb{I}_{A_i}(\omega^{***})$ and $\mathbb{I}_{C_i}(\omega^{***})$ is completely determined by $\omega^{***}(t)$, $t \in (w \cap [0, \max u])$. Therefore, it follows from Equations (D.5) and (D.11) that

$$\sum_{i \in S^*} \lambda_i \mathbb{I}_{C_i}(\omega^{***}) \big( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega^{***}) \big) \geq 0. \tag{D.14}$$

Similarly, for any $i \in S^{**}$, Equations (D.13) and (D.8) imply that $\mathbb{I}_{C_i}(\omega^{***}) = \mathbb{I}_{C_i^{**}}(\omega^{**})$, and Equations (D.2) and (D.12) imply that $\mathbb{I}_{A_i}(\omega^{***}) = \mathbb{I}_{A_i}(\omega^{**})$. Therefore, it follows from Equation (D.9) that

$$\sum_{i \in S^{**}} \lambda_i \mathbb{I}_{C_i}(\omega^{***}) \big( \tilde{P}(A_i | C_i) - \mathbb{I}_{A_i}(\omega^{***}) \big) \geq 0. \tag{D.15}$$

Consider now any $i \in \{1, \ldots, n\}$ such that $i \notin S^*$ and $i \notin S^{**}$. Since $i \notin S^*$, there is some $x_u \in \mathscr{X}_u$ such that $(A_i, C_i) \in \mathscr{C}_{x_u}$. If $x_u = x_u^*$, then since $i \notin S^{**}$, it follows from Equation (D.8) that $\mathbb{I}_{C_i^*}(\omega^*) = 0$, and therefore, Equation (D.13) implies that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. If $x_u \neq x_u^*$, then $(X_u = x_u) \cap (X_u = x_u^*) = \emptyset$, and therefore, since $(A_i, C_i) \in \mathscr{C}_{x_u}$ implies that $C_i \subseteq (X_u = x_u)$, it follows that $C_i \cap (X_u = x_u^*) = \emptyset$. Since it follows from Equations (D.10) and (D.12) that $\omega^{***}(t) = x_t^*$ for all $t \in u$, this implies that $\omega^{***} \notin C_i$, and therefore, we find that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. Hence, in all cases, we find that $\mathbb{I}_{C_i}(\omega^{***}) = 0$. Since this is true for any $i \in \{1, \ldots, n\}$ such that $i \notin S^*$ and $i \notin S^{**}$, it follows from Equations (D.14) and (D.15) that

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{C_i}(\omega^{***})\big(\tilde{P}(A_i|C_i) - \mathbb{I}_{A_i}(\omega^{***})\big) \geq 0. \tag{D.16}$$

We will now prove that $\omega^{***} \in C_0$. We consider two cases: $S^* \neq \emptyset$ and $S^* = \emptyset$. First assume that $S^* \neq \emptyset$. In that case, we have that $\omega^* \in C^*$, which implies that there is some $i \in S^*$ such that $\omega^* \in C_i$. It then follows from Equations (D.1) and (D.11) that $\omega^{***} \in C_i \subseteq C_0$. Next, assume that $S^* = \emptyset$. In that case, we have that $\omega^* \in C_0$, which implies that there is some $i \in \{1, \ldots, n\}$ such that $\omega^* \in C_i$. Since $(A_i, C_i) \in \mathscr{C}$ and $S^* = \emptyset$, there is some $x_u \in \mathscr{X}_u$ such that $(A_i, C_i) \in \mathscr{C}_{x_u}$ and, since Equation (D.2) implies that $x_t = \omega^*(t)$ for all $t \in u$, it follows that $x_u = x_u^*$. We conclude from this that $(A_i, C_i) \in \mathscr{C}_{x_u^*}$. Furthermore, since $\omega^* \in C_i \subseteq C_i^*$, we know that $\mathbb{I}_{C_i^*}(\omega^*) = 1$. Therefore, it follows from Equation (D.8) that $S^{**} \neq \emptyset$, which implies that $\omega^{**} \in C^{**}$. Hence, there is some $j \in S^{**}$ such that $\omega^{**} \in C_j^{**}$ and, since $j \in S^{**}$, we also know that $\mathbb{I}_{C_j^*}(\omega^*) = 1$, or equivalently, that $\omega^* \in C_j^*$. By combining this with Equation (D.13), it follows that $\omega^{***} \in C_j \subseteq C_0$. So, in all cases, we find that $\omega^{***} \in C_0$. By combining this with Equation (D.16), it follows that Equation (D.4) holds, and therefore, that $\tilde{P}$ is coherent.

Since $\tilde{P}$ is coherent, and because $\mathscr{C} \subseteq \mathscr{C}^{SP}$, it now follows from Theorem 4.2 and Definition 4.3 that $\tilde{P}$ can be extended to a stochastic process $P$. Furthermore, since $P$ coincides with $\tilde{P}$ on $\mathscr{C}$, it follows from Equation (D.3) that $P$ satisfies Equations (29) and (30). This concludes the first part of this proof.

In the remainder of this proof, we will show that $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W}$, as desired. To this end, let $P^*$ be any full conditional probability that coincides with $P$ on $\mathscr{C}^{SP}$; Corollary 4.4 implies that such a full conditional probability always exists.

Now observe that due to Equation (D.1), we have for all $x \in \mathscr{X}$ that $(X_0 = y, X_\emptyset = x_\emptyset) \in \mathscr{C}_\emptyset$. Therefore, and because of Equation (D.3), we find that

$$P(X_0 = y) = P(X_0 = y \mid X_\emptyset = x_\emptyset) = P_\emptyset(X_0 = y \mid X_\emptyset = x_\emptyset) = P_\emptyset(X_0 = y) \text{ for all } y \in \mathscr{X},$$

which together with the fact that $P_\emptyset \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W}$, implies that $P(X_0) \in \mathscr{M}$. Hence, in order to prove that $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W}$, it remains to show that $P$ is well-behaved as well as consistent with $\mathscr{Q}$.

In order to do this, we start by establishing an important equality. Consider any $w \in \mathscr{U}$ and $s \in \mathbb{R}_{\geq 0}$ such that $w < s$ and $u < s$. Then for all $x_w \in \mathscr{X}_w$ and $y \in \mathscr{X}$, we have that

$$
\begin{aligned}
P(X_s = y|X_w = x_w) &= P^*(X_s = y|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y, X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y|X_{u\setminus w} = x_{u\setminus w}, X_w = x_w)P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P^*(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} \tilde{P}(X_s = y|X_u = x_u, X_{w\setminus u} = x_{w\setminus u})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w) \\
&= \sum_{x_{u\setminus w} \in \mathscr{X}_{u\setminus w}} P_{x_u}(X_s = y|X_u = x_u, X_{w\setminus[0,\max u]} = x_{w\setminus[0,\max u]})P^*(X_{u\setminus w} = x_{u\setminus w}|X_w = x_w).
\end{aligned}
\tag{D.17}
$$

Using this equality, we will next show that for small enough $\Delta \in \mathbb{R}_{>0}$, the transition matrices $T_{t,x_v}^{t+\Delta}$ and $T_{t-\Delta,x_v}^{t}$ corresponding to $P$ can each be written as a (different) convex combination of transition matrices corresponding to processes in $\mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W}$.

Formally, we will show that for any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, there is some finite index set $\mathscr{I}$, some $v^* \in \mathscr{U}_{<t}$, a set of non-negative coefficients $(\lambda_i)_{i \in \mathscr{I}}$ that sum to one, a set of stochastic processes $({}^iP \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W})_{i \in \mathscr{I}}$ and a set of state instantiations $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i \in \mathscr{I}}$ such that

$$(\exists \delta > 0)\,(\forall 0 < \Delta < \delta)\quad T_{t,x_v}^{t+\Delta} = \sum_{i \in \mathscr{I}} \lambda_i\, {}^iT_{t,{}^ix_{v^*}}^{t+\Delta} \tag{D.18}$$

and, similarly, that for any $t > 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, there is some finite index set $\mathscr{I}$, some $v^* \in \mathscr{U}_{<t}$, a set of non-negative coefficients $(\lambda_i)_{i \in \mathscr{I}}$ that sum to one, a set of stochastic processes $({}^iP \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W})_{i \in \mathscr{I}}$ and a set of state instantiations $({}^ix_{v^*} \in \mathscr{X}_{v^*})_{i \in \mathscr{I}}$ such that

$$(\exists \delta > 0) \ (\forall 0 < \Delta < \delta) \quad T^t_{t-\Delta, x_v} = \sum_{i \in \mathscr{I}} \lambda_i \, {}^i T^t_{t-\Delta, {}^i x_{v^*}}. \tag{D.19}$$

We start by constructing the convex combination that satisfies Equation (D.18). So consider any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$. We distinguish between two cases: $t < \max u$ and $t \geq \max u$. If $t < \max u$, then for all $\Delta \in (0, \max u - \max v)$ and $x, y \in \mathscr{X}$, we see that $(X_{t+\Delta} = y, (X_t = x, X_v = x_v)) \in \mathscr{C}_\emptyset$, and therefore, since $P$ is an extension of $\tilde{P}$, it follows from Equation (D.3) that

$$P(X_{t+\Delta} = y | X_t = x, X_v = x_v) = P_\emptyset(X_{t+\Delta} = y | X_t = x, X_v = x_v).$$

Hence, if we let $\mathscr{I} \coloneqq \{i\}$, $v^* \coloneqq v$, $\lambda_i \coloneqq 1$, ${}^i P \coloneqq P_\emptyset$ and ${}^i x_{v^*} \coloneqq x_v$, Equation (D.18) is satisfied by choosing $\delta \coloneqq \max u - \max v$. If $t \geq \max u$, then for all $\Delta > 0$, it follows from Equation (D.17) (with $s \coloneqq t + \Delta$ and $w \coloneqq v \cup t$) that

$$P(X_{t+\Delta} = y | X_t = x, X_v = x_v)$$
$$= \sum_{x_{u \setminus (v \cup t)} \in \mathscr{X}_{u \setminus (v \cup t)}} P_{x_u}(X_{t+\Delta} = y | X_t = x, X_{(u \setminus t) \cup (v \setminus [0, \max u])} = x_{(u \setminus t) \cup (v \setminus [0, \max u])}) P^*(X_{u \setminus (v \cup t)} = x_{u \setminus (v \cup t)} | X_t = x, X_v = x_v).$$

Therefore, if we let $\mathscr{I} \coloneqq \mathscr{X}_{u \setminus (v \cup t)}$, $v^* \coloneqq (u \setminus t) \cup (v \setminus [0, \max u])$ and, for all $x_{u \setminus (v \cup t)} \in \mathscr{I}$,

$$\lambda_{x_{u \setminus (v \cup t)}} \coloneqq P^*(X_{u \setminus (v \cup t)} = x_{u \setminus (v \cup t)} | X_t = x, X_v = x_v),$$

${}^{x_{u \setminus (v \cup t)}} P = P_{x_u}$ and ${}^{x_{u \setminus (v \cup t)}} x_{v^*} \coloneqq x_{(u \setminus t) \cup (v \setminus [0, \max u])}$, Equation (D.18) is satisfied for any $\delta > 0$. Hence, Equation (D.18) can be satisfied both when $t < \max u$ and when $t \geq \max u$.

We will next construct the convex combination that satisfies Equation (D.19). So, consider any $t > 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$. We again distinguish between two cases: $t \leq \max u$ and $t > \max u$. If $t \leq \max u$, then for all $\Delta \in (0, t - \max v)$ and $x, y \in \mathscr{X}$, we see that $(X_t = y, (X_{t-\Delta} = x, X_v = x_v)) \in \mathscr{C}_\emptyset$, and therefore, since $P$ is an extension of $\tilde{P}$, it follows from Equation (D.3) that

$$P(X_t = y | X_{t-\Delta} = x, X_v = x_v) = P_\emptyset(X_t = y | X_{t-\Delta} = x, X_v = x_v).$$

Hence, if we let $\mathscr{I} \coloneqq \{i\}$, $v^* \coloneqq v$, $\lambda_i \coloneqq 1$, ${}^i P \coloneqq P_\emptyset$ and ${}^i x_{v^*} \coloneqq x_v$, Equation (D.19) is satisfied by choosing $\delta \coloneqq t - \max v$. If $t > \max u$, then for all $\Delta \in (0, t - \max(v \cup u))$, it follows from Equation (D.17) (with $s \coloneqq t$ and $w \coloneqq v \cup t - \Delta$) that

$$P(X_t = y | X_{t-\Delta} = x, X_v = x_v)$$
$$= \sum_{x_{u \setminus v} \in \mathscr{X}_{u \setminus v}} P_{x_u}(X_t = y | X_{t-\Delta} = x, X_{u \cup (v \setminus [0, \max u])} = x_{u \cup (v \setminus [0, \max u])}) P^*(X_{u \setminus v} = x_{u \setminus v} | X_{t-\Delta} = x, X_v = x_v).$$

Therefore, if we let $\mathscr{I} \coloneqq \mathscr{X}_{u \setminus v}$, $v^* \coloneqq u \cup (v \setminus [0, \max u])$ and, for all $x_{u \setminus v} \in \mathscr{I}$,

$$\lambda_{x_{u \setminus v}} \coloneqq P^*(X_{u \setminus v} = x_{u \setminus v} | X_{t-\Delta} = x, X_v = x_v),$$

${}^{x_{u \setminus v}} P = P_{x_u}$ and ${}^{x_{u \setminus v}} x_{v^*} \coloneqq x_{u \cup (v \setminus [0, \max u])}$, Equation (D.19) is satisfied by choosing $\delta \coloneqq t - \max(v \cup u)$. Hence, Equation (D.19) can be satisfied both when $t \leq \max u$ and when $t > \max u$.

Therefore, indeed, as claimed before, both $T^{t+\Delta}_{t, x_v}$ and $T^t_{t-\Delta, x_v}$ can be written as a convex combination of transition matrices corresponding to elements of $\mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}$—assuming that $\Delta$ is small enough. We will now use this fact to prove that $P$ is well-behaved and consistent with $\mathscr{Q}$.

We start by proving that $P$ is well-behaved. First consider any $t \geq 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, and consider the indexed set $\{{}^i P \in \mathbb{P}^W_{\mathscr{Q}, \mathscr{M}}\}_{i \in \mathscr{I}}$ of stochastic processes whose transition matrices appear in the convex combination that satisfies Equation (D.18). Then,

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T^{t+\Delta}_{t, x_v} - I \right\| = \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| \sum_{i \in \mathscr{I}} \lambda_i \, {}^i T^{t+\Delta}_{t, {}^i x_{v^*}} - I \right\| \leq \sum_{i \in \mathscr{I}} \lambda_i \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| {}^i T^{t+\Delta}_{t, {}^i x_{v^*}} - I \right\| < +\infty, \tag{D.20}$$

where the last inequality follows from Proposition 4.5 and the fact that every process ${}^i P$ that appears in this combination is well-behaved.

Similarly, for any $t > 0$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$, Equation (D.19) can be satisfied with a convex combination of transition matrices corresponding to well-behaved processes, which implies that

$$\limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| T^t_{t-\Delta, x_v} - I \right\| = \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| \sum_{i \in \mathscr{I}} \lambda_i \, {}^i T^t_{t-\Delta, {}^i x_{v^*}} - I \right\| \leq \sum_{i \in \mathscr{I}} \lambda_i \limsup_{\Delta \to 0^+} \frac{1}{\Delta} \left\| {}^i T^t_{t-\Delta, {}^i x_{v^*}} - I \right\| < +\infty. \tag{D.21}$$

Because the $t \in \mathbb{R}_{\geq 0}$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$ in Equation (D.20), and the $t \in \mathbb{R}_{>0}$, $v \in \mathscr{U}_{<t}$ and $x_v \in \mathscr{X}_v$ in Equation (D.21) are arbitrary, by invoking Proposition 4.5, it follows that $P$ is well-behaved.

We end by proving that $P$ is consistent with $\mathscr{Q}$. Assume *ex absurdo* that $P$ is not consistent with $\mathscr{Q}$, or equivalently, that there are $t \geq 0$, $v \in \mathscr{U}_{<t}$, $x_v \in \mathscr{X}_v$ and $Q^* \in \mathscr{R}$ such that $Q^* \in \overline{\partial} T^t_{t,x_v}$ and $Q^* \notin \mathscr{Q}$. We will show that this leads to a contradiction. We consider two (possibly overlapping) cases: $Q^* \in \overline{\partial}_+ T^t_{t,x_v}$ and $Q^* \in \overline{\partial}_- T^t_{t,x_v}$.

