52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Anticorrelated Discrete-Time Stochastic Simulation

Peter A. Maginnis, Matthew West and Geir E. Dullerud

Abstract— **We provide the first known rigorous theoretical analysis of previously published anticorrelated variance reduction techniques for tau-leaping systems. These algorithms provide a way to reduce the expected MSE of mean estimators by introducing local negative correlation between Monte Carlo sample paths. We prove a recursive equation governing the evolution of these covariances in both the nonlinear and linear cases. Further, we prove sufficient algebraic conditions for variance reduction in the linear rates case that require no stochastic simulation. Finally, we present an example system to illustrate both the application of these tests and to demonstrate their effectiveness.**

I. INTRODUCTION

The results of this paper pertain to previously published [11] anticorrelated stochastic simulation algorithms for variance reduction in discrete time tau-leaping Markov processes. This work focuses primarily on providing rigorous theoretical results to explain numerical behavior observed in the previous paper. Specifically, we provide a recursive, discrete-time equation governing the covariance of the algorithm, and, in considering the linear case, prove an explicit affine matrix equation to determine the behavior of the stochastic algorithm. We also prove an algebraic test to determine the potential benefits of the algorithm *a priori* to stochastic simulation in this linear case. While these algorithms, as shown in [11], are widely applicable to problems of estimation (including particle filters for nonlinear systems), here we focus on variance reduction of stochastic simulation.

Markov processes with transitions governed by a finite number of reaction channels are widely applicable, for instance in aerosol simulations [8] or gene regulation networks [12]. Such systems can be exactly simulated using Gillespie's stochastic simulation algorithm (SSA) [1], but certain applications driven by numerous fast reactions can render this method unwieldy. The tau-leaping method is a fast, discrete-time approximation for simulating such discrete Markov processes, where the number of events within a timestep are sampled from appropriate Poisson distributions. The method was introduced by Gillespie [2], and stability and convergence to the exact SSA are well established [5]. Significant work has been done to further reduce the cost of tau-leaping algorithms, including the development of adaptive step size selection [6], [7] and implicit variants [3]. The algorithms we focus on here are aimed at reducing the variance of ensembles of tau-leaping simulations in order to reduce the error of estimators.

P. A. Maginnis and M. West received funding from the National Science Foundation (NSF) under grant CMG-0934491. G. E. Dullerud was partially supported by AFOSR under grant FA9550-10-1-0573.

In section II we provide some background on anticorrelated sampling methods and context for the discrete-time Markov processes of interest. In section III we show how anticorrelated sample paths of these discrete-time processes can be simulated and provide novel representations of the evolution of their inter-ensemble covariances in both the general nonlinear and linear rate function cases. Next, in section IV, we prove sufficient conditions for these algorithms to reduce variance of mean estimates in the linear case. Finally, in section V we provide an example system to illustrate the application of the sufficient conditions, and provide supporting stochastic simulations that concur with these conditions. Note that this work provides the only known theoretical analysis of the effectiveness of these anticorrelated simulation algorithms for this example.

II. PRELIMINARIES

A. Anticorrelated Stochastic Sampling Algorithm

The usual method of simulating random variables is drawing a sequence of independent, identically distributed (i.i.d.) realizations. These samples can be used to compute statistics in order to estimate properties of some distribution, say its mean. One method to reduce the variance of such mean estimates is anticorrelated sampling, whereby the independence assumption is relaxed and samples are drawn in correlated batches such that the estimators they produce have lower variance. The three examples of this technique that are of interest to this work are antithetic, stratified and hybrid sampling. They are studied in detail in [9] and employed in a particle filtering context in [11]. Briefly, these techniques can be summarized as follows.

In each technique, M random variables are simulated simultaneously such that they are negatively correlated. This process can be repeated in an i.i.d. fashion to produce an arbitrarily large sequence of random variables. Let $F_\lambda(n)$ denote the CDF of the Poisson distribution with parameter λ and consider the formal inverse

$$F_\lambda^{-1}(u) := \inf\{n \in \mathbb{Z}_+ : F_\lambda(n) \geq u\}. \tag{1}$$

Then $S = F_\lambda^{-1}(U) \sim \text{Pois}(\lambda)$ if $U \sim \text{Unif}[0,1]$. In *antithetic* sampling, $M = 2$ as pairs of random variables are drawn simultaneously. To generate each pair, draw a single $U \sim \text{Unif}[0,1]$ and set $S_1 = F_\lambda^{-1}(U)$ and $S_2 = F_\lambda^{-1}(1-U)$. Clearly, S_1 is a $\text{Pois}(\lambda)$ random variable, but $1-U \sim \text{Unif}[0,1]$ as well, so S_2 is clearly also a Poisson random variable with parameter λ, and is negatively correlated with S_1. This process trivially generalizes to produce $S_1 \sim \text{Pois}(\lambda_1)$ and $S_2 \sim \text{Pois}(\lambda_2)$, negatively

correlated for any λ_1, λ_2, and this fact is used in the proof of Lemma 1.

Alternatively, *stratified* sampling draws M simultaneous correlated samples by first partitioning the unit interval $[0.1]$ into M intervals $\{A_r\}_{r=1}^M$. Here we stipulate the intervals be of uniform size, and draw $U_r \overset{\text{indep.}}{\sim} \text{Unif}(A_r), r \in \{1, \ldots, M\}$. Then we apply a uniformly distributed random permutation Π_M to produce $V_r = \Pi_M(U_\ell)$, so that $V_r \sim \text{Unif}[0, 1]$ and each V_{r_1} is negatively correlated to V_{r_2} for any $r_1 \neq r_2$ since neither can appear in the same stratum. Then we simply define $S_r = F_{\lambda_r}^{-1}(V_r)$ to produce stratified Poisson random variables.

