Imprecise Continuous-Time Markov Chains: Efficient Computational Methods with Guaranteed Error Bounds

Alexander Erreygers ALEXANDER.ERREYGERS@UGENT.BE
Ghent University, SMACS Research Group (Belgium)

Jasper De Bock JASPER.DEBOCK@UGENT.BE
Ghent University - imec, IDLab, ELIS (Belgium)

Abstract

Imprecise continuous-time Markov chains are a robust type of continuous-time Markov chains that allow for partially specified time-dependent parameters. Computing inferences for them requires the solution of a non-linear differential equation. As there is no general analytical expression for this solution, efficient numerical approximation methods are essential to the applicability of this model. We here improve the uniform approximation method of Krak et al. (2016) in two ways and propose a novel and more efficient adaptive approximation method. For ergodic chains, we also provide a method that allows us to approximate stationary distributions up to any desired maximal error.

Keywords: imprecise continuous-time Markov chain; lower transition operator; lower transition rate operator; approximation method; ergodicity; coefficient of ergodicity.

1. Introduction

Markov chains are a popular type of stochastic processes that can be used to model a variety of systems with uncertain dynamics, both in discrete and continuous time. In many applications, however, the core assumption of a Markov chain—i.e., the Markov property—is not entirely justified. Moreover, it is often difficult to exactly determine the parameters that characterise the Markov chain. In an effort to handle these modelling errors in an elegant manner, several authors have recently turned to imprecise probabilities (Škulj and Hable, 2013; Hermans and de Cooman, 2012; Škulj, 2015; Krak et al., 2016; De Bock, 2017).

As Krak et al. (2016) thoroughly demonstrate, making inferences about an imprecise continuous-time Markov chain—determining lower and upper expectations or probabilities—requires the solution of a non-linear vector differential equation. To the best of our knowledge, this differential equation cannot be solved analytically, at least not in general. Krak et al. (2016) proposed a method to numerically approximate the solution of the differential equation, and argued that it outperforms the approximation method that Škulj (2015) previously introduced. One of the main results of this contribution is a novel approximation method that outperforms that of Krak et al. (2016).

An important property—both theoretically and practically—of continuous-time Markov chains is the behaviour of the solution of the differential equation as the time parameter recedes to infinity. If regardless of the initial condition the solution converges, we say that the chain is ergodic. We show that in this case the approximation is guaranteed to converge as well. This constitutes the second main result of this contribution and serves as a motivation behind the novel approximation method. Furthermore, we also quantify a worst-case convergence rate for the approximation. This unites the work of Škulj (2015), who studied the rate of convergence for discrete-time Markov chains,

and De Bock (2017), who studied the ergodic behaviour of continuous-time Markov chains from a qualitative point of view. One of the uses of our worst-case convergence rate is that it allows us to approximate the limit value of the solution up to a guaranteed error.

In order to comply with the page limit, we do not provide any proofs for our statements. We refer the interested reader to the appendix of (Erreygers and De Bock, 2017), an extended version of this contribution that is available on arXiv.

2. Mathematical Preliminaries

Throughout this contribution, we denote the set of real, non-negative real and strictly positive real numbers by $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$, respectively. The set of natural numbers is denoted by $\mathbb{N}$, if we include zero we write $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For any set S, we let $|S|$ denote its cardinality. If a and b are two real numbers, we say that a is lower (greater) than b if $a \leq b$ ($a \geq b$), and that a is strictly lower (greater) than b if $a < b$ ($a > b$).

2.1 Gambles and Norms

We consider a finite *state space* $\mathcal{X}$, and are mainly concerned with real-valued functions on $\mathcal{X}$. All of these real-valued functions on $\mathcal{X}$ are collected in the set $\mathcal{L}(\mathcal{X})$, which is a vector space. If we identify the state space $\mathcal{X}$ with $\{1, \ldots, |\mathcal{X}|\}$, then any function $f \in \mathcal{L}(\mathcal{X})$ can be identified with a vector: for all $x \in \mathcal{X}$, the x-component of this vector is $f(x)$. A special function on $\mathcal{X}$ is the indicator $\mathbb{I}_A$ of an event A. For any $A \subseteq \mathcal{X}$, it is defined for all $x \in \mathcal{X}$ as $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ otherwise. In order not to obfuscate the notation too much, for any $y \in \mathcal{X}$ we write $\mathbb{I}_y$ instead of $\mathbb{I}_{\{y\}}$. If it is required from the context, we will also identify the real number $\gamma \in \mathbb{R}$ with the map γ from $\mathcal{X}$ to $\mathbb{R}$, defined as $\gamma(x) = \gamma$ for all $x \in \mathcal{X}$.

We provide the set $\mathcal{L}(\mathcal{X})$ of functions with the standard maximum norm $\|\cdot\|$, defined for all $f \in \mathcal{L}(\mathcal{X})$ as $\|f\| := \max\{|f(x)| : x \in \mathcal{X}\}$. A seminorm that captures the variation of $f \in \mathcal{L}(\mathcal{X})$ will also be of use; we therefore define the variation seminorm $\|f\|_v := \max f - \min f$. Since the value $\|f\|_v / 2$ occurs often in formulas, we introduce the shorthand notation $\|f\|_c := \|f\|_v / 2$.

2.2 Non-Negatively Homogeneous Operators

An operator A that maps $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{X})$ is *non-negatively homogeneous* if for all $\mu \in \mathbb{R}_{\geq 0}$ and all $f \in \mathcal{L}(\mathcal{X})$, $A(\mu f) = \mu A f$. The maximum norm $\|\cdot\|$ for functions induces an operator norm:

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

If for all $\mu \in \mathbb{R}$ and all $f, g \in \mathcal{L}(\mathcal{X})$, $A(\mu f + g) = \mu A f + Ag$, then the operator A is *linear*. In that case, it can be identified with a matrix of dimension $|\mathcal{X}| \times |\mathcal{X}|$, the (x, y)-component of which is $[A\mathbb{I}_y](x)$. The identity operator I is an important special case, defined for all $f \in \mathcal{L}(\mathcal{X})$ as $If := f$.

Two types of non-negatively homogeneous operators play a vital role in the theory of imprecise Markov chains: lower transition operators and lower transition rate operators.

