Continuous-time Markov chains as transformers of unbounded observables

Vincent Danos, Tobias Heindel, Ilias Garnier, Jakob Simonsen

Continuous-time Markov chains as transformers of unbounded observables

Vincent Danos[1], Tobias Heindel[2], Ilias Garnier[1], Jakob Grue Simonsen[2]

[1] École Normale Supérieure
[2] University of Copenhagen, DIKU

Abstract. This paper provides broad sufficient conditions for the computability of time-dependent averages of stochastic processes of the form $f(X_t)$ where X_t is a continuous-time Markov chain (CTMC), and f is a real-valued function (aka an observable). We consider chains with values in a *countable* state space $\mathbf{S}$, and possibly *unbounded* fs.
Observables are seen as generalised predicates on $\mathbf{S}$ and chains are interpreted as transformers of such generalised predicates, mapping each observable f to a new observable $P_t f$ defined as $(P_t f)(x) = \mathsf{E}_x(f(X_t))$, which represents the mean value of f at time time t as a function of the initial state x.
We obtain three results. First, the well-definedness of this operator interpretation is obtained for a large class of chains and observables by restricting P_t to judiciously chosen rescalings of the basic Banach space $C_0(\mathbf{S})$ of $\mathbf{S}$-indexed sequences which vanish at infinity. We prove, under appropriate assumptions, that the restricted family P_t forms a strongly continuous operator semigroup (equivalently the time evolution map $t \mapsto P_t$ is continuous w.r.t. the usual topology on bounded operators). The computability of the time evolution map follows by generic arguments of constructive analysis. A key point here is that the assumptions are flexible enough to accommodate unbounded observables, and we give explicit examples of such using stochastic Petri nets and stochastic string rewriting. Thirdly, we show that if the rate matrix (aka the q-matrix) of the CTMC is locally algebraic on a subspace containing f, the time evolution of projections $t \mapsto (P_t f)(x)$ is PTIME computable for each x. These results provide a functional analytic alternative to Monte Carlo simulation as test bed for mean-field approximations, moment closure, and similar techniques that are fast, but lack absolute error guarantees.

1 Introduction

The study of properties of stochastic processes is a currently highly active field of research in computer science, but has been avidly studied in mathematics even prior to Kolmogorov's development of the axiomatic approach to probability. For the particular case of continuous-time Markov chains (CTMCs), the crucial property of interest is *how it acts* on a function from its underlying state space to the reals – a "generalised predicate" in analogy to the usual notion of predicate transformers in computer science [Koz83]. Here, we are interested in the problem of computing time-dependent expectations of real-valued functions of the state space (so-called *observables*) on a continuous-time Markov chain X_t. Ultimately, one wants provably correct, practical algorithms that given some continuous-time Markov chain X_t and some observable f computes the *transient conditional mean* $t \mapsto (x \mapsto \mathsf{E}_x(f(X_t)))$, i.e., the average of f conditioned on the initial state as a function of time. However, it is not *a priori* clear that this function is even computable let alone practically so; hence, it must be established that there is an algorithm that computes the transient conditional mean to arbitrary desired precision. As the basic objects are typically (computable) real numbers, we will employ the type 2 theory of effectivity (TTE) for computability.

It is useful to recall the nature of the objects that we manipulate and want to compute in the finite case. A continuous-time Markov chain (CTMC) on a finite state space $\mathbf{S}$ is entirely captured by its q-matrix, which is an $\mathbf{S} \times \mathbf{S}$-indexed real matrix in which every row sums to zero and all negative entries lie on the diagonal. Any finite q-matrix Q induces a semigroup $t \mapsto P_t = e^{tQ}$ which describes the time evolution of the CTMC. For any π on $\mathbf{S}$ (viewed as a row vector), the map $t \mapsto \pi P_t$ corresponds to the probability of being at any given state at time t for the initial probabilistic state π. Therefore, if X_0 is distributed according to π, the associated stochastic process X_t is distributed according to πP_t. Dually, one can interpret a CTMC as a transformer of observables, using the discrete time setting [Koz83] as an analogy. For any $f \colon \mathbf{S} \to \mathbb{R}$ (seen as a column vector), $P_t f = (x \mapsto \sum_{y \in \mathbf{S}} p_{t,xy} f(y))$ is the vector of conditional means[3] of f at time t as a function of the initial state. The function we seek to compute is precisely $t \mapsto P_t f$. Formally, $P_t f$ is the (column) vector of conditional means (i.e., $(P_t f)(x) = \mathsf{E}_x(f(X_t))$ for $x \in \mathbf{S}$) where X_t is the CTMC with transition function P_t. For the finite state case, $t \mapsto P_t f = e^{tQ} f$ is the unique solution to the initial value problem

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} u_t &= Q u_t \\ u_0 &= f \end{aligned} \tag{1}$$

and this can be solved by any method for computing matrix exponentials or numerical solvers for finite ODEs. In the countably infinite state case, observables are possibly *unbounded*–this makes computing transient conditional means fundamentally harder than in the finite case and calls for more sophisticated mathematics, as we shall describe in § 3.2.

For motivation, we give two paradigmatic examples of transient means. The classic example of a stochastic process of the form $f(X_t)$, i.e., a pair of a CTMC X_t and an observable f, where X_t models a set of chemical reactions with the molecule count of a certain chemical species as observable: states of X_t are multisets over a finite set of species; then, we are interested in computing the evolution of the mean count of a certain species.

[3] By definition, $\mathsf{E}_x(f(X_t)) = \int_{\omega \in \Omega} f(X_t(\omega))\, dp_x$ for some suitable probability space (Ω, p), where p_x is the probability measure p on Ω conditioned by $X_0 = x$. Performing a change of variables, $\mathsf{E}_x(f(X_t)) = \int_{\mathbf{S}} f\, dX_t^*(p_x)$, where $X_t^*(p_x)$ is the image measure of p_x through X_t. Since $p_{t,x-} = X_t^*(p_x)(-)$, we deduce $P_t f(x) = \mathsf{E}_x(f(X_t))$.

Another natural example native to computer science is a stochastic interpretation of any string rewriting system as a CTMC X_t. An obvious class of observables for string rewriting are functions f that count the occurrence of a certain word as sub-string in each state of the CTMC X_t; note that this is different from counting "molecules", as there is always only a singe word! For example, consider the string rewriting system with the single rule $a \to aba$ and initial state a: the mean occurrence count of the letter a grows as the exponential function e^t while the mean occurrence count of the word aa is zero at all times; adding the rule $ba \to ab$ does not change the mean a-count but renders the mean count of the word aa non-trivial.

Note that these two classes of models are only meant as simple examples and the results of this paper are not geared towards any particular modelling language for CTMCs; the results apply equally well to Kappa models [DFF+10] or stochastic graph transformation.[4]

Our main result is a sufficient condition for $\mathsf{E}_x(f(X_t))$ to be computable, powerful enough to encompass many interesting unbounded observables. We proceed by construction of a suitable computable Banach space where to restrict our operators P_t, in such a way that they form a strongly continuous semigroup (SCSG). It follows that the initial value problems corresponding to observables in the domain of the generator of the semigroup admit computable solutions. The construction of our Banach space rests on functional analytic techniques recently developed in Ref. [Spi12,Spi15], which we combine with the work on SCSGs by Weihrauch & Zhong [WZ07].

We do not assume that the reader has some acquaintance with continuous-time Markov chains on a countable state space and recall the basic concepts of transition functions and q-matrices (Def. 5 and 6) in the preliminaries. Having said that, and both in order to save space and to keep the logic of the paper clearly apparent, most of the mathematical material used is postponed to a series of appendices.

2 Two motivating examples of CTMCs with observables

We illustrate our constructions with: (i) chemical reaction networks (CRN), aka stochastic Petri nets, and (ii) stochastic string rewriting as a simple example of (rule-based) modelling. In both cases, the construction of the q-matrix implied by a model is readily done, and so is the definition of a natural set of *unbounded* observables with clear relevance to the dynamics of a model: word occurrence counts for stochastic string rewriting (Def. 2) and multiset inclusions for Petri nets (Def. 3).

[4] In fact, string rewriting is simply the restriction of stochastic graph transformation [HLM06] to directed, connected, acyclic, edge labelled graphs with in and out degree of all nodes bounded by one, i.e., to graphs consisting of a unique maximal path.

2.1 Stochastic string rewriting and word occurrences

Stochastic string rewriting can be thought of as never ending, fair competition between all redexes of rules, "racing" for reduction; the formal definition is as follows, in perfect analogy to Ref. [HLM06] which covers the case of graphs.

Definition 1 (Stochastic string rewriting). *For each rule $\rho = l \to r \in \Sigma^+ \times \Sigma^+$ we define the* q-matrix of ρ, *denoted by Q_ρ, as the q-matrix $Q_\rho = (q^\rho_{uv})_{u,v \in \Sigma^+}$ on the state space of words Σ^+ with off-diagonal entries*

$$q^\rho_{uv} = \left|\left\{(w, w') \in \Sigma^* \times \Sigma^* \;\middle|\; u = wlw',\ v = wrw'\right\}\right|$$

for each pair of words $u, v \in \Sigma^+$ such that $u \neq v$, and diagonal entries, $q^\rho_{uu} = -\sum_{v \neq u} q^\rho_{uv}$ for all $u \in \Sigma^+$. For a finite set of rules $\mathcal{R} \subseteq \Sigma^+ \times \Sigma^+$, we define $Q_\mathcal{R} = \sum_{\rho \in \mathcal{R}} Q_\rho$, and with additional choices of rate constants $\mathsf{k}\colon \mathcal{R} \to \mathbb{Q}^+$, we define $Q_{\mathcal{R},\mathsf{k}} = \sum_{\rho \in \mathcal{R}} \mathsf{k}_\rho Q_\rho$.

For a given rule set $\mathcal{R}$, each entry $q^\mathcal{R}_{uv}$ of the q-matrix corresponds to the propensity to rewrite: it is just the number of ways in which u can be rewritten to v. We shall usually work without rate constants for the sake of readability. Note that the use of Σ^+ for the left and right hand side of rules is convenient to get string rewriting as a special case of graph transformation in a straightforward manner.

The occurrence counting function of a word as sub-string in the state of the CTMC of $\mathcal{R}$ is as follows.

Definition 2 (Word counting functions). *Let $w \in \Sigma^+$ be a word. The* w-counting function, *denoted by $\sharp_w \colon \Sigma^+ \to \mathbb{R}_{\geq 0}$, maps each word $x \in \Sigma^+$ to $\sharp_w(x) = |\{(u, v) \in \Sigma^* \times \Sigma^* \mid x = uwv\}|$.*

2.2 Stochastic Petri nets and sub-multiset occurrences

We recall the definition of stochastic Petri nets and occurrence counting of a multisets. Note that for the purposes of the present paper, places and species are synonymous.