While antithetic and stratified techniques are well known and used in some form, they perform reduce variance with different effectivenesses in different regimes of Poisson parameter λ. Knowing this, we proposed in [9] a hybridization that performs at least as well as either original technique in every regime. The *hybrid* technique can be summarized as follows: we partition the unit interval into some even number M of uniform intervals $A_r = [\frac{r-1}{M}, \frac{r}{M}), r \in \{1, \ldots, M\}$. Then sample

$$U_r \overset{\text{indep.}}{\sim} \text{Unif}(A_r), \quad r \in \{1, \ldots, M/2\} \tag{2}$$

$$U_r := 1 - U_{M-r+1}, \quad r \in \{(M/2+1), \ldots, M\} \tag{3}$$

$$V_r = \Pi_M(U_\ell) \tag{4}$$

$$S_r = F_{\lambda_r}^{-1}(V_r). \tag{5}$$

B. Continuous-Time Markov Processes and Tau-leaping

Consider a continuous time Markov processes $X(t)$ moving due a finite set of I reaction channels each with a corresponding rate functions $\rho^i(t, X(t))$. One way of representing such a system is the random time-change representation

$$X(t) = x_0 + \sum_{i=1}^{I} \Upsilon^i \left(\int_0^t \rho^i(s, X(s)) \, \mathrm{d}s \right) \zeta^i, \tag{6}$$

where $t \in [0, T]$ and for each i, $\Upsilon^i(t) \in \mathbb{R}$ is a unit-rate Poisson process and $\zeta^i \in \mathbb{R}^n$ is the vector change in state due to a single instance of the ith reaction A standard method of simulating this system is Gillespie's SSA [1]. However, when reactions occur very frequently, SSA becomes expensive and an approximate method known as tau-leaping, also due to Gillsepie [2], is commonly used. The tau-leaping method is a discrete time approximation where transitions due to the continuous time Poisson process are approximated by corresponding Poisson random variables. That is, for time increment τ, let $t_\ell = \ell\tau$ and $X_\ell \approx X(t_\ell)$ for $\ell \in \{0, \ldots, L\}$, where $L := \max\{\ell : t_\ell \leq T\}$. Then X_ℓ evolves via

$$X_{\ell+1} = X_\ell + \sum_{i=1}^{I} S_\ell^i \left(\rho^i\big(t_\ell, X_\ell\big)\tau \right) \zeta^i, \tag{7}$$

where $S_\ell^i(\lambda) \sim \text{Pois}(\lambda)$. Define $\lambda^i(\ell, X_\ell) = \rho^i\big(t_\ell, X_\ell\big)\tau$ and denote $S_\ell^i\left(\lambda^i(\ell, X_\ell)\right)$ by S_ℓ^i. Thus (7) becomes

$$X_{\ell+1} = X_\ell + \sum_{i=1}^{I} S_\ell^i \zeta^i. \tag{8}$$

III. ANTICORRELATED STOCHASTIC SIMULATION

We now focus our attention on variance reduced simulation of systems like (8). We will introduce correlation between some of the Poisson samples used to simulate these systems in order to produce ensembles of correlated paths. When we can extend the negative correlation of our Poisson samples to negatively correlated paths, the variance of pathwise mean-eastimators will go down. Allowing for a slight abuse of notation, consider an ensemble of M tau-leaping systems given by:

$$X_{r,t+1} = X_{r,t} + \sum_{i=1}^{I} S_{r,t}^i \zeta^i, \qquad r \in \{1, \ldots, M\}, \tag{9}$$

where, for each event channel $i \in \{1, \ldots, I\}$, the collection $\{S_{r,t}^i\}_{r=1}^M \subset \mathbb{R}$ are simulated as an anticorrelated ensemble, and each element is assigned as input to a different path. Since

$$S_{r,t}^i \overset{\text{corr.}}{\sim} \text{Pois}(\lambda^i(t, X_{r,t})), \tag{10}$$

each of these paths is a fair draw of the system. Now, all $S_{r,t}^i$ are mutually correlated, both with other members of their ensemble through the anticorrelated simulation technique, but also due to interactions between paths $X_{r,t}$ which may now also be correlated. Define

$$\widetilde{\Psi}_t^M := \sum_{r=1}^{M} \widetilde{X}_{r,t} \quad \text{s.t. } \widetilde{X}_{r,t} \overset{\text{i.i.d.}}{\sim} X_t \tag{11}$$

the naive mean estimator of the stochastic process and the corresponding anticorrelated mean estimator

$$\Psi_t^M := \sum_{r=1}^{M} X_{r,t} \quad \text{s.t. } X_{r,t} \overset{\text{corr.}}{\sim} X_t \tag{12}$$

simulated as above. To prove that the anticorrelated estimator has lower variance than the naive estimator in some sense (say, the MSE: $\operatorname{tr} \text{Cov}(\Psi_t^M) \leq \operatorname{tr} \text{Cov}(\widetilde{\Psi}_t^M)$ for each t), it is sufficient to show that $\operatorname{tr} \text{Cov}(X_{r_1,t}, X_{r_2,t}) \leq 0$ for any $r_1 \neq r_2 \in \{1, \ldots, M\}$. We now derive a recursive matrix equation to compute this pairwise covariance.