Definition 1 *An operator $\underline{T}$ from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{X})$ is called a* lower transition operator *if for all $f \in \mathcal{L}(\mathcal{X})$ and all $\mu \in \mathbb{R}_{\geq 0}$:*

$$\textit{L1: } \underline{T}f \geq \min f; \qquad \textit{L2: } \underline{T}(f+g) \geq \underline{T}f + \underline{T}g; \qquad \textit{L3: } \underline{T}(\mu f) = \mu \underline{T} f.$$

Every lower transition operator $\underline{T}$ has a conjugate upper transition operator $\overline{T}$, defined for all $f \in \mathcal{L}(\mathcal{X})$ as $\overline{T}f := -\underline{T}(-f)$.

Definition 2 *An operator $\underline{Q}$ from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{X})$ is called a* lower transition rate operator *if for any $f, g \in \mathcal{L}(\mathcal{X})$, any $\mu \in \mathbb{R}_{\geq 0}$, any $\gamma \in \mathbb{R}$ and any $x, y \in \mathcal{X}$ such that $x \neq y$:*

$$\textit{R1: } \underline{Q}\gamma = 0; \qquad \textit{R2: } \underline{Q}(f+g) \geq \underline{Q}f + \underline{Q}g; \qquad \textit{R3: } \underline{Q}(\mu f) = \mu \underline{Q}f; \qquad \textit{R4: } [\underline{Q}\mathbb{I}_x](y) \geq 0.$$

The conjugate lower transition rate operator $\overline{Q}$ is defined for all $f \in \mathcal{L}(\mathcal{X})$ as $\overline{Q}f := -\underline{Q}(-f)$.

As will become clear in Section 3, lower transition operators and lower transition rate operators are tightly linked. For instance, we can use a lower transition rate operator to construct a lower transition operator. One way is to use Eqn. (1) further on. Another one is given in the following proposition, which is a strengthened version of (De Bock, 2017, Proposition 5).

Proposition 3 *Consider any lower transition rate operator $\underline{Q}$ and any $\delta \in \mathbb{R}_{\geq 0}$. Then the operator $(I + \delta\underline{Q})$ is a lower transition operator if and only if $\delta \left\|\underline{Q}\right\| \leq 2$.*

We end this section with the first—although minor—novel result of this contribution. The norm of a lower transition rate operator is essential for all the approximation methods that we will discuss. The following proposition supplies us with an easy formula for determining it.

Proposition 4 *Let $\underline{Q}$ be a lower transition rate operator. Then $\left\|\underline{Q}\right\| = 2\max\{\left|[\underline{Q}\mathbb{I}_x](x)\right| : x \in \mathcal{X}\}$.*

Example 1 *Consider a binary state space $\mathcal{X} = \{0, 1\}$ and two closed intervals $[\underline{q}_0, \overline{q}_0] \subset \mathbb{R}_{\geq 0}$ and $[\underline{q}_1, \overline{q}_1] \subset \mathbb{R}_{\geq 0}$. Let*

$$\underline{Q}f := \min\left\{\begin{bmatrix} q_0(f(1) - f(0)) \\ q_1(f(0) - f(1)) \end{bmatrix} : q_0 \in [\underline{q}_0, \overline{q}_0], q_1 \in [\underline{q}_1, \overline{q}_1]\right\} \textit{ for all } f \in \mathcal{L}(\mathcal{X}).$$

Then one can easily verify that $\underline{Q}$ is a lower transition rate operator.

Krak et al. (2016) also consider a running example with a binary state space, but they let $\mathcal{X} := \{\texttt{healthy}, \texttt{sick}\}$. We here identify `healthy` *with* 0 *and* `sick` *with* 1. *In (Krak et al., 2016, Example 18), they propose the following values for the transition rates: $[\underline{q}_0, \overline{q}_0] := [1/52, 3/52]$ and $[\underline{q}_1, \overline{q}_1] := [1/2, 2]$. It takes Krak et al. a lot of work to determine the exact value of the norm of $\underline{Q}$, see (Krak et al., 2016, Example 19). We simply use Proposition 4: $\left\|\underline{Q}\right\| = 2\max\{3/52, 2\} = 4$.*

3. Imprecise Continuous-Time Markov Chains

For any lower transition rate operator $\underline{Q}$ and any $f \in \mathcal{L}(\mathcal{X})$, Škulj (2015) has shown that the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\underline{T}_t f = \underline{Q}\underline{T}_t f. \tag{1}$$

with initial condition $\underline{T}_0 f := f$ has a unique solution for all $t \in \mathbb{R}_{\geq 0}$. Later, De Bock (2017) proved that the time-dependent operator $\underline{T}_t$ itself satisfies a similar differential equation, and that it is a lower transition operator. Finding the unique solution of Eqn. (1) is non-trivial. Fortunately, we can approximate this solution, as by (De Bock, 2017, Proposition 10)

$$\underline{T}_t = \lim_{n\to\infty}\left(I + \frac{t}{n}\underline{Q}\right)^n. \tag{2}$$

Example 2 *In the simple case of Example 1, we can use Eqn. (2) to obtain analytical expressions for the solution of Eqn. (1). Assume that* $\underline{q}_0 + \overline{q}_1 > 0$ *and fix some* $t \in \mathbb{R}_{\geq 0}$*. Then*

$$[\underline{T}_t f](0) = f(0) + \underline{q}_0 h(t) \text{ and } [\underline{T}_t f](1) = f(1) - \overline{q}_1 h(t) \text{ for all } f \in \mathcal{L}(\mathcal{X}) \text{ with } f(0) \leq f(1),$$

where $h(t) := \|f\|_v \, (\underline{q}_0 + \overline{q}_1)^{-1} \big(1 - e^{-t(\underline{q}_0 + \overline{q}_1)}\big)$*. The case* $f(0) \geq f(1)$ *yields similar expressions.*

For a linear lower transition rate operator $\underline{Q}$—i.e., if it is a transition rate matrix Q—Eqn. (2) reduces to the definition of the matrix exponential. It is well-known—see (Anderson, 1991)—that this matrix exponential $T_t = e^{tQ}$ can be interpreted as the transition matrix at time t of a time-homogeneous or stationary continuous-time Markov chain: the (x, y)-component of T_t is the probability of being in state y at time t if the chain started in state x at time 0. Therefore, it follows that the expectation of the function $f \in \mathcal{L}(\mathcal{X})$ at time $t \in \mathbb{R}_{\geq 0}$ conditional on the initial state $x \in \mathcal{X}$, denoted by $\mathrm{E}(f(X_t)|X_0 = x)$, is equal to $[T_t f](x)$.

