2016 American Control Conference (ACC)
Boston Marriott Copley Place
July 6-8, 2016. Boston, MA, USA

Variance-Reduced Model Predictive Control of Markov Jump Processes

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

Abstract— **We present an algorithm for variance-reduced Monte Carlo estimates of the expected cost-to-go used in the stochastic model predictive control of Markov jump processes. Specifically, we extend previous work on antithetic stochastic simulation of Markov chains with a finite number of reaction classes to the approximate computation of an expected cost function of a controlled process. In the presence of strict constraints on number of available Monte Carlo samples, we demonstrate significant reduction in the number of Monte Carlo simulations required to achieve a particular cost, including a factor of two reduction in the small resource limit, for a simplified, nonlinear chemical reaction model.**

I. Introduction

Model predictive control (MPC) is an intuitive framework that is well suited for control of complex, large-scale systems under hard control constraints [14]. Of particular interest is the application of MPC to uncertain systems, where the use of feedback is essential. Within this large class, which includes robust MPC, we focus on stochastic MPC, which is usually characterized by a combination of stochastic dynamics (possibly characterized by stochastic noise) and/or probabilistic constraints [9]. Specifically, we are interested in techniques that leverage Monte Carlo simulation (often called scenario-based methods in this context) in service of approximating solutions to the finite-horizon open-loop stochastic optimization component of MPC implementation. We leverage previous work on anticorrelated simulation of Markov jump processes (a broad class of potentially non-linear and non-Gaussian Markov processes on a countable state-space) to reduce the variance of mean estimates of the finite-horizon expected cost. The goal is to allow for a reduced Monte Carlo budget to achieve the same or better performance of stochastic MPC.

The two main contributions of this paper are the presentation of an algorithm for variance-reduced stochastic model predictive control of Markov jump processes and its demonstration on a simple, non-linear chemical reaction network.

The use of simulation based techniques for the treatment of stochastic MPC has been studied in many different settings over the last fifteen years. Work on the optimal controller synthesis problem using sampling techniques from statistical learning theory [20] and subsequent refinements for linear systems with convex constraints [4] have successfully provided lower bounds on the sampling budget required to satisfy probabilistic constraints. Others have tackled the MPC problem constructed as linear systems with quadratic constraints [2] as well as general Markov processes on finite state spaces [15]. Approaches to solve the stochastic optimization problems related to the more general, Bayesian setting have employed both Markov chain Monte Carlo in the context of simulated annealing [11] and sequential Monte Carlo particle methods [8].

The stochastic dynamical setting we consider here is the Markov jump process, a continuous-time, countable state-space process that experiences transitions that can be classified by a finite number of reaction channels. This class is a broad collection of systems, and appears commonly in models for chemical reaction systems [18], gene regulation systems [3], and atmospheric aerosol simulation [19].

To simulate the model, Gillespie's stochastic simulation algorithm (SSA) [6] is frequently used. However, when the frequency of at least some of the reaction events is relatively large, SSA can quickly become expensive and impractical. In these cases, a discrete-time approximation method known as tau-leaping (also due to Gillespie [7]) is often used. The tau-leaping method involves simulating the system at discrete time intervals, where the number of transitions due to each continuous time Poisson process during that interval are approximated by the sampling of an appropriately chosen Poisson random variable. The convergence properties of the tau-leaping algorithm have been rigorously proven [17] and many variants exist, including implicit tau-leaping [16] and adaptive stepping methods [1], [5]. For this work, we restrict our attention to a commonly used explicit, fixed step size approach.

Negatively correlated ensembles of sample trajectories from such systems can be easily simulated [12], [13] in order to produce reduced-variance (and reduced-error) estimates of a particular expected value. In this paper, we show how such anticorrelated ensembles can be drawn for open loop simulations of the process during MPC to produce reduced error estimates of the true average cost. This is in an effort to reduce the number of Monte Carlo simulations required to accurately estimate the expected cost of a candidate sequence of control actions.