Proposition 1: For any collection of rate functions $\{\lambda^i\}_{i=1}^I$,

$$\begin{aligned}
\text{Cov}(X_{r_1,t+1}, X_{r_2,t+1}) &= \text{Cov}(X_{r_1,t}, X_{r_2,t}) \\
&+ \sum_{i=1}^{I} \text{Cov}(X_{r_1,t}, \lambda^i(t, X_{r_2,t}))\zeta^{i\top} \\
&+ \sum_{i=1}^{I} \zeta^i \, \text{Cov}(\lambda^i(t, X_{r_1,t}), X_{r_2,t}) \\
&+ \sum_{i_1=1}^{I} \sum_{i_2=1}^{I} \text{Cov}(\lambda^{i_1}(t, X_{r_1,t}), \lambda^{i_2}(t, X_{r_2,t}))\zeta^{i_1}\zeta^{i_2\top}. \\
&+ \sum_{i=1}^{I} c_{r_1,r_2,t}^i \zeta^i \zeta^{i\top},
\end{aligned} \tag{13}$$

where

$$c^i_{r_1,r_2,t} := \mathbb{E}\big[\big(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t})\big) \cdot \big(S^i_{r_2,t} - \lambda^i(t, X_{r_2,t})\big)\big] \tag{14}$$

is a real number.

Proof: First, we expand the covariance in terms of the last time step using the system definition and the bilinearity property of the covariance operator to obtain

$$\begin{aligned}
&\mathrm{Cov}(X_{r_1,t+1}, X_{r_2,t+1}) \\
&= \mathrm{Cov}\left(X_{r_1,t} + \sum_{i=1}^{I} S^i_{r_1,t}\zeta^i, X_{r_2,t} + \sum_{i=1}^{I} S^i_{r_2,t}\zeta^i\right) && (15)\\
&= \mathrm{Cov}(X_{r_1,t}, X_{r_2,t}) + \sum_{i=1}^{I} \mathrm{Cov}(X_{r_1,t}, S^i_{r_2,t}){\zeta^i}^\top \\
&\quad + \sum_{i=1}^{I} \zeta^i \,\mathrm{Cov}(S^i_{r_1,t}, X_{r_2,t}) \\
&\quad + \sum_{i_1=1}^{I}\sum_{i_2=1}^{I} \mathrm{Cov}(S^{i_1}_{r_1,t}, S^{i_2}_{r_2,t})\zeta^{i_1}{\zeta^{i_2}}^\top. && (16)
\end{aligned}$$

Observe that the cross term

$$\begin{aligned}
&\mathrm{Cov}(X_{r_1,t}, S^i_{r_2,t}) \\
&= \mathbb{E}\big[(X_{r_1,t} - \mathbb{E}[X_{r_1,t}])(S^i_{r_2,t} - \mathbb{E}[S^i_{r_2,t}])\big] && (17)\\
&= \mathbb{E}\Big[\mathbb{E}\big[(X_{r_1,t} - \mathbb{E}[X_{r_1,t}]) \\
&\qquad \cdot (S^i_{r_2,t} - \mathbb{E}[S^i_{r_2,t}])|X_{r_1,t}, X_{r_2,t}\big]\Big] && (18)\\
&= \mathbb{E}\Big[(X_{r_1,t} - \mathbb{E}[X_{r_1,t}]) \\
&\qquad \cdot \mathbb{E}\big[S^i_{r_2,t} - \mathbb{E}[S^i_{r_2,t}]|X_{r_1,t}, X_{r_2,t}\big]\Big] && (19)\\
&= \mathbb{E}\Big[(X_{r_1,t} - \mathbb{E}[X_{r_1,t}]) \\
&\qquad \cdot (\lambda^i(t, X_{r_2,t}) - \mathbb{E}[\lambda^i(t, X_{r_2,t})])\Big] && (20)\\
&= \mathrm{Cov}(X_{r_1,t}, \lambda^i(t, X_{r_2,t})), && (21)
\end{aligned}$$

where the second to last equality holds since $\mathbb{E}[S^i_{r_2,t}|X_{r_1,t}, X_{r_2,t}] = \lambda^i(t, X_{r_2,t})$ and $\mathbb{E}[S^i_{r_2,t}] = \mathbb{E}[\lambda^i(t, X_{r_2,t})]$. Similarly,

$$\mathrm{Cov}(S^i_{r_1,t}, X_{r_2,t}) = \mathrm{Cov}(\lambda^i(t, X_{r_1,t}), X_{r_2,t}). \tag{22}$$