As Eqn. (2) is a non-linear generalisation of the definition of the matrix exponential, we can interpret $\underline{T}_t$ as the non-linear generalisation of the matrix exponential $T_t = e^{tQ}$. Extending this parallel, we might interpret $\underline{T}_t$ as the non-linear generalisation of the transition matrix—i.e., as the lower transition operator—at time t of a generalised continuous-time Markov chain. In fact, Krak et al. (2016) prove that this is the case. They show that—under some conditions on $\underline{Q}$—$[\underline{T}_t f](x)$ can be interpreted as the tightest lower bound for $\mathrm{E}(f(X_t)|X_0 = x)$ with respect to a set of—not necessarily Markovian—stochastic processes that are consistent with $\underline{Q}$. Krak et al. (2016) argue that, just like a transition rate matrix Q characterises a (precise) continuous-time Markov chain, a lower transition rate operator $\underline{Q}$ characterises a so-called imprecise continuous-time Markov chain.

The main objective of this contribution is to determine $\underline{T}_t f$ for some $f \in \mathcal{L}(\mathcal{X})$ and some $t \in \mathbb{R}_{>0}$. Our motivation is that, from an applied point of view on imprecise continuous-time Markov chains, what one is most interested in are tight lower and upper bounds on expectations of the form $\mathrm{E}(f(X_t)|X_0 = x)$. As explained above, the lower bound is given by $\underline{\mathrm{E}}(f(X_t)|X_0 = x) = [\underline{T}_t f](x)$. Similarly, the upper bound is given by $\overline{\mathrm{E}}(f(X_t)|X_0 = x) = -[\underline{T}_t(-f)](x)$. Note that the lower (or upper) probability of an event $A \subseteq \mathcal{X}$ conditional on the initial state x is a special case of a lower (or upper) expectation: $\underline{\mathrm{P}}(X_t \in A|X_0 = x) = \underline{\mathrm{E}}(\mathbb{I}_A(X_t)|X_0 = x)$ and similarly for the upper probability. Hence, for the sake of generality we can focus on $\underline{T}_t f$ and forget about its interpretation. As in most cases analytically solving Eqn. (1) is infeasible or even impossible, we resort to methods that yield an approximation up to some guaranteed maximal error.

4. Approximation Methods

Škulj (2015) was, to the best of our knowledge, the first to propose methods that approximate the solution $\underline{T}_t f$ of Eqn. (1). He proposes three methods: one with a uniform grid, a second with an adaptive grid and a third that is a combination of the previous two. In essence, he determines a step size δ and then approximates $\underline{T}_{t+\delta} f$ with $e^{\delta Q} \underline{T}_t f$, where Q is a transition rate matrix determined from $\underline{Q}$ and $\underline{T}_t f$. One drawback of this method is that it needs the matrix exponential $e^{\delta Q}$, which—in general—needs to be approximated as well. Škulj (2015) mentions that his methods turn out to be quite computationally heavy, even if the uniform and adaptive methods are combined.

We consider two alternative approximation methods—one with a uniform grid and one with an adaptive grid—that both work in the same way. First, we pick a small step $\delta_1 \in \mathbb{R}_{\geq 0}$ and apply the operator $(I + \delta_1 \underline{Q})$ to the function $g_0 = f$, resulting in a function $g_1 := (I + \delta_1 \underline{Q}) f$. Recall from

Proposition 3 that if we want $(I + \delta_1 \underline{Q})$ to be a lower transition operator, then we need to demand that $\delta_1 \left\|\underline{Q}\right\| \leq 2$. Next, we pick a (possibly different) step $\delta_2 \in \mathbb{R}_{\geq 0}$ such that $\delta_2 \left\|\underline{Q}\right\| \leq 2$ and apply the lower transition operator $(I + \delta_2 \underline{Q})$ to the function g_1, resulting in a function $g_2 := (I + \delta_2 \underline{Q}) g_1$. If we continue this process until the sum of all the small steps is equal to t, then we end up with an approximation for $\underline{T}_t f$. More formally, let $s := (\delta_1, \ldots, \delta_k)$ denote a sequence in $\mathbb{R}_{\geq 0}$ such that, for all $i \in \{1, \ldots, k\}$, $\delta_i \left\|\underline{Q}\right\| \leq 2$. Using this sequence s we define the *approximating lower transition operator*

$$\Phi(s) := (I + \delta_k \underline{Q}) \cdots (I + \delta_1 \underline{Q}).$$

What we are looking for is a convenient way to determine the sequence s such that the error $\|\underline{T}_t f - \Phi(s) f\|$ is guaranteed to be lower than some desired maximal error $\epsilon \in \mathbb{R}_{>0}$.

4.1 Using a Uniform Grid

Krak et al. (2016) provide one way to determine the sequence s. They assume a uniform grid, in the sense that all elements of the sequence s are equal to δ. The step size δ is completely determined by the desired maximal error ϵ, the time t, the variation norm of the function f and the norm of $\underline{Q}$; (Krak et al., 2016, Proposition 8.5) guarantees that the actual error is lower than ϵ. Algorithm 1 provides a slightly improved version of (Krak et al., 2016, Algorithm 1). The improvement is due to Proposition 3: we demand that $n \geq t \left\|\underline{Q}\right\| /2$ instead of $n \geq t \left\|\underline{Q}\right\|$.

Algorithm 1: Uniform approximation

Data: A lower transition rate operator $\underline{Q}$, a function $f \in \mathcal{L}(\mathcal{X})$, a maximal error $\epsilon \in \mathbb{R}_{>0}$, and a time point $t \in \mathbb{R}_{\geq 0}$.

