Journal of Computational Physics 322 (2016) 400–414

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Variance-reduced simulation of lattice discrete-time Markov chains with applications in reaction networks

P.A. Maginnis *, M. West, G.E. Dullerud

Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green St., Urbana, IL, 61801, USA

ARTICLE INFO

Article history:
Received 3 October 2015
Received in revised form 23 May 2016
Accepted 13 June 2016
Available online 23 June 2016

Keywords:
Stochastic simulation
Variance reduction
Tau-leaping
Antithetic sampling
Monte Carlo
Reaction networks

ABSTRACT

We propose an algorithm to accelerate Monte Carlo simulation for a broad class of stochastic processes. Specifically, the class of countable-state, discrete-time Markov chains driven by additive Poisson noise, or lattice discrete-time Markov chains. In particular, this class includes simulation of reaction networks via the tau-leaping algorithm. To produce the speedup, we simulate pairs of fair-draw trajectories that are negatively correlated. Thus, when averaged, these paths produce an unbiased Monte Carlo estimator that has reduced variance and, therefore, reduced error. Numerical results for three example systems included in this work demonstrate two to four orders of magnitude reduction of mean-square error. The numerical examples were chosen to illustrate different application areas and levels of system complexity. The areas are: gene expression (affine state-dependent rates), aerosol particle coagulation with emission and human immunodeficiency virus infection (both with nonlinear state-dependent rates). Our algorithm views the system dynamics as a "black-box", i.e., we only require control of pseudorandom number generator inputs. As a result, typical codes can be retrofitted with our algorithm using only minor changes. We prove several analytical results. Among these, we characterize the relationship of covariances between paths in the general nonlinear state-dependent intensity rates case, and we prove variance reduction of mean estimators in the special case of affine intensity rates.

1. Introduction

The objective of this work is to reduce the computational cost of Monte Carlo mean estimation of lattice discrete-time Markov chains (lattice DTMCs), a class which includes countable state-space Markov processes governed by a finite number of reaction classes. This goal is achieved by construction of anticorrelated ensembles of sample path realizations of lattice DTMCs which produce unbiased, variance-reduced mean estimates of the lattice DTMC distribution. The benefit of reduced variance mean estimators is a reduction in the number of sample paths needed to produce equally accurate estimates, and is particularly necessary when simulating large systems. For example, we show two to four orders of magnitude mean-square error (MSE) reduction for mean estimates of the tau-leaping distributions for three systems in the numerical examples we present for high volume regimes. This corresponds to a one to two orders of magnitude decrease in the number of sample paths needed to estimate the mean behavior of these discrete-time systems at fixed accuracy. The proposed methods require

* Corresponding author.
E-mail addresses: maginni1@illinois.edu (P.A. Maginnis), mwest@illinois.edu (M. West), dullerud@illinois.edu (G.E. Dullerud).

http://dx.doi.org/10.1016/j.jcp.2016.06.019
0021-9991/

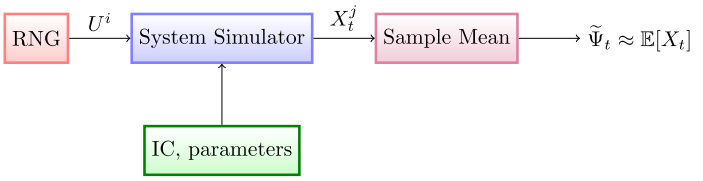

Fig. 1. Traditional Monte Carlo mean estimation using iid stochastic simulation. Here, iid sequences $\{U^1, U^2, \dots\}$ of standard uniform variates from a random number generator (RNG) drive a particular discrete-time process simulation. These sources of randomness, combined with initial conditions (IC) and system parameters, govern the system evolution, and are used to produce a sequence of iid sample paths $\{X_t^j\}_{j=1}^N$. This sequence is then used to compute the sample mean, $\widetilde{\Psi}_t = \frac{1}{N}\sum_{j=1}^N X_t^j$, in order to approximate $\mathbb{E}[X_t]$, the true mean of the process.

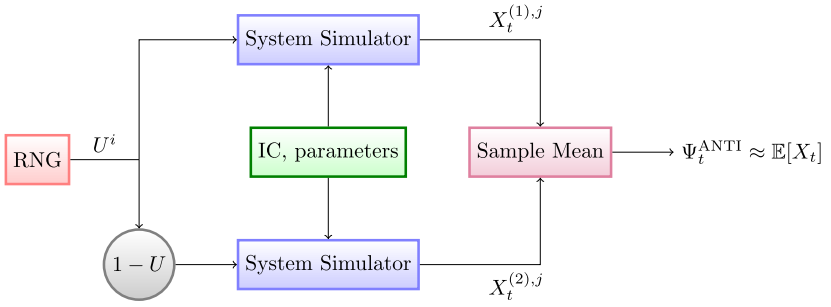

Fig. 2. Variance-reduced Monte Carlo mean estimation using antithetic stochastic simulation as an alternative to the iid Monte Carlo estimation architecture shown in Fig. 1. Again, a single stream of standard uniform variates $\{U^1, U^2, \dots\}$ drive a particular discrete-time process simulation to produce a sequence of iid sample paths $\{X_t^{(1),j}\}_{j=1}^N$. Additionally, however, the corresponding antithetic standard uniform sequence $\{1-U^1, 1-U^2, \dots\}$ (itself a sequence of uniform variables on $[0,1]$) is used as input for another realization of the same discrete-time process simulation with the same initial conditions and reaction rates to produce another iid sequence $\{X_t^{(2),j}\}_{j=1}^N$ of sample paths with identical marginal distribution. The most important feature of these sequences is that, for any j, $X_t^{(1),j}$ and $X_t^{(2),j}$ are correlated for each t. When this correlation is negative, their sample mean, $\Psi_t^{\text{ANTI}} = \frac{1}{2N}\sum_{j=1}^N \left(X_t^{(1),j} + X_t^{(2),j}\right)$, will be an unbiased estimator of the true mean $\mathbb{E}[X_t]$ and will have lower variance than $\widetilde{\Psi}_t$.

only the manipulation of uniform random inputs to the "black-box" system dynamics, and thus are easily implemented in any typical code. This feature of our approach is similar to that of other numerical algorithms that incorporate legacy simulation codes in a modular fashion, such as the recursive projection method [1] and equation-free methods [2]. The flow of traditional independent, identically distributed (iid) Monte Carlo simulation for mean estimation is shown in Fig. 1, and is contrasted with Monte Carlo driven using our antithetically paired sample paths, shown in Fig. 2.

We reduce the computational cost by simulating an ensemble of paths $\{X^{(r)}\}$ so that they are statistically dependent on each other, in such a way that they are identically distributed but are now negatively correlated. This contrasts with traditional Monte Carlo, where sample paths are required to be iid. Negative covariances (anticorrelation) between the simulations reduce the MSE (or variance) of the mean estimate. We produce negatively correlated ensembles of sample paths by generalizing the well established variance reduction technique of antithetic sampling [3], often used in the simulation of static random variates, to the dynamic, multi-dimensional realm of stochastic processes. Of particular interest are mean estimators that dominate traditional iid Monte Carlo estimates, meaning unbiased estimators whose MSE is lower than iid estimators over any operating parameters. Dominant mean estimators avoid the need to tune a given technique to a particular application and also provide performance guarantees. Over such a general class of models (with no apparent exploitable symmetries), a generalized variance reduction via anticorrelation is challenging; analytical guarantees are difficult to achieve for the same reason that Monte Carlo simulation is necessary, namely that such models are often analytically intractable.

Our approach differs from well-known variance reduction techniques that are interested in rare events, such as importance sampling [4] or restarting type methods [5], in that it is focused primarily on typical behavior as opposed to rare events and that it involves no change of measure. The marginal distributions of paths produced are identical to normal stochastic simulation of lattice DTMCs, and it is only correlation between some samples that is altered [6]. Two key benefits of this approach are: 1) it requires negligible computational overhead above generic tau-leaping to operate; and 2) it is compatible with existing codes in that it requires no change to the simulation processes. Both of these features are due to the method only requiring accounting for and manipulation of random uniform inputs to the process. Note also that, by passing the seed of the PRNG as an argument to each simulation call, we require effectively no additional memory and still maintain the so-called embarrassing parallelizability that is a key feature of traditional iid Monte Carlo simulation for large-scale applications.