We treat the last term of (16) by considering two different cases. When $i_1 \neq i_2$,

$$\begin{aligned}
&\mathrm{Cov}(S^{i_1}_{r_1,t}, S^{i_2}_{r_2,t}) \\
&= \mathbb{E}\big[(S^{i_1}_{r_1,t} - \mathbb{E}[S^{i_1}_{r_1,t}])(S^{i_2}_{r_2,t} - \mathbb{E}[S^{i_2}_{r_2,t}])\big] && (23)\\
&= \mathbb{E}\Big[\mathbb{E}\big[(S^{i_1}_{r_1,t} - \mathbb{E}[S^{i_1}_{r_1,t}]) \\
&\qquad \cdot (S^{i_2}_{r_2,t} - \mathbb{E}[S^{i_2}_{r_2,t}])|X_{r_1,t}, X_{r_2,t}\big]\Big] && (24)\\
&= \mathbb{E}\Big[\mathbb{E}\big[S^{i_1}_{r_1,t} - \mathbb{E}[S^{i_1}_{r_1,t}]|X_{r_1,t}, X_{r_2,t}\big] \\
&\qquad \cdot \mathbb{E}\big[S^{i_2}_{r_2,t} - \mathbb{E}[S^{i_2}_{r_2,t}]|X_{r_1,t}, X_{r_2,t}\big]\Big] && (25)\\
&= \mathbb{E}\Big[(\lambda^{i_1}(t, X_{r_1,t}) - \mathbb{E}[\lambda^{i_1}(t, X_{r_1,t})]) \\
&\qquad \cdot (\lambda^{i_2}(t, X_{r_2,t}) - \mathbb{E}[\lambda^{i_2}(t, X_{r_2,t})])\Big] && (26)\\
&= \mathrm{Cov}(\lambda^{i_1}(t, X_{r_1,t}), \lambda^{i_2}(t, X_{r_2,t})), && (27)
\end{aligned}$$

since, when conditioned on $(X_{r_1,t}, X_{r_2,t})$, $S^{i_1}_{r_1,t}$ and $S^{i_2}_{r_2,t}$ are independent. When $i_1 = i_2 = i$, however, $S^{i_1}_{r_1,t}$ and $S^{i_2}_{r_2,t}$ are conditionally dependent by construction. In this case, consider the equivalent form

$$\begin{aligned}
&\mathrm{Cov}(S^i_{r_1,t}, S^i_{r_2,t}) \\
&= \mathbb{E}\Big[\mathbb{E}\big[(S^i_{r_1,t} - \mathbb{E}[S^i_{r_1,t}]) \\
&\qquad \cdot (S^i_{r_2,t} - \mathbb{E}[S^i_{r_2,t}])|X_{r_1,t}, X_{r_2,t}\big]\Big] && (28)\\
&= \mathbb{E}\Bigg[\mathbb{E}\Big[\big(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t}) \\
&\qquad + \lambda^i(t, X_{r_1,t}) - \mathbb{E}[S^i_{r_1,t}]\big) \\
&\qquad \cdot \big(S^i_{r_2,t} - \lambda^i(t, X_{r_2,t}) \\
&\qquad + \lambda^i(t, X_{r_2,t}) - \mathbb{E}[S^i_{r_2,t}]\big)|X_{r_1,t}, X_{r_2,t}\Big]\Bigg] && (29)\\
&= \mathbb{E}\Bigg[\mathbb{E}\Big[\big(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t})\big) \cdot \big(S^i_{r_2,t} - \lambda^i(t, X_{r_2,t})\big) \\
&+ \big(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t})\big) \cdot \big(\lambda^i(t, X_{r_2,t}) - \mathbb{E}[S^i_{r_2,t}]\big) \\
&+ \big(S^i_{r_2,t} - \lambda^i(t, X_{r_2,t})\big) \cdot \big(\lambda^i(t, X_{r_1,t}) - \mathbb{E}[S^i_{r_1,t}]\big) \\
&+ \big(\lambda^i(t, X_{r_1,t}) - \mathbb{E}[S^i_{r_1,t}]\big) \cdot \big(\lambda^i(t, X_{r_2,t}) - \mathbb{E}[S^i_{r_2,t}]\big) \\
&\qquad |X_{r_1,t}, X_{r_2,t}\Big]\Bigg]. && (30)
\end{aligned}$$

Each of the two cross terms vanish, since $\big(\lambda^i(t, X_{r,t}) - \mathbb{E}[S^i_{r,t}]\big)$ is $X_{r,t}$-measurable, can thus be pulled out of the conditional expectation, and is mean zero. The last term is identical to the case when $i_1 \neq i_2$. Thus (30) can be written as

$$\begin{aligned}
&\mathrm{Cov}(S^i_{r_1,t}, S^i_{r_2,t}) \\
&\quad = c^i_{r_1,r_2,t} + \mathrm{Cov}(\lambda^i(t, X_{r_1,t}), \lambda^i(t, X_{r_2,t})). && (31)
\end{aligned}$$

The result follows by substituting (21),(22),(27), and (31) into (16). ■

For a general nonlinear collection of rates $\{\lambda^i(t, X_t)\}$, this equation is difficult to analyze or solve. Instead, consider the

case of linear rate functions. The next result proves a useful structural property in this case.

Corollary 1: If the rate functions are linear, i.e. $\lambda^i(t, X_t) = {\kappa^i}^\top X_t, \kappa^i \in \mathbb{R}^n$ for each i, then for $r_1 \neq r_2$ the recursive covariance equation is affine. Specifically

$$\begin{aligned}&\operatorname{Cov}(X_{r_1,t+1}, X_{r_2,t+1})\\ &\quad = \mathcal{L}(\operatorname{Cov}(X_{r_1,t}, X_{r_2,t})) + \sum_{i=1}^{I} c^i_{r_1,r_2,t} \zeta^i {\zeta^i}^\top, \end{aligned} \tag{32}$$

where $\mathcal{L}$ is a known, time-invariant linear map from the space of symmetric matrices to itself.