Result: $\underline{T}_t f \pm \epsilon$

$g_0 \leftarrow f$

if $\|f\|_c = 0$ **or** $\left\|\underline{Q}\right\| = 0$ **or** $t = 0$ **then** $(n, \delta) \leftarrow (0, 0)$

else

  $n \leftarrow \lceil \max\{t \left\|\underline{Q}\right\| /2, t^2 \left\|\underline{Q}\right\|^2 \|f\|_c /\epsilon\} \rceil$

  $\delta \leftarrow t/n$

  for $i = 0, \ldots, n-1$ **do**

    $g_{i+1} \leftarrow g_i + \delta \underline{Q} g_i$

return g_n

More formally, for any $t \in \mathbb{R}_{\geq 0}$ and any $n \in \mathbb{N}$ such that $t \left\|\underline{Q}\right\| \leq 2n$, we consider the *uniformly approximating lower transition operator*

$$\Psi_t(n) := \left(I + \frac{t}{n} \underline{Q}\right)^n.$$

As a special case, we define $\Psi_t(0) := I$. The following theorem then guarantees that the choice of n in Algorithm 1 results in an error $\|\underline{T}_t f - \Psi_t(n) f\|$ that is lower than the desired maximal error ϵ.

Theorem 5 *Let $\underline{Q}$ be a lower transition rate operator and fix some $f \in \mathcal{L}(\mathcal{X})$, $t \in \mathbb{R}_{\geq 0}$ and $\epsilon \in \mathbb{R}_{>0}$. If we use Algorithm 1 to determine n, δ and $g_0, \ldots, g_n$, then we are guaranteed that*

$$\|\underline{T}_t f - \Psi_t(n) f\| = \|\underline{T}_t f - g_n\| \leq \epsilon' := \delta^2 \left\|\underline{Q}\right\|^2 \sum_{i=0}^{n-1} \|g_i\|_c \leq \epsilon.$$

Table 1: Comparison of the presented approximation methods, obtained using a naive, unoptimised implementation of the algorithms in Python. N is the total number of iterations, D_ϵ ($D_{\epsilon'}$) is the average duration—in seconds, averaged over 50 independent runs—without (with) keeping track of ϵ', and ϵ_a is the actual error. The Python code is made available at github.com/alexander-e/ictmc.

Method	N	D_ϵ	$D_{\epsilon'}$	$\epsilon' \times 10^3$	$\epsilon_a \times 10^3$
Uniform	8,000	0.0345	0.0574	0.430	0.0335
Uniform	250	0.00171	0.0264	13.8	1.07
Adaptive with $m = 1$	3,437	0.0371	0.0428	1.000	0.108
Adaptive with $m = 20$	3,456	0.0143	0.0254	0.992	0.107
Uniform ergodic with $m = 1$	6,133	0.0264	0.0449	0.560	0.0437

Theorem 5 is an extension of (Krak et al., 2016, Proposition 8.5). We already mentioned that the demand $n \geq t \left\|\underline{Q}\right\|$ can be relaxed to $n \geq t \left\|\underline{Q}\right\| / 2$. Furthermore, it turns out that we can compute an upper bound ϵ' on the error that is (possibly) lower than the desired maximal error ϵ. If we want to determine this ϵ' while running Algorithm 1, we simply need to add $\epsilon' \leftarrow 0$ to line 1 and insert $\epsilon' \leftarrow \epsilon' + \delta^2 \left\|\underline{Q}\right\|^2 \left\|g_i\right\|_c$ just before line 7.

Example 3 *We again consider the simple case of Example 1 and illustrate the use of Theorem 5 with a numerical example based on (Krak et al., 2016, Example 20). Krak et al. (2016) use Algorithm 1 to approximate $\underline{T}_1 \mathbb{I}_1$, and find that $n = 8{,}000$ guarantees an error lower than the desired maximal error $\epsilon := 1 \times 10^{-3}$. As reported in Table 1, we use Theorem 5 to compute ϵ'. We find that $\epsilon' \approx 0.430 \times 10^{-3}$, which is approximately a factor two smaller than the desired maximal error ϵ.*

In this case, since we know the analytical expression for $\underline{T}_1 \mathbb{I}_1$ from Example 2, we can determine the actual error $\epsilon_a = \|\underline{T}_1 \mathbb{I}_1 - \Psi_1(8000)\mathbb{I}_1\|$. Quite remarkably, the actual error is approximately 3.35×10^{-5}, which is roughly 30 times smaller than the desired maximal error. This leads us to think that the number of iterations used by the uniform method is too high. In fact, we find that using as few as 250 iterations—roughly $8{,}000/30$—already results in an actual error that is approximately equal to the desired one: $\|\underline{T}_1 \mathbb{I}_1 - \Psi_1(250)\mathbb{I}_1\| \approx 1.07 \times 10^{-3}$.

4.2 Using an Adaptive Grid

In Example 3, we noticed that the maximal desired error was already satisfied for a uniform grid that was much coarser than that constructed by Algorithm 1. Because of this, we are led to believe that we can find a better approximation method than the uniform method of Algorithm 1.

To this end, we now consider grids where, for some integer m, every m consecutive time steps in the grid are equal. In particular, we consider a sequence $\delta_1, \ldots, \delta_n$ in $\mathbb{R}_{\geq 0}$ and some $k \in \mathbb{N}$ such that $1 \leq k \leq m$ and, for all $i \in \{1, \ldots, n\}$, $\delta_i \left\|\underline{Q}\right\| \leq 2$. From such a sequence, we then construct the *m-fold approximating lower transition operator*:

$$\Phi_{m,k}(\delta_1, \ldots, \delta_n) := (I + \delta_n \underline{Q})^k (I + \delta_{n-1} \underline{Q})^m \cdots (I + \delta_1 \underline{Q})^m,$$

where if $n = 1$ only $(I + \delta_1 \underline{Q})^k$ remains and if $n = 2$ only $(I + \delta_2 \underline{Q})^k (I + \delta_1 \underline{Q})^m$ remains.

The uniform approximation method of before is a special case of the m-fold approximating lower transition operator; a more interesting method to construct an m-fold approximation is Algorithm 2. In this algorithm, we re-evaluate the time step every m iterations, possibly increasing its length.