Definition 3 (Multisets and multiset occurrences). *A* multiset *over a finite set $\mathcal{P}$ of* places *is a function $x\colon \mathcal{P} \to \mathbb{N}$ that maps each place to the number of* tokens *in that place. Given a multiset, $x \in \mathbb{N}^\mathcal{P}$, the* x-occurrence counting *function $\sharp_x \colon \mathbb{N}^\mathcal{P} \to \mathbb{N}$ is defined by*

$$\sharp_x(y) = \begin{cases} \frac{y!}{(y-x)!} & x \leq y \\ 0 & \textit{otherwise} \end{cases}$$

where $z! = \prod_{p \in \mathcal{P}} z(p)!$ is the multiset factorial for all $z \in \mathbb{N}^\mathcal{P}$.

Definition 4 (Stochastic Petri net). *Let $\mathcal{P}$ be a finite set of* places. *A* stochastic Petri net *over $\mathcal{P}$ is a set*

$$\mathcal{T} \subseteq \mathbb{N}^\mathcal{P} \times \mathbb{R}_{>0} \times \mathbb{N}^\mathcal{P}$$

where $\mathbb{N}^{\mathcal{P}}$ is the set of multisets over $\mathcal{P}$, which are called markings *of the net; elements of the set $\mathcal{T}$ are called* transitions*. The q-matrix $Q_{l,k,r}$ on the set of markings for a transition $(l, k, r) \equiv l \to^k r \in \mathcal{T}$ has off-diagonal entries*

$$q^{l,k,r}_{xy} = \begin{cases} k \cdot \natural_l(x) & \natural_l(x) > 0, y = x - l + r \\ 0 & \textit{otherwise} \end{cases}$$

where addition and subtraction is extended pointwise to $\mathbb{N}^{\mathcal{P}}$. The q-matrix of $\mathcal{T}$ is $Q_{\mathcal{T}} = \sum_{(l,k,r)\in\mathcal{T}} Q_{l,k,r}$.

3 Preliminaries

For the remainder of the paper, we fix an at most countable set $\mathbf{S}$ as state space.

3.1 Transition functions and q-matrices

We first recall the basic definitions of transition functions and q-matrices. We make the usual assumptions [And91] one needs to work comfortably: namely that q-matrices are stable and conservative and that transition functions are standard and also minimal as described at the end of § 3.1 (and detailed in Appendix A).

With these assumptions in place, transition functions and q-matrices determine each other, and one can freely work with one or the other as is most convenient.

Definition 5 (Standard transition function [And91, p. 5f.]). *A* transition function *on $\mathbf{S}$ is a family $\{P_t\}_{t\in\mathbb{R}_{\geq 0}}$ of $\mathbf{S} \times \mathbf{S}$-matrices $P_t = (p_{t,xy})_{x,y\in\mathbf{S}}$ with non-negative, real entries $p_{t,xy}$ such that*

1. $\lim_{t\searrow 0} p_{t,xx} = 1$ *for all* $x \in \mathbf{S}$;
2. $\lim_{t\searrow 0} p_{t,xy} = 0$ *for all* $x, y \in \mathbf{S}$ *such that* $y \neq x$;
3. $P_{t+s} = P_t P_s = (\sum_{z\in\mathbf{S}} p_{s,xz} p_{t,zy})_{x,y\in\mathbf{S}}$ *for all* $s, t \in \mathbb{R}_{\geq 0}$; *and*
4. $\sum_{z\in\mathbf{S}} p_{t,xz} \leq 1$ *for all* $x \in \mathbf{S}$ *and* $t \in \mathbb{R}_{\geq 0}$.

Thus, each row of a transition function corresponds to a sub-probability measure, and transition functions converge entry-wise to the identity matrix at time zero. Unless stated otherwise, all transition functions in this paper are standard.

Taking entry-wise derivatives at time 0 gives a q-matrix.

Definition 6 (q-matrix). *A* q-matrix *on $\mathbf{S}$ is an $\mathbf{S} \times \mathbf{S}$-matrix $Q = (q_{xy})_{x,y\in\mathbf{S}}$ with real entries q_{xy} such that $q_{xy} \geq 0$ (if $x \neq y$), $q_{xx} \leq 0$, and $\sum_{z\in\mathbf{S}} q_{xz} = 0$ for all $x, y \in \mathbf{S}$.*

Conversely, for each q-matrix, there exists a unique entry-wise minimal transition function that solves Equation (2) [And91, Theorem 2.2],

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = Q P_t, \; P_0 = I \tag{2}$$

which is called *the* transition function of Q. From now on, we assume that all transition functions are minimal solutions to Equation (2) for some q-matrix Q (cf. Definition 15, see also [Nor98, p. 69]).

3.2 The Abstract Cauchy problem for $P_t f$

Abstract Cauchy problems (ACPs) in Banach spaces [Ein52], are the classic generalisation of finite-dimensional initial value problems. (See Appendices C and D for details.) Concretely, we want to obtain $P_t f$ as unique solution u_t of the following generalisation of our earlier equation (1):

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t} u_t &= \mathcal{Q} u_t \quad (t \geq 0) \\ u_0 &= f \end{aligned} \tag{ACP}$$

where f is an observable and $\mathcal{Q}$ is a linear operator which plays the role of the q-matrix.

There are a few points worth noting. First, the topological vector space of *all* observables $\mathbb{R}^{\mathbf{S}}$ cannot be equipped with a suitable complete norm. Therefore, one has to look for a Banach subspace $\mathcal{B} \subset \mathbb{R}^{\mathbf{S}}$ wherein to interpret the above equation. There are several ways to do this, and they are not equally interesting. Second, as $P_t f = u_t$ is the desired solution, and $P_0 = I$, it follows that $\frac{\mathrm{d}}{\mathrm{d}t} P_t f|_{t=0} = \mathcal{Q} f$. If this derivative does not exist, $\mathcal{Q} f$ is simply not defined. In fact, as is clear from the examples, we can only expect $\mathcal{Q}$ to be partially defined and an unbounded operator.[5] On the positive side, if P_t is a strongly continuous semigroup (Def. 20) on $\mathcal{B}$, meaning $\lim_{h \searrow 0} P_h f = f$ for all f in $\mathcal{B}$, and we take $\mathcal{Q}$ as its generator defined on a subspace of $\mathcal{B}$ (Def. 21) by the above formula, then $P_t f$ is the *unique* solution of (ACP), provided f is in the domain of $\mathcal{Q}$ [EN00, Proposition II.6.2].

Even better, in this case, not only does (ACP) have $P_t f$ as unique solution, but we get an explicit approximation scheme:

$$P_t f = \lim_{n \to \infty} e^{t A_n} f \tag{3}$$

where θ is a constant of the SCSG such that $nI - \mathcal{Q}$ is invertible for $n > \theta$ and $A_n = n\mathcal{Q}(nI - \mathcal{Q})^{-1}$ is called a *Yosida approximant*.[6] The Yosida approximants are the cornerstone of the generation theorems [EN00, Corollary 3.6] that allow one to pass from the generator $\mathcal{Q}$ to the corresponding SCSG. The constant θ also bounds the growth of the SCSG in norm, that is to say $\|P_t\| \leq M e^{\theta t}$ for some M. This should already make clear that Equation (3) is crucial to obtain error bounds for results on the computability of SCSGs. In fact it is the starting point of the proof of the main result on the computability of SCSGs [WZ07, Theorem 5.4.2,p. 521].

It remains to see whether we can find a Banach space to build ACPs which accomodate interesting (specifically unbounded) observables.

---

[5] Even when $\mathcal{Q} f$ is defined, one has to check $\mathcal{Q} f = Q f$, that is to say: $1/h(P_h f - f)$ converges to $Q f$ in the Banach space norm. But this will turn out to be easy compared to finding sufficient conditions for $\mathcal{Q} f$ to be defined.

[6] In case f does not belong to the domain of the infinitesimal generator $\mathcal{Q}$, (3) will generate a 'mild' solution that is a solution to the integral form of (ACP) which might not be differentiable everywhere.

<table>
<tr><td></td><td></td><td>solution
(finite $\mathbf{S}$)</td><td colspan="2">generalisation
(countably infinite $\mathbf{S}$)</td></tr>
<tr><td>IVP
transient
distributions</td><td>$\frac{\mathrm{d}}{\mathrm{d}t}\pi_t = \pi_t Q$
$\pi_0 = \pi$</td><td>$\pi_t = \pi e^{Qt}$</td><td colspan="2">$\pi_t = \pi P_t$
P_t SCSG on $L^1(\mathbf{S})$, in general</td></tr>
<tr><td rowspan="2">IVP
transient
conditional means</td><td rowspan="2">$\frac{\mathrm{d}}{\mathrm{d}t}u_t = Qu_t$
$u_0 = f$</td><td rowspan="2">$u_t = e^{Qt}f$</td><td>$\sup_{i\in\mathbf{S}} -q_{ii} < \infty$
or Feller</td><td>$\sup_{i\in\mathbf{S}} -q_{ii} = \infty$
not Feller</td></tr>
<tr><td>$u_t = P_t f$
P_t SCSG on
$L^\infty(\mathbf{S})$ or $C_0(\mathbf{S})$</td><td>[Spi12, Theorem 6.3]
or
open problem</td></tr>
</table>

Table 1. Transition functions acting on Banach spaces: state of the art

3.3 Banach space wanted!

Table 1 gives an overview of initial value problems for transient distributions (first row) and transient conditional means (second row). Transient distributions are summable sequences, and transition functions form SCSGs [Reu57] and therefore allow for a well-posed corresponding ACP. But the classic example of a Banach space to reason about conditional means [RR72] is the space $C_0(\mathbf{S})$ of functions vanishing at infinity, i.e., functions $f\colon \mathbf{S} \to \mathbb{R}$ such that for all $\epsilon > 0$, the set $\{x \in \mathbf{S} \mid f(x) \geq \epsilon\}$ is finite, equipped with the supremum norm (Def. 17). The corresponding processes are called Feller transition functions [And91, § 1.5] and verify a principle of finite velocity of information flow (for all t, y, $p_{t,xy}$ vanishes as x goes to infinity).

4 Spieksma's theorem

A solution is provided by a result of Spieksma [Spi12, Theorem 6.3], which constructs a Banach space for a given q-matrix Q and an observable f of interest. The formulation will be in terms of so-called drift functions.