Proof: Suppose all rate functions are linear. Then (13) becomes

$$\begin{aligned}\operatorname{Cov}(X_{r_1,t+1}, X_{r_2,t+1}) &= \operatorname{Cov}(X_{r_1,t}, X_{r_2,t})\\ &+ \sum_{i=1}^{I} \operatorname{Cov}(X_{r_1,t}, X_{r_2,t}) \kappa^i {\zeta^i}^\top\\ &+ \sum_{i=1}^{I} \zeta^i {\kappa^i}^\top \operatorname{Cov}(X_{r_1,t}, X_{r_2,t})\\ &+ \sum_{i_1=1}^{I}\sum_{i_2=1}^{I} \zeta^{i_1} {\kappa^{i_1}}^\top \operatorname{Cov}(X_{r_1,t}, X_{r_2,t}) \kappa^{i_2} {\zeta^{i_2}}^\top\\ &+ \sum_{i=1}^{I} c^i_{r_1,r_2,t} \zeta^i {\zeta^i}^\top. \end{aligned} \tag{33}$$

It is clear that (33) is an affine recursive matrix equation with unknown $\operatorname{Cov}(X_{r_1,t}, X_{r_2,t})$. In order to represent $\mathcal{L}$ more compactly, we define $A^{r_1,r_2}_t := \operatorname{Cov}(X_{r_1,t}, X_{r_2,t})$ and $Q := \sum_{i=1}^{I} \zeta^i {\kappa^i}^\top \in \mathbb{R}^{n\times n}$ so that

$$\begin{aligned}A^{r_1,r_2}_{t+1} &= A^{r_1,r_2}_t + A^{r_1,r_2}_t Q^\top + Q A^{r_1,r_2}_t\\ &\quad + Q A^{r_1,r_2}_t Q^\top + \sum_{i=1}^{I} c^i_{r_1,r_2,t} \zeta^i {\zeta^i}^\top. \end{aligned} \tag{34}$$

Note that the linear map that acts on A^{r_1,r_2}_t depends only on parameters of the system, namely the jump vectors and jump rates, and is thus time-invariant. ∎

Before proceeding, we require a simple result that $c^i_{r_1,r_2,t} \leq 0$.

Lemma 1: For any i, if $\{S^i_{r,t}\}_{r=1}^M$ is simulated using one of our three anticorrelated techniques (antithetic, stratified or hybrid), then, for $r_1 \neq r_2$

$$\begin{aligned}&c^i_{r_1,r_2,t}\\ &\quad := \mathbb{E}\big[\big(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t})\big) \cdot \big(S^i_{r_2,t} - \lambda^i(t, X_{r_2,t})\big)\big] \leq 0 \end{aligned} \tag{35}$$

Proof: We proceed by proving the antithetic case, then the stratified case, and the hybrid case follows as an immediate corollary. In the antithetic case, $M = 2$, and we first observe that for any $\lambda_1, \lambda_2 \in \mathbb{R}_+$, $\operatorname{Cov}(S_1, S_2) \leq 0$ where $S_i \sim \operatorname{Pois}(\lambda_i)$ and S_1 and S_2 and antithetically paired. Indeed, for some $U \sim \operatorname{Unif}[0,1]$,

$$\operatorname{Cov}(S_1, S_2) = \operatorname{Cov}(F^{-1}_{\lambda_1}(U), F^{-1}_{\lambda_2}(1-U)) \tag{36}$$

$$= -\operatorname{Cov}(F^{-1}_{\lambda_1}(U), -F^{-1}_{\lambda_2}(1-U)) \leq 0 \tag{37}$$

due to a result proven in [4] since $F^{-1}_{\lambda_1}(U)$ and $-F^{-1}_{\lambda_2}(1-U)$ are both non-decreasing functions of U. Now suppose $S_{r_i}(t,x) \sim \operatorname{Pois}(\lambda(t,x))$. Thus for any fixed $x_1, x_2 \in \mathbb{R}^n$,

$$0 \geq \operatorname{Cov}(S_{r_1}(t, x_1), S_{r_2}(t, x_2)) \tag{38}$$

$$= \mathbb{E}[(S_{r_1}(t,x_1) - \lambda(t,x_1))(S_{r_2}(t,x_2) - \lambda(t,x_2))]. \tag{39}$$

Since this is true for any x_i, it is also true that

$$\begin{aligned}0 \geq \mathbb{E}[&(S^i_{r_1,t} - \lambda^i(t, X_{r_1,t})\\ &\cdot (S^i_{r_2,t} - \lambda^i(t, X_{r_2,t}) | X_{r_1,t}, X_{r_2,t}], \end{aligned} \tag{40}$$

almost surely, since this is exactly the same integral (in U) for given $X_{r_1,t}, X_{r_2,t}$ random. Taking expectation of both sides, we get

$$0 \geq c^i_{r_1,r_2,t}. \tag{41}$$

To prove the stratified case, it is clear that, as above, we need only prove that the nonpositivity condition holds for arbitrary $\lambda_1, \lambda_2 \in \mathbb{R}_+$. Then the claim would hold after taking randomized $\lambda^i(t, X_{r_1,t})$ in a conditional expectation and integrating both sides. Suppose we have M strata $A_r, r \in \{1, \ldots, M\}$. Consider any pair of stratified Poisson samples $S_{r_1} = F^{-1}_{\lambda_1}(V_{r_1}) \sim \operatorname{Pois}(\lambda_1)$ and $S_{r_2} = F^{-1}_{\lambda_2}(V_{r_2}) \sim \operatorname{Pois}(\lambda_2)$, where $V_{r_1}, V_{r_2} \sim \operatorname{Unif}[0,1]$ are stratified uniform random variables. Let $A(v)$ denote the stratum containing $v \in [0,1]$. Then