The class of lattice CTMCs (e.g., discrete-space, continuous-time Markov processes governed by a finite number of reaction channels) includes a broad range of stochastic systems, and, in particular, includes models for stochastic chemical

systems [7], high dimensional fluid/particle mixtures [8], and gene expression systems [9]. Tau-leaping methods [10] are fast, discrete-time approximations of such lattice CTMCs, and are a primary example of the lattice DTMCs we study in this work. In tau-leaping algorithms, the number of events within a timestep are sampled from appropriate Poisson distributions; it is the exact analog of a deterministic Euler approximation in time of a lattice CTMC. The method was introduced by Gillespie [10], and is particularly desirable for the simulation of processes that have at least some reaction channels which experience many transitions on short time scales. The discrete-time approximation used in tau-leaping produces a biased lattice DTMC distribution with respect to the original lattice CTMC distribution, but stability and convergence to the stochastic simulation algorithm (SSA) [11,7] (which simulates lattice CTMCs exactly) have been proven [12] as step-size τ becomes small. Significant work has been done to produce enhanced tau-leaping algorithms, including the development of adaptive step size selection [13,14] and implicit variants [15]. Outside of the tau-leaping setting, the slow-scale SSA method [16] has been proposed to simulate fast and slow dynamics in a separate but coupled fashion. Other variance reduction techniques applied to the continuous simulation version of this class of processes include, for example the common random numbers and common reaction path methods proposed in Rathinam et al. [17] and an efficient finite-difference technique proposed by Anderson [18], used for the estimation of parameter sensitivities. Anderson and Higham [19] have also proposed a multilevel technique that reduces variance using a type of iterated control variates. For multiscale systems in this class, exact, reduced cost sampling techniques using binning strategies [20] compatible with the techniques of this paper have been shown to be effective.

The outline of the paper is as follows. In Section 2, we motivate the variance reduction problem, define the class of lattice DTMCs, and recall that tau-leaping approximations for stochastic simulation are a common source of such systems. Section 3 defines the sample paths used to produce variance-reduced estimators for lattice DTMC simulation. In Section 4, we prove several analytical results for this class of estimators, including that the variance-reduced lattice DTMC methods are still generating fair path samples of the original discrete-time stochastic system. Further, we examine the evolution of the covariance of these paths and provide sufficient conditions for variance reduction in the affine rates case. Three example systems are numerically simulated with the proposed sampling method in Section 5: an affine-state-dependent gene expression system and two nonlinear systems: particle coagulation under gravitational settling in the presence of particle emissions and an HIV infection system. In each case, the change in MSE as the system size increases is studied. The numerical results in the first system provide additional verification of the analytical results, and the latter two systems demonstrate dramatic performance improvement in the analytically intractable nonlinear case.

2. Motivation and definitions

Our setting of interest is the Monte Carlo estimation of the mean behavior μ_t of a discrete-time stochastic process X_t. The classical Monte Carlo approach is to draw M iid samples of the process to produce an estimator $\widetilde{\Psi}_t^M \approx \mu_t = \mathbb{E}[X_t]$ (hereafter, the tilde will denote the use of iid simulation). Throughout the paper, we will compare this iid approach to our anticorrelated approach, which produces an alternative mean estimator Ψ_t^M, constructed from identically distributed but non-independent sample paths. We define the iid mean-estimator $\widetilde{\Psi}_t^M$ of a discrete-time stochastic process $X_t \in \mathbb{R}^D$ to be the sample-mean of an iid collection of M sample paths $\widetilde{X}_t^{(r)}$,

$$\widetilde{\Psi}_t^M := \frac{1}{M}\sum_{r=1}^{M} \widetilde{X}_t^{(r)}. \tag{1}$$

It is well established [3] that this is an unbiased mean-estimator that achieves the minimum variance possible using iid samples. To reduce the MSE of estimators constructed from system samples, we must reduce estimator variance. Indeed, these two quantities nearly coincide, as MSE can be expressed as a scaled trace of the variance matrix. Specifically,

$$\begin{aligned}\mathrm{MSE}(\widetilde{\Psi}_t^M) &= \mathbb{E}[\|\widetilde{\Psi}_t^M - \mu_t\|_2^2] \\ &= \mathrm{tr}\,\mathrm{Var}(\widetilde{\Psi}_t^M) = \frac{1}{M}\mathrm{tr}\,\mathrm{Var}(X_t).\end{aligned} \tag{2}$$

Of course, the MSE is reduced as the number of samples M increases, but more stochastic simulation can come at a significant computational cost. Consider, however, our alternative ensemble $X_t^{(r)} \sim X_t$ for $r \in \{1, \ldots, M\}$ such that the sample paths are mutually correlated while still identically distributed. In this case, we may reduce the variance of the estimator Ψ_t^M by ensuring that samples that compose it are negatively correlated. Analytically, the variance of our proposed mean estimator,

$$\Psi_t^M := \frac{1}{M}\sum_{r=1}^{M} X_t^{(r)}, \tag{3}$$

can be expressed as

$$\mathrm{MSE}(\Psi_t^M) = \frac{1}{M}\operatorname{tr}\operatorname{Var}(X_t) + \frac{1}{M^2}\sum_{r\neq p}\operatorname{tr}\operatorname{Cov}(X_t^{(r)}, X_t^{(p)}) \tag{4}$$

$$= \mathrm{MSE}(\widetilde{\Psi}_t^M) + \frac{1}{M^2}\sum_{r\neq p}\operatorname{tr}\operatorname{Cov}(X_t^{(r)}, X_t^{(p)}). \tag{5}$$

Thus if the paths $\{X_t^{(r)}\}_{r=1}^M$ are simulated in such a way to ensure mutual negative correlation between different samples, the MSE of the correlated ensemble estimator will be less than the MSE of the iid estimator. Note that, to generate any sufficiently accurate mean estimator, we may produce multiple iid realizations of anticorrelated collections of M paths $\{X^{(r)}\}_{r=1}^M$. However, we need only analyze estimators produced from a single collection of anticorrelated paths. In particular, for the purposes of antithetic mean estimation, the inclusion of additional antithetic pairs in an ensemble will reduce the error of mean estimates similarly to iid Monte Carlo, since pairs $(X_t^{(r)}, X_t^{(r+1)})$ are iid in $r \in \{1, 3, 5, \dots\}$. This fact is proven in Lemma 2. The challenge, of course, is how to produce anticorrelated paths with correct marginal distributions without any foreknowledge of the particular parameters of a stochastic system in a large class. The key motivation for our approach lies in the random time-change representation of a lattice CTMC and its corresponding lattice DTMC.

2.1. Lattice discrete-time Markov chains

Consider the random time-change (or Kurtz) representation [21] of a lattice CTMC, and suppose w.l.o.g. that $X(t) \in \mathbb{Z}^D$, $t \in [0, T]$. The process has I event channels, each with propensity function $a^i(t, X(t))$, and is defined by

$$X(t) = x_0 + \sum_{i=1}^{I} Y^i\left(\int_0^t a^i(s, X(s))\,\mathrm{d}s\right)\nu^i. \tag{6}$$

Here, $\{Y^i\}_{i=1}^I$ are independent unit-rate Poisson processes and $\nu^i \in \mathbb{Z}^D$, $i = 1, \dots, I$, are the state jump vectors, so that $\nu^i = X(t^+) - X(t^-)$ if the ith event channel experiences a transition at time t. The evolution of such a process can be studied alongside a corresponding a lattice discrete-time Markov chain. For fixed timestep increment τ, consider the discrete-time approximation $X_k \approx X(\tau k)$ for $k \in \{0, \dots, K\}$, where $K := \max\{k : \tau k \leq T\}$. Then X_k evolves via

$$X_{k+1} = X_k + \sum_{i=1}^{I} S_k^i\left(a^i(\tau k, X_k)\tau\right)\nu^i, \tag{7}$$

where $S_k^i(\lambda) \sim \mathrm{Pois}(\lambda)$. For compactness, define $\lambda^i(k, X_k) = a^i(\tau k, X_k)\tau$ and denote $S_k^i\left(\lambda^i(k, X_k)\right)$ by S_k^i. Allowing for an abuse of notation, let t replace k for the discrete-time index (used as a subscript), and (7) becomes

$$X_{t+1} = X_t + \sum_{i=1}^{I} S_t^i\nu^i. \tag{8}$$

The technique presented in this paper demonstrates how to produce unbiased, anticorrelated ensembles of this discrete-time system (8) with respect to its own distribution. Variance-reduced simulation of the continuous-time system (6) is possible by using the tau-leaping method of Gillespie [10] to produce a corresponding DTMC, but this introduces bias with respect to the CTMC distribution. Unbiased, variance-reduced simulation of the lattice CTMC system is the subject of future work. The discrete time system evolves at each timestep via a collection of independent, marginally-Poisson random variable draws with stochastic-valued parameters. This structure is crucial in the construction of anticorrelated path ensembles drawn from (8), as defined in Algorithm 1 below. Due to the discrete time approximation, there is a nonzero probability of transition to a state outside of the domain of the continuous time system (e.g. a negative number of particles). We handle this using the method of Rathinam et al. [12] in the tau-leaping context, namely by truncating the state to zero instead if it would transition to a negative state. For clarity, we restrict our attention to explicit tau-leaping with fixed step size, though extension to implicit [15] and/or adaptive [13,14] variants are of future interest.