Algorithm 2: Adaptive approximation

Data: A lower transition rate operator $\underline{Q}$, a gamble $f \in \mathcal{L}(\mathcal{X})$, an integer $m \in \mathbb{N}$, a tolerance $\epsilon \in \mathbb{R}_{>0}$, and a time period $t \in \mathbb{R}_{\geq 0}$.

Result: $\underline{T}_t f \pm \epsilon$

$(g_{(0,m)}, \Delta, i) \leftarrow (f, t, 0)$
if $\|f\|_c = 0$ **or** $\|\underline{Q}\| = 0$ **or** $t = 0$ **then** $(n, k) \leftarrow (0, m)$
else
  while $\Delta > 0$ **and** $\|g_{(i,m)}\|_c > 0$ **do**
    $i \leftarrow i + 1$
    $\delta_i \leftarrow \min\{\Delta, 2/\|\underline{Q}\|, \epsilon/(t\|\underline{Q}\|^2 \|g_{(i-1,m)}\|_c)\}$
    if $m\delta_i > \Delta$ **then**
      $k_i \leftarrow \lceil \Delta/\delta_i \rceil$
      $\delta_i \leftarrow \Delta/k_i$
    else $k_i \leftarrow m$
    $g_{(i,0)} \leftarrow g_{(i-1,m)}, \Delta \leftarrow \Delta - k_i\delta_i$
    for $j = 0, \ldots, k_i - 1$ **do**
      $g_{(i,j+1)} \leftarrow g_{(i,j)} + \delta_i \underline{Q} g_{(i,j)}$
  $(n, k) \leftarrow (i, k_i)$
return $g_{(n,k)}$

From the properties of lower transition operators, it follows that for all $i \in \{2, \ldots, n-1\}$, $\|g_{(i-1,m)}\|_c \leq \|g_{(i-2,m)}\|_c$. Hence, the re-evaluated step size δ_i is indeed larger than (or equal to) the previous step size δ_{i-1}. The only exception to this is the final step size δ_n: it might be that the remaining time Δ is smaller than $m\delta_n$, in which case we need to choose k and δ_n such that $k\delta_n = \Delta$.

Theorem 6 guarantees that the adaptive approximation of Algorithm 2 indeed results in an actual error lower than the desired maximal error ϵ. Even more, it provides a method to compute an upper bound ϵ' of the actual error that is lower than the desired maximal error. Finally, it also states that the adaptive method of Algorithm 2 needs at most an equal number of iterations than the uniform method of Algorithm 1.

Theorem 6 *Let $\underline{Q}$ be a lower transition rate operator, $f \in \mathcal{L}(\mathcal{X})$, $t \in \mathbb{R}_{\geq 0}$, $\epsilon \in \mathbb{R}_{>0}$ and $m \in \mathbb{N}$. We use Algorithm 2 to determine n and k, and if applicable also k_i, δ_i and $g_{(i,j)}$. If $\|f\|_c = 0$, $\|\underline{Q}\| = 0$ or $t = 0$, then $\|\underline{T}_t f - g_{(n,k)}\| = 0$. Otherwise, we are guaranteed that*

$$\|\underline{T}_t f - \Phi_{m,k}(\delta_1 \ldots, \delta_n) f\| = \|\underline{T}_t f - g_{(n,k)}\| \leq \epsilon' := \sum_{i=1}^{n} \delta_i^2 \|\underline{Q}\|^2 \sum_{j=0}^{k_i - 1} \|g_{(i,j)}\|_c \leq \epsilon$$

and that the total number of iterations has an upper bound:

$$\sum_{i=1}^{n} k_i = (n-1)m + k \leq \left\lceil \max\left\{ \|\underline{Q}\| t/2, t^2 \|\underline{Q}\|^2 \|f\|_c / \epsilon \right\} \right\rceil.$$

Again, we can determine ϵ' while running Algorithm 2. An alternate—less tight—version of ϵ' can be obtained by replacing the sum of $\left\|g_{(i,j)}\right\|_c$ for j from 0 to $k_i - 1$ by $k_i \left\|g_{(i,0)}\right\|_c = k_i \left\|g_{(i-1,m)}\right\|_c$. Determining this alternative ϵ' while running Algorithm 2 adds negligible computational overhead compared to the ϵ' of Theorem 6, as $\left\|g_{(i-1,m)}\right\|_c$ is needed to re-evaluate the step size anyway.

The reason why we only re-evaluate the step size δ after every m iterations is twofold. First and foremost, all we currently know for sure is that for all $\delta \in \mathbb{R}_{\geq 0}$ such that $\delta \left\|\underline{Q}\right\| \leq 2$, all $m \in \mathbb{N}$ and all $f \in \mathcal{L}(\mathcal{X})$, $\left\|(I + \delta\underline{Q})^m f\right\|_c \leq \|f\|_c$. Re-evaluating the step size every m iterations is therefore only justified if a priori we are certain that $\left\|(I + \delta_i\underline{Q})^m g_{(i-1,m)}\right\|_c < \left\|g_{(i-1,m)}\right\|_c$. We come back to this in Section 5. A second reason is that there might be a trade-off between the time it takes to re-evaluate the step size and the time that is gained by the resulting reduction of the number of iterations. The following numerical example illustrates this trade off.

Example 4 *Recall that in Example 3 we wanted to approximate $\underline{T}_1\mathbb{I}_1$ up to a maximal desired error $\epsilon = 1 \times 10^{-3}$. Instead of using the uniform method of Algorithm 1, we now use the adaptive method of Algorithm 2 with $m = 1$. The initial step size is the same as that of the uniform method, but because we re-evaluate the step size we only need $3{,}437$ iterations, as reported in Table 1. We also find that in this case $\epsilon' = 1.00 \times 10^{-3}$, which is a coincidence. Nevertheless, the actual error of the approximation is 0.108×10^{-3}, which is about ten times smaller than what we were aiming for.*

However, fewer iterations do not necessarily imply a shorter duration of the computations. Qualitatively, we can conclude the following from Table 1. First, keeping track of ϵ' increases the duration, as expected. Second, the adaptive method is faster than the uniform method, at least if we choose m large enough. And third, both methods yield an actual error that is at least an order of magnitude lower than the desired maximal error.