$$\begin{aligned}&\operatorname{Cov}(S_{r_1}, S_{r_2})\\ &\quad = \int_{v_1=0}^{1} \int_{v_2 \in [0,1]\setminus A(v_1)} (F^{-1}_{\lambda_1}(v_1) - \lambda_1)\\ &\qquad\qquad \cdot (F^{-1}_{\lambda_2}(v_2) - \lambda_2)\, \mathrm{d}v_2\, \mathrm{d}v_1 \end{aligned} \tag{42}$$

$$\begin{aligned}&\quad = \int_{v_1=0}^{1} \int_{v_2=0}^{1} (F^{-1}_{\lambda_1}(v_1) - \lambda_1)\\ &\qquad\qquad \cdot (F^{-1}_{\lambda_2}(v_2) - \lambda_2)\, \mathrm{d}v_2\, \mathrm{d}v_1\\ &\quad - \int_{v_1=0}^{1} \int_{v_2 \in A(v_1)} (F^{-1}_{\lambda_1}(v_1) - \lambda_1)\\ &\qquad\qquad \cdot (F^{-1}_{\lambda_2}(v_2) - \lambda_2)\, \mathrm{d}v_2\, \mathrm{d}v_1 \end{aligned} \tag{43}$$

$$= -\operatorname{Cov}(S_{r_1}, \widehat{S}_2) \tag{44}$$

where $\widehat{S}_2 := F^{-1}_{\lambda_2}(\widehat{V}_2)$ for $\widehat{V}_2 \sim \operatorname{Unif} A(V_{r_1}) \sim \operatorname{Unif}[0,1]$. The last equality uses the fact that the first integral in (43) is precisely the covariance of independent Poisson variables and is thus equal to 0. Since V_{r_1} and $\widehat{V}_2$ are forced to be drawn from the same stratum but are otherwise independently sampled, it is trivial to see that $\operatorname{Cov}(S_{r_1}, \widehat{S}_2) \geq 0$ and the proof is complete. Using a result in [9], the covariance of the hybrid estimator is at most the covariance of either of the corresponding antithetic or stratified samplers. ∎

IV. SUFFICIENT CONDITIONS FOR VARIANCE REDUCTION

In order to develop a sufficient, testable condition for ensemble variance reduction of our algorithm, recall that $\operatorname{tr}\operatorname{Cov}(\Psi^M_t) \leq \operatorname{tr}\operatorname{Cov}(\widetilde{\Psi}^M_t)$ for each $t > 0$ if $\operatorname{tr} A^{r_1,r_2}_t \leq 0$ for each r_1, r_2, and $t > 0$. If our initial conditions $\{X_{r,0}\}$

are deterministic or independent, then we have $A_0^{r_1,r_2} = 0$ and thus

$$A_1^{r_1,r_2} = \sum_{i=1}^{I} c^i_{r_1,r_2,0} \zeta^i {\zeta^i}^\top \tag{45}$$

$$\implies \operatorname{tr} A_1^{r_1,r_2} \leq 0. \tag{46}$$

If we consider the set of $n \times n$-matrices as a vector space together with the Frobenius inner product ($\langle A, B\rangle := \operatorname{tr} AB^\top$), then we'd like the matrices $\{A_t^{r_1,r_2}\}_{t>0}$ to each remain in the half-space $\mathcal{H}_- := \{A \in \mathbb{R}^{n\times n} : \langle A, I_n\rangle \leq 0\}$. The second term of the affine equation (32) lies always in the negative cone $\mathcal{R} := \{\sum_{i=1}^I c^i \zeta^i {\zeta^i}^\top : c^i \leq 0 \text{ for each } i\} \subset \mathcal{H}_-$, which leads us to our first sufficient condition for variance reduction.

Proposition 2: Suppose $A_0^{r_1,r_2} = 0$. For any $R \in \mathcal{R}$, if $\mathcal{L}^\ell(R) \in \mathcal{H}_-$ for every $\ell \geq 1$, then $A_t^{r_1,r_2} \in \mathcal{H}_-$ for every $t \geq 0$. Further, if this condition holds for each $1 \leq r_1 \neq r_2 \leq M$, then $\operatorname{tr}\operatorname{Cov}(\Psi_t^M) \leq \operatorname{tr}\operatorname{Cov}(\widetilde{\Psi}_t^M)$, that is, the anticorrelated mean estimator has lower expected mean-square error than a naive estimator.

Proof: Define $R_t^{r_1,r_2} := \sum_{i=1}^I c^i_{r_1,r_2,t} \zeta^i {\zeta^i}^\top$. Then, for each $r_1 \neq r_2$ and $t \geq 0$, $R_t^{r_1,r_2} \in \mathcal{R}$. For each $t \geq 1$, a solution to (32) is given by

$$A_t^{r_1,r_2} := \sum_{k=0}^{t-1} \mathcal{L}^{t-k-1}(R_k^{r_1,r_2}). \tag{47}$$

Since $\mathcal{L}^{t-k-1}(R_k^{r_1,r_2}) \in \mathcal{H}_-$ and since the half-space $\mathcal{H}_-$ is closed under addition, $A_t^{r_1,r_2} \in \mathcal{H}_-$ for every $t \geq 0$. ∎

A less general but more easily verifiable condition is provided in the next result, in the case when there exists an eigenbasis $\{v_j\}_{j=1}^{n^2}$ for $\mathcal{L}$ in the space of $n \times n$ matrices.