If $Q^* \in \overline{\partial}_+ T^t_{t,x_v}$, it follows from Equation (21) that there is a sequence $\{\Delta_j\}_{j\in\mathbb{N}} \to 0^+$ such that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}(T^{t+\Delta_j}_{t,x_v} - I) = Q^*. \tag{D.22}$$

Consider the $\mathscr{I}$, $v^* \in \mathscr{U}_{<t}$, $\{\lambda_i\}_{i\in\mathscr{I}}$, $\{^iP \in \mathbb{P}^W_{\mathscr{Q},\mathscr{M}}\}_{i\in\mathscr{I}}$, $\{^ix_{v^*}\}_{i\in\mathscr{I}}$ and $\delta$ that satisfy Equation (D.18). Fix any $i \in \mathscr{I}$. Since $^iP$ is well-behaved, the sequence

$$\left\{ \frac{1}{\Delta_j}(^iT^{t+\Delta_j}_{t,\,^ix_{v^*}} - I) \right\}_{j\in\mathbb{N}}$$

is bounded, and therefore, the Bolzano–Weierstraß theorem implies that it has a convergent subsequence, of which we denote the limit by $Q_i$. Hence, without loss of generality—simply remove the indexes $j$ that do not correspond to the subsequence—we may assume that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}(^iT^{t+\Delta_j}_{t,\,^ix_{v^*}} - I) = Q_i. \tag{D.23}$$

Since we know from Proposition 3.3 that $Q_i$ is a limit of rate matrices, $Q_i$ is also a rate matrix, and therefore, it follows from Equation (21) that $Q_i \in \overline{\partial}_+ T^t_{t,\,^ix_{v^*}}$, which, since $^iP$ is consistent with $\mathscr{Q}$, implies that $Q_i \in \mathscr{Q}$. By repeating this argument for every other $i \in \mathscr{I}$, we obtain a set of rate matrices $\{Q_i \in \mathscr{Q}\}_{i\in\mathscr{I}}$ such that, without loss of generality, Equation (D.23) holds for every $i \in \mathscr{I}$. Additionally, since $\lim_{j\to\infty} \Delta_j = 0$, we may assume without loss of generality that $0 < \Delta_j < \delta$ for all $j \in \mathbb{N}$. Equations (D.18) and (D.23) now imply that

$$\lim_{j\to+\infty} \frac{1}{\Delta_j}(T^{t+\Delta_j}_{t,x_v} - I) = \lim_{j\to+\infty} \frac{1}{\Delta_j}(\sum_{i\in\mathscr{I}} \lambda_i\,^iT^{t+\Delta_j}_{t,\,^ix_{v^*}} - I) = \sum_{i\in I} \lambda_i Q_i,$$

which, because of Equation (D.22), implies that $Q^* = \sum_{i\in\mathscr{I}} \lambda_i Q_i$. Since $\mathscr{Q}$ is convex, this implies that $Q^* \in \mathscr{Q}$, a contradiction. Recall that this contradiction was derived from the assumption that $Q^* \in \overline{\partial}_+ T^t_{t,x_v}$.

If instead $Q^* \in \overline{\partial}_- T^t_{t,x_v}$, a completely analogous argument leads to the same contradiction: simply use Equations (22) and (D.19) instead of Equations (21) and (D.18), respectively, and adapt the rest of the argument accordingly. Since the two cases $Q^* \in \overline{\partial}_+ T^t_{t,x_v}$ and $Q^* \in \overline{\partial}_- T^t_{t,x_v}$ both lead to a contradiction, we conclude that our ex absurdo assumption must be false, or equivalently, that $P$ is consistent with $\mathscr{Q}$.

Hence, since we already know that $P$ is a well-behaved stochastic process such that $P(X_0) \in \mathscr{M}$, it follows that $P \in \mathbb{P}^W_{\mathscr{Q},\mathscr{M}}$. $\square$

**Proposition 6.4.** *Consider any non-empty bounded set of rate matrices $\mathscr{Q}$, and any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then,*

$$\underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] \leq \underline{\mathbb{E}}^{WM}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot] \leq \underline{\mathbb{E}}^{WHM}_{\mathscr{Q},\mathscr{M}}[\cdot\,|\,\cdot].$$

**Proof.** This result is an immediate consequence of Proposition 6.1 and Definition 6.5. $\square$

**Theorem 6.5.** *Let $\mathscr{Q}$ be an arbitrary non-empty, bounded, and convex set of rate matrices, and consider any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Then for any $u, v, w \in \mathscr{U}$ such that $u < v < w$ and any $f \in \mathscr{L}(\mathscr{X}_{u\cup v\cup w})$:*

$$\underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w)\,|\,X_u] = \underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}\Big[\underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w)\,|\,X_u, X_v]\,\Big|\,X_u\Big]. \tag{35}$$

**Proof.** Let $g(X_u, X_v) \coloneqq \underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w)|X_u, X_v]$ and fix any $x_u \in \mathscr{X}_u$. For any $P \in \mathbb{P}^W_{\mathscr{Q},\mathscr{M}}$, it then follows from the basic properties of expectations (which follow from F1–F4) and Definition 6.5 that

$$\mathbb{E}_P[f(X_u, X_v, X_w)|x_u] = \mathbb{E}_P\big[\mathbb{E}_P[f(X_u, X_v, X_w)|X_u, X_v]|x_u\big] \geq \mathbb{E}_P\big[g(X_u, X_v)|x_u\big] \geq \underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}\big[g(X_u, X_v)|x_u\big].$$

Since $P \in \mathbb{P}^W_{\mathscr{Q},\mathscr{M}}$ is arbitrary, this implies that

$$\underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}[f(X_u, X_v, X_w)|x_u] \geq \underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}\big[g(X_u, X_v)|x_u\big]. \tag{D.24}$$

Now fix any $\varepsilon \in \mathbb{R}_{>0}$. Then due to Definition 6.5, there is some $P_\emptyset \in \mathbb{P}^W_{\mathscr{Q},\mathscr{M}}$ such that

$$\mathbb{E}_{P_\emptyset}\big[g(x_u, X_v)|x_u\big] = \mathbb{E}_{P_\emptyset}\big[g(X_u, X_v)|x_u\big] < \underline{\mathbb{E}}^W_{\mathscr{Q},\mathscr{M}}\big[g(X_u, X_v)|x_u\big] + \varepsilon/2, \tag{D.25}$$

where the first equality follows from the basic properties of expectations. Similarly, for all $x_v \in \mathscr{X}_v$, there is some $P_{x_v} \in \mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$ such that

$$\mathbb{E}_{P_{x_v}}\left[f(X_u, X_v, X_w)|x_u, x_v\right] < \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|x_u, x_v\right] + \varepsilon/2 = g(x_u, x_v) + \varepsilon/2. \tag{D.26}$$

Since $\mathscr{Q}$ is non-empty, bounded, and convex, Theorem 6.3 now implies the existence of a process $P \in \mathbb{P}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}$ such that for all $x_v \in \mathscr{X}_v$ and $x_w \in \mathscr{X}_w$:

$$P(X_v = x_v \mid X_u = x_u) = P_\emptyset(X_v = x_v \mid X_u = x_u)$$

and

$$P(X_w = x_w \mid X_u = x_u, X_v = x_v) = P_{x_v}(X_w = x_w \mid X_u = x_u, X_v = x_v)$$

and therefore also, due to Equations (D.25) and (D.26),

$$\mathbb{E}_P[g(x_u, X_v)|x_u] = \mathbb{E}_{P_\emptyset}[g(x_u, X_v)|x_u] < \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[g(x_u, X_v)\big|x_u\right] + \varepsilon/2$$

and

$$\mathbb{E}_P[f(X_u, X_v, X_w)|x_u, x_v] = \mathbb{E}_{P_{x_v}}[f(X_u, X_v, X_w)|x_u, x_v] < g(x_u, x_v) + \varepsilon/2.$$

Hence, we find that

$$\mathbb{E}_P\left[\mathbb{E}_P\left[f(X_u, X_v, X_w)|x_u, X_v\right]|x_u\right] \leq \mathbb{E}_P\left[g(x_u, X_v)\big|x_u\right] + \varepsilon/2 < \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_u, X_v)\big|x_u\right] + \varepsilon.$$

Since $\varepsilon$ was arbitrary, and because

$$\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|x_u\right] \leq \mathbb{E}_P\left[f(X_u, X_v, X_w)|x_u\right] = \mathbb{E}_P\left[\mathbb{E}_P\left[f(X_u, X_v, X_w)|x_u, X_v\right]|x_u\right],$$

this implies that $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|x_u\right] \leq \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_u, X_v)\big|x_u\right]$ and therefore, because of Equation (D.24), that $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|x_u\right] = \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_u, X_v)\big|x_u\right]$. Since $x_u$ was arbitrary, this implies that $\underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|X_u\right] = \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[g(X_u, X_v)\big|X_u\right]$. The result is now immediate because $g(X_u, X_v) = \underline{\mathbb{E}}^{\mathsf{W}}_{\mathscr{Q},\mathscr{M}}\left[f(X_u, X_v, X_w)|X_u, X_v\right]$. □

## Appendix E. Proofs and lemmas for the results in Section 7

**Proposition 7.1.** *For any two lower transition operators $\underline{T}, \underline{S} \in \underline{\mathbb{T}}$, their composition $\underline{T}\,\underline{S}$ is again a lower transition operator.*

**Proof.** Consider any $f, g \in \mathscr{L}(\mathscr{X})$, $\lambda \in \mathbb{R}_{\geq 0}$ and $x \in \mathscr{X}$. Since $\underline{T}$ and $\underline{S}$ are lower transition operators, they both satisfy LT1, and therefore, we find that

$$[\underline{T}\,\underline{S}f](x) \geq \min\{[\underline{S}f](y): y \in \mathscr{X}\} \geq \min\{\min\{f(z): z \in \mathscr{X}\}: y \in \mathscr{X}\} = \min\{f(z): z \in \mathscr{X}\},$$

which implies that $\underline{T}\,\underline{S}$ satisfies LT1 as well. Similarly, $\underline{T}\,\underline{S}$ satisfies LT2 because

$$[\underline{T}\,\underline{S}(f + g)](x) \geq [\underline{T}(\underline{S}f + \underline{S}g)](x) \geq [\underline{T}\,\underline{S}f](x) + [\underline{T}\,\underline{S}g](x),$$

where the first inequality follows from LT5 and the fact that $\underline{S}$ satisfies LT2, and where the second inequality follows from the fact that $\underline{T}$ satisfies LT2. Finally, since $\underline{T}$ and $\underline{S}$ both satisfy LT3, it follows that $\underline{T}\,\underline{S}$ also satisfies LT3, because

$$[\underline{T}\,\underline{S}(\lambda f)](x) = [\underline{T}(\lambda\underline{S}f)](x) = \lambda[\underline{T}\,\underline{S}f](x).$$

We conclude that $\underline{T}\,\underline{S}$ satisfies LT1–LT3, and therefore, because of Definition 7.4, it is a lower transition operator. □

We require the following two results for the proof of Proposition 7.2.

**Lemma E.1.** *[27, Proposition 1] Consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{N}}$ of lower transition operators such that $\underline{T}f = \lim_{i\to\infty}\underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$. Then $\underline{T}$ is a lower transition operator.*

**Lemma E.2.** *[27, Proposition 2] Let $\underline{T}$ be a lower transition operator, and consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{N}}$ of lower transition operators. Then, $\underline{T} = \lim_{i\to\infty}\underline{T}_i$ if and only if $\underline{T}f = \lim_{i\to\infty}\underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$.*

**Proposition 7.2.** *The metric space $(\underline{\mathbb{T}}, d)$ is complete with respect to the metric $d$ that is induced by our usual norm $\|\cdot\|$.*

**Proof.** Consider any sequence $\{\underline{T}_i\}_{i\in\mathbb{N}}$ of lower transition operators that is Cauchy with respect to the operator norm $\|\cdot\|$. We will prove that $\{\underline{T}_i\}_{i\in\mathbb{N}}$ converges to a limit $\underline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ that is itself a lower transition operator.

Consider any $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$. For any $k, \ell \in \mathbb{N}$, (N11) then implies that

$$\left|[\underline{T}_k f](x) - [\underline{T}_\ell f](x)\right| \leq \left\|\underline{T}_k f - \underline{T}_\ell f\right\| = \left\|(\underline{T}_k - \underline{T}_\ell)f\right\| \leq \left\|\underline{T}_k - \underline{T}_\ell\right\| \|f\|.$$

Therefore, and because $\{\underline{T}_i\}_{i\in\mathbb{N}}$ is Cauchy with respect to the norm $\|\cdot\|$, it follows that $\{[\underline{T}_i f](x)\}_{i\in\mathbb{N}}$ is Cauchy with respect to the norm $|\cdot|$. Hence, since $\mathbb{R}$ is (well known to be) complete with respect to the topology that is induced by $|\cdot|$, we find that $\{[\underline{T}_i f](x)\}_{i\in\mathbb{N}}$ converges to a limit in $\mathbb{R}$, which we will denote by $[\underline{T}f](x)$. Let $\underline{T}f$ be the unique function in $\mathscr{L}(\mathscr{X})$ that has $[\underline{T}f](x)$, $x \in \mathscr{X}$, as its components. Then clearly, $\underline{T}f = \lim_{i\to+\infty} \underline{T}_i f$.

Let $\underline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$ be the unique operator that maps any $f \in \mathscr{L}(\mathscr{X})$ to $\underline{T}f$. It then follows from Lemma E.1 that $\underline{T}$ is a lower transition operator. Therefore, and because we already know that $\underline{T}f = \lim_{i\to+\infty} \underline{T}_i f$ for all $f \in \mathscr{L}(\mathscr{X})$, it now follows from Lemma E.2 that $\lim_{i\to+\infty} \underline{T}_i = \underline{T}$. $\square$

**Proposition 7.5.** *For any non-empty bounded set $\mathscr{Q} \subseteq \mathscr{R}$ of rate matrices, the corresponding operator $\underline{Q}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$, as defined by Equation* (38)*, is a lower transition rate operator.*

**Proof.** Consider any $Q \in \mathscr{Q}$. It then follows from Definition 3.2 that the matrix $Q$, when regarded as a map from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, satisfies LR1–LR4. Since each of these properties is preserved under taking lower envelopes, it follows that $\underline{Q}$ satisfies LR1–LR4, which means that $\underline{Q}$ is a lower transition rate operator. $\square$

**Proposition 7.6.** *Consider a lower transition rate operator $\underline{Q}$ and let $\mathscr{Q}_{\underline{Q}}$ be the corresponding set of dominating rate matrices, as defined by Equation* (39)*. Then $\mathscr{Q}_{\underline{Q}}$ is non-empty and bounded and, for all $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$.*

**Proof.** We will start by showing that, for every $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $Qf = \underline{Q}f$. To this end, fix any $f \in \mathscr{L}(\mathscr{X})$. Now choose $\Delta > 0$ small enough such that $0 \leq \Delta \|\underline{Q}\| \leq 1$ (this always possible because of property LR5) and define $\underline{T} \coloneqq I + \Delta\underline{Q}$. Since $\underline{Q}$ is a lower transition rate operator, it then follows from Proposition 7.3 that $\underline{T}$ is a lower transition operator. For any $x \in \mathscr{X}$, we now consider the operator $\underline{T}_x\colon \mathscr{L}(\mathscr{X}) \to \mathbb{R}$, as defined by Equation (37), which is a coherent lower prevision. Because of [22, Theorem 3.3.3(b)], this implies the existence of an expectation operator $\mathbb{E}_x$ on $\mathscr{L}(\mathscr{X})$—Reference [22] calls this a linear prevision on $\mathscr{L}(\mathscr{X})$—such that $\mathbb{E}_x g \geq \underline{T}_x g$ for all $g \in \mathscr{L}(\mathscr{X})$ and $\mathbb{E}_x f = \underline{T}_x f$. Let $P_x$ be the unique probability mass function that corresponds to $\mathbb{E}_x$. For all $x, y \in \mathscr{X}$, we now let $T(x, y) \coloneqq P_x(y) \coloneqq \mathbb{E}_x(\mathbb{I}_y)$. Then $T$ is clearly a transition matrix. Furthermore, for every $x \in \mathscr{X}$ and $g \in \mathscr{L}(\mathscr{X})$, we have that $(Tg)(x) = \mathbb{E}_x g$. Hence, it follows that $Tg \geq \underline{T}g$ for all $g \in \mathscr{L}(\mathscr{X})$ and that $Tf = \underline{T}f$. Now let $Q \coloneqq 1/\Delta(T - I)$, which, because of Proposition 3.3, is a rate matrix. Since $Tf = \underline{T}f$, it then follows that

$$Qf = \frac{1}{\Delta}(Tf - f) = \frac{1}{\Delta}(\underline{T}f - f) = \underline{Q}f.$$

Similarly, since $Tg \geq \underline{T}g$ for all $g \in \mathscr{L}(\mathscr{X})$, it follows that $Qg \geq \underline{Q}g$, or equivalently, since $Q$ is a rate matrix, that $Q \in \mathscr{Q}_{\underline{Q}}$. Since $f$ was arbitrary, this proves that, for all $f \in \mathscr{L}(\mathscr{X})$, there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $Qf = \underline{Q}f$. Since $\mathscr{L}(\mathscr{X})$ is non-empty, this clearly also implies that $\mathscr{Q}_{\underline{Q}}$ is non-empty.