II. Background

A. Markov Jump Processes and τ-leaping Simulation

We consider the class of continuous time Markov processes evolving via a finite set of I reactions, each with a rate function $\rho^i(t, X(t))$ that governs the frequency of its corresponding reaction. Such systems evolve on a countable state-space, and a useful representation of them is the random time-change due to Kurtz [10]

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

978-1-4673-8682-1/$31.00 ©2016 AACC

where, for each i, $\Upsilon^i \in \mathbb{R}$ is an independent, unit-rate Poisson process, $\zeta^i \in \mathbb{R}^n$ is the vector change in state due to a single occurrence of the ith reaction (i.e. $\zeta^i = X(t_+) - X(t_-)$ if an instance of the ith reaction occurs at t), and $x_0 \in \mathbb{R}^n$ is the initial condition, which can be either deterministic or random.

We will simulate realizations of such systems using the tau-leaping method. Recall, the tau-leaping method approximately samples the system at discrete time intervals; we approximate the number of transitions due to each continuous time Poisson process during that interval by simulating an appropriately chosen Poisson random variable. In particular, for time increment τ, let $t_\ell = \ell\tau$ and $X_\ell \approx X(t_\ell)$ for $\ell \in \mathbb{N}$. Then the evolution of X_ℓ is given by

$$X_{\ell+1} = X_\ell + \sum_{i=1}^{I} S_\ell^i(\rho^i(t_\ell, X_\ell)\tau)\zeta^i, \tag{2}$$

where $S_\ell^i(\lambda)$ is a Poisson random variable with mean λ. Effectively, then, tau-leaping is equivalent to a left Euler approximation of the integral in (1). Allowing for an abuse of notation, denote the discrete time index ℓ by t and we may, for the sake of compactness, write

$$X_{t+1} = X_t + \sum_{i=1}^{I} S_t^i \zeta^i, \tag{3}$$

where $S_t^i \sim \mathrm{Pois}(\lambda^i(t, X_t))$ and $\lambda^i := \rho^i\tau$.

B. Variance Reduced Stochastic Simulation

As shown in previous work [13] we can reduce the variance of stochastic process Monte Carlo by constructing sample paths with the exact marginal distribution of our tau-leaping system, but that are jointly negatively correlated. This results in unbiased and consistent mean estimators that have lower variance than their iid counterparts, and hence less samples are necessary to achieve the same precision in estimates of the expected value. While there are several different approaches to anticorrelated stochastic simulation, we restrict our attention here to the study of antithetic simulation and its performance relative to traditional iid Monte Carlo. We construct antithetic paths as follows.

Recall the tau-leaping system (3) for approximate simulation of Markov jump processes

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

We may simulate a sample path of this process by drawing independent Poisson random variables at each time t and for each reaction channel i, updating the reaction rates at each time step. Consequently, we can produce an iid ensemble of such paths by repeating this simulation process with new independent Poisson samples. Alternatively, we could simultaneously produce two correlated sample paths (with each being an exact realization of the system in (3) with all internal Poisson random variables being independent in t and i), which, when jointly used in an ensemble of Monte Carlo sample paths, would produce a reduced-variance mean estimate.

To accomplish this, recall the definition of the quantile function, or inverse cumulative distribution function with parameter λ:

$$F_\lambda^{-1}(u) := \inf\{n \in \mathbb{Z}_+ : F_\lambda(n) \geq u\}. \tag{4}$$

Recall that the quantile function, when evaluated on a unit uniform random variate produces a Poisson random variable with mean λ. That is, if $U \sim \mathrm{Unif}(0,1)$, then $F_\lambda^{-1}(U) \sim \mathrm{Pois}(\lambda)$. To produce an antithetic pair of Poisson variables $\{S_1, S_2\}$ with parameters λ_1 and λ_2, respectively, define

$$\begin{aligned} U &\sim \mathrm{Unif}(0,1) \\ S_1 &:= F_{\lambda_1}^{-1}(U) \\ S_2 &:= F_{\lambda_2}^{-1}(1-U). \end{aligned}$$

Since $1-U \sim \mathrm{Unif}(0,1)$, $S_2 \sim \mathrm{Pois}(\lambda_2)$ and by a standard result [21], $\mathrm{Cov}(S_1, S_2) \leq 0$.