3. Constructing antithetic sample paths for lattice DTMCs

We now define the Markov process samples used in the pathwise mean estimators. Define M iid Monte Carlo sample paths by

$$\widetilde{X}^{(r)} \overset{\text{iid}}{\sim} \widetilde{X}, \quad r \in \{1, \dots, M\} \tag{9}$$

so that the random value of the rth sample path at time t is denoted $\widetilde{X}_t^{(r)}$. Here, we explicitly construct antithetic pairs of stochastic paths $X^{(1)}$, $X^{(2)}$ as our anticorrelated method of choice; for the details in constructing stratified or hybrid

ensembles of paths, see Maginnis et al. [6]. In each case, the analysis of variance reduction ultimately hinges on the value of $\mathrm{Cov}(X_t^{(r_1)}, X_t^{(r_2)})$ for $r_1 \neq r_2$ and for $t > 0$.

To generate anticorrelated sample paths of the Markov chain, several adaptations of classical random variable variance reduction techniques [3] are necessary. Since, in general, the parameter $\lambda^{i,(r)}$ of the Poisson random variable used to simulate the ith reaction channel of the Markov chain at time t depend on the current state $X_t^{(r)}$, the Poisson variables used by different sample paths may have different parameters. To produce unbiased sample paths for the Markov jump process, we produce antithetically paired random inputs (which are Poisson when conditioned on the random state value) as shown in Algorithm 1. Here F_λ is the Poisson CDF with parameter λ and $F_\lambda^{-1}(u) := \inf\{q : F_\lambda(q) \geq u\}$. Using a result shown in Whitt [22], this scheme is optimal for pairs of random variables with the same marginal distribution. While these marginal Poisson samples will not have the same parameters in general, this scheme is still a reasonable choice since the parameters are unknown *a priori*.

Algorithm 1 Constructing antithetic paths for lattice DTMC systems (8).

Initialize: $X_0^{(j)} \leftarrow x_0$
for $t = 0$ to T **do**
  for $i = 1$ to I **do**
    sample iid $U_t^i \overset{\text{iid}}{\sim} \mathrm{Unif}(0,1)$
    $S_t^{i,(1)} \leftarrow F^{-1}_{\lambda^i(t,X_t^{(1)})}(U_t^i) \sim \mathrm{Pois}(\lambda^i(t, X_t^{(1)}))$
    $S_t^{i,(2)} \leftarrow F^{-1}_{\lambda^i(t,X_t^{(2)})}(1 - U_t^i) \sim \mathrm{Pois}(\lambda^i(t, X_t^{(2)}))$
  end for
  for $r \in \{1,2\}$ **do**
    $X_{t+1}^{(r)} \leftarrow X_t^{(r)} + \sum_{i=1}^I S_t^{i,(r)} \nu^i$
  end for
end for

3.1. Implementation architecture

Implementing Monte Carlo simulations of large-scale stochastic systems often requires consideration of both memory usage and parallelizability. While direct implementation of Algorithm 1 is sufficient for many applications, in some situations it might not be desirable to store multiple trajectory instances simultaneously in memory or to utilize only one processor. An important feature of lattice DTMC systems is that, for given system parameters and number of time steps, the number of uniform variates required to simulate a sample path is fixed. As a result, we may reproduce a complete sequence of uniform random numbers, and thus an entire trajectory, by storing or communicating the scalar state or seed of the pseudorandom number generator (PRNG) instead of an entire sequence of random numbers, as shown in Fig. 2. Furthermore, an antithetic pair of sample trajectories needn't be computed simultaneously. A single processor may compute an antithetic pair of sample paths in series by storing the seed of the PRNG, avoiding the need to simultaneously store two trajectories. Furthermore, an antithetic pair of trajectories can easily be computed on separate processors with minimal communication overhead by simply passing each processor the PRNG seed corresponding to the pair.

3.2. Lattice DTMC error quantification

We define the mean square error (MSE) of an estimation method at time t to be

$$\mathrm{MSE}\left(\Psi_t^M\right) = \mathbb{E}\left[\left\|\Psi_t^M - \mathbb{E}[\Psi_t^M]\right\|_2^2\right], \tag{10}$$

where $\|\cdot\|_2$ denotes the Euclidean vector norm. As shown above, this estimator error is the trace of its variance,

$$\mathrm{MSE}\left(\Psi_t^M\right) = \mathbb{E}\left[\left\|\Psi_t^M - \mathbb{E}[\Psi_t^M]\right\|_2^2\right] = \mathrm{tr}\,\mathrm{Var}\left(\Psi_t^M\right). \tag{11}$$

From this, we define our primary error measure, the pathwise MSE, given by

$$\mathrm{MSE}\left(\Psi^M\right) = \mathbb{E}\left[\left\|\Psi^M - \mathbb{E}[\Psi^M]\right\|^2\right], \tag{12}$$

where $\|\cdot\|$ denotes the Frobenius matrix norm. The stepwise MSE is related to the pathwise MSE by

$$\mathrm{MSE}\left(\Psi^M\right) = \sum_{t=0}^{T} \mathrm{MSE}\left(\Psi_t^M\right). \tag{13}$$

4. Analytical results for pathwise variance reduction algorithms

We now present several analytical results about anticorrelated mean estimators. For brevity, some of the proofs are omitted or condensed when they are straightforward. A simple yet important first result is that the local relaxation of the independence assumption in the law of large numbers does not sacrifice unbiasedness or consistency of a mean estimator. In the sequel, $X^{(r)}$ denotes the rth member of a possibly correlated ensemble and $\{X^{(r),j}\}_{r=1}^{M}$ is the jth iid realization of a complete M-element ensemble. In other words, the sequence of identically distributed sample paths $X^{(r),j} \sim X$, for $r \in \{1,\dots,M\}$ and $j \in \{1,\dots,N\}$ has the property that $X^{(r_1),j_1}$ and $X^{(r_2),j_2}$ are independent for $j_1 \neq j_2$ and are not necessarily independent for $j_1 = j_2$. Using such a sequence, we may construct a mean estimator

$$\Psi^{NM} = \frac{1}{NM}\sum_{j=1}^{N}\sum_{r=1}^{M} X^{(r),j} =: \frac{1}{N}\sum_{j=1}^{N}\Psi^{M,j}.$$

Lemma 1. *The mean estimator Ψ^{NM} is unbiased with respect to the lattice DTMC distribution (8) and, for fixed M, is consistent in N.*

Since such estimators are unbiased, reduction of their variance is tantamount to reduction of their MSE. For estimators formed from anticorrelated collections of samples as above, the number of such collections used to construct an estimator does not affect its performance relative to an iid estimator using the same number of samples, as shown in the following Lemma.

Lemma 2. *For Ψ^{NM} constructed as above,*

$$\frac{\mathrm{MSE}\left(\Psi^{NM}\right)}{\mathrm{MSE}\left(\widetilde{\Psi}^{NM}\right)} = \frac{\mathrm{MSE}\left(\Psi^{M}\right)}{\mathrm{MSE}\left(\widetilde{\Psi}^{M}\right)}.$$

In particular, we may analyze any antithetic mean estimator by examining only the estimator constructed from a single antithetic pair, $\{X^{(1)}, X^{(2)}\}$. To quantify variance reduction of such estimators, one must first characterize the evolution of the covariance of correlated paths, which can evolve in time in nonlinear, recursive fashion, as shown below.