We end this proof by showing that $\mathscr{Q}_{\underline{Q}}$ is bounded. Consider any $x \in \mathscr{X}$. Then for all $Q \in \mathscr{Q}_{\underline{Q}}$, we have that $Q(x, x) = [Q\,\mathbb{I}_x](x) \geq [\underline{Q}\,\mathbb{I}_x](x)$, which implies that

$$\inf\left\{Q(x, x)\colon Q \in \mathscr{Q}_{\underline{Q}}\right\} \geq [\underline{Q}\,\mathbb{I}_x](x) > -\infty.$$

Since $x \in \mathscr{X}$ is arbitrary, Proposition 3.5 now guarantees that $\mathscr{Q}_{\underline{Q}}$ is bounded. $\square$

**Proposition 7.7.** *Consider a lower transition rate operator $\underline{Q}$ and let $\mathscr{Q}_{\underline{Q}}$ be the corresponding set of dominating rate matrices, as defined by Equation* (39)*. Then $\mathscr{Q}_{\underline{Q}}$ is closed and convex, and has separately specified rows.*

**Proof.** We start by showing that $\mathscr{Q}_{\underline{Q}}$ is closed, or equivalently, that for any converging sequence $\{Q_i\}_{i\in\mathbb{N}}$ in $\mathscr{Q}_{\underline{Q}}$, the limit $Q \coloneqq \lim_{i\to+\infty} Q_i$ is again an element of $\mathscr{Q}_{\underline{Q}}$. Since $\{Q_i\}_{i\in\mathbb{N}}$ belongs to the bounded set of rate matrices $\mathscr{Q}_{\underline{Q}}$—see Proposition 7.6—we know that the limit $Q$ is a real-valued matrix, and therefore, since R1 and R2 are clearly both preserved under taking limits, $Q$ is a rate matrix. Now, assume *ex absurdo* that $Q \notin \mathscr{Q}_{\underline{Q}}$. Then, by Equation (39), there is some $f \in \mathscr{L}(\mathscr{X})$ and some $x \in \mathscr{X}$ such that $[Qf](x) < [\underline{Q}f](x)$, and therefore some $\varepsilon \in \mathbb{R}_{>0}$ such that $[Qf](x) + \varepsilon < [\underline{Q}f](x)$. Hence, since $\lim_{i\to+\infty} Q_i = Q$, there is some $i^* \in \mathbb{N}$ such that $[Q_{i^*}f](x) < [Qf](x) + \varepsilon < [\underline{Q}f](x)$. Since $Q_{i^*} \in \mathscr{Q}_{\underline{Q}}$, this is a contradiction. Therefore, $Q \in \mathscr{Q}_{\underline{Q}}$, and because the converging sequence $\{Q_i\}_{i\in\mathbb{N}}$ was arbitrary, this proves that $\mathscr{Q}_{\underline{Q}}$ is closed.

Next, we show that $\mathscr{Q}_{\underline{Q}}$ is convex, or equivalently, that for any two rate matrices $Q_1, Q_2 \in \mathscr{Q}_{\underline{Q}}$, and any $\lambda \in [0, 1]$, the matrix $Q_\lambda \coloneqq \lambda Q_1 + (1 - \lambda)Q_2$ is again an element of $\mathscr{Q}_{\underline{Q}}$. It is easily verified from Definition 3.2 that $Q_\lambda$ is a rate matrix. Furthermore, for any $f \in \mathscr{L}(\mathscr{X})$, we find that

$$Q_\lambda f = \lambda Q_1 f + (1-\lambda)Q_2 f \geq \lambda \underline{Q} f + (1-\lambda)\underline{Q} f = \underline{Q} f,$$

where the inequality holds because $Q_1$ and $Q_2$ belong to $\mathscr{Q}_{\underline{Q}}$. Hence, it follows from Equation (39) that $Q_\lambda \in \mathscr{Q}_{\underline{Q}}$.

We finally show that $\mathscr{Q}_{\underline{Q}}$ has separately specified rows. For all $x \in \mathscr{X}$, let $\mathscr{Q}_x := \{Q(x,\cdot) : Q \in \mathscr{Q}_{\underline{Q}}\}$. Consider now any rate matrix $Q$ such that, for all $x \in \mathscr{X}$, $Q(x,\cdot) \in \mathscr{Q}_x$ and assume *ex absurdo* that $Q \notin \mathscr{Q}_{\underline{Q}}$. Equation (39) then implies the existence of some $f \in \mathscr{L}(\mathscr{X})$ and $x \in \mathscr{X}$ such that $[Qf](x) < [\underline{Q}f](x)$. Since $Q(x,\cdot) \in \overline{\mathscr{Q}}_x$, this in turn implies that there is some $Q' \in \mathscr{Q}_{\underline{Q}}$ such that $[Q'f](x) < [\underline{Q}f](x)$, a contradiction. Hence we find that $Q \in \mathscr{Q}_{\underline{Q}}$. □

**Lemma E.3.** *Let $\mathscr{Q}$ be a non-empty, bounded, closed and convex set of rate matrices. Then, for all $x \in \mathscr{X}$, $\mathscr{Q}_x := \{Q(x,\cdot) : Q \in \mathscr{Q}\}$ is a non-empty, bounded, closed, and convex subset of $\mathscr{L}(\mathscr{X})$.*

**Proof.** Fix any $x \in \mathscr{X}$. The non-emptiness, boundedness and convexity of $\mathscr{Q}_x$ then follows trivially from the fact that $\mathscr{Q}$ is non-empty, bounded and convex. It remains to show that $\mathscr{Q}_x$ is closed. Consider therefore any sequence $\{Q_i\}_{i\in\mathbb{N}}$ in $\mathscr{Q}$ such that $\{Q_i(x,\cdot)\}_{i\in\mathbb{N}}$ converges to a limit $Q_\infty(x,\cdot)$. We need to prove that $Q_\infty(x,\cdot) \in \mathscr{Q}_x$.

Since $\{Q_i\}_{i\in\mathbb{N}}$ belongs to the—closed and bounded and therefore—compact set $\mathscr{Q}$, it has a convergent subsequence $\{Q_{i_k}\}_{k\in\mathbb{N}}$ whose limit $Q^*$ belongs to $\mathscr{Q}$. Since $\{Q_i(x,\cdot)\}_{i\in\mathbb{N}}$ converges to $Q_\infty(x,\cdot)$, it follows that $Q^*(x,\cdot) = Q_\infty(x,\cdot)$, which, since $Q^* \in \mathscr{Q}$, implies that $Q_\infty(x,\cdot) \in \mathscr{Q}_x$. □

**Proposition 7.8.** *Consider any non-empty, bounded, closed and convex set of rate matrices $\mathscr{Q} \subseteq \mathscr{R}$ with separately specified rows that has $\underline{Q}$ as its lower envelope. Then $\mathscr{Q} = \mathscr{Q}_{\underline{Q}}$.*

**Proof.** Since $\mathscr{Q}$ has $\underline{Q}$ as its lower envelope, it follows from Equation (39) that $\mathscr{Q} \subseteq \mathscr{Q}_{\underline{Q}}$. Assume *ex absurdo* that $\mathscr{Q} \subset \mathscr{Q}_{\underline{Q}}$, or equivalently, that there is some $Q \in \mathscr{Q}_{\underline{Q}}$ such that $Q \notin \mathscr{Q}$. Consider now any such $Q$.

Because $\mathscr{Q}$ has separately specified rows, it follows from $Q \notin \mathscr{Q}$ that there is some $x \in \mathscr{X}$ such that $Q(x,\cdot) \notin \mathscr{Q}_x := \{Q'(x,\cdot) : Q' \in \mathscr{Q}\}$. Furthermore, since $\mathscr{Q}$ is non-empty, bounded, closed and convex, it follows from Lemma E.3 that $\mathscr{Q}_x$ is a non-empty, bounded, closed and convex subset of $\mathscr{L}(\mathscr{X})$, and therefore, since $\mathscr{L}(\mathscr{X})$ is a normed linear space and $Q(x,\cdot) \notin \mathscr{Q}_x$, it follows from the separating hyperplane theorem [52, Chapter 14, Corollary 25] that there is a linear operator $\psi : \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ such that

$$\psi(Q(x,\cdot)) < \inf\{\psi(q) : q \in \mathscr{Q}_x\} =: C. \tag{E.1}$$

Now let $f \in \mathscr{L}(\mathscr{X})$ be defined by $f(y) := \psi(\mathbb{I}_y)$ for all $y \in \mathscr{X}$. Then for any $\varepsilon > 0$, since $\mathscr{Q}$ has $\underline{Q}$ as its lower envelope, and because of Equation (38), there is some $Q^* \in \mathscr{Q}$ such that $[Q^*f](x) \leq [\underline{Q}f](x) + \varepsilon$, and therefore also

$$C \leq \psi(Q^*(x,\cdot)) = \sum_{y\in\mathscr{X}} Q^*(x,y)\psi(\mathbb{I}_y) = \sum_{y\in\mathscr{X}} Q^*(x,y)f(y) = [Q^*f](x) \leq [\underline{Q}f](x) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies that $C \leq [\underline{Q}f](x)$ and, by combining this with Equation (E.1), it follows that

$$[Qf](x) = \sum_{y\in\mathscr{X}} Q(x,y)f(y) = \sum_{y\in\mathscr{X}} Q(x,y)\psi(\mathbb{I}_y) = \psi(Q(x,\cdot)) < C \leq [\underline{Q}f](x).$$

This implies that $Qf \not\geq \underline{Q}f$, and therefore, that $Q \notin \mathscr{Q}_{\underline{Q}}$, a contradiction. □

In order to prove the norm inequality stated by Proposition 7.9, we first give a number of norm inequalities that are more conveniently proved separately. These inequalities are stated by Lemmas E.4–E.6 and Corollary E.7 below.

We start with the following result, which states that the distance between two composed lower transition operators $\underline{T}_1\underline{T}_2\cdots\underline{T}_n$ and $\underline{S}_1\underline{S}_2\cdots\underline{S}_n$ is bounded from above by the sum of the distances $\|\underline{T}_i - \underline{S}_i\|$ between their components. Note that this result is a generalised version of Lemma A.2, which until this point has remained unproven. To see that the result below indeed implies Lemma A.2, simply recall from Section 7.1 that every transition matrix $T$ is also a lower transition operator.

**Lemma E.4.** *Consider two finite sequences $\underline{T}_1, \ldots, \underline{T}_n$ and $\underline{S}_1, \ldots, \underline{S}_n$ of lower transition operators. Then*

$$\left\| \prod_{i=1}^n \underline{T}_i - \prod_{i=1}^n \underline{S}_i \right\| \leq \sum_{i=1}^n \|\underline{T}_i - \underline{S}_i\|. \tag{E.2}$$

**Proof.** We provide a proof by induction. Clearly, Equation (E.2) holds for $n = 1$. Suppose that it holds for $n - 1$. We show that it then also holds for $n$.

$$\left\| \prod_{i=1}^{n} \underline{T}_i - \prod_{i=1}^{n} \underline{S}_i \right\| = \left\| \prod_{i=1}^{n} \underline{T}_i - \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n + \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n - \prod_{i=1}^{n} \underline{S}_i \right\|$$

$$\leq \left\| \prod_{i=1}^{n} \underline{T}_i - \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n \right\| + \left\| \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n - \prod_{i=1}^{n} \underline{S}_i \right\|$$

$$= \left\| \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{T}_n - \left( \prod_{i=1}^{n-1} \underline{T}_i \right) \underline{S}_n \right\| + \left\| \left( \prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right) \underline{S}_n \right\|$$

$$\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \left\| \prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right\| \left\| \underline{S}_n \right\|$$

$$\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \left\| \prod_{i=1}^{n-1} \underline{T}_i - \prod_{i=1}^{n-1} \underline{S}_i \right\|$$

$$\leq \left\| \underline{T}_n - \underline{S}_n \right\| + \sum_{i=1}^{n-1} \left\| \underline{T}_i - \underline{S}_i \right\| = \sum_{i=1}^{n} \left\| \underline{T}_i - \underline{S}_i \right\| .$$

Here, in the second inequality, we applied Proposition 7.1 and properties N10 and LT7. In the third inequality, we used property LT4. In the final inequality, we used the induction hypothesis. $\square$

**Lemma E.5.** *Consider a lower transition rate operator $\underline{Q}$ and, for all $i \in \{1, \ldots, n\}$, some $\Delta_i \geq 0$ such that $\Delta_i \left\| \underline{Q} \right\| \leq 1$. Let $\Delta \coloneqq \sum_{i=1}^{n} \Delta_i$. Then*

$$\left\| \prod_{i=1}^{n} (I + \Delta_i \underline{Q}) - (I + \Delta \underline{Q}) \right\| \leq \Delta^2 \left\| \underline{Q} \right\|^2 .$$

**Proof.** We provide a proof by induction. For $n = 1$, the result is trivial. So consider the case $n \geq 2$ and assume that the result is true for $n - 1$.

For all $i \in \{2, \ldots, n\}$, since $\Delta_i \left\| \underline{Q} \right\| \leq 1$, it follows from Proposition 7.3 that $I$ and $(I + \Delta_i \underline{Q})$ are lower transition operators. Therefore,

$$\left\| \prod_{i=1}^{n} (I + \Delta_i \underline{Q}) - (I + \Delta \underline{Q}) \right\| = \left\| \prod_{i=2}^{n} (I + \Delta_i \underline{Q}) + \Delta_1 \underline{Q} \prod_{i=2}^{n} (I + \Delta_i \underline{Q}) - (I + \sum_{i=2}^{n} \Delta_i \underline{Q}) - \Delta_1 \underline{Q} \right\|$$

$$\leq \left\| \prod_{i=2}^{n} (I + \Delta_i \underline{Q}) - (I + \sum_{i=2}^{n} \Delta_i \underline{Q}) \right\| + \left\| \Delta_1 \underline{Q} \prod_{i=2}^{n} (I + \Delta_i \underline{Q}) - \Delta_1 \underline{Q} \right\|$$

$$\leq (\sum_{i=2}^{n} \Delta_i)^2 \left\| \underline{Q} \right\|^2 + 2\Delta_1 \left\| \underline{Q} \right\| \left\| \prod_{i=2}^{n} (I + \Delta_i \underline{Q}) - I \right\|$$

$$\leq (\sum_{i=2}^{n} \Delta_i)^2 \left\| \underline{Q} \right\|^2 + 2\Delta_1 \left\| \underline{Q} \right\| \sum_{i=2}^{n} \left\| (I + \Delta_i \underline{Q}) - I \right\|$$

$$= (\sum_{i=2}^{n} \Delta_i)^2 \left\| \underline{Q} \right\|^2 + \left( 2\Delta_1 \sum_{i=2}^{n} \Delta_i \right) \left\| \underline{Q} \right\|^2 \leq \left( \Delta_1 + \sum_{i=2}^{n} \Delta_i \right)^2 \left\| \underline{Q} \right\|^2 = \Delta^2 \left\| \underline{Q} \right\|^2 ,$$

where the second inequality follows from the induction hypothesis and property LR6, and the third inequality follows from Lemma E.4. $\square$

**Lemma E.6.** *For any $k \in \{1, \ldots, n\}$, consider a sequence $\Delta_{k,i} \geq 0$, $i \in \{1, \ldots, n_k\}$, and let $\Delta_k \coloneqq \sum_{i=1}^{n_k} \Delta_{k,i}$. Let $C \coloneqq \sum_{k=1}^{n} \Delta_k$ and let $\delta \coloneqq \max_{k=1}^{n} \Delta_k$. Then for any lower transition rate operator $\underline{Q}$ such that $\delta \left\| \underline{Q} \right\| \leq 1$:*

$$\left\| \prod_{k=1}^{n} \left( \prod_{i=1}^{n_k} (I + \Delta_{k,i} \underline{Q}) \right) - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) \right\| \leq \delta C \left\| \underline{Q} \right\|^2 .$$

**Proof.** For any $k \in \{1, \ldots, n\}$, we know that $\Delta_{k,i} \|\underline{Q}\| \leq \Delta_k \|\underline{Q}\| \leq \delta \|\underline{Q}\| \leq 1$ for all $i \in \{1, \ldots, n_k\}$, and therefore, it follows from Propositions 7.3 and 7.1 that $\prod_{i=1}^{n_k} (I + \Delta_{k,i}\underline{Q})$ and $(I + \Delta_k \underline{Q})$ are lower transition operators. Hence,

$$\left\| \prod_{k=1}^{n} \left( \prod_{i=1}^{n_k} (I + \Delta_{k,i}\underline{Q}) \right) - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) \right\| \leq \sum_{k=1}^{n} \left\| \left( \prod_{i=1}^{n_k} (I + \Delta_{k,i}\underline{Q}) \right) - (I + \Delta_k \underline{Q}) \right\|$$

$$\leq \sum_{k=1}^{n} \Delta_k^2 \|\underline{Q}\|^2 \leq \sum_{k=1}^{n} \delta \Delta_k \|\underline{Q}\|^2 = \delta C \|\underline{Q}\|^2,$$

where the first inequality follows from Lemma E.4 and the second inequality follows from Lemma E.5.  □