Definition 7 (Drift function). *Let Q be a q-matrix on $\mathbf{S}$, and let $c \in \mathbb{R}$. A function $W\colon \mathbf{S} \to \mathbb{R}_{>0}$ is called a* c-drift function *for Q if for all $x \in \mathbf{S}$ $(QW)(x) := \sum_{y\in\mathbf{S}} q_{xy}W(y) \leq cW(x)$.*

We will say that W is a *drift function* for Q if there exists $c \in \mathbb{R}$ such that it is a c-drift function for Q. One can show that $P_tW \leq e^{ct}W$ in this case. Thus, drift functions control their own growth under the transition function.

We also need to work with weighted variants of $C_0(\mathbf{S})$.

Definition 8 (Weighted $C_0(\mathbf{S})$-spaces). *Let $\mathbf{S}$ be a set and let $W\colon \mathbf{S} \to \mathbb{R}_{>0}$ be a positive real-valued function, referred to as a* weight*. The Banach space $C_0(\mathbf{S}, W)$ consists of functions $f\colon \mathbf{S} \to \mathbb{R}$ such that ${}^{f}\!/_{W}$ vanishes at infinity, where $({}^{f}\!/_{W})(x) = {}^{f(x)}\!/_{W(x)}$. The norm $\|\_\|_W$ on such functions f is $\|f\|_W = \sup_{x\in\mathbf{S}} |{}^{f(x)}\!/_{W(x)}|$.*

As $C_0(\mathbf{S}, W)$ is isometric to $C_0(\mathbf{S})$ it is indeed a Banach space. It is also a closed subspace of $L_\infty(\mathbf{S}, W)$, the set of functions such that ${}^f/_W$ is bounded. We will use later the fact that:

Lemma 1. *Finite linear combinations of indicator functions,*[7] *form a dense subset of* $C_0(\mathbf{S})$

Indeed, suppose $\mathbf{S} = \mathbb{N}$ for convenient notations, and define $g_n = f$ truncated after n. It is easy to see that g_n converges to f in the $\|\_\|_W$ norm. In fact, $\|g_n - f\|_W = \sup_{x>n} |f(x)|/W(x) \to 0$ iff f/W vanishes at infinity.

Now the idea is to find positive drift functions V, W for Q such that $V \in C_0(\mathbf{S}, W)$, i.e., such that ${}^V/_W$ vanishes at infinity. Thus, both functions can grow at most exponentially in mean and V is negligible compared to W at infinity. Intuitively, functions on the order of V are as good as functions vanishing at infinity, in the case of Feller processes [RR72], i.e., CTMCs such that transition functions induce SCSGs on $C_0(\mathbf{S})$.

Theorem 1. *Let P_t be a transition function on $\mathbf{S}$ with q-matrix Q, let $V, W : \mathbf{S} \to \mathbb{R}_{>0}$ be drift functions for Q such that $V \in C_0(\mathbf{S}, W)$. Then:*

1. *P_t induces an SCSG on $C_0(\mathbf{S}, W)$.*
2. *For all $f \in C_0(\mathbf{S}, W)$, $P_t f$ is given by Equation* (3) *in $C_0(\mathbf{S}, W)$ where $\mathcal{Q}$ is the generator of P_t.*

The first part of the theorem is proved in [Spi12, Theorem 6.3]; the second part follows from the general theory of ACPs. Note that f does not need to be in the domain of $\mathcal{Q}$, in which case we only obtain a 'mild' solution to the ACP, i.e., a solution to its integral form which might not be everywhere differentiable. If on the other hand, f is in the domain of $\mathcal{Q}$, then $P_t f$ is continuously differentiable.

We have now covered the mathematical ground needed to make sense of conditional means of not necessarily bounded observables. This, however, does not immediately yield an *algorithm* for computing transient means, as it does not yet ensure that transient means are computable even in presence of a complete specification of the underlying system. Even transient conditional probabilities can fail to be computable [AFR11]! Before we proceed to computability questions, let us return to our two classes of examples.

4.1 Applications: string rewriting and Petri nets

We now give examples of drift functions for stochastic string rewriting and Petri nets. The former case is well-behaved since the mean letter count grows at most exponentially. The case of Petri nets will be more subtle and we shall give an example of an explosive Petri net such that we can nevertheless reason about conditional means of unbounded observables.

For string rewriting, we have canonical drift functions.

[7] The indicator function $\mathbb{1}_x$ is defined as usual as $\mathbb{1}_x(y) = \delta_{xy}$.

Lemma 2 (Powers of length are drift functions). *Let $\mathcal{R} \subseteq \Sigma^+ \times \Sigma^+$ be a finite string rewriting system and let $n \in \mathbb{N}^+$ be a positive natural number. There exists a constant $c_n \in \mathbb{R}_{>0}$ such that $|\_|^n : \Sigma^+ \to \mathbb{R}_{\geq 0}$ is a c_n-drift function.*

Proof. The proof is deferred to Appendix E.

Now, we can apply Spieksma's method to get a Banach space for reasoning about conditional means and moments of word counting functions.

Corollary 1 (Stochastic string rewriting). *Let $\mathcal{R}$ be a finite string rewriting system, let $n \in \mathbb{N} \setminus \{0\}$, and let $|\_| : \Sigma^+ \to \mathbb{N}$ be the word length function. The transition function P_t of q-matrix $Q_{\mathcal{R}}$ is a SCSG on $C_0(\Sigma^+, |\_|^n)$.*

Thus, all higher conditional moments of word counting functions can be accommodated in a suitable Banach space. The case of Petri nets is more subtle, since, in general, the (weighted) token count is not a drift function.

Example 1. Consider the Petri net with single transition $2A \to^1 3A$ and with one place A. The token count $\sharp_A$ is not a drift function. In fact, the corresponding CTMC is explosive (by [Spi15, Theorem 2.1]).

Our final example is an extension of the previous explosive CTMC with a new species whose count can nevertheless be treated using Theorem 1.

Example 2 (Unobserved explosion). Consider the Petri net with transitions

$$\{2A \to^1 3A, B \to^1 2B\}.$$

The underlying CTMC is explosive, and we cannot apply Theorem 1 to compute the transient conditional mean of the A-count for the exact same reason as in Example 1. However, we can do so for the B-count, using the weight function $W = \sharp_B{}^2$ and observable $f = \sharp_B$. Putting $V = f$ allows to apply Spieksma's recipe (ruling out states with B-count 0 for convenience). The conditional mean $\mathsf{E}_{2A+B}(\sharp_B(X_t))$ can be best understood by adding a coffin state, on which both the A- and B-count are zero and in which the Markov chain resides after (the first and only) explosion.

5 Computability

We follow the school of type-2 theory of effectivity, as reviewed in Appendix G following [BHW08]. A real number x is computable iff there is a Turing machine that on input $d \in \mathbb{N}$ (the desired precision), outputs a rational number r with $|r - x| < 2^{-d}$. Next, a function $g : \mathbb{R} \to \mathbb{R}$ is computable if there is a Turing machine that, for each $x \in \mathbb{R}$, takes an arbitrary Cauchy sequence with limit x as input and generates a Cauchy sequence that converges to $g(x)$–where convergence has to be sufficiently rapid, e.g., by using the dyadic representation of the reals.

Computability extends naturally to any Banach space $\mathcal{B}$ other than $\mathbb{R}$. We only need a recursively enumerable dense subset on which the norm, addition and scalar multiplication are computable, thus making $\mathcal{B}$ a *computable* Banach space (Definition 30); usually, the dense subset is induced by a basis of a dense subspace. For weighted C_0-spaces (with computable weight functions) and their duals, we fix an arbitrary enumeration of all *rational* linear combinations of indicator functions $\mathbb{1}_x$; for the Banach space of bounded linear operators on weighted C_0-spaces we use the standard construction for continuous function spaces [WZ07, Lemma 3.1]. A SCSG P_t is computable if the function $t \mapsto P_t$ from the reals to the Banach space of bounded linear operators is computable. The computable SCSGs correspond to those obtained from CTMCs through Theorem 1.

For simplicity, motivated by the observation that we do not lose any of the intended applications to rule-based modeling, we restrict to row- and column-finite q-matrices with rational entries in our main result.

Theorem 2 (Computability of CTMCs as observation transformers). *Let Q be a q-matrix on $\mathbf{S}$, let $V, W\colon \mathbf{S} \to \mathbb{R}_{\geq 0}$ be positive drift functions for Q such that $V \in C_0(\mathbf{S}, W)$. If*

- *the q-matrix Q is row- and column-finite, consists of rational entries, and is computable as a function $Q\colon \mathbf{S}^2 \to \mathbb{Q}$, and*
- *$W\colon \mathbf{S} \to \mathbb{Q}$ is computable,*

the following hold.

1. *The SCSG P_t, i.e., the function $t \mapsto P_t$, is computable.*
2. *The evolution of conditional means $(t, f) \mapsto P_t f$ is computable.*
3. *The evolution of means $(\pi, t, f) \mapsto \pi P_t f$ is computable for all $\pi\colon \mathbf{S} \to \mathbb{R}$ such that $\sum_{x \in \mathbf{S}} W(x)|\pi(x)| < \infty$.*

Proof. We will apply a result by Weihrauch & Zhong on the computability of SCSGs [WZ07, Theorem 5.4]. Applying this result requires some extra information:

1. the SCSG P_t must be bounded in norm by $e^{\theta t}$ for some positive constant θ;
2. we must have a recursive enumeration of a dense subset of the graph of the infinitesimal generator of the SCSG P_t.

We first show that the constant θ, featuring in Theorem 1, i.e., the witness that W is a θ-drift function for Q, satisfies $\|P_t\| \leq e^{\theta t}$ (using the first part of the proof of Theorem 6.3 of Ref. [Spi12]). Next, we obtain a recursive enumeration of a dense subset $A \subseteq \mathcal{Q}$ of the domain of the generator $\mathcal{Q}$ by applying Q to all rational linear combinations of indicator functions $\mathbb{1}_x$. Note that for the latter, we use that indicator functions belong to the domain of the generator and $\mathcal{Q}\mathbb{1}_x = Q\mathbb{1}_x$ (see Proposition 2).

Now, by Theorem 5.4.2 of Ref. [WZ07], we obtain the first two computability results, as $\theta, A, 1$ is a so-called piece of type IG-information [WZ07, p. 513]. Finally, the third point amounts to showing computability of the duality pairing

$$\langle \_, \_ \rangle\colon L^1(\mathbf{S}, W) \times C_0(\mathbf{S}, W) \to \mathbb{R}$$

where $L^1(\mathbf{S}, W)$ is (isomorphic to) the dual of $C_0(\mathbf{S}, W)$ (Lemma 3).