Proposition 3: Suppose the eigenvectors $\{v_j\}_{j=1}^{n^2}$ of $\mathcal{L}$ form a basis for $\mathbb{R}^{n\times n}$. For each $i \in \{1, \ldots, M\}$, define $R^i := \zeta^i {\zeta^i}^\top$ and let ρ_j^i denote the jth coordinate of R^i with respect to this basis. If, for each $j \in \{1, \ldots, n^2\}$, $\rho_j^i \langle v_j, I_n\rangle \geq 0$ and $\rho_j^i \neq 0 \implies \lambda_j \geq 0$, then $A_t^{r_1,r_2} \in \mathcal{H}_-$ for every $t \geq 0$ and $r_1 \neq r_2$.

Proof: Any $R \in \mathcal{R}$ can be expressed as

$$R = \sum_{i=1}^{I} c^i R^i = \sum_{i=1}^{I} c^i \Big(\sum_{j=1}^{n^2} \rho_j^i v_j \Big), \tag{48}$$

and for any $\ell \geq 0$,

$$\mathcal{L}^\ell(R) = \sum_{i=1}^{I} c^i \Big(\sum_{j=1}^{n^2} \lambda_j^\ell \rho_j^i v_j \Big), \tag{49}$$

If $\rho_j^i \langle v_j, I_n\rangle \geq 0$ and $\rho_j^i \neq 0 \implies \lambda_j \geq 0$, then $c^i {\lambda_j}^\ell \rho_j^i v_j \in \mathcal{H}_-$ for any $c^i \leq 0$. Thus, by closure of the half-space under addition, $\mathcal{L}^\ell(R) \in \mathcal{H}_-$ for any $\ell \geq 0$ and the conditions of Propositon 2 are satisfied for any $r_1 \neq r_2$. ∎

Thus passage of a simple algebraic test is sufficient to prove variance reduction. Note that this condition depends only on parameters of the system (ζ^i and κ^i) and thus requires no dynamic checking or even stochastic simulation to verify.

V. EXAMPLE SYSTEM

Suppose we consider a 2-dimensional system with particle objects A and B. Suppose the reaction channels are given by

$$A \xrightarrow{\lambda^1} B \tag{50}$$

$$2A \xleftarrow{\lambda^2} 2B, \tag{51}$$

where the state $X_t \in \mathbb{Z}^2$ is the number of particles of each species. Suppose that the rate functions are linear, and the parameters of the system are given by:

$$\begin{aligned} \zeta^1 &= \begin{pmatrix} -1 \\ 1 \end{pmatrix} & \kappa^1 &= \begin{pmatrix} 0.03 \\ 0.01 \end{pmatrix} \\ \zeta^2 &= \begin{pmatrix} 2 \\ -2 \end{pmatrix} & \kappa^2 &= \begin{pmatrix} 0.01 \\ 0.03 \end{pmatrix} \end{aligned} \tag{52}$$

Note specifically here that this system is not meant to represent chemical reactions and is decidedly not governed by mass action type rate functions. It is easy to compute the quantities Q and R^i to be

$$R^1 = \zeta^1 {\zeta^1}^\top = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{53}$$

$$R^2 = \zeta^2 {\zeta^2}^\top = \begin{pmatrix} 4 & -4 \\ -4 & 4 \end{pmatrix} \tag{54}$$

$$Q = \sum_{i=1}^{2} \zeta^i {\kappa^i}^\top = \begin{pmatrix} -0.01 & 0.05 \\ 0.01 & -0.05 \end{pmatrix} \tag{55}$$

To verify the conditions of Proposition 3, consider the vectorized form of 34:

$$\begin{aligned} \operatorname{vec}(A_{t+1}^{r_1,r_2}) = \Big(& I_{n^2} + Q \otimes I_n + I_n \otimes Q + Q \otimes Q \\ & - \sum_{i=1}^{I} \zeta^i {\kappa^i}^\top \otimes \zeta^i {\kappa^i}^\top \Big) \operatorname{vec}(A_t^{r_1,r_2}) \\ & + \sum_{i=1}^{I} c^i_{r_1,r_2,t} \operatorname{vec}(R^i). \end{aligned} \tag{56}$$

In these coordinates, it is easy to express $\mathcal{L}$ as left multiplication by a 4×4 matrix L, specifically

$$L \approx \begin{pmatrix} 0.989 & 0.048 & 0.048 & -0.001 \\ 0.011 & 0.942 & 0.002 & 0.051 \\ 0.011 & 0.002 & 0.942 & 0.051 \\ -0.001 & 0.008 & 0.008 & 0.899 \end{pmatrix} \tag{57}$$

with eigenvalues $\approx \{1, 0.94, 0.882, 0.94\}$ and corresponding eigenbasis

$$[v_1, v_2, v_3, v_4] \approx \begin{pmatrix} 1 & 43.3 & 0 & -5.3 \\ -1 & 9.6 & -1 & 4.3 \\ -1 & 9.6 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}. \tag{58}$$