**Corollary E.7.** *Let $\underline{Q}$ be a lower transition rate operator. Consider any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and any $u \in \mathscr{U}_{[t,s]}$ such that $\sigma(u) \|\underline{Q}\| \leq 1$. Then for all $u' \in \mathscr{U}_{[t,s]}$ such that $u \subseteq u'$:*

$$\|\Phi_u - \Phi_{u'}\| \leq \sigma(u)(s-t) \|\underline{Q}\|^2.$$

**Proof.** This result is trivial if $s = t$, because then $u = u' = \{t\} = \{s\}$. Hence, without loss of generality, we assume that $s > t$, which implies that $u = (t_0, \ldots, t_n)$, with $n \in \mathbb{N}$, $t_0 = t$ and $t_n = s$. Since $u \subseteq u'$, we know that, for all $k \in \{1, \ldots, n\}$, there is some sequence $\Delta_{k,i} > 0$, $i \in \{1, \ldots, n_k\}$, with $n_k \in \mathbb{N}$ and $\Delta_k = \sum_{i=1}^{n_k} \Delta_{k,i} \leq \sigma(u)$, such that $\sum_{k=1}^{n} \Delta_k = s - t$,

$$\Phi_{u'} := \prod_{k=1}^{n} \left( \prod_{i=1}^{n_k} (I + \Delta_{k,i}\underline{Q}) \right) \quad \text{and} \quad \Phi_u := \prod_{k=1}^{n} (I + \Delta_k \underline{Q}).$$

Because of Lemma E.6, this implies that $\|\Phi_{u'} - \Phi_u\| \leq \sigma(u)(s-t) \|\underline{Q}\|^2$.  □

**Proposition 7.9.** *Consider any $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, any $\delta \in \mathbb{R}_{>0}$ such that $\delta \|\underline{Q}\| \leq 1$, and any $u, u^* \in \mathscr{U}_{[t,s]}$ such that $\sigma(u) \leq \delta$ and $\sigma(u^*) \leq \delta$. Let $C := s - t$. Then*

$$\|\Phi_u - \Phi_{u^*}\| \leq 2\delta C \|\underline{Q}\|^2.$$

**Proof.** Consider any $u' \in \mathscr{U}_{[t,s]}$ that is finer than $u$ and $u^*$, meaning that the timepoints it consists of contain the timepoints in $u$ and the timepoints in $u^*$. For example, let $u'$ be the ordered union of the timepoints in $u$ and $u^*$. Corollary E.7 then implies that $\|\Phi_u - \Phi_{u'}\| \leq \delta C \|\underline{Q}\|^2$ and $\|\Phi_{u^*} - \Phi_{u'}\| \leq \delta C \|\underline{Q}\|^2$, and therefore, it follows that

$$\|\Phi_u - \Phi_{u^*}\| \leq \|\Phi_u - \Phi_{u'}\| + \|\Phi_{u'} - \Phi_{u^*}\| \leq 2\delta C \|\underline{Q}\|^2.$$  □

**Corollary 7.10.** *For every sequence $\{u_i\}_{i \in \mathbb{N}}$ in $\mathscr{U}_{[t,s]}$ such that $\lim_{i \to \infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i \in \mathbb{N}}$ is Cauchy.*

**Proof.** By definition of a Cauchy sequence, we need to show that

$$(\forall \varepsilon > 0)(\exists n \in \mathbb{N})(\forall i, j \geq n) \|\Phi_{u_i} - \Phi_{u_j}\| < \varepsilon.$$

Therefore, fix any $\varepsilon \in \mathbb{R}_{>0}$, and choose $\delta \in \mathbb{R}_{>0}$ such that $2\delta(s-t) \|\underline{Q}\|^2 < \varepsilon$ and $\delta \|\underline{Q}\| \leq 1$ [this is clearly always possible]. Because $\lim_{i \to \infty} \sigma(u_i) = 0$, there is some $n \in \mathbb{N}$ such that, for all $i \geq n$, $\sigma(u_i) \leq \delta$. Consider any $i, j \geq n$. It then follows from Proposition 7.9 that

$$\|\Phi_{u_i} - \Phi_{u_j}\| \leq 2\delta(s-t) \|\underline{Q}\|^2 < \varepsilon.$$  □

**Corollary 7.11.** *For every sequence $\{u_i\}_{i \in \mathbb{N}}$ in $\mathscr{U}_{[t,s]}$ such that $\lim_{i \to \infty} \sigma(u_i) = 0$, the corresponding sequence $\{\Phi_{u_i}\}_{i \in \mathbb{N}}$ converges to a lower transition operator.*

**Proof.** Since $\lim_{i \to \infty} \sigma(u_i) = 0$, and due to Propositions 7.3 and 7.1 and property LR5, there is some index $n$ such that the sequence $\Phi_{u_n}, \Phi_{u_{n+1}}, \ldots$ consists of lower transition operators. Due to Corollary 7.10, this sequence is Cauchy and therefore, because of Proposition 7.2, this sequence has a limit that is also a lower transition operator. Since the limit starting from $n$ and the limit starting from 1 are identical (initial elements do not influence the limit), we find that the sequence $\{\Phi_{u_i}\}_{i \in \mathbb{N}}$ has a limit, and that this limit is a lower transition operator.  □

**Theorem 7.12.** *For any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$ and any lower transition rate operator $\underline{Q}$, there is a unique lower transition operator $\underline{T} \in \mathbb{T}$ such that*

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall u \in \mathscr{U}_{[t,s]} : \sigma(u) \leq \delta)\ \|\underline{T} - \Phi_u\| \leq \varepsilon. \tag{43}$$

**Proof.** Let $\{u_i\}_{i\in\mathbb{N}}$ be any sequence in $\mathscr{U}_{[t,s]}$ such that $\lim_{i\to\infty} \sigma(u_i) = 0$. Because of Corollary 7.11, the sequence $\{\Phi_{u_i}\}_{i\in\mathbb{N}}$ then converges to a lower transition operator, which we denote by $\underline{T}$.

Let $C \coloneqq s - t$, fix any $\varepsilon > 0$, and choose any $\delta > 0$ such that $4\delta C \left\|\underline{Q}\right\|^2 < \varepsilon$ and $\delta \left\|\underline{Q}\right\| \le 1$ [this is clearly always possible]. Since $\lim_{i\to+\infty} \Phi_{u_i} = \underline{T}$ and $\lim_{i\to\infty} \sigma(u_i) = 0$, there is some $i^* \in \mathbb{N}$ such that

$$\sigma(u_{i^*}) \le \delta \text{ and } \left\|\underline{T} - \Phi_{u_{i^*}}\right\| \le \frac{\varepsilon}{2}. \tag{E.3}$$

Consider now any $u \in \mathscr{U}_{[t,s]}$ such that $\sigma(u) \le \delta$. Then

$$\left\|\underline{T} - \Phi_u\right\| \le \left\|\underline{T} - \Phi_{u_{i^*}}\right\| + \left\|\Phi_{u_{i^*}} - \Phi_u\right\| \le \frac{\varepsilon}{2} + 2\delta C \left\|\underline{Q}\right\|^2 \le \varepsilon,$$

where the second inequality follows from Equation (E.3) and Proposition 7.9.

It remains to show that $\underline{T}$ is unique. Therefore, let $\underline{T}'$ be any lower transition operator that satisfies Equation (43). Then clearly, for any $\varepsilon > 0$, there is some $u \in \mathscr{U}_{[t,s]}$ such that $\left\|\underline{T} - \Phi_u\right\| \le \varepsilon$ and $\left\|\underline{T}' - \Phi_u\right\| \le \varepsilon$, and therefore also $\left\|\underline{T} - \underline{T}'\right\| \le 2\varepsilon$. Since $\varepsilon > 0$ is arbitrary, this implies that $\left\|\underline{T} - \underline{T}'\right\| = 0$, or equivalently, that $\underline{T} = \underline{T}'$. Because this holds for every $\underline{T}'$ that satisfies Equation (43), it follows that $\underline{T}$ is the unique operator satisfying this equation. $\square$

We following lemma restates the norm inequality given by Corollary E.7, so that it can be used with lower transition operators $\underline{L}_t^s$. In effect, it therefore provides a bound on how well we can approximate the operator $\underline{L}_t^s$ using an operator $\Phi_u$ constructed using a finite partition $u \in \mathscr{U}_{[t,s]}$ of the interval $[t,s]$. In particular, this approximation improves as we take $u$ to be increasingly finer, that is, as we decrease $\sigma(u)$.

**Lemma E.8.** *Let $\underline{Q}$ be a lower transition rate operator. Then for any $t, s \in \mathbb{R}_{\ge 0}$ such that $t \le s$ and any $u \in \mathscr{U}_{[t,s]}$ such that $\sigma(u)\left\|\underline{Q}\right\| \le 1$:*

$$\left\|\underline{L}_t^s - \Phi_u\right\| \le \sigma(u)(s-t)\left\|\underline{Q}\right\|^2.$$

**Proof.** Fix any $\varepsilon > 0$. Because of Definition 7.4, there is some $u' \in \mathscr{U}_{[t,s]}$ such that $u \subseteq u'$ and $\left\|\underline{L}_t^s - \Phi_{u'}\right\| \le \varepsilon$. By combining this with Corollary E.7, it follows that

$$\left\|\underline{L}_t^s - \Phi_u\right\| \le \left\|\underline{L}_t^s - \Phi_{u'}\right\| + \left\|\Phi_{u'} - \Phi_u\right\| \le \varepsilon + \sigma(u)(s-t)\left\|\underline{Q}\right\|^2.$$

Since $\varepsilon > 0$ is arbitrary, the result is now immediate. $\square$

**Proposition 7.13.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, r, s \in \mathbb{R}_{\ge 0}$ such that $t \le r \le s$, it holds that*

$$\underline{L}_t^s = \underline{L}_t^r \underline{L}_r^s.$$

*Furthermore, for all $t \in \mathbb{R}_{\ge 0}$, we have that $\underline{L}_t^t = I$.*

**Proof.** Consider any $t \in \mathbb{R}_{\ge 0}$. Then $\mathscr{U}_{[t,t]}$ contains only a single degenerate sequence of time points $u$, which consists of the single time point $t$, and for this sequence $u$, we trivially have that $\Phi_u = I$ and $\sigma(u) = 0$. Therefore, it follows from Definition 7.4 that $\underline{L}_t^t = I$.

Consider now any $t, r, s \in \mathbb{R}_{\ge 0}$ such that $t \le r \le s$. It remains to show that $\underline{L}_t^s = \underline{L}_t^r \underline{L}_r^s$. If $t = r$ or $r = s$, this follows trivially from the first part of this proof. Hence, without loss of generality, we may assume that $t < r < s$.

Fix any $\delta > 0$ such that $\delta \left\|\underline{Q}\right\| \le 1$. Since $r > t$, there is some $0 < \Delta_1 < \delta$ and $n_1 \in \mathbb{N}$ such that $\Delta_1 n_1 = r - t$. Similarly, since $s > r$, there is some $0 < \Delta_2 < \delta$ and $n_2 \in \mathbb{N}$ such that $\Delta_2 n_2 = s - r$. Furthermore, because of Propositions 7.3 and 7.1, $(I + \Delta_1 \underline{Q})^{n_1}$ and $(I + \Delta_1 \underline{Q})^{n_1}$ are lower transition operators. Hence, we find that

$$\begin{aligned}
\left\|\underline{L}_t^s - \underline{L}_t^r \underline{L}_r^s\right\| &\le \left\|\underline{L}_t^s - (I + \Delta_1 \underline{Q})^{n_1}(I + \Delta_2 \underline{Q})^{n_2}\right\| + \left\|(I + \Delta_1 \underline{Q})^{n_1}(I + \Delta_2 \underline{Q})^{n_2} - \underline{L}_t^r \underline{L}_r^s\right\| \\
&\le \left\|\underline{L}_t^s - (I + \Delta_1 \underline{Q})^{n_1}(I + \Delta_2 \underline{Q})^{n_2}\right\| + \left\|(I + \Delta_1 \underline{Q})^{n_1} - \underline{L}_t^r\right\| + \left\|(I + \Delta_2 \underline{Q})^{n_2} - \underline{L}_r^s\right\| \\
&\le \max\{\Delta_1, \Delta_2\}(s-t)\left\|\underline{Q}\right\|^2 + \Delta_1(r-t)\left\|\underline{Q}\right\|^2 + \Delta_2(s-r)\left\|\underline{Q}\right\|^2 \\
&\le \delta(s-t)\left\|\underline{Q}\right\|^2 + \delta(r-t)\left\|\underline{Q}\right\|^2 + \delta(s-r)\left\|\underline{Q}\right\|^2 = 2\delta(s-t)\left\|\underline{Q}\right\|^2,
\end{aligned}$$

where the second inequality follows from Lemma E.4 and the third inequality follows from Lemma E.8. Since $\delta > 0$ can be chosen arbitrarily small, this implies that $\left\|\underline{L}_t^s - \underline{L}_t^r \underline{L}_r^s\right\| = 0$, or equivalently, that $\underline{L}_t^s = \underline{L}_t^r \underline{L}_r^s$. $\square$

**Proposition 7.14.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\ge 0}$ such that $t \le s$, we have that $\underline{L}_t^s = \underline{L}_0^{s-t}$.*

**Proof.** Consider any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any $\Delta \in \mathbb{R}_{\geq 0}$, and any sequence $\{u_i\}_{i \in \mathbb{N}}$ in $\mathscr{U}_{[t,s]}$ such that $\lim_{i \to \infty} \sigma(u_i) = 0$. For any $i \in \mathbb{N}$, with $u_i = t_0, \ldots, t_n$, we now define the sequence $u_i^* = (t_0 + \Delta), (t_1 + \Delta), \ldots, (t_n + \Delta)$. Then clearly, $\lim_{i \to \infty} \sigma(u_i^*) = 0$ and, for all $i \in \mathbb{N}$, $u_i^* \in \mathscr{U}_{[t+\Delta, s+\Delta]}$ and, due to Equation (42), $\Phi_{u_i} = \Phi_{u_i^*}$. We now have that

$$L_t^s = \lim_{i \to \infty} \Phi_{u_i} = \lim_{i \to \infty} \Phi_{u_i^*} = L_{t+\Delta}^{s+\Delta},$$

where the first and last equality follow from Theorem 7.12.  □

The following lemma provides a bound on how well the operator $L_t^{t+\Delta}$ can be approximated using the much simpler operator $(I + \Delta \underline{Q})$. Note that this lemma states the general version of Lemma A.3, which until this point has remained unproven. To see that the result below indeed implies Lemma A.3, simply recall from Section 7.2 that any rate matrix $Q$ is also a lower transition rate operator. Since it follows from our comments in Section 7.4 that the lower transition operator $L_t^{t+\Delta}$ corresponding to such a $\underline{Q} = Q$ satisfies $L_t^{t+\Delta} = e^{Q\Delta}$, Lemma A.3 clearly follows from the following result.

**Lemma E.9.** *Let $\underline{Q}$ be a lower transition rate operator. Then for any $t, \Delta \in \mathbb{R}_{\geq 0}$:*

$$\left\| L_t^{t+\Delta} - (I + \Delta \underline{Q}) \right\| \leq \Delta^2 \left\| \underline{Q} \right\|^2.$$

**Proof.** Fix any $\varepsilon > 0$. Because of Definition 7.4, there is some $u \in \mathscr{U}_{[t,t+\Delta]}$ such that $\sigma(u) \left\| \underline{Q} \right\| \leq 1$ and $\left\| L_t^s - \Phi_u \right\| \leq \varepsilon$. By combining this with Lemma E.5, it follows that

$$\left\| L_t^s - (I + \Delta \underline{Q}) \right\| \leq \left\| L_t^s - \Phi_u \right\| + \left\| \Phi_u - (I + \Delta \underline{Q}) \right\| \leq \varepsilon + \Delta^2 \left\| \underline{Q} \right\|^2.$$

Since $\varepsilon > 0$ is arbitrary, the result is now immediate.  □

The final lemma of this section states how well we can approximate the operator $L_t^{t+\Delta}$ using the identity matrix $I$. Put differently, since we know from Proposition 7.13 that $L_t^t = I$, this tells us how quickly $L_t^{t+\Delta}$ changes as we increase $\Delta$. This result is a generalised version of Lemma A.4, which has so far remained unproven. The reason why Lemma A.4 is a special case is again because for $\underline{Q} = Q$, we have that $L_t^{t+\Delta} = e^{Q\Delta}$.