Theorem 3. *If $X^{(1)}, X^{(2)} \in \mathbb{Z}^D \times \mathbb{N}$ are two realizations that satisfy (8) and are constructed using Algorithm 1, then their mutual covariance satisfies*

$$\begin{aligned}
\mathrm{Cov}(X_{t+1}^{(1)}, X_{t+1}^{(2)}) = \mathrm{Cov}(X_t^{(1)}, X_t^{(2)}) &+ \sum_{i=1}^{I} \nu^i \,\mathrm{Cov}(\lambda^i(t, X_t^{(1)}), X_t^{(2)}) + \sum_{i=1}^{I} \mathrm{Cov}(X_t^{(1)}, \lambda^i(t, X_t^{(2)}))\nu^{i\top} \\
&+ \sum_{i_1=1}^{I}\sum_{i_2=1}^{I} \nu^{i_1}\nu^{i_2\top}\,\mathrm{Cov}(\lambda^{i_1}(t, X_t^{(1)}), \lambda^{i_2}(t, X_t^{(2)})) \\
&+ \sum_{i=1}^{I} \nu^i\nu^{i\top}\mathbb{E}\left[\left(S_t^{i,(1)} - \lambda^i(t, X_t^{(1)})\right)\cdot\left(S_t^{i,(2)} - \lambda^i(t, X_t^{(2)})\right)\right].
\end{aligned} \tag{14}$$

Proof. First, consider the simplified case of a single event channel. In this case, the system dynamics are given by

$$X_{t+1} = X_t + S_t\nu.$$

Draw any two anticorrelated sample paths $X^{(1)}$ and $X^{(2)}$, simulated as above, and consider their mutual covariance which can be expanded using the system definition and bilinearity as

$$\mathrm{Cov}(X_{t+1}^{(1)}, X_{t+1}^{(2)}) = \mathrm{Cov}(X_t^{(1)}, X_t^{(2)}) + \nu\,\mathrm{Cov}(S_t^{(1)}, X_t^{(2)}) + \mathrm{Cov}(X_t^{(1)}, S_t^{(2)})\nu^\top + \nu\nu^\top\,\mathrm{Cov}(S_t^{(1)}, S_t^{(2)}).$$

The treatment of the last term is informative for how to simplify the other terms.

$$\begin{aligned}
\mathrm{Cov}(S_t^{(1)}, S_t^{(2)}) &= \mathbb{E}\Big[(S_t^{(1)} - \mathbb{E}[S_t^{(1)}])\cdot(S_t^{(2)} - \mathbb{E}[S_t^{(2)}])\Big] \\
&= \mathbb{E}\Big[\left(S_t^{(1)} - \lambda(t, X_t^{(1)})\right)\cdot\left(S_t^{(2)} - \lambda(t, X_t^{(2)})\right) \\
&\quad + \left(S_t^{(1)} - \lambda(t, X_t^{(1)})\right)\cdot\left(\lambda(t, X_t^{(2)}) - \mathbb{E}[S_t^{(2)}]\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \left(S_t^{(2)} - \lambda(t, X_t^{(2)})\right) \cdot \left(\lambda(t, X_t^{(1)}) - \mathbb{E}[S_t^{(1)}]\right) \\
&+ \left(\lambda(t, X_t^{(1)}) - \mathbb{E}[S_t^{(1)}]\right) \cdot \left(\lambda(t, X_t^{(2)}) - \mathbb{E}[S_t^{(2)}]\right)\Big] \\
&= \mathbb{E}\Big[\left(S_t^{(1)} - \lambda(t, X_t^{(1)})\right) \cdot \left(S_t^{(2)} - \lambda(t, X_t^{(2)})\right)\Big] \\
&\quad + \mathbb{E}\Big[\mathbb{E}\Big[S_t^{(1)} - \lambda(t, X_t^{(1)}) | X_t^{(1)}, X_t^{(2)}\Big]\left(\lambda(t, X_t^{(2)}) - \mathbb{E}[S_t^{(2)}]\right)\Big] \\
&\quad + \mathbb{E}\Big[\mathbb{E}\Big[S_t^{(2)} - \lambda(t, X_t^{(2)}) | X_t^{(1)}, X_t^{(2)}\Big]\left(\lambda(t, X_t^{(1)}) - \mathbb{E}[S_t^{(1)}]\right)\Big] \\
&\quad + \mathrm{Cov}\left(\lambda(t, X_t^{(1)}), \lambda(t, X_t^{(2)})\right) \\
&= \mathbb{E}\Big[\left(S_t^{(1)} - \lambda(t, X_t^{(1)})\right) \cdot \left(S_t^{(2)} - \lambda(t, X_t^{(2)})\right)\Big] \\
&\quad + \mathrm{Cov}\left(\lambda(t, X_t^{(1)}), \lambda(t, X_t^{(2)})\right),
\end{aligned}
$$

where the last equality follows since the conditional expectations in the middle two terms are zero. It is easy to see, again using the law of total expectation, that the first term can only be non-zero when $S_t^{(1)}$ and $S_t^{(2)}$ are associated with the same event channel. Indeed, suppose $i_1 \neq i_2$. Then

$$
\begin{aligned}
&\mathbb{E}\Big[\left(S_t^{i_1,(1)} - \lambda^{i_1}(t, X_t^{(1)})\right) \cdot \left(S_t^{i_2,(2)} - \lambda^{i_2}(t, X_t^{(2)})\right)\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\left(S_t^{i_1,(1)} - \lambda^{i_1}(t, X_t^{(1)})\right) \cdot \left(S_t^{i_2,(2)} - \lambda^{i_2}(t, X_t^{(2)})\right) | X_t^{(1)}, X_t^{(2)}\Big]\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[S_t^{i_1,(1)} - \lambda^{i_1}(t, X_t^{(1)}) | X_t^{(1)}, X_t^{(2)}\Big] \cdot \mathbb{E}\Big[S_t^{i_2,(2)} - \lambda^{i_2}(t, X_t^{(2)}) | X_t^{(1)}, X_t^{(2)}\Big]\Big] \\
&\quad = 0.
\end{aligned}
$$

Thus (14) follows immediately for the multi event channel case. □

The following result proves a useful connection between the properties of an anticorrelated Poisson variable and the analogous lattice DTMC constructed using Algorithm 1. First, however, we require a lemma regarding the antithetic Poisson sampling used in Algorithm 1.

Lemma 4. *For any $\lambda_1, \lambda_2 \in \mathbb{R}_+$, if $S^{(r)} \sim \mathrm{Pois}(\lambda_r)$, $r \in \{1, 2\}$ are simulated using the antithetic sampling technique used in steps 3 and 4 of Algorithm 1, then* $\mathrm{Cov}(S^{(1)}, S^{(2)}) \leq 0$.

The proof is omitted here, but uses a result in Whitt [22], namely that composition with non-decreasing functions cannot change the sign of the covariance between two random variables. Using Lemma 4, we prove a result that guarantees the non-positivity of the final term of (14).

Theorem 5. *Suppose that $X^{(1)}$, $X^{(2)}$ are two realizations simulated using Algorithm 1. Then, for any event channel i and for each time $t \geq 0$,*

$$
\mathbb{E}\Big[\left(S_t^{i,(1)} - \lambda^i(t, X_t^{(1)})\right) \cdot \left(S_t^{i,(2)} - \lambda^i(t, X_t^{(2)})\right)\Big] \leq 0. \tag{15}
$$

Proof. Suppose $S^{(1)}(t, x), S^{(2)}(t, x) \sim \mathrm{Pois}(\lambda(t, x))$ are simulated using such an anticorrelated technique. Fix any $x_1, x_2 \in \mathbb{R}^D$, and, by Lemma 4,

$$
\begin{aligned}
0 &\geq \mathrm{Cov}(S^{(1)}(t, x_1), S^{(2)}(t, x_2)) \\
&= \mathbb{E}\Big[\left(S^{(1)}(t, x_1) - \lambda(t, x_1)\right) \cdot \left(S^{(2)}(t, x_2) - \lambda(t, x_2)\right)\Big].
\end{aligned}
$$

Since this is true for any $x_i \in \mathbb{R}^D$, it is necessarily true that

$$
0 \geq \mathbb{E}\Big[\left(S_t^{i,(1)} - \lambda^i(t, X_t^{(1)})\right) \cdot \left(S_{r_2,t}^{i,(2)} - \lambda^i(t, X_t^{(2)})\right) | X_t^{(1)}, X_t^{(2)}\Big], \tag{16}
$$

almost surely, since this is exactly the same integral (in U) for given $X_t^{(1)}$, $X_t^{(2)}$ random. Taking expectation of both sides, we get (15). □

We remark here that, as noted above, anticorrelated sampling schemes other than antithetic sampling may be used in the framework of Algorithm 1, such as stratified or hybrid antithetic/stratified sampling [6]. To prove a result equivalent

to Theorem 5 for these methods, we need only prove a corresponding version of Lemma 4 for any $X^{(r_1)}$, $X^{(r_2)}$ for $r_1, r_2 \in \{1, \dots, M\}$.

The following corollary refines the previous results when more is required of the intensity functions $\lambda^i(t, X_t)$ beyond nonnegativity, namely that they be affine in the state X_t. The first condition can be used to greatly simplify (14).