5. Ergodicity

Let $\Phi_{m,k}(\delta_1, \ldots, \delta_n)f$ be an approximation constructed using the adaptive method of Algorithm 2. Re-evaluating the step size is then only justified if a priori we are sure that

$$1/2 \left\|(I + \delta_i\underline{Q})^m \Phi_{i-1} f\right\|_v = \left\|g_{(i,m)}\right\|_c < \left\|g_{(i-1,m)}\right\|_c = 1/2 \left\|\Phi_{i-1} f\right\|_v \text{ for all } i \in \{1, \ldots, n-1\},$$

where $\Phi_0 \coloneqq I$ and $\Phi_i \coloneqq (I + \delta_i\underline{Q})^m \Phi_{i-1}$. As $(\Phi_{i-1}f) \in \mathcal{L}(\mathcal{X})$, this is definitely true if we require that

$$(\forall \delta \in \{\delta_1, \ldots, \delta_{n-1}\})(\forall f \in \mathcal{L}(\mathcal{X})) \ \left\|(I + \delta\underline{Q})^m f\right\|_v < \|f\|_v . \tag{3}$$

In fact, since this inequality is invariant under translation or positive scaling of f, it suffices if

$$(\forall \delta \in \{\delta_1, \ldots, \delta_{n-1}\})(\forall f \in \mathcal{L}(\mathcal{X}) \colon 0 \leq f \leq 1) \ \left\|(I + \delta\underline{Q})^m f\right\|_v < 1.$$

Readers that are familiar with (the ergodicity of) imprecise discrete-time Markov chains—see (Hermans and de Cooman, 2012) or (Škulj and Hable, 2013)—will probably recognise this condition, as it states that the (weak) coefficient of ergodicity of $(I + \delta\underline{Q})^m$ should be strictly smaller than 1. For all lower transition operators $\underline{T}$, Škulj and Hable (2013) define this (weak) *coefficient of ergodicity* as

$$\rho(\underline{T}) \coloneqq \max\left\{\|\underline{T}f\|_v : f \in \mathcal{L}(\mathcal{X}), 0 \leq f \leq 1\right\}. \tag{4}$$

5.1 Ergodicity of Lower Transition Rate Operators

As will become apparent, whether or not combinations of $m \in \mathbb{N}$ and $\delta \in \mathbb{R}_{\geq 0}$ exist such that $\delta \left\|\underline{Q}\right\| \leq 2$ and $\rho((I + \delta\underline{Q})^m) < 1$ is tightly connected with the behaviour of $\underline{T}_t f$ for large t. De Bock (2017) proved that for all lower transition rate operator $\underline{Q}$ and all $f \in \mathcal{L}(\mathcal{X})$, the limit $\lim_{t\to\infty} \underline{T}_t f$ exists. An important case is when this limit is a constant function for all f.

Definition 7 (Definition 2 of (De Bock, 2017)) *The lower transition rate operator $\underline{Q}$ is* ergodic *if for all $f \in \mathcal{L}(\mathcal{X})$, $\lim_{t\to\infty} \underline{T}_t f$ exists and is a constant function.*

As shown by De Bock (2017), ergodicity is easily verified in practice: it is completely determined by the signs of $[\overline{Q}\mathbb{I}_x](y)$ and $[\underline{Q}\mathbb{I}_A](z)$, for all $x, y \in \mathcal{X}$ and certain combinations of $z \in \mathcal{X}$ and $A \subset \mathcal{X}$. It turns out that an ergodic lower transition rate operator $\underline{Q}$ does not only induce a lower transition operator $\underline{T}_t$ that converges, it also induces discrete approximations—of the form $(I + \delta_k\underline{Q}) \cdots (I + \delta_1\underline{Q})$—with special properties. The following theorem, which we consider to be one of the main results of this contribution, highlights this.

Theorem 8 *The lower transition rate operator $\underline{Q}$ is ergodic if and only if there is some $n < |\mathcal{X}|$ such that $\rho(\Phi(\delta_1, \ldots, \delta_k)) < 1$ for one (and then all) $k \geq n$ and one (and then all) sequence(s) $\delta_1, \ldots, \delta_k$ in $\mathbb{R}_{>0}$ such that $\delta_i \left\|\underline{Q}\right\| < 2$ for all $i \in \{1, \ldots, k\}$.*

5.2 Ergodicity and the Uniform Approximation Method

Theorem 8 guarantees that the conditions that were discussed at the beginning of this section are satisfied. In particular, if the lower transition rate operator is ergodic, then there is some $n < |\mathcal{X}|$ such that $\rho((I + \delta\underline{Q})^m) < 1$ for all $m \geq n$ and all $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\|\underline{Q}\right\| < 2$. Consequently, if we choose $m \geq |\mathcal{X}| - 1$ then re-evaluating the step size δ will—except maybe for the last re-evaluation—result in a new step size that is strictly greater than the previous one. Therefore, we conclude that if the lower transition rate operator is ergodic, then using the adaptive method of Algorithm 2 is certainly justified; it will result in fewer iterations, provided we choose a large enough m.

Another nice consequence of the ergodicity of a lower transition rate operator $\underline{Q}$ is that we can prove an alternate a priori guaranteed upper bound for the error of uniform approximations.

Proposition 9 *Let $\underline{Q}$ be a lower transition rate operator and fix some $f \in \mathcal{L}(\mathcal{X})$, $m, n \in \mathbb{N}$ and $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\|\underline{Q}\right\| < 2$. If $\beta \coloneqq \rho((I + \delta\underline{Q})^m) < 1$, then*

$$\|\underline{T}_t f - \Psi_t(n)\| \leq \epsilon_e \coloneqq m\delta^2 \left\|\underline{Q}\right\|^2 \|f\|_c \frac{1 - \beta^k}{1 - \beta} \leq \epsilon_d \coloneqq \frac{m\delta^2 \left\|\underline{Q}\right\|^2 \|f\|_c}{1 - \beta},$$

where $t \coloneqq n\delta$ and $k \coloneqq \lceil n/m \rceil$. The same is true for $\beta = \rho(\underline{T}_{m\delta})$.