**Lemma E.10.** *Let $\underline{Q}$ be a lower transition rate operator. Then for any $t, \Delta \in \mathbb{R}_{\geq 0}$:*

$$\left\| L_t^{t+\Delta} - I \right\| \leq \Delta \left\| \underline{Q} \right\|.$$

**Proof.** Fix any $n \in \mathbb{N}$ and let $\Delta_n := \Delta/n$. Propositions 7.13 and 7.14 then imply that

$$L_t^{t+\Delta} = L_0^\Delta = L_0^{n\Delta_n} = L_0^{\Delta_n} L_{\Delta_n}^{2\Delta_n} \cdots L_{(n-2)\Delta_n}^{(n-1)\Delta_n} L_{(n-1)\Delta_n}^{n\Delta_n} = L_0^{\Delta_n} L_0^{\Delta_n} \cdots L_0^{\Delta_n} L_0^{\Delta_n} = (L_0^{\Delta_n})^n,$$

and therefore, it follows from Lemma E.4 that

$$\left\| L_t^{t+\Delta} - I \right\| = \left\| (L_0^{\Delta_n})^n - I^n \right\| \leq n \left\| L_0^{\Delta_n} - I \right\| \leq n \left\| L_0^{\Delta_n} - (I + \Delta_n \underline{Q}) \right\| + n\Delta_n \left\| \underline{Q} \right\|,$$

which, when combined with Lemma E.9, implies that

$$\left\| L_t^{t+\Delta} - I \right\| \leq n\Delta_n^2 \left\| \underline{Q} \right\|^2 + n\Delta_n \left\| \underline{Q} \right\| = \frac{1}{n}\Delta^2 \left\| \underline{Q} \right\|^2 + \Delta \left\| \underline{Q} \right\|.$$

Since $n \in \mathbb{N}$ is arbitrary, the result is now immediate.  □

**Proposition 7.15.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, it holds that*[12]

$$\frac{\partial}{\partial t} L_t^s = -\underline{Q} \, L_t^s \quad \text{and} \quad \frac{\partial}{\partial s} L_t^s = \underline{Q} \, L_t^s.$$

---

[12] If $0 = t < s$, the derivative with respect to $t$ is taken to be a right derivative. If $t = s$, the derivative with respect to $s$ is taken to be a right derivative and the derivative with respect to $t$ is taken to be a left derivative (or becomes meaningless if $t = 0$).

**Proof.** We first prove that $\frac{\partial}{\partial t} L_t^s = -\underline{Q}\, L_t^s$, or equivalently, that

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall \Delta : 0 < |\Delta| < \delta,\; 0 \le t + \Delta \le s)\; \left\| \frac{L_{t+\Delta}^s - L_t^s}{\Delta} + \underline{Q}\, L_t^s \right\| < \varepsilon. \tag{E.4}$$

Fix any $\varepsilon \in \mathbb{R}_{>0}$, and choose any $\delta > 0$ such that $3\delta \left\| \underline{Q} \right\|^2 < \varepsilon$. Consider now any $\Delta \in \mathbb{R}$ such that $0 < |\Delta| < \delta$ and $0 \le t + \Delta \le s$. We will show that the inequality in Equation (E.4) holds. Let $t' := \max\{t, t+\Delta\}$. We then find that

$$\left\| L_{t+\Delta}^s - L_t^s + \Delta\, \underline{Q}\, L_t^s \right\| = \left\| L_{t'}^s - L_{t'-|\Delta|}^s + |\Delta|\, \underline{Q}\, L_t^s \right\| = \left\| L_{t'}^s - L_{t'-|\Delta|}^{t'} L_{t'}^s + |\Delta|\, \underline{Q}\, L_t^{t'} L_{t'}^s \right\|,$$

where the last equality follows from Proposition 7.13 and because $t \le t' \le s$. Therefore, and because of N10 and LT4, we find that

$$\left\| L_{t+\Delta}^s - L_t^s + \Delta\, \underline{Q}\, L_t^s \right\| \le \left\| I - L_{t'-|\Delta|}^{t'} + |\Delta|\, \underline{Q}\, L_t^{t'} \right\| \left\| L_{t'}^s \right\| \le \left\| I - L_{t'-|\Delta|}^{t'} + |\Delta|\, \underline{Q}\, L_t^{t'} \right\|,$$

which implies that

$$\left\| L_{t+\Delta}^s - L_t^s + \Delta\, \underline{Q}\, L_t^s \right\| \le \left\| I + |\Delta|\, \underline{Q} - L_{t'-|\Delta|}^{t'} \right\| + |\Delta| \left\| \underline{Q}\, L_t^{t'} - \underline{Q} \right\|$$

$$\le \Delta^2 \left\| \underline{Q} \right\|^2 + |\Delta|\, 2 \left\| \underline{Q} \right\| \left\| L_t^{t'} - I \right\| \le \Delta^2 \left\| \underline{Q} \right\|^2 + 2\,|\Delta|\,(t'-t) \left\| \underline{Q} \right\|^2,$$

using Lemma E.9 and LR6 for the second inequality and Lemma E.10 for the third inequality. Hence, since $0 \le (t'-t) \le |\Delta|$, we find that $\left\| L_{t+\Delta}^s - L_t^s + \Delta\, \underline{Q}\, L_t^s \right\| \le 3\Delta^2 \left\| \underline{Q} \right\|^2$, which implies that indeed, as required,

$$\left\| \frac{L_{t+\Delta}^s - L_t^s}{\Delta} + \underline{Q}\, L_t^s \right\| = \frac{1}{|\Delta|} \left\| L_{t+\Delta}^s - L_t^s + \Delta\, \underline{Q}\, L_t^s \right\| \le 3\,|\Delta| \left\| \underline{Q} \right\|^2 \le 3\delta \left\| \underline{Q} \right\|^2 < \varepsilon.$$

Next, we prove that $\frac{\partial}{\partial s} L_t^s = \underline{Q}\, L_t^s$, or equivalently, that

$$(\forall \varepsilon > 0)\,(\exists \delta > 0)\,(\forall \Delta : 0 < |\Delta| < \delta,\; t \le s + \Delta)\; \left\| \frac{L_t^{s+\Delta} - L_t^s}{\Delta} - \underline{Q}\, L_t^s \right\| < \varepsilon. \tag{E.5}$$

Fix any $\varepsilon > 0$ and $\alpha > 0$ and let $t^* := t + \alpha$ and $s^* := s + \alpha$. It then follows from Equation (E.4) that there is some $\delta^* > 0$ such that

$$(\forall \Delta^* : 0 < |\Delta^*| < \delta^*,\; 0 \le t^* + \Delta^* \le s^*)\; \left\| \frac{L_{t^*+\Delta^*}^{s^*} - L_{t^*}^{s^*}}{\Delta^*} + \underline{Q}\, L_{t^*}^{s^*} \right\| < \varepsilon. \tag{E.6}$$

Let $\delta := \min\{\delta^*, \alpha\}$. Consider now any $\Delta \in \mathbb{R}$ such that $0 < |\Delta| < \delta$ and $t \le s + \Delta$. We will prove that the inequality in Equation (E.5) holds. Let $\Delta^* := -\Delta$. We then have that $0 < |\Delta^*| < \delta \le \delta^*$ and that $t^* + \Delta^* = t + \alpha - \Delta \ge t + \alpha - \delta \ge t \ge 0$ and $t^* + \Delta^* = t + \alpha - \Delta \le s + \Delta + \alpha - \Delta = s^*$, and therefore, we find that indeed, as required,

$$\left\| \frac{L_t^{s+\Delta} - L_t^s}{\Delta} - \underline{Q}\, L_t^s \right\| = \left\| \frac{L_{t^*+\Delta^*}^{s^*} - L_{t^*}^{s^*}}{-\Delta^*} - \underline{Q}\, L_{t^*}^{s^*} \right\| = \left\| \frac{L_{t^*+\Delta^*}^{s^*} - L_{t^*}^{s^*}}{\Delta^*} + \underline{Q}\, L_{t^*}^{s^*} \right\| < \varepsilon,$$

where the first equality follows from Proposition 7.14 and the final inequality follows from Equation (E.6). $\square$

## Appendix F. Proofs and lemmas for the results in Section 8

Before giving the proof of Proposition 8.1, we first prove some crucial steps of the proof separately, in the lemmas below.

First of all, recall from Proposition 4.7 that, for any $\varepsilon \in \mathbb{R}_{>0}$, there is some $\delta > 0$ such that, for all $0 < \Delta < \delta$, the transition matrix $T_{t,x_u}^{t+\Delta}$ of a fixed stochastic process $P \in \mathbb{P}_{\mathcal{Q},\mathcal{M}}^{\mathrm{W}}$ can be approximated by $(I + \Delta Q)$ with an error of at most $\Delta\varepsilon$, using a rate matrix $Q \in \overline{\partial} T_{t,x_u}^t \subseteq \mathcal{Q}$. Quite similarly, the following lemma states that, for any given time interval $[t,s]$, there exists a finite partition $u \in \mathcal{U}_{[t,s]}$, $u = t_0, \dots, t_n$, such that the transition matrices $T_{t_i,x_u}^{t_{i+1}}$ can all be approximated by $(I + \Delta_{i+1} Q_{i+1})$, for some $Q_{i+1} \in \mathcal{Q}$ and with $\Delta_{i+1} = t_{i+1} - t_i$.

The reason that this result does not follow trivially from Proposition 4.7 is because the $\delta$—and hence also $\Delta$—in Proposition 4.7 depends on the particular time point that is considered. For this reason, the intuitive idea of using Proposition 4.7 to first find some $\Delta_0$ and $Q_0$ such that $T_{t_0,x_u}^{t_1} = T_{t_0,x_u}^{t_0+\Delta}$ can be approximated by $I + \Delta_1 Q_1$, and to then continue in this way to find some $\Delta_2$ and $Q_2$, and then some $\Delta_3$ and $Q_3$, and so on, is not feasible, because this process may continue indefinitely if $\sum_{i=1}^{\infty} \Delta_i$ is finite. In order to make this work, we need some kind of guarantee that it suffices to consider a finite number of $\Delta_i$, and this is exactly what the following lemma establishes.

**Lemma F.1.** *Consider any $P \in \mathbb{P}^{\mathrm{W}}_{\mathcal{Q}, \mathcal{M}}$, any $0 \le t < s$, any $u \in \mathcal{U}_{<t}$ and any $x_u \in \mathcal{X}_u$. Then for all $\varepsilon > 0$ and $\delta > 0$, there is some $v \in \mathcal{U}_{[t,s]}$ such that $\sigma(v) < \delta$ and, for all $i \in \{0, \ldots, n-1\}$:*

$$(\exists Q \in \mathcal{Q}) \left\| T^{t_{i+1}}_{t_i, x_u} - (I + \Delta_{i+1} Q) \right\| < \Delta_{i+1} \varepsilon.$$

**Proof.** Fix any $\varepsilon > 0$ and $\delta > 0$. It then follows from Proposition 4.7 and Definition 6.1 that there is some $0 < \delta^* < \min\{\delta, 1/2(s-t)\}$ such that, for all $0 < \Delta < \delta^*$:

$$(\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta}(T^{t+\Delta}_{t, x_u} - I) - Q \right\| < \varepsilon \quad \text{and} \quad (\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta}(T^s_{s-\Delta, x_u} - I) - Q \right\| < \varepsilon. \tag{F.1}$$

Let $t^* := t + \delta^*$ and $s^* := s - \delta^*$. Then clearly, $t < t^* < s^* < s$. For any $r \in [t^*, s^*]$, it follows from Proposition 4.7 and Definition 6.1 that there is some $0 < \delta_r < \delta^*$ such that, for all $0 < \Delta < \delta_r$:

$$(\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta}(T^{r+\Delta}_{r, x_u} - I) - Q \right\| < \varepsilon \quad \text{and} \quad (\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta}(T^r_{r-\Delta, x_u} - I) - Q \right\| < \varepsilon. \tag{F.2}$$

Let $U_r := (r - \delta_r, r + \delta_r)$. Then the set $C := \{U_r : r \in [t^*, s^*]\}$ is an open cover of $[t^*, s^*]$. By the Heine–Borel theorem, $C$ contains a finite subcover $C^*$ of $[t^*, s^*]$. Without loss of generality, we can take this subcover to be minimal, in the sense that if we remove any of its elements, it is no longer a cover. Let $m$ be the cardinality of $C^*$ and let $r_1 < r_2 < \cdots < r_m$ be the ordered sequence of the midpoints of the intervals in $C^*$.

We will now prove that

$$r_i - \delta_{r_i} < r_j - \delta_{r_j} \quad \text{and} \quad r_i + \delta_{r_i} < r_j + \delta_{r_j} \quad \text{for all } 1 \le i < j \le m. \tag{F.3}$$

Assume *ex absurdo* that this statement is not true. Then this implies that there are $1 \le i < j \le m$ such that either $r_i - \delta_{r_i} \ge r_j - \delta_{r_j}$ or $r_i + \delta_{r_i} \ge r_j + \delta_{r_j}$. If $r_i - \delta_{r_i} \ge r_j - \delta_{r_j}$, then since $i < j$ implies that $r_i < r_j$, it follows that $\delta_{r_j} \ge \delta_{r_i} + r_j - r_i > \delta_{r_i}$ and therefore, that $r_j + \delta_{r_j} > r_i + \delta_{r_i}$. Hence, we find that $U_{r_i} \subseteq U_{r_j}$. Since $C^*$ was taken to be a minimal cover, this is a contradiction. Similarly, if $r_i + \delta_{r_i} \ge r_j + \delta_{r_j}$, then since $i < j$ implies that $r_i < r_j$, it follows that $\delta_{r_i} \ge \delta_{r_j} + r_j - r_i > \delta_{r_j}$ and therefore, that $r_i - \delta_{r_i} < r_j - \delta_{r_i}$. Hence, we find that $U_{r_j} \subseteq U_{r_i}$. Since $C^*$ was taken to be a minimal cover, this is again a contradiction. From these two contradictions, it follows that Equation (F.3) is indeed true.

Next, we prove that

$$r_{k+1} - \delta_{r_{k+1}} < r_k + \delta_{r_k} \quad \text{for all } k \in \{1, \ldots, m-1\}. \tag{F.4}$$

Assume *ex absurdo* that this statement is not true or, equivalently, that there is some $k \in \{1, \ldots, m-1\}$ such that $r_k + \delta_{r_k} \le r_{k+1} - \delta_{r_{k+1}}$. For all $i \in \{k+1, \ldots, m\}$, it then follows from Equation (F.3) that $r_k + \delta_{r_k} \le r_i - \delta_{r_i}$, which implies that $r_k + \delta_{r_k} \notin U_{r_i}$. Furthermore, for all $i \in \{1, \ldots, k\}$, it follows from Equation (F.3) that $r_i + \delta_{r_i} \le r_k + \delta_{r_k}$, which again implies that $r_k + \delta_{r_k} \notin U_{r_i}$. Hence, for all $i \in \{1, \ldots, m\}$, we have found that $r_k + \delta_{r_k} \notin U_{r_i}$. Since $C^*$ is a cover of $[t^*, s^*]$, this implies that $r_k + \delta_{r_k} \notin [t^*, s^*]$, which, since $r_k \in [t^*, s^*]$, implies that $r_k + \delta_{r_k} > s^*$. Hence, since we know from Equation (F.3) that $r_k - \delta_{r_k} < r_m - \delta_{r_m}$, it follows that $U_{r_k} \cap [t^*, s^*] \subseteq U_{r_m} \cap [t^*, s^*]$. This contradicts the fact that $C^*$ was taken to be a minimal cover, and therefore, Equation (F.4) must indeed be true.

For all $k \in \{1, \ldots, m-1\}$, we now define $q_k := 1/2(r_k + \delta_{r_k} + r_{k+1} - \delta_{r_{k+1}})$. Using Equation (F.3), it then follows that

$$q_k < \frac{r_{k+1} + \delta_{r_{k+1}} + r_{k+1} - \delta_{r_{k+1}}}{2} = r_{k+1} \quad \text{and} \quad q_k > \frac{r_k + \delta_{r_k} + r_k - \delta_{r_k}}{2} = r_k,$$

and Equation (F.4) trivially implies that $r_{k+1} - \delta_{r_{k+1}} < q_k < r_k + \delta_{r_k}$. Hence,

$$r_k < q_k < r_k + \delta_{r_k} \quad \text{and} \quad r_{k+1} - \delta_{r_{k+1}} < q_k < r_{k+1}.$$

Due to Equation (F.2), and with $\Delta^*_k := q_k - r_k$ and $\Delta^{**}_k := r_{k+1} - q_k$, this implies that

$$(\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta^*_k}(T^{q_k}_{r_k, x_u} - I) - Q \right\| < \varepsilon \quad \text{and} \quad (\exists Q \in \mathcal{Q}) \left\| \frac{1}{\Delta^{**}_k}(T^{r_{k+1}}_{q_k, x_u} - I) - Q \right\| < \varepsilon. \tag{F.5}$$

For all $k \in \{1, \ldots, m\}$, we now let $t_{2k} := r_k$ and, for all $k \in \{1, \ldots, m-1\}$, we let $t_{2k+1} := q_k$. For the resulting sequence $t_2 < t_3 < \cdots < t_{2m-1} < t_{2m}$, it then follows from Equation (F.5) and N9 that, for all $i \in \{2, \ldots, 2m-1\}$:

$$(\exists Q \in \mathcal{Q}) \left\| T^{t_{i+1}}_{t_i, x_u} - (I + \Delta_{i+1} Q) \right\| < \Delta_{i+1} \varepsilon, \tag{F.6}$$

with $\Delta_{i+1} := t_{i+1} - t_i < \delta$.