Corollary 1. *Suppose that*

$$\lambda^i(t, X_t) = \left(a^i(t) + {\kappa^i}^\top X_t\right)\tau. \tag{17}$$

If the conditions of Theorem 3 are satisfied, then the following recursion is satisfied

$$\begin{aligned}\operatorname{Cov}(X_{t+1}^{(1)}, X_{t+1}^{(2)}) = \operatorname{Cov}(X_t^{(1)}, X_t^{(2)}) &\\ + \sum_{i=1}^{I} \tau \nu^i {\kappa^i}^\top \operatorname{Cov}(X_t^{(1)}, X_t^{(2)}) + \sum_{i=1}^{I} \tau \operatorname{Cov}(X_t^{(1)}, X_t^{(2)}) \kappa^i {\nu^i}^\top &\\ + \sum_{i_1=1}^{I}\sum_{i_2=1}^{I} \tau^2 \nu^{i_1} {\kappa^{i_1}}^\top \operatorname{Cov}(X_t^{(1)}, X_t^{(2)}) \kappa^{i_2} {\nu^{i_2}}^\top &\\ + \sum_{i=1}^{I} \nu^i {\nu^i}^\top \mathbb{E}\left[\left(S_t^{i,(1)} - \lambda^i(t, X_t^{(1)})\right) \cdot \left(S_t^{i,(2)} - \lambda^i(t, X_t^{(2)})\right)\right].\end{aligned} \tag{18}$$

This expression can be more compactly written as

$$\operatorname{Cov}(X_{t+1}^{(1)}, X_{t+1}^{(2)}) = \mathcal{L}(\operatorname{Cov}(X_t^{(1)}, X_t^{(2)})) + \sum_{i=1}^{I} c_t^i \nu^i {\nu^i}^\top,$$

where $\mathcal{L}$ is a time invariant linear operator on the space of symmetric matrices and

$$c_t^i := \mathbb{E}\left[\left(S_t^{i,(1)} - \lambda^i(t, X_t^{(1)})\right) \cdot \left(S_t^{i,(2)} - \lambda^i(t, X_t^{(2)})\right)\right], \quad t \in \{1, \dots, T\}$$

are sequences of reals that depend on x_0 and $\{\lambda^i\}_{i=1}^I$. If, in addition, the conditions of Theorem 5 are satisfied, then $c_t^i \le 0$ for every i and t.

In the affine rates case, we can derive a sufficient, testable condition for ensemble variance reduction of our algorithm. First, we require an algebraic proposition regarding matrix invariance in a half-space. Consider the set S_D of $D \times D$ symmetric matrices as a vector space together with the field $\mathbb{R}$ and the Frobenius inner product $\langle A, B\rangle := \operatorname{tr} AB^\top$. Define the half-space $\mathcal{H}_- := \{A \in S_D : \langle A, I_D\rangle \le 0\}$. Define the cone $\mathcal{R} := \{\sum_{i=1}^I c^i \nu^i {\nu^i}^\top : c^i \le 0 \text{ for each } i\} \subset \mathcal{H}_-$, and consider the following sufficient condition.

Proposition 1. *Suppose*

(i) *the sequence $A_t \in S_D$ evolves according to*

$$A_{t+1} = \mathcal{L}(A_t) + \sum_{i=1}^{I} c_t^i \nu^i {\nu^i}^\top, \tag{19}$$

where $\mathcal{L}$, c_t^i and ν^i are all defined as in Corollary 1, and $A_0 = 0$;

(ii) *that for any $R \in \mathcal{R}$ and for every $t \ge 1$, $\mathcal{L}^t(R) \in \mathcal{H}_-$.*

Then $A_t \in \mathcal{H}_-$ for every $t \ge 0$.

Proof. Define $R_t := \sum_{i=1}^I c_t^i \nu^i {\nu^i}^\top$. Then, for each $t \ge 0$, $R_t \in \mathcal{R}$. For each $t \ge 1$, a solution to (18) is given by

$$A_t := \sum_{\ell=0}^{t-1} \mathcal{L}^{t-\ell-1}(R_\ell). \tag{20}$$

Since $\mathcal{L}^{t-\ell-1}(R_\ell) \in \mathcal{H}_-$ and since the half-space $\mathcal{H}_-$ is closed under addition, $A_t \in \mathcal{H}_-$ for every $t \ge 0$. □

The final corollary shows that the above conditions are sufficient to prove the dominance of the antithetic estimator Ψ^M over the iid estimator $\widetilde{\Psi}^M$ in this affine rates setting.

Corollary 2. *Suppose that $X^{(r)}$ satisfy the conditions of Corollary 1, that $\mathcal{L}$ satisfies the conditions of Proposition 1, and that $\mathrm{Cov}(X_0^{(1)}, X_0^{(2)}) = 0$. Then*

$$\mathrm{MSE}(\Psi^M) \leq \mathrm{MSE}(\widetilde{\Psi}^M). \tag{21}$$

Proof. It is easy to see that $\mathrm{tr}\,\mathrm{Cov}(\Psi_t^M) \leq \mathrm{tr}\,\mathrm{Cov}(\widetilde{\Psi}_t^M)$ for each $t \geq 0$ if $\mathrm{tr}\,\mathrm{Cov}(X_t^{(r_1)}, X_t^{(r_2)}) \leq 0$ for each $r_1 \neq r_2 \in \{1, \ldots, M\}$ and $t \geq 0$. Since the evolution equation (18) of $\mathrm{Cov}(X_t^{(1)}, X_t^{(2)})$ is identical to (19), we have by Proposition 1 that $\mathrm{Cov}(X_t^{(1)}, X_t^{(2)}) \in \mathcal{H}_-$ for every $t \geq 0$. That is, $\mathrm{tr}\,\mathrm{Cov}(X_t^{(r_1)}, X_t^{(r_2)}) \leq 0$ for every $t \geq 0$. Thus $\mathrm{tr}\,\mathrm{Cov}(\Psi_t^M) \leq \mathrm{tr}\,\mathrm{Cov}(\widetilde{\Psi}_t^M)$ for each $t \geq 0$, and the claim holds by (5). □

5. Numerical results for lattice DTMC samplers

We now introduce three example stochastic systems for numerical study of Algorithm 1, that simultaneously illustrate the above analytical results and provide further intuition for its efficacy in more general cases. These systems are drawn from the literature and have been specifically chosen to exhibit increasingly complex rate functions. The first is a simple model of gene expression which appears in Briat and Khammash [9], modeling the production and decay of mRNA and protein molecules. This system has rate functions which are affine in the state variables, and thus corresponds to Corollary 1, and we prove that it satisfies the sufficient conditions of Corollary 2. The second system is a simplified model of coagulation of water molecules via gravitational settling. As presented in Seinfeld and Pandis [23], the system is composed of two sizes of water molecules, large and small, where coagulations within sizes are rare but between sizes are frequent and are specified by a nonlinear propensity function. Finally, we present a seven-dimensional model of HIV infection with 19 reaction channels and nonlinear rate functions, as found in Banks et al. [24]. The latter two systems are chosen because their complexity extends beyond the scope of our analytical results. While we cannot yet provide analytical guarantees of improved performance using anticorrelated simulation for such models, the numerical results do demonstrate significant computational savings for each system. Thus, we provide promising evidence of the wider applicability of variance-reduced tau-leaping. First, we briefly introduce each of the three systems in Subsections 5.1, 5.2 and 5.3, then we study the performance of Algorithm 1 in all three settings via a parametric study of the scale of each system in Subsection 5.4.

5.1. Affine gene expression system

Consider a simple gene expression system, where mRNA is produced and decays, and it produces a protein which also decays. This simple model is quite commonly studied, the specific formulation and parameter values appear here as in Briat and Khammash [9], and are taken to be unitless. That is,

$$\begin{aligned}
\emptyset &\xrightarrow{k_r} \mathrm{mRNA} \\
\mathrm{mRNA} &\xrightarrow{\gamma_r} \emptyset \\
\mathrm{mRNA} &\xrightarrow{k_p} \mathrm{protein} + \mathrm{mRNA} \\
\mathrm{protein} &\xrightarrow{\gamma_p} \emptyset.
\end{aligned}$$

Define the state of the system to be number of mRNA and protein particles, respectively, as a vector $X \in \mathbb{Z}^2$, with initial condition $X_0 = V \cdot [1.0\ 0.5]^\top$ (where V is a system volume scaling parameter) with $I = 4$ reaction channels, given by:

$$\begin{aligned}
\nu^1 &= [1\ 0]^\top & a^1(X_t) &= k_r V \\
\nu^2 &= [-1\ 0]^\top & a^2(X_t) &= \gamma_r X_{t,1} \\
\nu^3 &= [0\ 1]^\top & a^3(X_t) &= k_p X_{t,1} \\
\nu^4 &= [0\ -1]^\top & a^4(X_t) &= \gamma_p X_{t,2},
\end{aligned}$$

for $k_r, \gamma_r, k_p, \gamma_p > 0$, and where $X_{t,d}$ denotes the dth component of the state vector at time t. The corresponding discrete time approximation is simulated using the tau-leaping approximation (7) with $\tau = 1$ (i.e. $\lambda^i = a^i$) and run from time $t = 0$ to time $t = T = 100$. A pair of antithetic sample trajectories are shown in Fig. 3. Our primary interest is the normalized estimator MSE for the gene expression estimator, shown in Fig. 5; this shows the degree of MSE reduction of antithetic simulation compared to iid simulation over a large range of system scales. This behavior will be discussed in detail in 5.4. As mentioned above, the number of antithetic pairs used in the variance reduced mean estimator is irrelevant for comparison