Interestingly enough, the upper bound ϵ_d is not dependent on t (or n) at all! This is a significant improvement on the upper bound of Theorem 5, as that upper bound is proportional to t^2.

By Theorem 8, there always is an $m < |\mathcal{X}|$ such that $\rho((I + \delta\underline{Q})^m) < 1$ for all $\delta \in \mathbb{R}_{>0}$ such that $\delta \left\|\underline{Q}\right\| < 2$. Thus, given such an m, we can easily improve Algorithm 1. After we have determined n and δ with Algorithm 1, we can simply determine the upper bound of Proposition 9. If $m(1 - \beta^k) < n(1 - \beta)$ (or $m < n(1 - \beta)$), then this upper bound is smaller than the desired maximal error ϵ, and we have found a tighter upper bound on the actual error. We can even go the extra mile and replace line 4 with a method that looks for the smallest possible $n \in \mathbb{N}$ that yields

$$m\delta^2 \left\|\underline{Q}\right\|^2 \|f\|_c (1-\beta^k) \leq (1-\beta)\epsilon,$$

where $k = \lceil n/m \rceil$ and $\delta = t/n$—and therefore also β—are dependent of n. This method could yield a smaller n, but the time we gain by having to execute fewer iterations does not necessarily compensate the time lost by looking for a smaller n. In any case, to actually implement these improvements we need to be able to compute $\beta := \rho((I+\delta\underline{Q})^m)$.

Example 5 *For the simple case of Example 1, we can derive an analytical expression for $\rho((I+\delta\underline{Q}))$ that is valid for all $\delta \in \mathbb{R}_{\geq 0}$ such that $\delta \left\|\underline{Q}\right\| \leq 2$. Therefore, we can use Proposition 9 to a priori determine an upper bound for the error. If we choose $m = 1$, then $\epsilon_e = 0.767 \times 10^{-3}$ and $\epsilon_d = 1.79 \times 10^{-3}$. Note that $\epsilon_e < \epsilon$, so we can probably decrease the number of iterations n. As reported in Table 1, we find that $n = 6{,}133$ still suffices, and that this results in an approximation correct up to $\epsilon' = 0.560 \times 10^{-3}$, roughly two times smaller than the desired maximal error ϵ. The actual error is 0.0437×10^{-3}, roughly ten times smaller than ϵ.*

5.3 Approximating the Coefficient of Ergodicity

Unfortunately, determining the exact value of $\rho((I+\delta\underline{Q})^m)$—and of $\rho(\underline{T})$ in general—turns out to be non-trivial and is often even impossible. Nevertheless, the following theorem gives some—actually computable—lower and upper bounds for the coefficient of ergodicity.

Theorem 10 *Let $\underline{T}$ be a lower transition operator. Then*

$$\rho(\underline{T}) \leq \max\left\{ \max\{[\overline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon x, y \in \mathcal{X}\} \colon \emptyset \neq A \subset \mathcal{X}\right\}, \tag{5}$$

$$\rho(\underline{T}) \geq \max\left\{ \max\{[\underline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon x, y \in \mathcal{X}\} \colon \emptyset \neq A \subset \mathcal{X}\right\}. \tag{6}$$

The upper bound in Theorem 10 is particularly useful in combination with Proposition 9, as it allows us to replace $\beta := \rho((I+\delta\underline{Q})^m)$ with a guaranteed upper bound. Of course, this only makes sense if this upper bound is strictly smaller than one. The following proposition guarantees that, for ergodic lower transition rate operators $\underline{Q}$, this is always the case.

Proposition 11 *Let $\underline{Q}$ be an ergodic lower transition rate operator. Then there is some $n < |\mathcal{X}|$ such that, for all $k \geq n$ and $\delta_1, \ldots, \delta_k$ in $\mathbb{R}_{>0}$ such that $\delta_i \left\|\underline{Q}\right\| < 2$ for all $i \in \{1, \ldots, k\}$, the upper bound for $\rho(\Phi(\delta_1, \ldots, \delta_k))$ that is given by Eqn. (5) is strictly smaller than one.*

5.4 Approximating Limit Values

The results that we have obtained earlier in this section naturally lead to a method to approximate $\underline{T}_\infty f := \lim_{t\to\infty} \underline{T}_t f$ up to some maximal error. This is an important problem in applications; for instance, Troffaes et al. (2015) try to determine $\underline{T}_\infty f$ for an ergodic lower transition rate operator that arises in their specific reliability analysis application. The method they use is rather ad hoc: they pick some t and n and then determine the uniform approximation $\Psi_t(n)f$. As $\|\Psi_t(n)f\|_v$ is small, they suspect that they are close to the actual limit value. They also observe that $\Psi_{2t}(4n)f$ only differs from $\Psi_t(n)f$ after the fourth significant digit, which they regard as further empirical evidence for the correctness of their approximation. While this ad hoc method seemingly works, the initial values for t and n have to be chosen somewhat arbitrarily. Also, this method provides no guarantee that the actual error is lower than some desired maximal error.

Theorem 8, Proposition 9, Theorem 10 and the following stopping criterion allow us to propose a method that corrects these two shortcomings.

Proposition 12 *Let $\underline{Q}$ be an ergodic lower transition rate operator and let $f \in \mathcal{L}(\mathcal{X})$, $t \in \mathbb{R}_{\geq 0}$ and $\epsilon \in \mathbb{R}_{>0}$. Let s denote a sequence $\delta_1, \ldots, \delta_k$ in $\mathbb{R}_{\geq 0}$ such that $\sum_{i=1}^{k} \delta_i = t$ and, for all $i \in \{1, \ldots, k\}$, $\delta_i \left\|\underline{Q}\right\| \leq 2$. If $\|\underline{T}_t f - \Phi(s) f\| \leq \epsilon/2$ and $\|\Phi(s) f\|_c \leq \epsilon/2$, then for all $\Delta \in \mathbb{R}_{\geq 0}$:*

$$\left|\underline{T}_{t+\Delta} f - \frac{\max \Phi(s) + \min \Phi(s)}{2}\right| \leq \epsilon \quad \text{and} \quad \left|\underline{T}_{\infty} f - \frac{\max \Phi(s) + \min \Phi(s)}{2}\right| \leq \epsilon.$$

Without actually stating it, we mention that a similar—though less useful—stopping criterion can be proved for non-ergodic transition rate matrices as well.