Next, since $C^*$ is a minimal cover, and because of Equation (F.3), we know that $r_1 - \delta_{r_1} < t^* \le r_1 = t_2$ and, since $\delta_{r_1} < \delta^*$, we also know that $r_1 - \delta_{r_1} > t$. Therefore, it follows that there is some $t_1 \in \mathbb{R}$ such that $t < r_1 - \delta_{r_1} < t_1 < t^* \le r_1$. If we now let $t_0 := t$, then $\Delta_1 := t_1 - t_0 < \delta^*$ and $\Delta_2 := t_2 - t_1 = r_1 - t_1 < \delta_{r_1}$, and therefore, it follows from Equations (F.1) and (F.2) and N9 that Equation (F.6) is also true for $i = 0$ and $i = 1$.

Finally, again since $C^*$ is a minimal cover and because of Equation (F.3), we know that $t_{2m} = r_m \leq s^* < r_m + \delta_{r_m}$ and, since $\delta_{r_m} < \delta^*$, we also know that $r_m + \delta_{r_m} < s$. Therefore, it follows that there is some $t_{2m+1} \in \mathbb{R}$ such that $t_{2m} = r_m \leq s^* < t_{2m+1} < r_m + \delta_{r_m} < s$. If we now let $t_{2m+2} \coloneqq s$, then $\Delta_{2m+2} \coloneqq t_{2m+2} - t_{2m+1} < \delta^*$ and $\Delta_{2m+1} \coloneqq t_{2m+1} - t_{2m} = t_{2m+1} - r_m < \delta_{r_m}$, and therefore, it follows from Equations (F.1) and (F.2) and N9 that Equation (F.6) is also true for $i = 2m$ and $i = 2m + 1$.

Hence, we conclude that Equation (F.6) holds for all $i \in \{0, 1, \ldots, 2m, 2m + 1\}$. The result now follows by letting $n \coloneqq 2m + 2$. $\quad\square$

The following lemma provides a decomposition property for the expectation operator $\mathbb{E}_P$ that corresponds to a stochastic processes $P \in \mathbb{P}$. For notational convenience, we express it in terms of transition matrices $T^s_{t, x_u}$, using Remark 4.1. This result will end up being useful in the proof of Proposition 8.1.

Roughly speaking, the result establishes that the (history-dependent) transition matrix $T^s_{t, x_u}$ of a stochastic process $P \in \mathbb{P}$ can be decomposed into a number of other transition matrices corresponding to this $P$, such that—crucially—these individual transition matrices do not depend on the exact way that the decomposition was performed—that is, the other transition matrices in the decomposition—despite $P$ not necessarily being a Markov process.

**Lemma F.2.** *Consider any $P \in \mathbb{P}^{\mathrm{W}}$, any $t, s \in \mathbb{R}_{\geq 0}$ such that $t < s$, any $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ and any sequence $t = t_0 < t_1 < \cdots < t_n = s$, with $n \in \mathbb{N}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $x_t \in \mathscr{X}$:*

$$[T^s_{t, x_u} f](x_t) = \left[ T^{t_1}_{t_0, x_u} \left( \prod_{i=2}^{n} T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} \right) f \right](x_t).$$

**Proof.** We provide a proof by induction. For $n = 1$, the result holds trivially. So consider now any $n > 1$ and assume that the result is true for $n - 1$. For any $g \in \mathscr{L}(\mathscr{X})$, Remark 4.1 then implies that

$$[T^{t_2}_{t_0, x_u} g](x_t) = \mathbb{E}_P[g(X_{t_2}) | X_t = x_t, X_u = x_u]$$
$$= \sum_{x_{t_1} \in \mathscr{X}} \mathbb{E}_P[g(X_{t_2}) | X_{t_1} = x_{t_1}, X_t = x_t, X_u = x_u] P(X_{t_1} = x_{t_1} | X_t = x_t, X_u = x_u)$$
$$= \sum_{x_{t_1} \in \mathscr{X}} [T^{t_2}_{t_1, x_{u \cup \{t\}}} g](x_{t_1}) P(X_{t_1} = x_{t_1} | X_t = x_t, X_u = x_u)$$
$$= [T^{t_1}_{t_0, x_u} T^{t_2}_{t_1, x_{u \cup \{t\}}} g](x_t),$$

and therefore, it follows that

$$[T^s_{t, x_u} f](x_t) = \left[ T^{t_2}_{t_0, x_u} \left( \prod_{i=3}^{n} T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} \right) f \right](x_t)$$
$$= \left[ T^{t_1}_{t_0, x_u} T^{t_2}_{t_1, x_{u \cup \{t\}}} \left( \prod_{i=3}^{n} T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} \right) f \right](x_t) = \left[ T^{t_1}_{t_0, x_u} \left( \prod_{i=2}^{n} T^{t_i}_{t_{i-1}, x_{u \cup \{t\}}} \right) f \right](x_t),$$

using the induction hypothesis for the first equality. $\quad\square$

The final two results that we require have to do with the connection between $\mathscr{Q}$ and $\underline{Q}$. Due to the definition of the lower transition rate operator $\underline{Q}$ corresponding to a given $\mathscr{Q}$—see Equation (38)—it holds for all $\Delta \in \mathbb{R}_{\geq 0}$, all $Q \in \mathscr{Q}$, and all $f \in \mathscr{L}(\mathscr{X})$ that $(I + \Delta Q) f \geq (I + \Delta \underline{Q}) f$. Lemma F.4 establishes that this inequality extends to compositions of such operators $(I + \Delta Q)$ and $(I + \Delta \underline{Q})$, provided that we impose an upper bound on $\Delta$. In order to prove this Lemma F.4, we also require Lemma F.3, which states that $\|Q\|$ is bounded above by $\|\underline{Q}\|$.

**Lemma F.3.** *Consider a non-empty bounded set $\mathscr{Q}$ of rate matrices and let $\underline{Q}$ be the corresponding lower transition rate operator. Then for all $Q \in \mathscr{Q}$, we have that $\|Q\| \leq \|\underline{Q}\|$.*

**Proof.** Consider any $f \in \mathscr{L}(\mathscr{X})$ such that $\|f\| = 1$. It then follows from Equation (38) that $Qf \geq \underline{Q} f$ and, due to the linearity of $Q$, also that $Qf = -Q(-f) \leq -\underline{Q}(-f)$. Hence, we find that $\underline{Q} f \leq Qf \leq -\underline{Q}(-f)$, which implies that

$$\|Qf\| \leq \max\{\|\underline{Q} f\|, \|-\underline{Q}(-f)\|\} = \max\{\|\underline{Q} f\|, \|\underline{Q}(-f)\|\}. \tag{F.7}$$

Since $\|f\| = 1$, it follows from Equation (1) that $\|\underline{Q} f\| \leq \|\underline{Q}\|$, and similarly, since $\|-f\| = \|f\| = 1$, we also find that $\|\underline{Q}(-f)\| \leq \|\underline{Q}\|$. By combining this with Equation (F.7), we find that $\|Qf\| \leq \|\underline{Q}\|$, and since this is true for every $f \in \mathscr{L}(\mathscr{X})$ such that $\|f\| = 1$, it now follows from Equation (1) that $\|Q\| \leq \|\underline{Q}\|$. $\quad\square$

**Lemma F.4.** *Consider a non-empty bounded set $\mathscr{Q}$ of rate matrices, let $\underline{Q}$ be the corresponding lower transition rate operator, and consider any $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\| \underline{Q} \right\| \leq 1$. Now fix any $n \in \mathbb{N}$ and, for all $i \in \{1, \ldots, n\}$, consider some $0 \leq \Delta_i \leq \delta$ and $Q_i \in \mathscr{Q}$. Then for any $f \in \mathscr{L}(\mathscr{X})$:*

$$\prod_{i=1}^{n}(I + \Delta_i Q_i)f \geq \prod_{i=1}^{n}(I + \Delta_i \underline{Q})f.$$

**Proof.** We provide a proof by induction. For $n = 1$, the result follows trivially from Equation (38). Consider now any $n > 1$ and assume that the result is true for $n - 1$. Since Lemma F.3 implies that $\Delta_1 \|Q_1\| \leq \Delta_1 \left\| \underline{Q} \right\| \leq \delta \left\| \underline{Q} \right\| \leq 1$, it then follows from Proposition 3.2 that $I + \Delta_1 Q_1$ is a transition matrix, and therefore, as noted in Section 7.1, also a lower transition matrix, which therefore satisfies LT5. We now find that

$$\prod_{i=1}^{n}(I + \Delta_i Q_i)f \geq (I + \Delta_1 Q_1)\prod_{i=2}^{n}(I + \Delta_i \underline{Q})f \geq \prod_{i=1}^{n}(I + \Delta_i \underline{Q})f,$$

where the first inequality follows from the induction hypothesis and LT5, and the second inequality follows from Equation (38). $\square$

**Proposition 8.1.** *Consider a non-empty bounded set of rate matrices $\mathscr{Q}$ whose corresponding lower transition rate operator is $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then, for any $P \in \mathbb{P}_{\mathscr{Q}}^{\mathsf{W}}$, any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any $u \in \mathscr{U}_{<t}$, any $x_t \in \mathscr{X}$ and $x_u \in \mathscr{X}_u$, and any $f \in \mathscr{L}(\mathscr{X})$:*

$$\mathbb{E}_P[f(X_s) \mid X_t = x_t, X_u = x_u] \geq [\underline{L}_t^s f](x_t).$$

**Proof.** This result is trivial if $t = s$. Hence, without loss of generality, we may assume that $t < s$. Fix any $\varepsilon > 0$ and let $C \coloneqq (s - t)$. Choose any $\varepsilon_1 > 0$ such that $\varepsilon_1 \|f\| < \varepsilon/2$ and any $\varepsilon_2 > 0$ such that $\varepsilon_2 C \|f\| < \varepsilon/2$.

Due to Theorem 7.12, there is some $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\| \underline{Q} \right\| \leq 1$ and

$$(\forall v \in \mathscr{U}_{[t,s]} : \sigma(v) \leq \delta) \left\| \underline{L}_t^s - \Phi_v \right\| \leq \varepsilon_1, \tag{F.8}$$

with $\Phi_v$ as in Equation (42). Since $P \in \mathbb{P}_{\mathscr{Q}}^{\mathsf{W}}$, it now follows from Proposition 4.7 that there is some $0 < \Delta_1 < \min\{\delta, C\}$ such that

$$(\exists Q_1 \in \overline{\partial}_+ T_{t, x_u}^t \subseteq \mathscr{Q}) \left\| T_{t_0, x_u}^{t_1} - (I + \Delta_1 Q_1) \right\| = \left\| T_{t, x_u}^{t + \Delta_1} - (I + \Delta_1 Q_1) \right\| < \Delta_1 \varepsilon_2,$$

with $t_0 \coloneqq t$ and $t_1 \coloneqq t + \Delta_1$. Furthermore, since $P \in \mathbb{P}_{\mathscr{Q}}^{\mathsf{W}}$, and because $\Delta_1 < C$ implies that $t_1 t + \Delta_1 < s$, it follows from Lemma F.1 that there is some $v \in \mathscr{U}_{[t_1,s]}$ such that $\sigma(v) < \delta$, with $v = t_1, \ldots, t_n$ and $t_n = s$, and such that for all $i \in \{2, \ldots, n\}$, with $\Delta_i \coloneqq t_i - t_{i-1}$:

$$(\exists Q_i \in \mathscr{Q}) \left\| T_{t_{i-1}, x_{u \cup \{t\}}}^{t_i} - (I + \Delta_i Q_i) \right\| < \Delta_i \varepsilon_2.$$

Since $\Delta_1 < \delta$ and, for all $i \in \{2, \ldots, n\}$, $\Delta_i \leq \sigma(v) < \delta$, we know that, for all $i \in \{1, \ldots, n\}$, $\Delta_i < \delta$ and therefore also, using Lemma F.3, that $\Delta_i \|Q_i\| \leq \delta \left\| \underline{Q} \right\| \leq 1$. Therefore, we find that

$$
\begin{aligned}
\left| [T_{t, x_u}^s f](x_t) - \left[ \left( \prod_{i=1}^{n}(I + \Delta_i Q_i) \right) f \right](x_t) \right| &= \left| \left[ T_{t_0, x_u}^{t_1} \left( \prod_{i=2}^{n} T_{t_{i-1}, x_{u \cup \{t\}}}^{t_i} \right) f \right](x_t) - \left[ \left( \prod_{i=1}^{n}(I + \Delta_i Q_i) \right) f \right](x_t) \right| \\
&\leq \left\| T_{t_0, x_u}^{t_1} \left( \prod_{i=2}^{n} T_{t_{i-1}, x_{u \cup \{t\}}}^{t_i} \right) - \prod_{i=1}^{n}(I + \Delta_i Q_i) \right\| \|f\| \\
&\leq \left\| T_{t_0, x_u}^{t_1} - (I + \Delta_1 Q_1) \right\| \|f\| + \sum_{i=2}^{n} \left\| T_{t_{i-1}, x_{u \cup \{t\}}}^{t_i} - (I + \Delta_i Q_i) \right\| \|f\| \\
&< \sum_{i=1}^{n} \Delta_i \varepsilon_2 \|f\| = C \varepsilon_2 \|f\| < \frac{\varepsilon}{2},
\end{aligned}
$$

where the equality follows from Lemma F.2, the first inequality follows from the properties of $\|\cdot\|$, and the second inequality follows from Lemma A.2 and Proposition 3.2 and, we also find that

$$\left| [\underline{L}_t^s f](x_t) - \left[ \left( \prod_{i=1}^{n}(I + \Delta_i \underline{Q}) \right) f \right](x_t) \right| \leq \left\| \underline{L}_t^s - \prod_{i=1}^{n}(I + \Delta_i \underline{Q}) \right\| \|f\| \leq \varepsilon_1 \|f\| < \frac{\varepsilon}{2},$$

using Equation (F.8) to establish the second inequality. Hence, Lemma F.4 implies that

$$[L_t^s f](x_t) < \left[ \left( \prod_{i=1}^{n}(I + \Delta_i \underline{Q}) \right) f \right](x_t) + \frac{\varepsilon}{2} \leq \left[ \left( \prod_{i=1}^{n}(I + \Delta_i Q_i) \right) f \right](x_t) + \frac{\varepsilon}{2} < [T_{t, x_u}^s f](x_t) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this allows us to infer that $[L_t^s f](x_t) \leq [T_{t, x_u}^s f](x_t)$. The result now follows because $[T_{t, x_u}^s f](x_t) = \mathbb{E}_P[f(X_s)|X_t = x_t, X_u = x_u]$, as noted in Remark 4.1. $\square$

The following lemma is required for the proof of Proposition 8.2. Recall from the definition of the lower envelope $\underline{Q}$ of a given set $\mathscr{Q}$ of rate matrices, that for any $f \in \mathscr{L}(\mathscr{X})$, and $x \in \mathscr{X}$ and any $\varepsilon > 0$, there is some $Q \in \mathscr{Q}$ such that $[Q f](x) < [\underline{Q} f](x) + \varepsilon$. In other words, $[\underline{Q} f](x)$ can be approximated arbitrarily closely using the elements of $\mathscr{Q}$. The following lemma establishes that the *entire* function $\underline{Q} f$ can be approximated arbitrarily closely by a single $Q \in \mathscr{Q}$, whenever $\mathscr{Q}$ has separately specified rows.

**Lemma F.5.** *Let $\mathscr{Q}$ be an arbitrary non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon \in \mathbb{R}_{>0}$, there exists a $Q \in \mathscr{Q}$ such that*

$$\left\| \underline{Q} f - Q f \right\| < \varepsilon.$$

**Proof.** This is immediate from the definition of the lower envelope of $\mathscr{Q}$, as given by Equation (38), and the fact that $\mathscr{Q}$ has separately specified rows. $\square$

**Proposition 8.2.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $f \in \mathscr{L}(\mathscr{X})$, and all $\varepsilon \in \mathbb{R}_{>0}$, there is a well-behaved Markov chain $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ such that*

$$\left| \mathbb{E}_P[f(X_s) \mid X_t = x_t] - [L_t^s f](x_t) \right| < \varepsilon \text{ for all } x_t \in \mathscr{X}.$$

**Proof.** Fix any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, any $f \in \mathscr{L}(\mathscr{X})$ and any $\varepsilon \in \mathbb{R}_{>0}$. If $t = s$, the result is trivial. Hence, without loss of generality, we may assume that $t < s$. Let $C := (s - t)$, choose any $\varepsilon^* > 0$ such that $\varepsilon^* C < \varepsilon$, choose any $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ such that $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 < \varepsilon^*$, choose any $\delta > 0$ such that $\delta \|\mathscr{Q}\|^2 \|f\| < \varepsilon_3$ and $\delta \|\underline{Q}\|^2 \|f\| < \varepsilon_1$ (this is always possible because of LR5), and consider any $u \in \mathscr{U}_{[t,s]}$ such that $\sigma(u) < \delta$, with $u = t_0, \ldots, t_n$ and $n \in \mathbb{N}$.