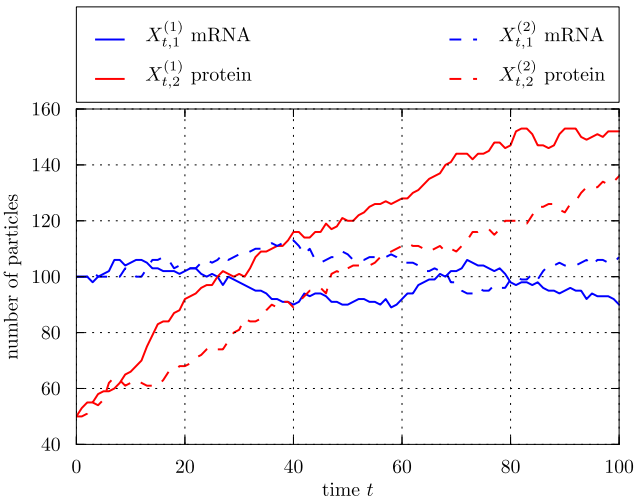

Fig. 3. Antithetic pair of sample path trajectories of the gene expression system using timestep $\tau = 1$ and rate parameters $(k_r, \gamma_r, k_p, \gamma_p) = (0.01, 0.03, 0.06, 0.0066)$ and initial condition $X_0 = [100\ 50]^\top$ (i.e. volume parameter $V = 100$) plotted versus time.

to iid estimators using an equal number of sample paths, since both decay at the same rate as proven in Lemma 2, so we consider mean estimators composed of a pair of antithetic paths.

Note that this system satisfies the conditions of Corollary 1 and Proposition 1. Indeed, the rate functions are affine in the state variables, and, with respect to (17),

$$\kappa^1 = [0\ 0]^\top$$
$$\kappa^2 = [\gamma_r\ 0]^\top$$
$$\kappa^3 = [k_p\ 0]^\top$$
$$\kappa^4 = [0\ \gamma_p]^\top.$$

It is easy to verify then that, if we vectorize the 2×2 covariance matrix objects as vectors in $\mathbb{R}^4$, then we may identify the linear operator $\mathcal{L}$ with left multiplication by a matrix $L \in \mathbb{R}^{4\times 4}$ given by

$$L = \begin{pmatrix} (1-\gamma_r\tau)^2 & 0 & 0 & 0 \\ k_p\tau(1-\gamma_r\tau) & (1-\gamma_r\tau)(1-\gamma_p\tau) & 0 & 0 \\ k_p\tau(1-\gamma_r\tau) & 0 & (1-\gamma_r\tau)(1-\gamma_p\tau) & 0 \\ k_p^2\tau^2 & k_p\tau(1-\gamma_p\tau) & k_p\tau(1-\gamma_p\tau) & (1-\gamma_p\tau)^2 \end{pmatrix},$$

a lower triangular matrix. Furthermore, the negative cone $\mathcal{R}$ can be identified with the set

$$\{(c^1,\ 0,\ 0,\ c^2)^\top : c^1, c^2 \le 0\} \subset \mathbb{R}^4, \tag{22}$$

so for any $R \in \mathcal{R}$, and for any $t \ge 1$, $\mathcal{L}^t(R)$ can be identified with $L^t\,\mathrm{vec}(R)$, which, if $\gamma_r\tau \le 1$ and $\gamma_p\tau \le 1$, is the product of a lower triangular matrix with non-negative entries and the standard vectorization of an element of $\mathcal{R}$. Therefore, the first and fourth components of the product will have the form

$$\begin{aligned} \left(L^t\,\mathrm{vec}(R)\right)_1 &= \left(L^t\right)_{1,1} c^1 && \le 0 \\ \left(L^t\,\mathrm{vec}(R)\right)_4 &= \left(L^t\right)_{4,1} c^1 + \left(L^t\right)_{4,4} c^2 && \le 0 \end{aligned}$$

for every $t \ge 1$, where subscripts are used here to denote vector and matrix components. So $\mathcal{L}^t(R) \in \mathcal{H}_-$ for every $t \ge 1$, and thus $\mathrm{tr}\,\mathrm{Cov}(X_t^{(1)}, X_t^{(2)}) \le 0$ for every $t \ge 0$, as long as $\gamma_r\tau, \gamma_p\tau \le 1$. Thus the antithetic mean estimator is a dominant mean estimator for the lattice DTMC distribution.

Because the propensity functions in this example are affine with respect to the state, the exact mean evolution of the corresponding continuous-time system is obtainable. Following the approach used in Engblom [25], the mean evolution of the lattice CTMC corresponding to the affine gene expression system is given by the solution to:

$$\dot{m}(t) = \nu A m(t) + \nu B, \qquad m(0) = X_0 \tag{23}$$

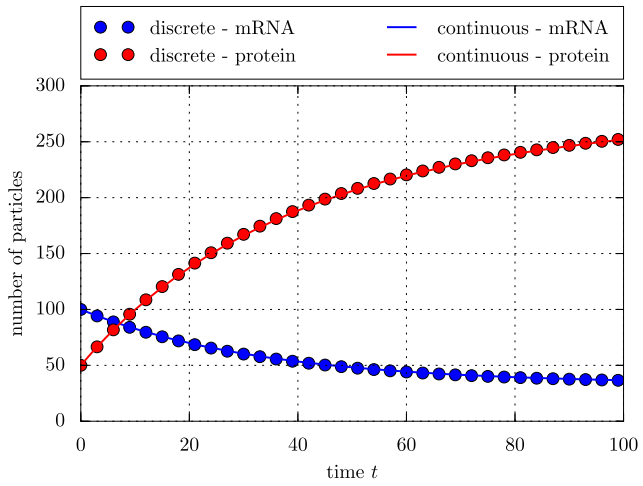

Fig. 4. Mean trajectory of the lattice CTMC gene expression system obtained by directly solving the master equation and the mean trajectory of the lattice DTMC system (with timestep $\tau = 1$) obtained using Monte Carlo simulation plotted versus time. Both systems used parameters parameters $(k_r, \gamma_r, k_p, \gamma_p) = (0.01, 0.03, 0.06, 0.0066)$ and initial condition $X_0 = [100\ 50]^\top$ (i.e. volume parameter $V = 100$).

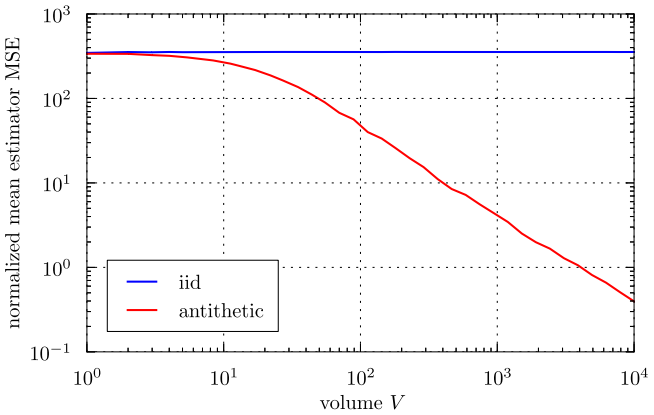

Fig. 5. Normalized pathwise MSE$(\Psi^M)M/V$ of an $M = 4$ sample estimator of the mean $\mathbb{E}[X_t]$ of the gene expression system using the iid and antithetic sampling techniques plotted versus volume scale V, for timestep $\tau = 1$ s, rate parameters $(k_r, \gamma_r, k_p, \gamma_p) = (0.01, 0.03, 0.06, 0.0066)$, initial condition $X_0 = [V\ V/2]^\top$ and timesteps from $t = 0$ to $t = 100$. Pathwise MSE calculated from 10^6 sample paths, error bars omitted.

where $m(t) := \mathbb{E}[X(t)]$, $\nu = [\nu^1\ \nu^2\ \nu^3\ \nu^4] \in \mathbb{R}^{2\times 4}$, and, using the propensities defined above, $a(X_t) := [a^i(X_t)]_{i=1}^4 = AX_t + B$ for $A \in \mathbb{R}^{4\times 2}$ and $B \in \mathbb{R}^{4\times 1}$. The solution of the above ordinary differential equation is

$$m(t) = e^{\nu A t} m(0) + \int_0^t e^{\nu A(t-\tau)} \nu B \, \mathrm{d}\tau,$$

and it is plotted together with the mean of the discrete-time version obtained via Monte Carlo simulation in Fig. 4. Our antithetic mean estimates are unbiased with respect to the discrete-time distribution, which, in turn, is biased with respect to the continuous-time distribution due to the tau-leaping approximation, though this bias is small for the values of τ and V shown in the plot.