To produce an antithetic pair of sample paths $\{X_{1,t}, X_{2,t}\}_{t=0}^T$, every time a Poisson random variable $S_t^i(\lambda^i(t, X_t))$ would be simulated for the iid path, instead simulate an antithetic pair of Poisson random variables, $\{S_{1,t}^i(\lambda^i(t, X_{1,t})), S_{2,t}^i(\lambda^i(t, X_{2,t}))\}$ and use each of these samples as inputs to their respective path. Thus every Poisson sample used in a single path will be independent of all others used in that path, but each Poisson variable used in $X_{1,t}$ will be negatively correlated with a Poisson variable used in $X_{2,t}$. The resulting paths can be proven to be negatively correlated in some cases and have been numerically shown to dramatically decrease Monte Carlo error in several example systems [13]. This construction is summarized in Algorithm 1.

Algorithm 1 Antithetic τ-leaping

```
Initialize: $X_{j,0} \leftarrow x_{j,0}$
for $t = 0$ to $T$ do
    for $i = 1$ to $I$ do
        simulate antithetic Poisson pair:
            $\{S_{1,t}^i(\lambda^i(t, X_{1,t})), S_{2,t}^i(\lambda^i(t, X_{2,t}))\}$
    end for
    $X_{j,t+1} \leftarrow X_{j,t} + \sum_{i=1}^I S_{j,t}^i \zeta^i$, $j \in \{1, 2\}$
end for
```

III. Variance Reduced Stochastic MPC

Model predictive control (MPC) is a control strategy which seeks to approximate optimal, infinite time horizon feedback control via optimal solution of open loop, finite time horizon problems [14]. The control at time t takes in information about the current state and past control actions to simulate the cost of taking a given set of control actions over a finite time window $[t, t+H-1]$. From these simulations, an optimal control action over this time window can be found, and the *first* of these actions is implemented as the current control action. The state information is then updated, and a new optimal open loop solution is found for the next window $[t+1, t+H]$ and so forth.

In this context, we will focus on control of a perfectly observed Markov process on a countable state space where we attempt to minimize the cumulative sum of a cost function $g(x, u)$. Suppose we have a Markov decision process X described by

$$X_{t+1} = f(X_t, u_t) \tag{5}$$

where u_t is a particular control action at time t. Suppose we want to find an optimal policy $u_t = \mu(x_t) \in \mathcal{U}$ such that

$$\mu \in \underset{m \in \mathcal{F}(\mathbb{R}^n, \mathcal{U})}{\operatorname{argmin}} \; \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t g(X_t, m(X_t))\right], \tag{6}$$

where $\mathcal{U}$ is some admissible set of control actions, $\mathcal{F}(\mathbb{R}^n, \mathcal{U})$ is the set of measurable functions from $\mathbb{R}^n$ to $\mathcal{U}$, and $\beta \in (0, 1)$ is a discount factor to ensure boundedness of the sum. This problem is of course challenging for most Markov processes X, and often impossible to solve in closed form. We attempt, however to find an approximate realization of this policy along a particular trajectory by implementing MPC. Specifically, at time t, suppose that our controlled process $X_t = x_t$. We will obtain

$$u^{t,H} \in \underset{\tilde{u} \in \mathcal{U}_t^H}{\operatorname{argmin}} \, \mathbb{E}\left[\sum_{s=t}^{t+H-1} g(X_s, \tilde{u}_s) | X_t = x_t\right], \tag{7}$$

where $u^{t,H}$ is an H-vector of control actions over the finite horizon, and $\mathcal{U}_t^H$ is the admissible set of such sequences at time t, and we consider β close to 1. This optimization problem is over a much smaller space; even naive optimization strategies will suffice for small problems. We then set our current control action to be the first element of $u^{t,H} = (u_t^{t,H}, \ldots, u_{t+H-1}^{t,H})$:

$$u_t = \mu^{\mathrm{MPC}}(x_t) := u_t^{t,H}, \tag{8}$$

ignoring the rest of the finite horizon optimizer. Time can then be updated to $t+1$, and the control window shifted to $[t+1, t+H]$ to solve for $\mu^{\mathrm{MPC}}(x_{t+1})$. Note here that we never solve for an approximation of the actual optimal policy μ for every state in our countable state space. Instead we solve for an approximation $\mu^{\mathrm{MPC}}(x_t)$ of $\mu(x_t)$, i.e. the evaluation of μ at a particular point on our controlled trajectory. In other words, the algorithm approximately *implements* the optimal policy rather than solving for it in a closed form.