This theorem immediately gives computability of the CTMCs and (conditional) means for stochastic string rewriting and Petri nets discussed in Sect. 4. Note that the theorem does not assume that V itself is computable; its role is to establish that the transition functions is a SCSG, but V plays no role in the actual computation of the solution. Note also that the algorithms that compute the functions $t \mapsto P_t$, $(t, f) \mapsto P_t f$, and $(\pi, t, f) \mapsto \pi P_t f$ push the responsibility to give arbitrarily good approximations of the respective input parameters π, t and f to the user. This however is no problem for any of our examples or rule-based models in general: t is typically rational, f is computable and even to the natural numbers, and π is often finitely supported or a Gaussian.

Remark 1 (Domain of the generator). The difficulty in working with SCSGs is to find a tractable description of the domain of their generator. Such a description is important for two reasons: first, as said before the theorem, for computability, we need a recursively enumerable dense subset of the graph of the domain of the generator; second, as the unique solution to the ACP is also differentiable, it is useful for numerical integration.

Computability ensures existence of algorithms computing transient means, but yields no guarantees of the efficiency of such algorithms. We now proceed to a special case that (i) encompasses a number of well-known examples, including context-free string rewriting, and (ii) leads to PTIME computability, by reducing the problem of transient conditional means to solving finite linear ODEs.

6 The finite dimensional case and PTIME via ODEs

We now turn to the special case where we can restrict to finite dimensional subspaces $\mathcal{B} \subseteq \mathbb{R}^{\mathbf{S}}$.[8] The prime example will be word counting functions and context-free string rewriting systems. Hyperedge replacement systems [DKH97], the context-free systems of graph transformation, can be handled *mutatis mutandis*. The main result is PTIME computability of conditional means. For convenience, we extend the usage of the term locally algebraic as follows.

Definition 9 (Locally algebraic). *We call a q-matrix Q on* $\mathbf{S}$ locally algebraic *for an observable $f \in \mathbb{R}^{\mathbf{S}}$ if the set $\{Q^n f \mid n \in \mathbb{N}\}$, containing all multiple applications of the q-matrix Q to the observable f, is linearly dependent, i.e., if there exists a number $N \in \mathbb{N}$ such that the application $Q^N f$ of the N-th power is a linear combination $\sum_{i=0}^{N-1} \alpha_i Q^i f$ of lower powers of Q applied to f.*

Using local algebraicity of a q-matrix Q for an observable f, one can generate a *finite* ODE with one variable for each conditional mean $\mathsf{E}[Q^n f(X_t) \mid X_0 = x]$ (as detailed in the proof of Theorem 3); then, recent results from computable analysis [PG16] entail PTIME complexity.

[8] Every such finite dimensional subspace is a Banach space.

Theorem 3 (PTIME complexity of conditional means). *Let Q be a q-matrix on $\mathbf{S}$, let $x \in \mathbf{S}$, let $f\colon \mathbb{R} \to \mathbf{S}$ be a function such that $f(x)$ is a computable real and $Q^N f = \sum_{i=0}^{N-1} \alpha_i Q^i f$ for some $N \in \mathbb{N}$ and computable coefficients α_i.*

The time evolution of the conditional mean $P_t f(x)$, i.e., the function $t \mapsto P_t f(x)$, is computable in polynomial time.

Proof. Consider the N-dimensional ODE with one variable E_n for each $n \in \{0, \ldots, N-1\}$ with time derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} E_n(t) = \begin{cases} E_{n+1}(t) & \text{if } n < N-1 \\ \sum_{i=0}^{N-1} \alpha_i E_i(t) & \text{if } n = N-1 \end{cases}$$

and initial condition $E_n(0) = Q^n f(x)$. Solving this ODE is in PTIME [PG16] (even over all of $\mathbb{R}_{\geq 0}$, using the "length" of the solution curve as implicit input). Finally, $E_i(t) = P_t Q^i f(x)$ is the unique solution.

Note that the linear ODE that we construct has a companion matrix as evolution operator, which allows to use special techniques for matrix exponentiation [TR03,BCO+07].

Proposition 1 (Local algebraicity of context-free string rewriting). *Let $\mathcal{R}$ be a string rewriting system, let $w \in \Sigma^+$, let $m \in \mathbb{N}$. The q-matrix $Q_{\mathcal{R}}$ of the string rewriting system $\mathcal{R}$ is locally algebraic for the m-th power of w-occurrence counting ${\sharp_w}^m$ if $\mathcal{R} \subseteq \Sigma \times \Sigma^+$.*

Proof. For every product of word counting functions $\prod_{i=1}^{m} \sharp_{w_i}$, applying the q-matrix $Q_{\mathcal{R}}$ to this product yields the observable $Q_{\mathcal{R}} \prod_{i=1}^{m} \sharp_{w_i}$. Using previous work on graph transformation [DHHZS15], restricted to acyclic, finite, edge labelled graphs that have a unique maximal path (with at least one edge), the observable $Q_{\mathcal{R}} \prod_{i=1}^{m} \sharp_{w_i}$ is a linear combination $\sum_{j=1}^{k} \alpha_j \prod_{l=1}^{k_j} \sharp_{w_{j,l}}$ of word counting functions $\sharp_{w_{j,l}}$ (with all $k_j \leq m$). Moreover, if $\mathcal{R}$ is context-free ($\mathcal{R} \subseteq \Sigma \times \Sigma^+$), we have $\sum_{l=1}^{k_j} |w_{j,l}| \leq \sum_{i=1}^{m} |w_i|$ for all $j \in \{1, \ldots, k\}$. Thus, we stay in a subspace that is spanned by a finite number of products of word counting functions.

Corollary 2. *For context-free string rewriting, conditional means and moments of word occurrence counts are computable in polynomial time.*

We conclude with a remark on lower bounds for the complexity.

Remark 2. The complexity of computing transient means, even for context-free string rewriting, is at least as hard as computing the exponential function. This becomes clear if we consider the rule $a \to aa$, and the observable of a-counts $\sharp_a$. Now, the time evolution of the $\sharp_a$-mean conditioned on the initial state to be a, i.e., the function $t \mapsto \mathsf{E}_a(\sharp_a(X_t))$, is exactly the exponential function e^t. Tight lower complexity bounds for the exponential function are a longstanding open problem [Ahr99].

7 Conclusion

The main result is computability of transient (conditional) means of Markov chains X_t "observed" by a function f, i.e., stochastic processes of the form $f(X_t)$. For this, we have described conditions under which a CTMC, specified by its q-matrix, induces a continuous-time *transformer* P_t that acts on observation functions. In analogy to predicate transformer semantics for programs, this should be called *observation transformer semantics* for CTMCs; formally, P_t is a strongly continuous semigroup on a suitable function space. Finally, motivated by important examples of context-free systems – be it the well-known class from Chomsky's hierarchy or the popular preferential attachment process (covered in previous work [DHHZS15]) – we have considered the special case of locally finite q-matrices. For this special case, we obtain a first complexity result, namely PTIME computability of transient conditional means.

The obvious next step is to implement our theoretical results since one cannot expect that the general algorithms of Weihrauch & Zhong [WZ07] perform well for every SCSG on a computable Banach space. For example, the Gauss-Jordan algorithm for infinite Matrices [Par12] should already be more practicable for inverting the operator $nI - \mathcal{Q}$ from Equation (3) compared to the brute force approach used by Weihrauch & Zhong [WZ07]. Computability ensures existence of algorithms for computing transient means, but yields no guarantees of the efficiency of such algorithms.

Even if it should turn out that efficient algorithms are a pipe dream – after all, transient probabilities $p_{t,xy}$ are a special case of transient conditional means – we expect that already implementations that are slow but to arbitrary desired precision will be useful for gauging the quality of approximations of the "mean-field" of a Markov process, especially in the area of social networks [Gle13], but possibly also for chemical systems [SSG15]. Theoretically, they are a valid alternative to Monte Carlo simulation, or even preferable.

References

AFR11. Nathanael Leedom Ackerman, Cameron E. Freer, and Daniel M. Roy. Noncomputable conditional distributions. In *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science, LICS 2011, June 21-24, 2011, Toronto, Ontario, Canada*, pages 107–116, 2011.

Ahr99. Timm Ahrendt. Fast computations of the exponential function. In *Proceedings of STACS'99*, STACS'99, pages 302–312, Berlin, Heidelberg, 1999. Springer-Verlag.

And91. William J. Anderson. *Continuous-Time Markov Chains*. Springer-Verlag New York, 1991.

Aus55. Donald G. Austin. On the existence of the derivative of Markoff transition probability functions. *Proceedings of the National Academy of Sciences of the United States of America*, 41(4):224–226, 1955.

BCO+07. A. Bostan, F. Chyzak, F. Ollivier, B. Salvy, É. Schost, and A. Sedoglavic. Fast computation of power series solutions of systems of differential equations. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on*

Discrete Algorithms, SODA '07, pages 1012–1021, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

BHW08. Vasco Brattka, Peter Hertling, and Klaus Weihrauch. *New Computational Paradigms: Changing Conceptions of What is Computable*, chapter A Tutorial on Computable Analysis, pages 425–491. Springer New York, 2008.

DFF+10. Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer, and Jean Krivine. Abstracting the differential semantics of rule-based models: Exact and automated model reduction. In *Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11-14 July 2010, Edinburgh, United Kingdom*, pages 362–381, 2010.

DHHZS15. Vincent Danos, Tobias Heindel, Ricardo Honorato-Zimmer, and Sandro Stucki. Moment semantics for reversible rule-based systems. In Jean Krivine and Jean-Bernard Stefani, editors, *Reversible Computation 2015*, pages 3–26. Springer International Publishing, 2015.

DKH97. Frank Drewes, Hans-Jörg Kreowski, and Annegret Habel. *Hyperedge Replacement, Graph Grammars*, pages 95–162. World Scientific, 1997.

Ein52. Hille Einar. A note on Cauchy's problem. *Annales de la Société Polonaise de Mathématique*, 25:56–68, 1952.

EN00. Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer-Verlag New York, 2000.

Gle13. James P. Gleeson. Binary-state dynamics on complex networks: Pair approximation and beyond. *Physical Review X*, 3:021004, Apr 2013.

HLM06. Reiko Heckel, Georgios Lajios, and Sebastian Menge. Stochastic graph transformation systems. *Fundamenta Informaticae*, 74(1):63–84, January 2006.