In this example, it's easy to compute the coordinates of R^1, R^2 with respect to this basis. For $j \in \{2, 3, 4\}$, the

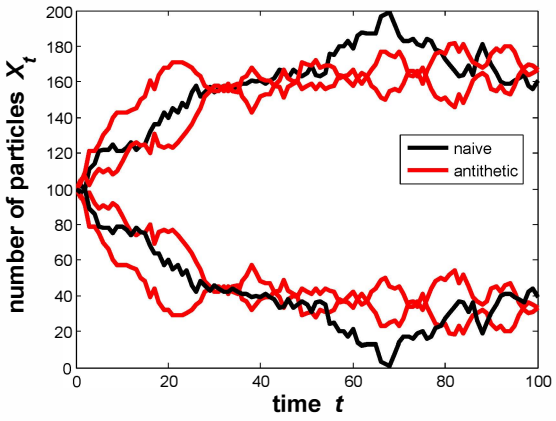

Fig. 1. A single naive sample path (components of X_t plotted separately) and two antithetically paired sample paths. Particle scale in this case is determined by the initial condition of each species, here 100 particles.

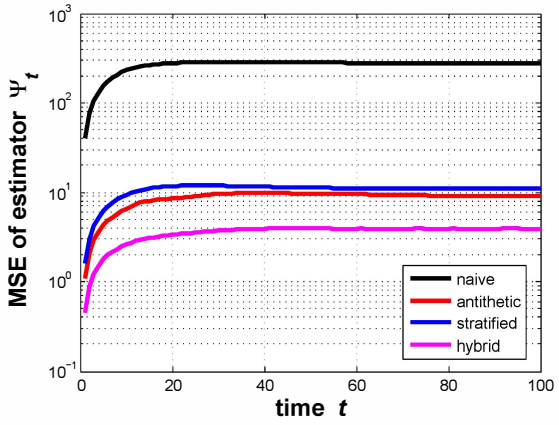

Fig. 2. The MSE of a 4-sample mean estimator Ψ_t^4 plotted versus time for each of the simulation algorithms. Standard Monte Carlo is shown in black. Note the dramatic decrease in estimator variance, as well as the fact that, as predicted, the hybrid technique enjoys the lowest MSE. MSE computed via sample variance for 10^5 samples.

component $\rho_j^i = 0$, and $\rho_1^1 = 1$ and $\rho_1^2 = 4$. The conditions of the Proposition are satisfied since $\lambda_1 = 1 \geq 0$ and

$$\rho_1^i v_1^\top \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \geq 0, \tag{59}$$

for both $i = 1$ and $i = 2$. Note that algebraic tests of this type are very cheap and need be performed at most n^2 times.

A. Stochastic Simulation Example

As an additional demonstration, the system of the previous example was simulated stochastically in order to directly compare the naive technique to the three anticorrelated sampling methods proposed. A 4 sample mean estimator was constructed for each anticorrelated technique as well as for the naive i.i.d. method. For illustrative purposes, a single naive sample path and a pair of antithetic paths are plotted versus time in Figure 1. The mean square error of each mean-estimator (i.e. $\operatorname{tr}\operatorname{Cov}(\Psi_t^4)$) is also plotted versus time in Figure 2. Note the confirmation of the predicted reduction in variance, in this case up to 2 orders of magnitude reduction in MSE for the hybrid anticorrelated algorithm.

REFERENCES

[1] D. T. Gillespie, "An exact method for numerically simulating the stochastic coalescence process in a cloud", *Journal of the Atmospheric Sciences*, vol. 32, 1975, pp 1977-1989.

[2] D. T. Gillespie, "Approximate accelerated stochastic sImulation of chemically reacting systems", *J. Chem. Phys.*, vol. 115(4), 2001, pp 1716-1733.

[3] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, "Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method", *J. Chem. Phys.*, 119:12784, 2003.

[4] K. D. Schmidt, On the covariance of monotone functions of a random variable, unpublished note, 2003.

[5] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, "Consistency and stability of tau-leaping schemes for chemical reaction systems", *Multiscale Modeling and Simulation*, vol. 4(3):867895, 2005.

[6] Y. Cao, D. T. Gillespie, and L. R. Petzold, "Efficient stepsize selection for the tau-leaping simulation method". *J. Chem. Phys.*, 124:044109, 2006.

[7] David F. Anderson, "Incorporating postleap checks in tau-leaping", *J. Chem. Phys.*, vol. 128(5):054103, 2008.

[8] N. Riemer, M. West, R. A. Zaveri, and R. C. Easter, "Simulating the evolution of soot mixing state with a particle-resolved aerosol model", *Journal of Geophysical Research*, vol. 114, 2009.

[9] P. A. Maginnis, "Variance reduction for Poisson and Markov jump processes", Master's thesis, UIUC, 2011.

[10] P. A. Maginnis, M. West, and G. E. Dullerud, Variance-reduced tau-leaping using anticorrelated sample paths, Submitted.

[11] P. A. Maginnis, M. West, G. Dullerud, "Application of variance reduction techniques for tau-leaping systems to particle filters", *IEEE Conference on Decision and Control*, Maui, HI, 2012.

[12] C. Briat and M. Khammash, "Computer control of gene expression: Robust setpoint tracking of protein mean and variance using integral feedback", *IEEE Conference on Decision and Control*, Maui, HI, 2012.