5.2. Nonlinear coagulation via gravitational settling

Following the treatment in Seinfeld and Pandis [23], Chapter 13, consider a system of water particles falling in the atmosphere under the influence of gravity. The system falls in a control volume and is made up of two classes of particles: large and small. The system evolves via the coagulation of a large particle and a small particle (the coagulation rate is driven by differences in terminal settling velocity, so particles of similar size are unlikely to coagulate) or by the emission of new small particles into the volume V. We may specify the state of the system as (N_s, M_s, N_l, M_l), or the number of small particles, total mass of small particles, the number of large particles, and the total mass of the large particles, respectively.

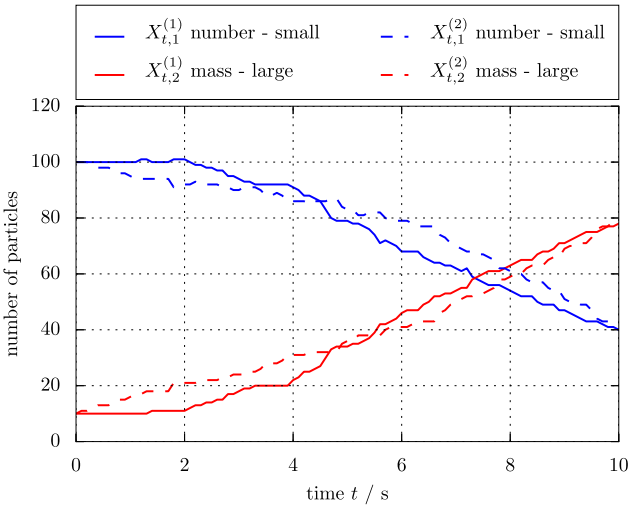

Fig. 6. Antithetically paired sample trajectories for the nonlinear coagulation system with small particle mass $m = 1$, proportionality constant $\alpha = 5 \cdot 10^{-4}$ into a control volume with $V = 1$, using timestep $\tau = 0.1$ s, and timesteps from $t = 0$ to $T = 100$.

The probability rate at which a single small particle coagulates with a single large particle in volume V is given by (13.A.4) in Seinfeld and Pandis [23]:

$$K_{sl}^{\mathrm{GS}} = \frac{\pi}{4}\frac{1}{V}(D_l + D_s)^2 |v_l - v_s|$$

where D_s and D_l are the diameters of the small and large particles, respectively, and v_s and v_l are the terminal settling velocities of the small and large particles, respectively. For simplicity, we take the collision efficiency to be 1. By Stokes' Law,

$$v_s = \frac{1}{18}\frac{(\rho_p - \rho_f)}{\mu} g {D_s}^2,$$

where ρ_p is the density of the particle, ρ_f is the density of the fluid, μ is the viscosity of the fluid, and g is the acceleration due to gravity. A similar equation holds for the terminal settling velocity v_l of the large particles. For simplicity, we consider the case where the number of large particles, N_l scales directly with the volume V and where the mass of each small particle is uniformly fixed as m. The state of the system then becomes (N_s, mN_s, V, M_l), and we need only track the smaller state $X = (N_s, M_l)$. Simplifying the master equation (13.81) in Seinfeld and Pandis [23], and using the fact that $D_s \propto \sqrt[3]{m}$ and $D_l \propto \sqrt[3]{\frac{M_l}{N_l}}$, we can specify the reaction channels of this system by:

$$\nu^1 = [m\ 0]^\top \qquad a^1(X_t) = V$$
$$\nu^2 = [-m\ m]^\top \qquad a^2(X_t) = \alpha K_{sl}^{\mathrm{GS}} V X_{t,1}$$

where α is a proportionality constant and, for simplicity

$$K_{sl}^{\mathrm{GS}} = \frac{1}{V}\left(\sqrt[3]{X_{t,2}/V} + \sqrt[3]{m}\right)^3 \left(\sqrt[3]{X_{t,2}/V} - \sqrt[3]{m}\right).$$

The state is initialized from $X_0 = V \cdot [100, 10]$ and a corresponding lattice DTMC system is obtained via the tau-leaping approximation (7) for $\tau = 0.1$ s, and is simulated for 10 s from timestep $t = 0$ to timestep $T = 100$. We take $\alpha = 5 \cdot 10^{-4}$ and for simplicity we take $m = 1$ so that the state $X_t \in \mathbb{Z}^2$. Concentration sample trajectories are shown in Fig. 6. The estimator MSE of the coagulation system plotted versus the system scale parameter V shown in Fig. 7. This behavior will be discussed in 5.4.

5.3. Nonlinear HIV infection system

Following Banks et al. [24], consider a model for HIV infection with state $X \in \mathbb{R}^7$ representing concentrations of uninfected and infected activated CD4+ T-cells ($X_{t,1}$ and $X_{t,2}$, respectively), uninfected and infected resting CD4+ T-cells ($X_{t,3}$ and $X_{t,4}$, respectively), infectious free virus ($X_{t,5}$), and HIV-specific effector and memory CD8+ T-cells ($X_{t,6}$ and $X_{t,7}$, respectively). The state is initialized from $X_0 = V \cdot [5\ 1\ 1400\ 1\ 10\ 5\ 1]^\top$ with $I = 19$ reaction channels with nonlinear rates [24], save that we maintain the scaling parameter V, the volume of the system. This system is discretized with the tau-leaping approximation (7) with $\tau = 0.005$ days and simulated from timestep $t = 0$ to timestep $t = T = 10\,000$. Sample trajectories are shown in Fig. 8 and normalized estimator MSE is shown in Fig. 9.

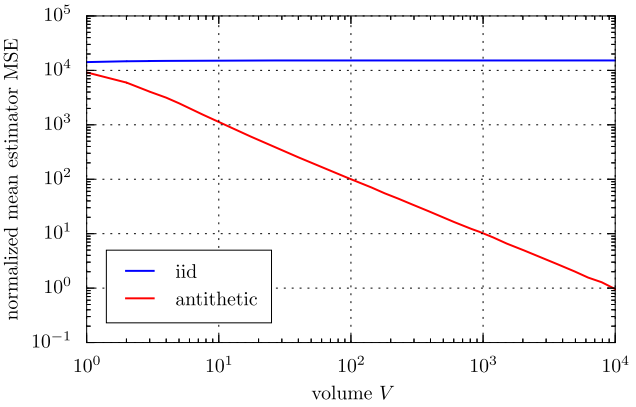

Fig. 7. Normalized pathwise MSE(Ψ^M)M/V of an $M = 4$ sample mean estimator for the nonlinear coagulation system versus source area from $V = 10^0$ to $V = 10^4$. Computation of pathwise MSE from 10^6 simulations, where each simulation uses timestep $\tau = 0.1$ s, and timesteps from $t = 0$ to $t = 100$. Error bars are small and are thus omitted.

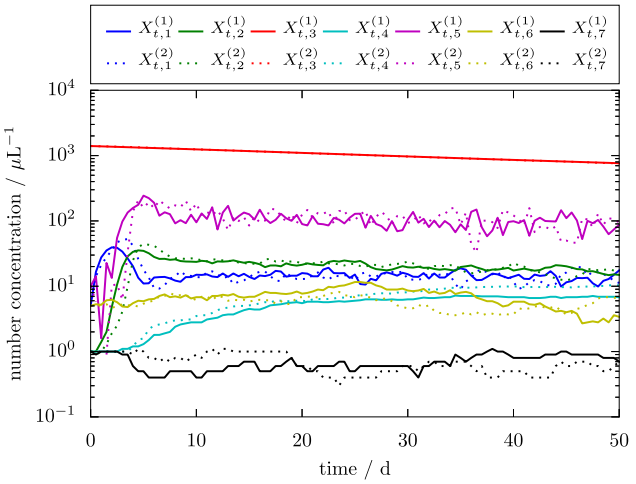

Fig. 8. Antithetic pair of sample trajectories of cell concentration for the HIV infection system for volume $V = 10$ µL, timestep $\tau = 0.005$ days, and timesteps from $t = 0$ to $t = 10\,000$ plotted versus time. All other parameter values taken from Banks et al. [24].