Kol51. Andrey Nikolaevich Kolmogorov. On the differentiability of the transition probabilities in stationary Markov processes with a denumberable number of states. *Moskovskogo Gosudarstvennogo Universiteta Učenye Zapiski Matematika*, 148:53–59, 1951.

Koz83. Dexter Kozen. A probabilistic pdl. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, pages 291–297, New York, NY, USA, 1983. ACM.

Nor98. James R. Norris. *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1998.

Par12. Alexandros G. Paraskevopoulos. The infinite gauss-jordan elimination on row-finite $\omega \times \omega$ matrices. *Arxiv preprint math*, 2012.

PG16. Amaury Pouly and Daniel S. Graça. Computational complexity of solving polynomial differential equations over unbounded domains. *Theoretical Computer Science*, 626:67–82, 2016.

Reu57. Gerd Edzard Harry Reuter. Denumerable markov processes and the associated contraction semigroups on l. *Acta Mathematica*, 97(1):1–46, 1957.

RR72. Gerd Edzard Harry Reuter and P. W. Riley. The Feller property for Markov semigroups on a countable state space. *Journal of the London Mathematical Society*, s2-5(2):267–275, August 1972.

Spi12. Flora M. Spieksma. Kolmogorov forward equation and explosiveness in countable state markov processes. *Annals of Operations Research*, pages 1–20, 2012.

Spi15. Flora M. Spieksma. Countable state markov processes: non-explosiveness and moment function. *Probability in the Engineering and Informational Sciences*, pages 1–15, 2015.

SSG15. David Schnoerr, Guido Sanguinetti, and Ramon Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 143(18), 2015.

TR03. Rajae Ben Taher and Mustapha Rachidi. On the matrix powers and exponential by the r-generalized fibonacci sequences methods: the companion matrix case. *Linear Algebra and its Applications*, 370:341–353, 2003.

WZ07. Klaus Weihrauch and Ning Zhong. Computable analysis of the abstract Cauchy problem in a Banach space and its applications I. *Mathematical Logic Quarterly*, 53(4-5):511–531, 2007.

A From Transition functions to q-matrices and back

Transition functions are usually a derived concept of (homogeneous) continuous-time Markov chains (CTMC) with countable state space $\mathbf{S}$ (see Definitions 13 and 14). However, taking them as first class entity (following Anderson's textbook [And91]) is more convenient for the exposition of the present paper as it allows one to work without any non-trivial measure theoretic prerequisites.[9] We have a countable state space $\mathbf{S}$, as usual.

Definition 10 (Transition function [And91, p. 5f.]). *An $\mathbb{R}_{\geq 0}$-indexed family $P_t = (p_{t,xy})_{x,y\in\mathbf{S}}$ of $\mathbf{S}\times\mathbf{S}$-matrices with non-negative, real entries $p_{t,xy}$ is called a* transition function on $\mathbf{S}$ *if*

1. $P_0 = I := (\delta_{xy})_{x,y\in\mathbf{S}}$*;*
2. $P_{t+s} = P_t P_s = \left(\sum_{z\in\mathbf{S}} p_{s,xz} p_{t,zy}\right)_{x,y\in\mathbf{S}}$ *for all* $s, t \in \mathbb{R}_{\geq 0}$*;*
3. $\sum_{z\in\mathbf{S}} p_{t,xz} \leq 1$ *for all* $x \in \mathbf{S}$ *and* $t \in \mathbb{R}_{\geq 0}$.

A transition function P_t is called honest *if* $\sum_{z\in\mathbf{S}} p_{t,xz} = 1$ *for all* $x \in \mathbf{S}$ *and* $t \in \mathbb{R}_{\geq 0}$ *and it is called* dishonest *if it fails to be honest; finally, it is called* standard *if* $\lim_{t\searrow 0} p_{t,xx} = 1$ *holds for all* $x \in \mathbf{S}$ *(and thus* $\lim_{t\searrow 0} p_{t,xy} = \delta_{xy}$ *holds for all* $x, y \in \mathbf{S}$*).*

The first two properties in Definition 10 make transition functions monoids of real-valued $\mathbf{S}\times\mathbf{S}$-matrices equipped with matrix multiplication; the last property requires that rows induce sub-probability measures.[10] Thus, an honest transition function P_t consists of proper probability matrices, i.e., the entries of every row sum up to 1. Unless explicitly allowed to be non-standard, we shall assume transition functions to be standard, i.e., entry-wise continuous at time 0.

[9] The reader familiar with CTMCs might object that in shifting focus from Markov chains to pairs of initial distributions and transition functions, we lose information. However, "[a]ll the probabilistic information about the process, insofar as it concerns only countably many time instants, is contained in the transition function and initial distribution. One could almost say that the transition function is the Markov chain." [And91, p. 4].

[10] A sub-probability measure on an at most countable set $\mathbf{S}$ (equipped with the σ-algebra of all subsets) can be safely identified with a function to the non-negative reals $f\colon \mathbf{S} \to \mathbb{R}_{\geq 0}$ that sums up to at most 1, i.e., such that $\sum_{x\in\mathbf{S}} f(x) \leq 1$ holds.

Example 3 (Yule-Furry process). The Yule-Furry process of parameter $\beta > 0$, the archetypal example of a branching process, can be thought of as a simplistic model of population growth; for $n_0 \leq n$, its transition probabilities have the following closed form.

$$p_{t,n_0\,n} = \binom{n-1}{n-n_0}(e^{-\beta t})^{n_0}(1-e^{-\beta t})^{n-n_0}$$

Using the metaphor of population growth, if we start with a population of size 1, the probability at time t to have grown to exactly size $n \geq 1$ is $e^{-\beta t}(1-e^{-\beta t})^{n-1}$.

Typically, transition functions of CTMCs do not have simple closed expressions for their entries; the Yule-Furry process is one of the few exceptions. In contrast, taking (time) derivatives of transition functions (at time zero) typically yields natural functions, intuitively corresponding to a flow of probability mass; for a Yule process of parameter β with initial population n_0, we think of n_0 independent individuals that can replicate at any given moment in time, and thus the outflow of probability mass is $-\beta n_0$.

Transition functions are always entry-wise differentiable [Kol51,Aus55] and the derivatives at time zero $p'_{0,xy}$ form the so-called *q-matrix* of the transition function [And91, p. 14]. Provided that all states are *stable*, i.e., $p'_{0,xx} \neq -\infty$ for all $x \in \mathbf{S}$, q-matrices of transition functions can safely be identified with stable q-matrices according to the following definition [And91, Theorem 2.2].

Definition 11 (q-Matrix [And91, p. 64]). *A* q-matrix *is an* $\mathbf{S} \times \mathbf{S}$*-matrix* $Q = (q_{xy})_{x,y\in\mathbf{S}}$ *with entries real numbers or* $-\infty$ *such that*

$$\infty > q_{xy} \geq 0 \text{ (if } x \neq y), \qquad -\infty \leq q_{xx} \leq 0, \text{ and } \qquad \textstyle\sum_{z\in\mathbf{S}} q_{xz} \leq 0$$

hold for all $x, y \in \mathbf{S}$*; it is* stable *(resp.* conservative*) if* $q_x := -q_{xx} < \infty$ *(resp.* $\sum_{z\in\mathbf{S}} q_{xz} = 0$*) holds for all* $x \in \mathbf{S}$.

In this paper, we assume q-matrices to be stable and conservative. For each q-matrix $Q = (q_{xy})_{x,y\in\mathbf{S}}$ on an at most countable state space $\mathbf{S}$, the *Kolmogorov backward equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}P_t = QP_t,\ P_0 = I \tag{4}$$

has a unique minimal non-negative solution $P_t = (p_{t,xy})_{x,y\in\mathbf{S}}$ (see, e.g., Theorem 2.2 of Ref. [And91]), which is a possibly dishonest transition function.

Definition 12 (Transition functions of q-matrices). *For a stable q-matrix Q on* $\mathbf{S}$*,* the transition function of q-matrix Q*, is the unique minimal transition function* P_t *solving the Kolmogorov backward equation, i.e., any other transition function* $\tilde{P}_t = (\tilde{p}_{t,xy})_{x,y\in\mathbf{S}}$ *that solves Equation* (4) *satisfies* $\tilde{p}_{t,xy} \geq p_{t,xy}$ *for all* $(t, x, y) \in \mathbb{R}_{\geq 0} \times \mathbf{S} \times \mathbf{S}$.

In case of a finite state space, all the above is trivial: the transition function P_t is just the matrix exponential e^{Qt}, and similarly, if the entries of the q-matrix are bounded. In general, the hard part is to find for a given q-matrix a corresponding transition function, while the opposite direction, passing from a transition function to its q-matrix, is always via differentiation [Kol51,Aus55]; however, with the usual assumptions on CTMCs (as in Norris's standard textbook [Nor98]), q-matrices are stable and conservative, which allows to pass back and forth between transition functions and q-matrices without complications [And91, Theorem 2.2].

B Markov chains

For completeness sake, we recall the classic definitions of continuous-time Markov chains in terms of a stochastic process satisfying the Markov property. However, note that all proofs rely on the functional analytic perspective described in Appendix A, which does not hinge on any non-trivial measure theory. The following definitions are based on [And91].

Definition 13 (Continuous-time Markov chain (CTMC)). *A stochastic process* $\{X_t\}_{t\in\mathbb{R}_{\geq 0}}$*, defined on a probability space* $(\Omega, \mathcal{F}, \Pr)$*, with values in a countable set* $\mathbf{S}$ *(to be called the* state space *of the process), is called a* continuous-time Markov chain (CTMC) *if for any finite set* $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq t_{n+1}$ *of "times", and corresponding set* $i_1, i_2, \ldots, i_{n-1}, i, j$ *of states in* $\mathbf{S}$ *we have*

$$\Pr\Big(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_{n-i}} = i_{n-1}, \ldots X_{t_1} = i_1\Big) = \Pr\Big(X_{t_{n+1}} = j \mid X_{t_n} = i\Big)$$

whenever

$$\Pr\Big(X_{t_n} = i, X_{t_{n-i}} = i_{n-1}, \ldots X_{t_1} = i_1\Big) > 0.$$

If for all $0 \leq s \leq t$ *and all* $i, j \in \mathbf{S}$ *the equation*

$$\Pr\Big(X_t = j \mid X_s = i\Big) = \Pr\Big(X_{t-s} = j \mid X_0 = i\Big)$$

holds, the stochastic process $\{X_t\}_{t\in\mathbb{R}_{\geq 0}}$ *is* homogeneous*.*

As described by Anderson [And91, p. 4], all probabilistic information about a CTMC is given by their transition function.