Fix any $i \in \{1, \ldots, n\}$ and let $g_i := L_{t_i}^{t_n} f$. It then follows from Lemma F.5 that there is some $Q_i \in \mathscr{Q}$ such that $\left\| \underline{Q} g_i - Q_i g_i \right\| < \varepsilon_2$ and, due to N11 and LT4, we also know that $\|g_i\| = \left\| L_{t_i}^{t_n} f \right\| \leq \left\| L_{t_i}^{t_n} \right\| \|f\| \leq \|f\|$. Hence, we find that

$$\begin{aligned} \left\| L_{t_{i-1}}^{t_i} g_i - e^{Q_i \Delta_i} g_i \right\| &\leq \left\| L_{t_{i-1}}^{t_i} g_i - [I + \Delta_i \underline{Q}] g_i \right\| + \left\| \Delta_i \underline{Q} g_i - \Delta_i Q_i g_i \right\| + \left\| [I + \Delta_i Q_i] g_i - e^{Q_i \Delta_i} g_i \right\| \\ &\leq \left\| L_{t_{i-1}}^{t_i} - [I + \Delta_i \underline{Q}] \right\| \|g_i\| + \Delta_i \left\| \underline{Q} g_i - Q_i g_i \right\| + \left\| I + \Delta_i Q_i - e^{Q_i \Delta_i} \right\| \|g_i\| \\ &< \left\| L_{t_{i-1}}^{t_i} - [I + \Delta_i \underline{Q}] \right\| \|f\| + \Delta_i \varepsilon_2 + \left\| I + \Delta_i Q_i - e^{Q_i \Delta_i} \right\| \|f\| \\ &\leq \Delta_i^2 \left\| \underline{Q} \right\|^2 \|f\| + \Delta_i \varepsilon_2 + \Delta_i^2 \|Q_i\|^2 \|f\| \\ &\leq \Delta_i (\delta \left\| \underline{Q} \right\|^2 \|f\| + \varepsilon_2 + \delta \|\mathscr{Q}\|^2 \|f\|) < \Delta_i (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) < \Delta_i \varepsilon^*, \end{aligned} \tag{F.9}$$

where the second inequality holds because of N11, the third inequality holds because $\|g_i\| \leq \|f\|$ and $\left\| \underline{Q} g_i - Q_i g_i \right\| < \varepsilon_2$, the fourth inequality holds because of Lemmas E.9 and A.3 and where the fifth inequality holds because $\Delta_i \leq \sigma(u) < \delta$.

Let $Q_0$ and $Q_{n+1}$ be two arbitrary elements of $\mathscr{Q}$ and, for all $i \in \{0, \ldots, n+1\}$, let $\mathscr{T}_{Q_i}$ denote the transition matrix system corresponding to $Q_i$, as in Definition 3.5. Then, by Proposition 6.2, there is some $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{\mathrm{WM}}$ with transition matrix system $\mathscr{T}_P$, such that

$$\mathscr{T}_P = \mathscr{T}_{Q_0}^{[0, t_0]} \otimes \mathscr{T}_{Q_1}^{[t_0, t_1]} \otimes \cdots \otimes \mathscr{T}_{Q_n}^{[t_{n-1}, t_n]} \otimes \mathscr{T}_{Q_n}^{[t_n, \infty)}.$$

Due to Equation (F.9), we know that the transition matrices of this process $P$ satisfy

$$(\forall i \in \{1, \ldots, n\}) \left\| L_{t_{i-1}}^{t_i} g_i - T_{t_{i-1}}^{t_i} g_i \right\| = \left\| L_{t_{i-1}}^{t_i} g_i - e^{Q_i \Delta_i} g_i \right\| < \Delta_i \varepsilon^*. \tag{F.10}$$

Furthermore, we also know that

$$\left\| L_{t_0}^{t_n} g_n - T_{t_0}^{t_n} g_n \right\| = \left\| L_{t_0}^{t_{n-1}} L_{t_{n-1}}^{t_n} g_n - T_{t_0}^{t_{n-1}} L_{t_{n-1}}^{t_n} g_n + T_{t_0}^{t_{n-1}} (L_{t_{n-1}}^{t_n} g_n - T_{t_{n-1}}^{t_n} g_n) \right\|$$

$$\leq \left\| L_{t_0}^{t_{n-1}} L_{t_{n-1}}^{t_n} g_n - T_{t_0}^{t_{n-1}} L_{t_{n-1}}^{t_n} g_n \right\| + \left\| T_{t_0}^{t_{n-1}} \right\| \left\| L_{t_{n-1}}^{t_n} g_n - T_{t_{n-1}}^{t_n} g_n \right\|$$

$$\leq \left\| L_{t_0}^{t_{n-1}} g_{n-1} - T_{t_0}^{t_{n-1}} g_{n-1} \right\| + \left\| L_{t_{n-1}}^{t_n} g_n - T_{t_{n-1}}^{t_n} g_n \right\|,$$

using Proposition 7.13 and Equation (25) for the first inequality, and LT4 for the second one. Similarly, we also find that

$$\left\| L_{t_0}^{t_{n-1}} g_{n-1} - T_{t_0}^{t_{n-1}} g_{n-1} \right\| \leq \left\| L_{t_0}^{t_{n-2}} g_{n-2} - T_{t_0}^{t_{n-2}} g_{n-2} \right\| + \left\| L_{t_{n-2}}^{t_{n-1}} g_{n-1} - T_{t_{n-2}}^{t_{n-1}} g_{n-1} \right\|.$$

By continuing in this way—applying induction—we eventually find that

$$\left\| L_t^s f - T_t^s f \right\| = \left\| L_{t_0}^{t_n} g_n - T_{t_0}^{t_n} g_n \right\| \leq \sum_{i=1}^{n} \left\| L_{t_{i-1}}^{t_i} g_i - T_{t_{i-1}}^{t_i} g_i \right\| \leq \sum_{i=1}^{n} \Delta_i \varepsilon^* = C \varepsilon^* < \varepsilon,$$

using Equation (F.10) to establish the second inequality. The result now follows directly from Remark 4.1, which states that $[T_t^s f](x_t) = \mathbb{E}_P[f(X_s)|X_t = x_t]$ for all $x_t \in \mathscr{X}$. $\square$

**Corollary 8.3.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$ and $x_t \in \mathscr{X}$, and all $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{W}[f(X_s) \,|\, X_t = x_t, X_u = x_u] = \underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{WM}[f(X_s) \,|\, X_t = x_t, X_u = x_u] = \left[ L_t^s f \right](x_t).$$

**Proof.** Fix any $\varepsilon > 0$. It then follows from Proposition 8.2 that there is some $P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{WM}$ such that $\mathbb{E}_P[f(X_s)|X_t = x_t] < [L_t^s f](x_t) + \varepsilon$, which implies that

$$\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{WM}[f(X_s) \,|\, X_t = x_t, X_u = x_u] \leq \mathbb{E}_P[f(X_s) \,|\, X_t = x_t, X_u = x_u] = \mathbb{E}_P[f(X_s)|X_t = x_t] < [L_t^s f](x_t) + \varepsilon,$$

using Equation (32) for the first inequality and the Markov property of $P$ for the equality. Since $\varepsilon > 0$ is arbitrary, it follows that $\underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{WM}[f(X_s)|X_t = x_t, X_u = x_u] \leq [L_t^s f](x_t)$. This implies the result because we also have that

$$\left[ L_t^s f \right](x_t) \leq \inf \left\{ \mathbb{E}_P[f(X_s) \,|\, X_t = x_t, X_u = x_u] : P \in \mathbb{P}_{\mathscr{Q}, \mathscr{M}}^{W} \right\}$$

$$= \underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{W}[f(X_s) \,|\, X_t = x_t, X_u = x_u] \leq \underline{\mathbb{E}}_{\mathscr{Q}, \mathscr{M}}^{WM}[f(X_s) \,|\, X_t = x_t, X_u = x_u],$$

where the first inequality follows from Proposition 8.1 and the last inequality follows from Proposition 6.4. $\square$

**Theorem 8.4.** *Let $\underline{Q}$ be an arbitrary lower transition rate operator, with $\mathscr{Q}_{\underline{Q}}$ its set of dominating rate matrices, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then the largest set of stochastic processes $\mathscr{P}$ for which the corresponding conditional lower expectation operator $\underline{\mathbb{E}}[\cdot \,|\, \cdot]$—as defined in Equation (31)—satisfies*

$$\underline{\mathbb{E}}[f(X_s) \,|\, X_t = x_t, X_u = x_u] = [L_t^s f](x_t)$$

*for all $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, all $u \in \mathscr{U}_{<t}$, all $x_t \in \mathscr{X}$ and $x_u \in \mathscr{X}_u$, and every $f \in \mathscr{L}(\mathscr{X})$, is the set $\mathbb{P}_{\mathscr{Q}_{\underline{Q}}}^{W}$.*

**Proof.** Clearly, it suffices to prove that for any $P \in \mathbb{P}$ that is not well-behaved or not consistent with $\mathscr{Q}_{\underline{Q}}$, there are $t, s \in \mathbb{R}_{\geq 0}$ with $t \leq s$, $u \in \mathscr{U}_{<t}$, $x_u \in \mathscr{X}_u$, $x_t \in \mathscr{X}$ and $f \in \mathscr{L}(\mathscr{X})$ such that

$$\mathbb{E}_P[f(X_s) \,|\, X_t = x_t, X_u = x_u] < [L_t^s f](x_t). \tag{F.11}$$

We start with the case that $P$ is not well-behaved. Fix any $\varepsilon > 0$ and let $C := \|\underline{Q}\| + \varepsilon$. It then follows from Proposition 4.5 that there are $t, s \in \mathbb{R}_{\geq 0}$ with $t < s$, $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ such that $1/\Delta \|T_{t, x_u}^s - I\| > C$ and $\Delta \|\underline{Q}\|^2 < \varepsilon$, with $\Delta := s - t > 0$. Let $Q := 1/\Delta(T_{t, x_u}^s - I)$. Since $\|Q\| > C$, it then follows that there is some $f' \in \mathscr{L}(\mathscr{X})$ such that $\|f'\| = 1$ and $\|Qf'\| > C$, which in turn implies that there is some $x_t \in \mathscr{X}$ such that $|[Qf'](x_t)| > C$. If $[Qf'](x_t) < 0$, we let $f := f'$, and if $[Qf'](x_t) > 0$, we let $f := -f'$. Clearly, this implies that $\|f\| = 1$ and $[Qf](x_t) < -C$. From $\|f\| = 1$, it furthermore follows that $[\underline{Q}f](x_t) \geq -\|\underline{Q}f\| \geq -\|\underline{Q}\|$, and therefore, we find that $[Qf](x_t) < -\|\underline{Q}\| - \varepsilon \leq [\underline{Q}f](x_t) - \varepsilon$, which implies that $[(I + \Delta Q)f](x_t) \leq [(I + \Delta\underline{Q})f](x_t) - \Delta\varepsilon$. Hence, since we also know that

$$\left| [L_t^s f](x_t) - [(I + \Delta\underline{Q})f](x_t) \right| \leq \left\| L_t^s - (I + \Delta\underline{Q}) \right\| \leq \Delta^2 \|\underline{Q}\|^2 < \Delta\varepsilon,$$

where we use $\|f\| = 1$ for the first inequality and Lemma E.9 for the second inequality, it follows that

$$[T_{t, x_u}^s f](x_t) = [(I + \Delta Q)f](x_t) \leq [(I + \Delta\underline{Q})f](x_t) - \Delta\varepsilon < [L_t^s f](x_t),$$

which, because of Remark 4.1, implies that Equation (F.11) holds.

Next, we consider the case that $P$ is well behaved, but not consistent with $\mathscr{Q}_{\underline{Q}}$. In that case, it follows from Definition 6.1 that there are $t^* \in \mathbb{R}_{\geq 0}$, $u \in \mathscr{U}_{<t}$ and $x_u \in \mathscr{X}_u$ such that $\overline{\partial} T^{t^*}_{t^*, x_u} \not\subseteq \mathscr{Q}_{\underline{Q}}$. Since we know that $\overline{\partial} T^{t^*}_{t^*, x_u}$ is a non-empty set of rate matrices because of Proposition 4.6, this implies the existence of a rate matrix $Q^* \in \overline{\partial} T^{t^*}_{t^*, x_u}$ such that $Q^* \notin \mathscr{Q}_{\underline{Q}}$. Furthermore, since $Q^* \notin \mathscr{Q}_{\underline{Q}}$, Equation (39) implies that there are $f' \in \mathscr{L}(\mathscr{X})$ and $x_t \in \mathscr{X}$ such that $[Q^* f'](x_t) < [\underline{Q} f'](x_t)$. Clearly, this implies that $f' \neq 0$, and therefore, that $\|f'\| > 0$. If we now let $f := 1/\|f'\| f'$, then $\|f\| = 1$, and furthermore, because of the linearity of $Q^*$ and the non-negative homogeneity of $\underline{Q}$, it follows that $[Q^* f](x_t) = 1/\|f'\| [Q^* f'](x_t) < 1/\|f'\| [\underline{Q} f'](x_t) = [\underline{Q} f](x_t)$. Consider now any $\varepsilon > 0$ such that $[Q^* f](x_t) \leq [\underline{Q} f](x_t) - 2\varepsilon$ (this is now clearly possible). Since $Q^* \in \overline{\partial} T^{t^*}_{t^*, x_u}$, it then follows from Definition 4.8 that there are $t, s \in \mathbb{R}_{\geq 0}$ such that $u < t < s$, $\|1/\Delta (T^s_{t, x_u} - I) - Q^*\| \leq \varepsilon$ and $\Delta \|\underline{Q}\|^2 < \varepsilon$, with $\Delta := s - t > 0$. Let $Q := 1/\Delta (T^s_{t, x_u} - I)$. Since $\|Q - Q^*\| \leq \varepsilon$ and $\|f\| = 1$, it then follows that $[Qf](x_t) \leq [Q^* f](x_t) + \varepsilon \leq [\underline{Q} f](x_t) - \varepsilon$. Hence, using the exact same argument as in the first part of this proof, we find that Equation (F.11) holds. $\square$

**Proposition 8.5.** *Let $\underline{Q}$ be a lower transition rate operator, choose any $t, s \in \mathbb{R}_{\geq 0}$ such that $t \leq s$, and let $\underline{L}^s_t$ be the lower transition operator corresponding to $\underline{Q}$. Then for any $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon \in \mathbb{R}_{>0}$, if we choose any $n \in \mathbb{N}$ such that*

$$ n \geq \max \left\{ (s-t) \left\| \underline{Q} \right\|, \frac{1}{2\varepsilon} (s-t)^2 \left\| \underline{Q} \right\|^2 \|f\|_v \right\}, $$

*with $\|f\|_v := \max f - \min f$, we are guaranteed that*

$$ \left\| \underline{L}^s_t f - \prod_{i=1}^{n} (I + \Delta \underline{Q}) f \right\| \leq \varepsilon, $$

*with $\Delta := (s-t)/n$.*

**Proof.** Define $h := f - \min f - 1/2 \|f\|_v$. Then

$$ \max h = \max f - \min f - 1/2 \|f\|_v = \|f\|_v - 1/2 \|f\|_v = 1/2 \|f\|_v $$

and

$$ \min h = \min f - \min f - 1/2 \|f\|_v = -1/2 \|f\|_v, $$

and therefore,

$$ \|h\| := \max\{|h(x)| : x \in \mathscr{X}\} = 1/2 \|f\|_v. $$

Now let $u \in \mathscr{U}_{[t,s]}$ be such that $u = t_0, t_1, \ldots, t_n$, where, for all $i \in \{0, 1, \ldots, n\}$, $t_i := t + i\Delta$. Since $\Delta = (s-t)/n$, we then have that $t_0 = t$, $t_n = s$, $\sigma(u) = \Delta$ and $\Phi_u = \prod_{i=1}^{n} (I + \Delta \underline{Q})$. Furthermore, since $n \geq (s-t) \left\| \underline{Q} \right\|$, we also know that $\Delta \left\| \underline{Q} \right\| = (s-t)/n \left\| \underline{Q} \right\| \leq 1$. Hence, we find that

$$ \left\| \underline{L}^s_t h - \prod_{i=1}^{n} (I + \Delta \underline{Q}) h \right\| = \left\| \underline{L}^s_t h - \Phi_u h \right\| \leq \left\| \underline{L}^s_t - \Phi_u \right\| \|h\| \leq \sigma(u)(s-t) \left\| \underline{Q} \right\|^2 \|h\| = \Delta(s-t) \left\| \underline{Q} \right\|^2 \frac{\|f\|_v}{2} $$

$$ = \frac{\varepsilon}{n} \frac{1}{2\varepsilon} (s-t)^2 \left\| \underline{Q} \right\|^2 \|f\|_v \leq \varepsilon, $$

where the first inequality follows from Property N11, the second inequality follows from Lemma E.8 and the final inequality follows from our lower bound on $n$. The result is now immediate because $\underline{L}^s_t$ and $\prod_{i=1}^{n} (I + \Delta \underline{Q})$ are both lower transition operators—for the latter, this follows from Propositions 7.3 and 7.1—which implies that

$$ \left\| \underline{L}^s_t h - \prod_{i=1}^{n} (I + \Delta \underline{Q}) h \right\| = \left\| \underline{L}^s_t f - \prod_{i=1}^{n} (I + \Delta \underline{Q}) f \right\| $$

because of LT6. $\square$

## Appendix G. Proofs for the results in Section 9

**Corollary 9.1.** *Let $\mathscr{M}$ be an arbitrary non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be an arbitrary non-empty bounded set of rate matrices that has separately specified rows, with corresponding lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_{\underline{Q}}$ be the corresponding lower transition operator system. Then, for all $s \in \mathbb{R}_{>0}$, all $u \in \mathscr{U}_{<s}$ such that $u \neq \emptyset$ and all $f \in \mathscr{L}(\mathscr{X}_{u \cup s})$:*

$$ \underline{\mathbb{E}}^{W}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u] = \underline{\mathbb{E}}^{WM}_{\mathscr{Q}, \mathscr{M}}[f(X_u, X_s) \mid X_u] = [\underline{L}^s_{t_n} f](X_u). $$