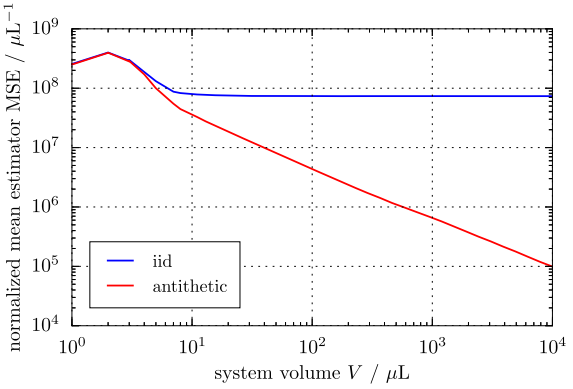

Fig. 9. Normalized pathwise MSE(Ψ^M)M/V of an $M = 4$ sample mean estimator for the expected path $\mathbb{E}[X_t]$ of the HIV infection system plotted versus system volume V from 1 µL to 10 000 µL. Simulations use a timestep $\tau = 0.005$ days, and timestep from $t = 0$ to $t = 10\,000$. All other parameter values taken from Banks et al. [24]. Pathwise MSE computed using 10^4 estimator samples, error bars omitted.

5.4. Parameter variations

Figs. 5, 7 and 9 illustrate the dependence of estimator error on a parameter that governs the number of particles of each system. Each figure plots a pathwise estimator's MSE divided by a scaling parameter V that governs the "volume" (number of computational "particles" in a typical simulation), versus the same scaling parameter. This parameter controls the "speed" of the system, i.e., the number of reactions that occur at each timestep; equivalently it governs the typical mean Poisson parameter sampled at each timestep. Each mean estimator is constructed using $M = 4$ samples, either 4 iid samples or 2 antithetic pairs to ensure a fair comparison. It is easy to see that using this normalization removes from consideration any reduction of MSE gained by generating more iid samples or pairs of samples for sufficiently large-scale systems, and thus there is no need to draw any more samples than a small M for comparison.

The volume scale V is roughly proportional to the typical Poisson parameter used to generate random samples, though in nonlinear cases such as the HIV system, this relationship will also tend to be nonlinear, as we observe quite clearly for the small volume region of Fig. 9. Since the variance of a mean estimator scales with the variance of an individual sample path, the cost of iid Monte Carlo increases roughly linearly for large V without bound (or remains constant if simulations are normalized as concentrations). So increasing V corresponds to both increased resolution of simulations as well as increased cost. However, as V becomes large and reactions occur more frequently, we expect greater relative gains in MSE for antithetic sampling by analogy to the Poisson variable case [26]. Indeed, we observe antithetic MSE that is nearly constant even for large V (or, in the particle concentration (normalized) setting, linearly inversely proportional to V), as opposed to linear growth for the iid estimator (or constant behavior when normalized). Thus when many jumps are typically observed in each timestep, the gains produced via antithetic simulation are most dramatic. It is worth noting that this is precisely the same operating regime that maximizes the desirability of tau-leaping over exact continuous-time simulation, via say SSA [10].

Figs. 5, 7 and 9, each show that the antithetic mean estimator has equal or lower MSE than the iid mean estimator for all observed values of V. Note that the affine gene expression estimator satisfies the sufficient conditions of Corollary 2 as shown above, and thus provably dominates the iid Monte Carlo mean estimator. In this case, we also know that a study of any other parameter and range will show that MSE is lower for the antithetic estimator. However, the nonlinear coagulation and HIV infection estimators also produce greatly reduced MSE over observed values of V, despite having nonlinear state-dependent propensity functions which make potential proof of dominance more difficult. These numerical results show that the affine conditions, while sufficient, are certainly not necessary to produce non-increased pathwise MSE over at least this parameter range. Thus we posit that these variance reduction techniques can be effective in a much larger class of models, even if analytical proof is not yet available.

6. Conclusions

We proposed Algorithm 1, an antithetic simulation technique for lattice-valued discrete-time Markov chain processes for variance-reduced Monte Carlo mean estimation. Via manipulation of uniform process inputs to the lattice DTMC simulation, we showed how to produce antithetically paired, identically distributed simulation paths with little additional computational effort and little alteration of code. Analytically, we studied the evolution of these covariances over time, and proved the dominance of our estimators in certain affine rates cases. Further, we implemented these techniques in three different computational models found in the literature, and demonstrated multiple orders of magnitude reduction in MSE over a wide range of scaling parameters. In light of these results, the main contribution of this paper is that, as constructed here for a very broad class of models, antithetic simulation is a low cost, easily implementable variance reduction strategy for mean estimation of stochastic paths, that can produce dramatic computational speed up. Additionally, we proved analytical guarantees in some important classes; a future challenge is broadening these results and extending them to other anticorrelated techniques such as stratification which may enjoy different performance characteristics.

Acknowledgement

This work was partially supported by NSF CPS grant 1329991.

References

[1] G.M. Shroff, H.B. Keller, Stabilization of unstable procedures: the recursive projection method, SIAM J. Numer. Anal. 30 (4) (1993) 1099–1120, http://dx.doi.org/10.1137/0730057.
[2] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runberg, C. Theodoropoulos, Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks, Commun. Math. Sci. 1 (4) (2003) 715–762.
[3] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd edn., Springer, 2004.
[4] P.W. Glynn, D.L. Iglehart, Importance sampling for stochastic simulation, Manag. Sci. 35 (1989) 1367–1392, http://dx.doi.org/10.1287/mnsc.35.11.1367.
[5] M. Villen-Altamirano, Rare event simulation: the RESTART methods, in: Proceedings of the International Conference on High Performance Computing and Simulation, 2012, pp. 32–41.
[6] P.A. Maginnis, M. West, G.E. Dullerud, Anticorrelated discrete-time stochastic simulation, in: Proceedings of the IEEE Conference on Decision and Control, 2013, pp. 618–623.

[7] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (1976) 403–434.
[8] N. Riemer, M. West, R.A. Zaveri, R.C. Easter, Simulating the evolution of soot mixing state with a particle-resolved aerosol model, J. Geophys. Res. 114 (2009) D09202, http://dx.doi.org/10.1029/2008JD011073.
[9] C. Briat, M. Khammash, Computer control of gene expression: robust setpoint tracking of protein mean and variance using integral feedback, in: Proceedings of the IEEE Conference on Decision and Control, 2012.
[10] D.T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, J. Chem. Phys. 115 (4) (2001) 1716–1733.
[11] D.T. Gillespie, An exact method for numerically simulating the stochastic coalescence process in a cloud, J. Atmos. Sci. 32 (1975) 1977–1989.
[12] M. Rathinam, L.R. Petzold, Y. Cao, D.T. Gillespie, Consistency and stability of tau-leaping schemes for chemical reaction systems, Multiscale Model. Simul. 4 (3) (2005) 867–895, http://dx.doi.org/10.1137/040603206.
[13] Y. Cao, D.T. Gillespie, L.R. Petzold, Efficient stepsize selection for the tau-leaping simulation method, J. Chem. Phys. 124 (2006) 044109, http://dx.doi.org/10.1063/1.2159468.
[14] D.F. Anderson, Incorporating postleap checks in tau-leaping, J. Chem. Phys. 128 (5) (2008) 054103.
[15] M. Rathinam, L.R. Petzold, Y. Cao, D.T. Gillespie, Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method, J. Chem. Phys. 119 (2003) 12784, http://dx.doi.org/10.1063/1.1627396.
[16] Y. Cao, D.T. Gillespie, L.R. Petzold, The slow-scale stochastic simulation algorithm, J. Chem. Phys. 122 (1) (2005) 014116, http://dx.doi.org/10.1063/1.1824902.
[17] M. Rathinam, P.W. Sheppard, M. Khammash, Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks, J. Chem. Phys. 132 (3) (2010) 034103.
[18] D.F. Anderson, An efficient finite difference method for parameter sensitivities of continuous time Markov chains, SIAM J. Numer. Anal. 50 (5) (2012) 2237–2258.
[19] D.F. Anderson, D.H. Higham, Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics, Multiscale Model. Simul. 10 (1) (2012) 146–179.
[20] M.D. Michelotti, M.T. Heath, M. West, Binning for efficient stochastic multiscale particle simulations, Multiscale Model. Simul. 11 (4) (2013) 1071–1096, http://dx.doi.org/10.1137/130908038.
[21] S.N. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, John Wiley and Sons, 1986.
[22] W. Whitt, Bivariate distributions with given marginals, Ann. Stat. 4 (6) (1976) 1280–1289.
[23] J.H. Seinfeld, S.N. Pandis, Atmospheric Chemistry and Physics, John Wiley and Sons, Hoboken, NJ, 2006.
[24] H.T. Banks, S. Hu, M. Joyner, A. Broido, B. Canter, K. Gayvert, K. Link, A comparison of computational efficiencies of stochastic algorithms in terms of two infection models, Math. Biosci. Eng. 9 (3) (2012) 487–526.
[25] S. Engblom, Computing the moments of high dimensional solutions of the master equation, Appl. Math. Comput. 180 (2) (2006) 498–515.
[26] P.A. Maginnis, Variance reduction for Poisson and Markov jump processes, Master's thesis, University of Illinois at Urbana-Champaign, 2011.