Definition 14 (Transition function of a CTMC). *Let* $\{X_t\}_{t\in\mathbb{R}_{\geq 0}}$ *be a homogeneous CTMC. The t-dependent matrix* $P_t = (p_{t,ij})_{i,j\in\mathbf{S}}$ *with*

$$p_{t,ij} = \Pr\big(X_t = j \mid X_0 = i\big)$$

for all $i, j \in \mathbf{S}$ *and* $t \in \mathbb{R}_{\geq 0}$ *is the* transition function *of the CTMC.*

We only consider homogeneous CTMCs that are stable, standard[11] and minimal (cf. Assumption 2.1 of Ref. [Spi12]).

Definition 15 (Stable, standard, minimal). *Let* $\{X_t\}_{t\in\mathbb{R}_{\geq 0}}$ *be a homogeneous CTMC. It is* stable *if the q-matrix of its transition function is stable, it is* standard *if its transition function is standard, and it is* minimal *if its transition function is the minimal transition function that solves the Kolmogorov (backward) equation* (4)*.*

[11] Note that Anderson only considers standard transition functions [And91, p. 6].

C Linear operators on Banach spaces

In the present paper, we only consider vector spaces over the field $\mathbb{R}$. We recall some core definitions (see also [EN00, Appendices A and B]).

Definition 16 (Banach space). *A Banach space is a complete normed vector space.*

We now give examples of Banach sequence spaces, i.e., Banach spaces consisting of functions $f\colon \mathbf{S} \to \mathbb{R}$ from a fixed set $\mathbf{S}$, which will always be the state space of some CTMC is this paper.

Definition 17 (C_0-spaces). *A function $f\colon \mathbf{S} \to \mathbb{R}$* vanishes at infinity *if for all $\epsilon > 0$, $f(x) < \epsilon$ holds for almost all $x \in \mathbf{S}$, i.e., the set $\{x \in \mathbf{S} \mid f(x) \geq \epsilon\}$ is finite. The Banach space $C_0(\mathbf{S})$ is the Banach space that consists of all functions $f\colon \mathbf{S} \to \mathbb{R}$ that vanish at infinity and its norm $\|\_\|$ is given by $\|f\| = \sup_{x\in\mathbf{S}} |f(x)|$.*

The duals of C_0-spaces exists and are isomorphic to L^1-spaces; however, C_0-spaces are in general not isomorphic to their double dual (which is isomorphic to the corresponding L^∞-space).

Definition 18 (L^1-spaces). *A function $f\colon \mathbf{S} \to \mathbb{R}$ is* summable *if the sum $\sum_{x\in\mathbf{S}} |f(x)|$ converges (for some order of summation). The Banach space $L^1(\mathbf{S})$ consists of summable functions and its norm $\|\_\|$ is given by $\|f\| = \sum_{x\in\mathbf{S}} |f(x)|$.*

Definition 19 (Linear operators, bounded linear operators). *Let $\mathcal{B}$ be a Banach space. A* linear operator *on $\mathcal{B}$ is a pair $(A, \mathcal{D}(A))$ where $\mathcal{D}(A) \subseteq \mathcal{B}$ is a linear subspace, called* domain*, and A is a linear map $A\colon \mathcal{D}(A) \to \mathcal{B}$. A linear operator on $\mathcal{B}$ is* bounded *if there exists a constant $K \in \mathbb{R}_{\geq 0}$ such that $\|A(u)\| \leq K\|u\|$ holds for all $u \in \mathcal{D}(A)$. We denote by $\mathcal{L}(\mathcal{B})$ the set of all linear operators $(A, \mathcal{D}(A))$ on $\mathcal{B}$ that are bounded and defined on the whole space, i.e., satisfy $\mathcal{D}(A) = \mathcal{B}$.*

D Abstract Cauchy problems of SCSGs

Every transition function $P_t = (p_{t,xy})_{x,y\in\mathbf{S}}$ (of some CTMC X_t with state space $\mathbf{S}$) induces a partial function on the vector space $\mathbb{R}^{\mathbf{S}}$ that maps a function $f\colon \mathbf{S} \to \mathbb{R}$ to the function $P_t f := (x \mapsto \sum_{y\in\mathbf{S}} p_{t,xy} f(y))$ (of the same "type" $\mathbf{S} \to \mathbb{R}$) whenever the respective sums converge absolutely for all states $x \in \mathbf{S}$, and it is undefined on f otherwise. This function is indeed always a linear operator on the vector space of bounded functions [And91, Sect. 1, Lemma 4.4.]. However, we are interested in a stronger property (Theorem 1).

Definition 20 (Strongly continuous semigroup). *Let $\mathcal{B}$ be a Banach space, and let $\mathcal{L}(\mathcal{B})$ be the set of bounded linear operators on $\mathcal{B}$ with $\mathcal{B}$ as domain of definition. A strongly continuous semi-group (SCSG) on $\mathcal{B}$ is a family $\{P_t\}_{t\in\mathbb{R}_{\geq 0}}$ of bounded linear operators $P_t \in \mathcal{L}(\mathcal{B})$ satisfying*

1. $P_0 = I_{\mathcal{B}}$, *the identity on* $\mathcal{B}$;
2. $P_{t+s} = P_t P_s$ *for all* $s, t \in \mathbb{R}_{\geq 0}$; *and*
3. $\lim_{h \searrow 0} \|P_h f - f\| = 0$ *holds for all* $f \in \mathcal{B}$.[12]

In the finite state case, each transition functions P_t can be obtained from their q-matrix Q – simply by taking the matrix exponential $P_t = e^{tQ}$. In general, a SCSG is only the limit of a sequence of matrix exponentials of bounded operators, the so-called Yosida-approximants (see, e.g. Ref. [EN00]). Part of the intricacy is explained by the fact that the time derivative $\frac{\mathrm{d}}{\mathrm{d}t} P_t f$ at 0, i.e., the limits $\lim_{h \searrow 0} {}^1\!/\!{}_h(P_h f - f)$, do not exists for all elements f of the Banach space. This brings us to a new object, which is typically related but in general not identical to the q-matrix (cf. [Reu57, Theorem 2]).

Definition 21 (Infinitesimal generator). *The* infinitesimal generator $\mathcal{Q}$ *of a strongly continuous semigroup* P_t *on a Banach space* $\mathcal{B}$ *is defined by*

$$\mathcal{Q}f = \lim_{h \searrow 0} {}^1\!/\!{}_h(P_h f - f)$$

for all $f \in \mathrm{dom}(\mathcal{Q}) = \left\{ f \in \mathcal{B} \;\middle|\; \textit{The limit } \lim_{h \searrow 0} {}^1\!/\!{}_h(P_h f - f) \textit{ exists.} \right\}$.

Infinitesimal generators determine the semigroup uniquely [EN00, Theorem 1.4]. To understand why, note first that it is sufficient to describe each P_t on a dense subset of the Banach space; the domain of definition of the infinitesimal generator is simply the canonical such subset. Finally, for each f in the domain of the domain of the generator, the function $t \mapsto P_t f$ is characterised as the unique differentiable solution to the abstract Cauchy problem [Ein52] in Equation (5).

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u_t &= \mathcal{Q} u_t \quad (t \geq 0) \\ u_0 &= f \end{aligned} \tag{5}$$

E Drift functions for observables

Powers of length are drift functions (Lemma 2)

> Let $\mathcal{R} \subseteq \Sigma^+ \times \Sigma^+$ be a finite string rewriting system and let $n \in \mathbb{N}^+$ be a positive natural number. There exists a constant $c_n \in \mathbb{R}_{>0}$ such that $|\_|^n \colon \Sigma^+ \to \mathbb{R}_{\geq 0}$ is a c_n-drift function.

Proof. Let $K = \max_{l \to r \in \mathcal{R}} |r|$; it is a (rough) bound for the maximal length "growth" of rules in $\mathcal{R}$. Define $c_n = (2K)^n |\mathcal{R}|$. For all words $v \in \Sigma^+$ and all rules $l \to r \in \mathcal{R}$,

$$(Q_{l \to r}|n|)(v) = \sum_{w \in \Sigma^+} q^{l \to r}_{vw} |w|^n \qquad \text{(by definition of } Q_{l \to r}|n|\text{)}$$

[12] This last condition, in presence of the first two, is equivalent to continuity of the function $t \mapsto P_t$, from $\mathbb{R}_{\geq 0}$ to $\mathcal{L}(\mathcal{B})$, if we consider $\mathcal{L}(\mathcal{B})$ with the strong operator topology.

$$
\begin{aligned}
&= \sum_{w \in \Sigma^+ \setminus \{v\}} q_{vw}^{l \to r} |w|^n - \underbrace{\left(\sum_{w \in \Sigma^+ \setminus \{v\}} q_{vw}^{l \to r} \right)}_{q_v^{l \to r}} |v|^n \quad \text{(by conservativity)} \\
&= \sum_{w \in \Sigma^+ \setminus \{v\}} q_{vw}^{l \to r} (|w|^n - |v|^n) \\
&\leq \sum_{w \in \Sigma^+ \setminus \{v\}} q_{vw}^{l \to r} ((|v| + K)^n - |v|^n) \qquad \text{(by definition of } K\text{)} \\
&\leq ((|v| + K)^n - |v|^n) \sum_{w \in \Sigma^+ \setminus \{v\}} q_{vw}^{l \to r} \\
&\leq |v|((|v| + K)^n - |v|^n) \qquad (\leq |v| \text{ redexes of } l \to r \text{ in } v) \\
&= |v| \sum_{k=1}^{n} \binom{n}{k} |v|^{n-k} K^k \\
&\qquad \text{(by Binomial Theorem and cancelling } -|v|^n\text{)} \\
&\leq |v| \sum_{k=1}^{n} \binom{n}{k} |v|^{n-k} K^k \\
&\leq |v| \sum_{k=1}^{n} \binom{n}{k} |v|^{n-k} K^k |v|^{k-1} K^{n-k} \\
&= |v| \sum_{k=1}^{n} \binom{n}{k} |v|^{n-1} K^n \\
&\leq 2^n K^n |v|^n
\end{aligned}
$$

The desired follows since $Q_{\mathcal{R}} = \sum_{l \to r \in \mathcal{R}} Q_{l \to r}$.

F Spieksma, *corrigendum*

The statement of Theorem 6.3 of Ref. [Spi12] contains an error (corrected in our version) that can be spotted if one tries to fill in the details that are left out at the end of Spieksma's proof in *op. cit.* A counterexample is given in Appendix I. The following proposition gives a useful and correct weakening of her statement, on a non-empty subset of the domain of the generator of the relevant SCSGs.