**Proof.** Because of Equations (53) and (54), this result is an immediate consequence of Corollary 8.3. □

**Corollary 9.2.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$, let $\mathscr{Q}$ be a non-empty, bounded and convex set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$, and let $\underline{\mathscr{T}}_Q$ be the corresponding lower transition operator system. Then, for any $u = t_0, \ldots, t_n$ and $v = s_0, \ldots, s_m$ in $\mathscr{U}_\emptyset$ such that $u < v$, and any $f \in \mathscr{L}(\mathscr{X}_{u \cup v})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_u, X_v) \mid X_u] = L_{t_n}^{s_0} L_{s_0}^{s_1} \cdots L_{s_{m-1}}^{s_m} f. \tag{57}$$

**Proof.** As explained in the main text of Section 9.2, this result is an immediate consequence of Theorem 6.5 and Corollary 9.1. □

**Proposition 9.3.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$ and let $\mathscr{Q}$ be a non-empty bounded set of rate matrices. Then for all $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}[f(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[f].$$

**Proof.** Fix $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon > 0$. It then follows from Equation (32) that there is a stochastic process $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ such that $\mathbb{E}_P[f(X_0)] \le \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_0)] + \varepsilon$. Furthermore, since $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$, it follows from Definitions 6.2 and 6.4 that $P(X_0) \in \mathscr{M}$, and therefore, that

$$\mathbb{E}_P[f(X_0)] = \sum_{x \in \mathscr{X}} P(X_0 = x) f(x) \ge \underline{\mathbb{E}}_{\mathscr{M}}[f].$$

Because of Equation (62), we also know that there is a probability mass function $p \in \mathscr{M}$ such that $\sum_{x \in \mathscr{X}} p(x) f(x) \le \underline{\mathbb{E}}_{\mathscr{M}}[f] + \varepsilon$. Consider now any $Q \in \mathscr{Q}$. It then follows from Corollary 5.3 that there is a unique well-behaved homogeneous Markov chain $P' \in \mathbb{P}^{\mathsf{M}}$, with transition matrix $Q$, and with $P'(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$. Since $P'$ clearly belongs to $\mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}$, this implies that

$$\sum_{x \in \mathscr{X}} p(x) f(x) = \sum_{x \in \mathscr{X}} P'(X_0 = x) f(x) = \mathbb{E}_{P'}[f(X_0)] \ge \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}[f(X_0)].$$

Hence, we conclude that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_0)] + \varepsilon \ge \mathbb{E}_P[f(X_0)] \ge \underline{\mathbb{E}}_{\mathscr{M}}[f] \ge \sum_{x \in \mathscr{X}} p(x) f(x) - \varepsilon \ge \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}[f(X_0)] - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies that $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_0)] \ge \underline{\mathbb{E}}_{\mathscr{M}}[f] \ge \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WHM}}[f(X_0)]$, and therefore, the result follows from Proposition 6.4. □

**Proposition 9.4.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$ and let $\mathscr{Q}$ be a non-empty bounded set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$. Then for any $s \in \mathbb{R}_{\ge 0}$ and any $f \in \mathscr{L}(\mathscr{X})$:*

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)] = \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_s)] = \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f].$$

**Proof.** Fix $s \in \mathbb{R}_{\ge 0}$, $f \in \mathscr{L}(\mathscr{X})$ and $\varepsilon > 0$. It then follows from Equation (32) that there is some $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ such that $\mathbb{E}_P[f(X_s)] \le \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)] + \varepsilon$. Since $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$, Equation (32) now implies that, for all $x \in \mathscr{X}$,

$$\mathbb{E}_P[f(X_s)|X_0 = x] \ge \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)|X_0 = x] = [L_0^s f](x), \tag{G.1}$$

where the last equality follows from Corollary 8.3. Hence, we find that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[f(X_s)] + \varepsilon \ge \mathbb{E}_P[f(X_s)] = \mathbb{E}_P[\mathbb{E}_P[f(X_s)|X_0]] \ge \mathbb{E}_P[[L_0^s f](X_0)] \ge \underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}[[L_0^s f](X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f],$$

where the first equality follows from Equation (34), the second inequality follows from Equation (G.1), the third inequality follows from $P \in \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{W}}$ and Equation (32), and the last equality follows from Proposition 9.3. Next, because of Equation (62), we know that there is a probability mass function $p \in \mathscr{M}$ such that $\sum_{x \in \mathscr{X}} p(x)[L_0^s f](x) \le \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f] + \varepsilon$. Let $\mathscr{M}^* \coloneqq \{p\}$. Proposition 8.2 then implies that there is a well-behaved Markov chain $P^* \in \mathbb{P}_{\mathscr{Q},\mathscr{M}^*}^{\mathsf{WM}} \subseteq \mathbb{P}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$ such that

$$\mathbb{E}_{P^*}[f(X_s) \mid X_0] \le [L_0^s f](X_0) + \varepsilon.$$

It now follows from the definition of $\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}$ that

$$\underline{\mathbb{E}}_{\mathscr{Q},\mathscr{M}}^{\mathsf{WM}}[f(X_s)] \le \mathbb{E}_{P^*}[f(X_s)] = \mathbb{E}_{P^*}[\mathbb{E}_{P^*}[f(X_s)|X_0]] \le \mathbb{E}_{P^*}[[L_0^s f](X_0)] + \varepsilon,$$

using Equation (34) for the equality. Furthermore, since $P^* \in \mathbb{P}^{WM}_{\mathscr{Q},\mathscr{M}^*}$ implies that $P^*(X_0 = x) = p(x)$ for all $x \in \mathscr{X}$, we also know that

$$\mathbb{E}_{P^*}[[L_0^s f](X_0)] = \sum_{x \in \mathscr{X}} P^*(X_0 = x)[L_0^s f](x) = \sum_{x \in \mathscr{X}} p(x)[L_0^s f](x) \leq \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f] + \varepsilon.$$

Combining all of the above inequalities, we find that

$$\underline{\mathbb{E}}^{WM}_{\mathscr{Q},\mathscr{M}}[f(X_s)] - 2\varepsilon \leq \mathbb{E}_{P^*}[[L_0^s f](X_0)] - \varepsilon \leq \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f] \leq \underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_s)] + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies that $\underline{\mathbb{E}}^{WM}_{\mathscr{Q},\mathscr{M}}[f(X_s)] \leq \underline{\mathbb{E}}_{\mathscr{M}}[L_0^s f] \leq \underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_s)]$. The result now follows from Proposition 6.4. □

**Proposition 9.5.** *Let $\mathscr{M}$ be a non-empty set of probability mass functions on $\mathscr{X}$ and let $\mathscr{Q}$ be a non-empty, bounded, convex set of rate matrices that has separately specified rows, with lower transition rate operator $\underline{Q}$. Then for any $u = t_0, \ldots, t_n$ in $\mathscr{U}_\emptyset$ such that $t_0 = 0$, and any $f \in \mathscr{L}(\mathscr{X}_u)$:*

$$\underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_u)] = \underline{\mathbb{E}}_{\mathscr{M}}[L_{t_0}^{t_1} L_{t_1}^{t_2} \cdots L_{t_{n-1}}^{t_n} f].$$

**Proof.** Consider any $u = t_0, \ldots, t_n$ in $\mathscr{U}_\emptyset$ such that $t_0 = 0$ and any $f \in \mathscr{L}(\mathscr{X}_u)$. Let $\hat{f} \in \mathscr{L}(\mathscr{X})$ be defined by

$$\hat{f}(x) \coloneqq \underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_u) \mid X_0 = x] = \left[ L_{t_0}^{t_1} L_{t_1}^{t_2} \cdots L_{t_{n-1}}^{t_n} f \right](x) \quad \text{for all } x \in \mathscr{X}, \tag{G.2}$$

using Corollary 9.2 to establish the last equality. It then follows from Theorem 6.5 and Proposition 9.3 that

$$\underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_u)] = \underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[\underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[f(X_u) \mid X_0]] = \underline{\mathbb{E}}^{W}_{\mathscr{Q},\mathscr{M}}[\hat{f}(X_0)] = \underline{\mathbb{E}}_{\mathscr{M}}[\hat{f}].$$

Combined with Equation (G.2), this implies the result. □

## References

[1] S. Asmussen, Applied Probability and Queues, Springer Science & Business Media, 2008.
[2] G. Bolch, S. Greiner, H. de Meer, K.S. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, John Wiley & Sons, 2006.
[3] R.J. Elliott, J. van der Hoek, Default times in a continuous time Markov chain economy, Appl. Math. Finance 20 (5) (2013) 450–460.
[4] T. Rolski, H. Schmidli, V. Schmidt, J. Teugels, Stochastic Processes for Insurance and Finance, John Wiley & Sons, 2009.
[5] J. Sass, U.G. Haussmann, Optimizing the terminal wealth under partial information: the drift process as a continuous time Markov chain, Finance Stoch. 8 (4) (2004) 553–577.
[6] S.W. Duffy, H.-H. Chen, L. Tabar, N.E. Day, Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase, Stat. Med. 14 (14) (1995) 1531–1543.
[7] C.H. Jackson, L.D. Sharples, S.G. Thompson, S.W. Duffy, E. Couto, Multistate Markov models for disease progression with classification error, J. R. Stat. Soc., Ser. D, Stat. 52 (2) (2003) 193–209.
[8] P. Lemey, M. Suchard, A. Rambaut, Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1pdm, PLoS Curr. Influenza (2009).
[9] F. Besnard, L. Bertling, An approach for condition-based maintenance optimization applied to wind turbine blades, IEEE Trans. Sustain. Energy 1 (2) (2010) 77–83.
[10] S.S. Gokhale, M.R. Lyu, K.S. Trivedi, Analysis of software fault removal policies using a non-homogeneous continuous time Markov chain, Softw. Qual. J. 12 (3) (2004) 211–230.
[11] D. Wang, K.S. Trivedi, Reliability analysis of phased-mission system with independent component repairs, IEEE Trans. Reliab. 56 (3) (2007) 540–551.
[12] G. Yin, Q. Zhang, Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach, Springer Science & Business Media, 2012.
[13] D. Škulj, Efficient computation of the bounds of continuous time imprecise Markov chains, Appl. Math. Comput. 250 (C) (2015) 165–180.
[14] M.C.M. Troffaes, J. Gledhill, D. Škulj, S. Blake, Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair, in: ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications, 2015, pp. 287–294.
[15] A. Rindos, S. Woolet, I. Viniotis, K. Trivedi, Exact methods for the transient analysis of nonhomogeneous continuous time Markov chains, in: Computations with Markov Chains, Springer, 1995, pp. 121–133.
[16] O.O. Aalen, S. Johansen, An empirical transition matrix for non-homogeneous Markov chains based on censored observations, Scand. J. Stat. 5 (3) (1978) 141–150.
[17] J.T. Johnson, G.R. Luecke, Nonhomogeneous, continuous-time Markov chains defined by series of proportional intensity matrices, Stoch. Process. Appl. 32 (1) (1989) 171–181.
[18] B. Harlamov, Continuous Semi-Markov Processes, John Wiley & Sons, New York, 2008.
[19] D. Insua, F. Ruggeri, M. Wiper, Bayesian Analysis of Stochastic Process Models, John Wiley & Sons, 2012.
[20] X. Guo, O. Hernández-Lerma, Continuous-Time Markov Decision Processes: Theory and Applications, Stochastic Modelling and Applied Probability, vol. 62, Springer, Berlin, 2009.
[21] X. Guo, O. Hernández-Lerma, Continuous-time controlled Markov chains, Ann. Appl. Probab. 13 (1) (2003) 363–388.
[22] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.
[23] M.C.M. Troffaes, G. de Cooman, Lower Previsions, Wiley, 2014.
[24] T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes (Eds.), Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.
[25] C. Rottondi, A. Erreygers, G. Verticale, J. De Bock, Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains, in: Proceedings of DRCN 2017, 2017, pp. 39–46.

[26] T. Krak, J. De Bock, A. Siebes, Imprecise continuous-time Markov chains, Extended ArXiv tech-report arXiv:1611.05796.

[27] J. De Bock, The limit behaviour of continuous-time imprecise Markov chains, J. Nonlinear Sci. 27 (1) (2017) 159–196.

[28] J.R. Norris, Markov Chains, Cambridge University Press, 1998.

[29] C.F. Van Loan, A Study of the Matrix Exponential, Numerical Analysis Report No. 10, University of Manchester, Manchester, UK, August 1975, Reissued as MIMS EPrint 2006.397, Manchester Institute for Mathematical Sciences, The University of Manchester, UK.

[30] L.E. Dubins, Finitely additive conditional probabilities, conglomerability and disintegrations, Ann. Probab. 3 (1) (1975) 89–99.

[31] P. Berti, E. Regazzini, P. Rigo, Coherent statistical inference and Bayes theorem, Ann. Stat. 19 (1) (1991) 366–381.

[32] B. De Finetti, Theory of Probability, John Wiley & Sons, New York, 1974.

[33] E. Regazzini, Finitely additive conditional probabilities, Rend. Semin. Mat. Fis. Milano 55 (1) (1985) 69–89.

[34] P.M. Williams, Notes on Conditional Previsions, Tech. rep. School of Mathematical and Physical Science, University of Sussex, 1975.

[35] P.M. Williams, Notes on conditional previsions, Int. J. Approx. Reason. 44 (3) (2007) 366–383.

[36] P. Berti, P. Rigo, On coherent conditional probabilities and disintegrations, Ann. Math. Artif. Intell. 35 (1–4) (2002) 71–82.

[37] P. Vicig, M. Zaffalon, F.G. Cozman, Notes on "Notes on conditional previsions", Int. J. Approx. Reason. 44 (3) (2007) 358–365.

[38] L.C.G. Rogers, D. Williams, Diffusions, Markov Processes, and Martingales. Volume 1. Foundations, Cambridge Mathematical Library, Cambridge University Press, 2000, http://opac.inria.fr/record=b1099160.

[39] T.M. Liggett, Continuous Time Markov Processes: An Introduction, American Mathematical Society, 2010, https://books.google.be/books?id=ClW_cKYSON0C.

[40] A. Goldsztejn, A. Neumaier, On the exponentiation of interval matrices, Reliab. Comput. 20 (2014) 52–72.

[41] E.P. Oppenheimer, A.N. Michel, Application of interval analysis techniques to linear systems. II. The interval matrix exponential function, IEEE Trans. Circuits Syst. 35 (10) (1988) 1230–1242.

[42] P.J. Huber, Robust Statistics, John Wiley & Sons, New York, 1981.

[43] J.C.F. da Rocha, F.G. Cozman, Inference with Separately Specified Sets of Probabilities in Credal Networks, Morgan Kaufmann Publishers Inc., 2002.

[44] F.G. Cozman, Credal networks, Artif. Intell. 120 (2) (2000) 199–233.

[45] J. De Bock, Credal Networks Under Epistemic Irrelevance: Theory and Algorithms, Ph.D. thesis, Ghent University, 2015.

[46] G. de Cooman, F. Hermans, A. Antonucci, M. Zaffalon, Epistemic irrelevance in credal nets: the case of imprecise Markov trees, Int. J. Approx. Reason. 51 (9) (2010) 1029–1052.

[47] A. Antonucci, C.P. de Campos, M. Zaffalon, Probabilistic graphical models, in: T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes (Eds.), Introduction to Imprecise Probabilities, John Wiley & Sons, Chichester, 2014, pp. 207–229.

[48] T. Krak, J. De Bock, A. Siebes, Efficient computation of updated lower expectations for imprecise continuous-time hidden Markov chains, in: ISIPTA '17: Proceedings of the Tenth International Symposium on Imprecise Probability: Theory and Applications, in: PMLR: Proceedings of Machine Learning Research, vol. 62, SIPTA, 2017, pp. 193–204.

[49] S. Lopatatzidis, J. De Bock, G. de Cooman, S. Vuyst, J. Walraevens, Robust queueing theory: an initial study using imprecise probabilities, Queueing Syst. 82 (1) (2015) 75–101, http://dx.doi.org/10.1007/s11134-015-9458-6.

[50] A. Benavoli, M. Zaffalon, E. Miranda, Robust filtering through coherent lower previsions, IEEE Trans. Autom. Control 56 (7) (2011) 1567–1581.

[51] J. De Bock, G. de Cooman, An efficient algorithm for estimating state sequences in imprecise hidden Markov models, J. Artif. Intell. Res. 50 (2014) 189–233.

[52] H.L. Royden, P.M. Fitzpatrick, Real Analysis, 4th edition, Prentice Hall, 2010.