Our method for determining $\underline{T}_{\infty} f$ is now relatively straightforward. Let $\underline{Q}$ be an ergodic lower transition rate operator and fix some $f \in \mathcal{L}(\mathcal{X})$. We can then approximate $\underline{T}_{\infty} f$ up to any desired maximal error $\epsilon \in \mathbb{R}_{>0}$ as follows. First, we look for some $m \in \mathbb{N}$ and some—preferably large—$\delta \in \mathbb{R}_{>0}$ such that $\delta \left\|\underline{Q}\right\| < 2$ and

$$2m\delta^2 \left\|\underline{Q}\right\|^2 \|f\|_c \leq (1-\beta)\epsilon,$$

where $\beta := \rho((I + \delta\underline{Q})^m)$. From Theorem 8, we know that a possible starting point for m is $|\mathcal{X}| - 1$. If we do not have an analytical expression for $\rho((I + \delta\underline{Q})^m)$, then we know from Proposition 11 that we can instead use the guaranteed upper bound of Theorem 10. If no such m and δ exist—for instance because the guaranteed upper bound on β is too conservative—then this method does not work. If on the other hand we do find such an m and δ, then we can keep on running the iterative step (line 7) of Algorithm 1 until we reach the first index $i \in \mathbb{N}$ such that $\|g_i\|_c \leq \epsilon/2$. By Propositions 9 and 12, we are now guaranteed that $(\max g_i + \min g_i)/2$ is an approximation of $\underline{T}_{\infty} f$ up to a maximal error ϵ.

Alternatively, we can fix a step size δ ourselves and use the method of Theorem 5 to compute ϵ'. In that case, we simply need to run the iterative scheme until we reach the first index i such that $\|g_i\|_c \leq \epsilon'$. By Proposition 12, we are then guaranteed that the error $(\max g_i + \min g_i)/2$ is an approximation of $\underline{T}_{\infty} f$ up to a maximal error $\epsilon = 2\epsilon'$. The same is true if we replace ϵ' by the error ϵ_e that is used in Proposition 9.

Example 6 *Using the analytical expressions of Example 2, we obtain $\underline{T}_{\infty}\mathbb{I}_1 \approx 9.5238095 \times 10^{-3}$.*

We want to approximate $\underline{T}_{\infty}\mathbb{I}_1$ up to a maximum error $\epsilon := 1 \times 10^{-6}$. We observe that $m = 1$ and $\delta \approx 3.485 \times 10^{-8}$ yield an ϵ_d that is lower than $\epsilon/2$. After 196,293,685 iterations, the norm of the approximation is sufficiently small, resulting in the approximation $\underline{T}_{\infty}\mathbb{I}_1 = (9.524 \pm 0.001) \times 10^{-3}$. Alternatively, choosing $\delta = 1 \times 10^{-7}$ and continuing until $\|g_i\|_c \leq \epsilon'$ yields the approximation $\underline{T}_{\infty}\mathbb{I}_1 = (9.5242 \pm 0.0008) \times 10^{-3}$ after only 69,572,154 iterations.

Mimicking Troffaes et al. (2015), we also tried the heuristic method of increasing t and n until we observe empirical convergence. After some trying, we find that $t = 7$ and $n = 7 \cdot 250 = 1750$ already yield an approximation with sufficiently small error: $\|\underline{T}_{\infty}\mathbb{I}_1 - \Psi_7(1750)\mathbb{I}_1\| \approx 7 \times 10^{-7} < \epsilon$. Note however that for non-binary examples, where $\underline{T}_{\infty} f$ cannot be computed analytically, this heuristic approach is unable to provide a guaranteed bound.

6. Conclusion

We have improved an existing method and proposed a novel method to approximate $\underline{T}_t f$ up to any desired maximal error, where $\underline{T}_t f$ is the solution of the non-linear differential equation (1) that plays an essential role in the theory of imprecise continuous-time Markov chains. As guaranteed by our

theoretical results, and as verified by our numerical examples, our methods outperform the existing method by Krak et al. (2016), especially if the lower transition rate operator is ergodic. For these ergodic lower transition rate operators, we also proposed a method to approximate $\lim_{t\to\infty} \underline{T}_t f$ up to any desired maximal error.

For the simple case of a binary state space, we observed in numerical examples that there is a rather large difference between the theoretically required number of iterations and the number of iterations that are empirically found to be sufficient. Similar differences can—although this falls beyond the scope of our present contribution—also be observed for the lower transition rate operator that is studied in (Troffaes et al., 2015). The underlying reason for these observed differences remains unclear so far. On the one hand, it could be that our methods are still on the conservative side, and that further improvements are possible. On the other hand, it might be that these differences are unavoidable, in the sense that guaranteed theoretical bounds come at the price of conservatism. We leave this as an interesting line of future research. Additionally, the performance of our proposed methods for systems with a larger state space deserves further inquiry.

Acknowledgments

Jasper De Bock is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO) and wishes to acknowledge its financial support. His work was also partially supported by the H2020-MSCA-ITN-2016 UTOPIAE, grant agreement 722734. Finally, the authors would like to express their gratitude to three anonymous reviewers, for their time, effort and constructive feedback.

References

W. J. Anderson. *Continuous-Time Markov Chains*. Springer-Verlag New York, 1991.

J. De Bock. The limit behaviour of imprecise continuous-time Markov chains. *Journal of Nonlinear Science*, 27(1):159–196, 2017.

A. Erreygers and J. De Bock. Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. 2017. arXiv Report 1702.07150 [math.PR].

F. Hermans and G. de Cooman. Characterisation of ergodic upper transition operators. *International Journal of Approximate Reasoning*, 53(4):573 – 583, 2012.

T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. 2016. arXiv Report 1611.05796 [math.PR].

D. Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250:165–180, 2015.

M. C. M. Troffaes, J. Gledhill, D. Skulj, and S. Blake. Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. In *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 287–294, 2015.

D. Škulj and R. Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2013.