Proposition 2. *Let P_t be a transition function on $\mathbf{S}$ with q-matrix Q such that there exists a positive function $W\colon \mathbf{S} \to \mathbb{R}_{>0}$ and a constant θ such that $QW \leq \theta W$, (i.e., $(QW)(x) \leq \theta W(x)$ for all $x \in \mathbf{S}$). Let the family $\{P_t\}_{t \in \mathbb{R}_{\geq 0}}$ be a SCSG on $C_0(\mathbf{S}, W)$, witnessed by $V \in C_0(\mathbf{S}, W)$ that is positive and a drift function for Q.*

For all $f \in C_0(\mathbf{S}, W)$ that satisfy $\|f\|_V < \infty$ and $Qf \in \mathcal{D}(Q)$, we have

$$Qf = \mathcal{Q}f = \lim_{h \searrow 0} {}^1\!/\!_h (P_t f - f), \tag{6}$$

i.e., the latter limit exists in $C_0(\mathbf{S}, W)$, and in particular $f \in \mathcal{D}(\mathcal{Q})$.

G Computable analysis

We collect the main definitions and terminology from [BHW08], as required in [WZ07].

G.1 Naming schemes for sets

The central point of computable analysis in the tradition of the type-2 theory of effectivity (aka TTE) are so-called representations of elements of sets (that later will always come equipped with a topology).

Definition 22 (Notations, representations, naming systems). *Let X be a set, and let Σ be finite alphabet. A* notation *of X is a surjective partial function $\nu :\subseteq \Sigma^* \to X$, and, similarly, a* representation *is a surjective partial function $\delta :\subseteq \Sigma^\omega \to X$; finally, a* naming system *for X is a notation or a representation of X. Given an element $x \in X$ and a naming system of X, a* γ-name *is an element of $\gamma^{-1}(\{x\})$ (which thus is a finite or infinite string over Σ that is mapped to x).*

G.2 Computable reals

There are several equivalent definitions of computable real numbers [BHW08, Theorem 3.2].

Definition 23 (Computable real [BHW08, Definition 3.1]). *A real number $x \in \mathbb{R}$ is* computable *if $\{q \in \mathbb{Q} \mid q < x\}$ is a decidable set (of rational numbers).*

Definition 24 (Computable sequence of reals [BHW08, Definition 3.8]). *A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is called* computable *if there exists a computable sequence $(q_k)_{k\in\mathbb{N}}$ of rational numbers such that $|x_n - q_{\langle n,i\rangle}| < 2^{-i}$ for all $n, i \in \mathbb{N}$.*

Definition 25 ([BHW08, Definition 3.11.]).

1. *If $(r_n)_{n\in\mathbb{N}}$ is a convergent sequence of real numbers with limit x, and $m : \mathbb{N} \to \mathbb{N}$ is a function such that $i \geq m(n)$ implies $|x - r_i| < 2^{-n}$ (for all $i, n \in \mathbb{N}$), we call the function m a* modulus of convergence *of the sequence $(r_n)_{n\in\mathbb{N}}$.*
2. *A sequence $(r_n)_{n\in\mathbb{N}}$ of real numbers* converges computably *if it has a computable modulus of convergence.*

G.3 Computable functions on infinite sets

A function $f :\subseteq \mathbb{R}^n \to \mathbb{R}$ is *computable* if there is an oracle Turing machine that, given any $k \in \mathbb{N}$, may ask for arbitrarily good rational approximations of the input $x \in \mathrm{dom}(f)$;[13] after finitely many steps, it writes a rational number q on the output tape with $|f(x) - q| < 2^{-k}$ [BHW08, Definition 4.1.]. However, for the purposes of computability, instead of using oracle Turing machines, it suffices to consider a simpler model, namely Turing machines with infinite I/O, i.e., Turing machines equipped with one infinite input and one infinite output tape that are read only and one-way write only, respectively. This computation model allows for a natural notion of computability on infinite strings.

Definition 26 (Computable stream functions [BHW08, Definition 4.7]). *Let Σ be an alphabet. A function $F :\subseteq \Sigma^\omega \to \Sigma^\omega$ is* computable *if there exists a Turing machine that, given a stream $p \in \mathrm{dom}(F)$ on the input tape, writes the output stream $F(p)$ on the one-way output tape, and that, given a stream $p \in \Sigma^\omega$ $\mathrm{dom}(F)$ does not write infinitely many symbols on the output tape (and instead moves to the halting state at some point).*

If X is a space, a (computable) representation *of X is a (computable) map $\delta_X :\subseteq \Sigma^\omega \to X$.*

A particular representation that we shall invariably use for $\mathbb{R}$ with the usual (Euclidean metric) is ρ, given as follows: Let, for each $n \in \mathbb{N}$, $\mathbb{Q}_n = \{m2^- - n : m \in \mathbb{Z}\}$, and let $\mathbb{Q}_D = \bigcup_n \mathbb{Q}_n$ be the set of dyadic numbers. Let $\nu_D : \mathbb{N} \to \mathbb{Q}_D$ be any computable bijective numbering such that if $\langle \cdot, \cdot, \cdot \rangle : \mathbb{N}^3 \to \mathbb{N}$ is a computable pairing function, we have $\nu_D(\langle i, j, k \rangle) = (i - j)2^{-k}$. The representation $\rho : \{0,1\}^* \to \mathbb{R}$ is now defined by

$$\mathrm{dom}(\rho) = \{p \in \{0,1\}^* : \forall k.(\nu_D(p(k) \in \mathbb{Q}_k \wedge |\nu_D(p(k)) - \nu_D(p(k+1)| <^{-(k+1)})))\}$$

and

$$\rho(p) = \lim_n \nu_D(p(n)$$

Note that ρ is merely a representation that uses rapidly converging Cauchy sequences and that the choice of $\{0,1\}^*$ is immaterial: we could have chosen a larger alphabet, or even a unary alphabet if need be.

The above definitions carry over with ease to computation between arbitrary sets w.r.t. suitable representations.

Definition 27 (Computable function [BHW08, Definition 4.7]). *Let Σ be an alphabet, and let $f :\subseteq X \to Y$ be a function; moreover, let $\delta_X :\subseteq \Sigma^\omega \to X$ and $\delta_Y :\subseteq \Sigma^\omega \to Y$ be representations. A function $F :\subseteq \Sigma^\omega \to \Sigma^\omega$ is called a (δ_X, δ_Y)-*realizer *of f if $\delta_Y F(p) = f\delta_X(p)$ holds for all $p \in \mathrm{dom}(f\delta_X)$, i.e., such that for any δ_X-name of some $x \in \mathrm{dom}(f)$, the value $F(p)$ is a δ_Y-name of $f(x)$. The function f is (δ_X, δ_Y)-*computable *if a computable (δ_X, δ_Y)-realizer of f exists.*

[13] Thus, the oracle Turing machine may ask finitely many questions of the kind "Give me a vector $p \in \mathbb{Q}^n$ of rational numbers with $d(x, p) < 1/2^i$," where the exponent i may depend on the answers to the previous questions.

G.4 Computable metric spaces

Definition 28 (Computable metric space [BHW08, Definition 7.1]). *A triple $\mathcal{B} = (X, d, \alpha)$ is called a* computable metric space*, if*

1. *$d\colon X \times X \to \mathbb{R}$ is a metric on X,*
2. *$\alpha\colon \mathbb{N} \to X$ is a sequence such that the set $\{\alpha(n) \mid n \in \mathbb{N}\}$ is dense in X,*
3. *$d \circ (\alpha \times \alpha)\colon \mathbb{N}^2 \to \mathbb{R}$ is a computable map (with $\mathbb{R}$ equipped with the representation ρ).*

Observe in the above definition that countable metric spaces are separable spaces and they come with a choice of a dense subset via an enumeration.

Definition 29 (Cauchy representation, rapid convergence [BHW08, Definition 7.2]). *Let (X, d, α) be a computable metric space. Then we define the* Cauchy representation *$\delta_X :\subseteq \Sigma^\omega \to X$ by*

$$\delta_X(01^{n_0+1}01^{n_1+1}01^{n_2+1}\ldots) := \lim_{i\to\infty} \alpha(n_i)$$

for any sequence $(n_i)_{i\in\mathbb{N}}$ such that $(\alpha(n_i))_{i\in\mathbb{N}}$ converges rapidly (and $\delta_X(p)$ is undefined for all other input sequences $p \in \Sigma^\omega$) where a sequence $(x_i)_{i\in\mathbb{N}}$ converges rapidly*, if it converges and $d(x_i, \lim_{n\to\infty} x_n) < 2^{-i}$ for all $i \in \mathbb{N}$.*

It is easily seen that the earlier representation ρ of $\mathbb{R}$ can easily be made into a Cauchy representation in the above sense.

Definition 30 (Computable Banach space). *A tuple $\mathcal{B} = (B, \|\_\|, \alpha)$ is called a* computable Banach *space if*

1. *$(B, \|\_\|)$ is a Banach space,*
2. *(B, d, α) such that $d(x, y) = \|x - y\|$ is a computable metric space (using the representation ρ of $\mathbb{R}$ and the Cauchy representation δ_B of B*
3. *addition $(x, y) \mapsto x + y$ and scalar multiplication $(a, x) \mapsto ax$ are computable operations (with the above representations).*

H Computability of duality pairing

Weighted C_0-spaces $C_0(\mathbf{S}, W)$ and their duals $L^1(\mathbf{S}, W)$ with recursively enumerable $\mathbf{S}$ will be equipped with a recursive enumeration α of all finite rational linear combinations of indicator functions $\mathbb{1}_x$ of states $x \in \mathbf{S}$ (observe that the set of such finite linear combinations is countable and can obviously be enumerated by an algorithm). As the indicator functions themselves are computable, and the map $W \to \mathbb{Q}^+$ is computable, it follows easily that both spaces are computable Banach spaces. Moreover, we equip the product space $L^1(\mathbf{S}, W) \times C_0(\mathbf{S}, W)$ with the sum of the norms of the underlying spaces, i.e., $\|(\pi, f)\| = \|\pi\|_W^* + \|f\|_W$ where $\|f\|_W = \sup_{x\in\mathbf{S}} |f(x)|/W(x)$ and $\|\pi\|_W^* = \sum_{x\in\mathbf{S}} W(x)|\pi(x)|$. Finally, we choose a recursive enumeration of all linear combinations of the set

$$\bigcup_{x\in\mathbf{S}} \{(\mathbb{1}_x, 0), (0, \mathbb{1}_x)\}$$

to make the product space a computable Banach space.

Lemma 3 (Computability of duality pairing). *Let* $\mathbf{S}$ *be a recursively enumerable set, and let* $W\colon \mathbf{S} \to \mathbb{Q}^+$ *be a computable function. The dual pairing*

$$\langle , \rangle \colon L^1(\mathbf{S}, W) \times C_0(\mathbf{S}, W) \to \mathbb{R},$$

mapping (π, f) *to* $\sum_{x\in\mathbf{S}} \pi(x)f(x)$*, is computable.*

Proof. For any two elements $(\pi, f), (\pi', f') \in L^1(\mathbf{S}, W) \times C_0(\mathbf{S}, W)$ of the product space, if π' is finitely supported, i.e., the set of states $K = \{x \in \mathbf{S} \mid \pi'(x) \neq 0\}$ is finite, we have the following estimate for the distance of $\langle \pi', f'\rangle$ and $\langle \pi, f\rangle$ in $\mathbb{R}$.

$$\begin{aligned}
|\langle \pi', f'\rangle - \langle \pi, f\rangle| &= |\sum_{x\in\mathbf{S}} \pi'(x)f'(x) - \pi(x)f(x)| && (7)\\
&\leq \sum_{x\in\mathbf{S}} |\pi'(x)f'(x) - \pi(x)f(x)| \\
&= \sum_{x\in K} |\pi'(x)f'(x) - \pi(x)f(x)| \\
&\quad + \sum_{x\in\mathbf{S}\setminus K} |0 - \pi(x)f(x)| && \text{(by definition of } K)\\
&= \sum_{x\in K} |\pi'(x)f'(x) - \pi(x)f(x)| \\
&\quad + \sum_{x\in\mathbf{S}\setminus K} \left| W(x)\pi(x)\frac{f(x)}{W(x)} \right| \\
&\leq \sum_{x\in K} |\pi'(x)f'(x) - \pi(x)f(x)| \\
&\quad + \|f\|_W \sum_{x\in\mathbf{S}\setminus K} |W(x)\pi(x)| && \text{(by definition of } \|f\|_W)\\
&\leq \sum_{x\in K} |\pi'(x)f'(x) - \pi(x)f(x)| \\
&\quad + \|f\|_W \sum_{x\in\mathbf{S}\setminus K} |W(x)(\pi(x) - \underbrace{\pi'(x)}_{0})| \\
&\leq \sum_{x\in K} |\pi'(x)f'(x) - \pi(x)f(x)| && (8)\\
&\quad + \|f\|_W \|\pi - \pi'\|_W^* && \text{(by definition of } \|_\|_W^*)
\end{aligned}$$

Moreover, putting $e_x = f(x) - f'(x)$ and $\phi_x = \pi(x) - \pi'(x)$,

$$\begin{aligned}
\pi(x)f(x) &= \pi(x)(f'(x) + e_x) \\
&= \pi(x)e_x + \pi(x)f'(x) \\
&= \pi(x)e_x + (\phi_x + \pi'(x))f'(x) \\
&= \pi(x)e_x + \phi_x f'(x) + \pi'(x)f'(x). && (9)
\end{aligned}$$

Next, we describe an algorithm that for each desired precision $\epsilon = 2^{-d}$ as input, first, makes three successive requests for approximations (π', f') of (π, f) of ascending precision based on the estimate from (8) and Equation (9), followed by computation of a rational approximation r of $\langle \pi, f\rangle$, before finally, giving the result r as output by writing up to $d+1$ digits of precision in binary representation.

1. Request (π_1, f_1) as approximation of (π, f) up to precision 1. Store $F' = 2\|f_1\|_W + 1$; now we have $F' > \|f\|_W$ as a save over-approximation of the norm $\|f\|_W$ of f.
2. Request (π_2, f_2) as approximation of (π, f) up to precision $\min\{\epsilon/3F', 1\}$). This ensures $\|f_2\|_W \|\pi - \pi_2\|^*_W < \epsilon/3$ (cf. (8)). Moreover, with $\psi_x = \pi(x) - \pi_2(x)$,

$$\begin{aligned}\sum_{x\in\mathbf{S}} |\psi_x f_2(x)| &= \sum_{x\in\mathbf{S}} W(x)|\psi_x|\ |f_2(x)|/W(x) \\ &\leq \|f_2\|_W \sum_{x\in\mathbf{S}} |W(x)\psi_x| \\ &= \|f_2\|_W\|\pi - \pi_2\|^*_W \\ &< \epsilon/3 \end{aligned} \tag{10}$$

3. Compute the support of π_2, i.e., $K_2 = \{x \in \mathbf{S} \mid \pi_2(x) \neq 0\}$ using the representation of (π_2, f_2). Compute $M = \max_{x\in K_2} W^{-1}(x)$.
4. Request (π_3, f_3) up to precision $\min\{\epsilon/3M2\|\pi_2\|^*_W, \epsilon/3F', 1\}$.

$$\begin{aligned}\sum_{x\in K_2} |\pi(x)(f(x) - f_3(x))| &\leq \|f - f_3\|_W \sum_{x\in K_2} |\pi(x)| \\ &\leq \epsilon/3M2\|\pi_2\|^*_W \sum_{x\in K_2} |W(x)\pi(x)/W(x)| \\ &\leq \epsilon/3M2\|\pi_2\|^*_W M \sum_{x\in K_2} |W(x)\pi(x)| \\ &\leq \epsilon/3M2\|\pi_2\|^*_W M \sum_{x\in K_2} W(x)|\pi(x)| \\ &\leq \epsilon/3M2\|\pi_2\|^*_W M\|\pi\|^*_W \\ &\leq \epsilon/3M2\|\pi_2\|^*_W M2\|\pi_2\|^*_W \\ &= \epsilon/3 \end{aligned} \tag{11}$$

5. Compute

$$r = \langle \pi_2, f_3\rangle$$

   thus, the second approximation of π_2 to keep the support K_2, and the best approximation of f_3.
6. Return enough digits of r in the dyadic representation, $d+1$ after the dyadic point.

To show that the algorithm is correct, we derive as in (8)

$$\begin{aligned}|r - \langle \pi, f\rangle| &= |\langle \pi_2, f_3\rangle - \langle \pi, f\rangle| \\ &\leq \textstyle\sum_{x\in K} |\pi_2(x) f_3(x) - \pi(x) f(x)| \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^*\end{aligned}$$

and then, as in 9, we obtain

$$\begin{aligned}&\leq \textstyle\sum_{x\in K} \Big|\pi_2(x) f_3(x) - \\ &\quad \Big(\pi(x)(f_3(x) - f(x)) + \psi_x f_3(x) + \pi_2(x) f_3(x)\Big)\Big| \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^* \\ &= \textstyle\sum_{x\in K} |-\pi(x)(f_3(x) - f(x)) - \psi_x f_3(x)| \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^* \\ &\leq \textstyle\sum_{x\in K} |\pi(x)(f_3(x) - f(x))| + |\psi_x f_3(x)| \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^* \\ &\leq \textstyle\sum_{x\in K} |\pi(x)(f_3(x) - f(x))| + |\psi_x f_3(x)| \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^*\end{aligned}$$

and then as in (10) and (11)

$$\begin{aligned}&\leq {}^{\epsilon}\!/_{3M2\|\pi_2\|_W^*} M2\|\pi_2\|_W^* + \|f_3\|_W \|\pi - \pi_2\|_W^* \\ &\qquad + \|f\|_W \|\pi - \pi_2\|_W^* \\ &\leq {}^{\epsilon}\!/_{3} + \|f_3\|_W {}^{\epsilon}\!/_{3F'} + \|f\|_W {}^{\epsilon}\!/_{3F'} < \epsilon.\end{aligned}$$

I Counterexample [Spi12, Theorem 6.3]

Example 4 (Counterexample to Theorem 6.3 of Ref. [Spi12]). Consider the positive natural numbers $\mathbb{N} \setminus \{0\}$ as state space of a CTMC with q-matrix Q on $\mathbf{S} = \mathbb{N} \setminus \{0\}$ given by

$$q_{xy} = \begin{cases} x^2 & y = x - 1 \\ -x^2 & y = x \neq 1 \\ 0 & y = x = 1 \end{cases} \tag{12}$$

and consider the function $W(x) = x^2$, which is positive as $x > 0$. Now, W is a drift function since, for $x > 1$, we have

$$\begin{aligned}QW(x) &= \sum_{y\in\mathbf{S}} q_{xy} W(y) \\ &= \sum_{y=x-1}^{x} q_{xy} W(y) \\ &= q_{x(x-1)}\big(W(x-1) - W(x)\big) \\ &= q_{x(x-1)}\big((x-1)^2 - x^2\big) \\ &< 0 \leq W(x)\end{aligned}$$

and $QW(1) = 0 \leq W(1)$. Thus, Q is a drift function.
Next, the identity function

$$V(x) = x$$

on $\mathbf{S}$ belongs to $C_0(\mathbf{S}, W)$ and is positive. Moreover,

$$\begin{aligned} QV(x) &= q_{x(x-1)}\big(V(x-1) - Vx)\big) \\ &= q_{x(x-1)}\big(x - 1 - x)\big) \\ &= -q_{x(x-1)} \\ &= -x^2 \end{aligned}$$

and $QV(1) = 0$, whence

$$QV(x) = -(x)^2 \cdot \mathbb{1}_{\mathbf{S}\setminus\{1\}}(x).$$

This implies that V is also a drift function and hence the minimal transition function P_t with q-matrix Q is a SCSG on $C_0(\mathbf{S}, W)$, applying the first part of Theorem 6.3 of Ref. [Spi12].

Finally, we have $\|V\|_V = 1 < \infty$ an also $\|QV\|_W = \|-W\|_W = 1 < \infty$. Thus, by the second part of Theorem 6.3 of Ref. [Spi12], we conclude that V belongs to the domain $\mathcal{D}(\mathcal{Q})$ of the generator $\mathcal{Q}$. This means by definition that the limit

$$\lim_{h \searrow 0} \frac{P_h V - V}{h}$$

exists in $C_0(\mathbf{S}, W)$, thus $V \in \mathcal{D}(\mathcal{Q})$, and $\mathcal{Q}V \in C_0(\mathbf{S}, W)$. Now $QV = \mathcal{Q}V$ in analogy to Sect. 1.4 of Ref [And91] which leads to $QV \in C_0(\mathbf{S}, W)$, which is false: $QV = -(\cdot)^2 \cdot \mathbb{1}_{\mathbf{S}\setminus\{1\}}$ and $-(\cdot)^2 \cdot \mathbb{1}_{\mathbf{S}\setminus\{1\}} \notin C_0(\mathbf{S}, (\cdot)^2)$!