Regardless of the optimization routine used, some approximation of the expectation in (7) will be required in order to find a minimizing control action over the finite horizon. Given that our selected control action will depend on minimizing this expectation, errors in approximating it can result in selecting a less optimal policy, producing worse performance in the model predictive controller. Typically this is done via a Monte Carlo ensemble of a large number sample paths initialized at x_t where we sum the cost for each trajectory, and average these costs to accurately approximate the expectation. For complex, noisy or large systems, this repeated simulation can become very costly for accurate estimates, and often actual run-time requirements will impose strict constraints on the available number of Monte Carlo sample paths.

To mitigate this problem, we propose implementing anti-correlated stochastic simulation of the finite horizon window to produce accurate estimates of the expected cost of a control sequence while using fewer Monte Carlo sample paths than traditional iid Monte Carlo simulation. By simulating process paths using Algorithm 1, we may immediately improve estimates of the desired expectation, and as we will show in the next section, this results in improved expected cost incurred by the resulting MPC policy. Algorithm 2 summarizes this approach for available Monte Carlo resources of N sample paths.

Algorithm 2 Variance Reduced MPC at time t

```
input: $x_t$
for $\tilde{u} \in \mathcal{U}_t^H$ do
    for $k = 1$ to $N/2$ do
        simulate $\{X_{1,s}^k, X_{2,s}^k\}$ for
        $s \in [t, t+H-1]$, $X_{j,t}^k = x_t$, and $u = \tilde{u}$ using
        Algorithm 1
    end for
    compute sample mean:
    $C(\tilde{u}) \leftarrow \frac{1}{N}\sum_{k=1}^{N/2}\sum_{s=t}^{t+H-1}[g(X_{1,s}^k, \tilde{u}_s) + g(X_{2,s}^k, \tilde{u}_s)]$
end for
select $u^{t,H}$ that minimizes $C(\tilde{u})$
$\mu_t^{\mathrm{MPC}}(x_t) \leftarrow u_t^{t,H}$
```

IV. Numerical Results

Consider the following simple, nonlinear chemical reaction system:

$$\emptyset \overset{\rho_{\mathrm{A}}}{\rightarrow} \mathrm{A}$$
$$\mathrm{A} + \mathrm{A} \overset{\rho_{\mathrm{R}}}{\rightarrow} \mathrm{B}$$
$$\mathrm{B} \overset{\rho_{\mathrm{B}}}{\rightarrow} \emptyset$$

where the reaction rates ρ_{R} and ρ_{B} are given by mass action kinetics

$$\rho_{\mathrm{R}}(x) = \frac{1}{2}\kappa_{\mathrm{R}} x^{\mathrm{A}}(x^{\mathrm{A}} - 1)$$
$$\rho_{\mathrm{B}}(x) = \kappa_{\mathrm{B}} x^{\mathrm{B}},$$

and $\rho_{\mathrm{A}}(u) = \kappa_{\mathrm{A}} u$ is the control input. For simplicity, take $\mathcal{U} = \{u_{\mathrm{LO}} = 10\,\text{molecules/s}, u_{\mathrm{HI}} = 100\,\text{molecules/s}\}$ to be binary. Let the state $X_t = (X_t^{\mathrm{A}}, X_t^{\mathrm{B}})^\top$ denote the number of particles of each species at time t. Consider the τ-leaping simulation of this system

$$X_{t+1} = X_t + \sum_{i=1}^{I} S_t^i \zeta^i, \tag{9}$$

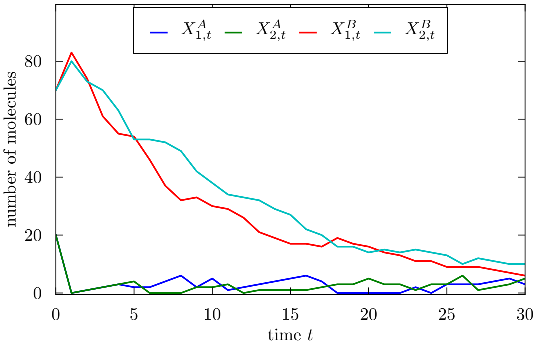

Fig. 1. Two anticorrelated sample paths of the chemical reaction system with a constant input of $u_{\mathrm{LO}} = 10$ molecules/s.

where $S_t^i \sim \mathrm{Pois}(\lambda^i(X_t, u_t))$ and

$$\begin{array}{llll}
\zeta^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \lambda^1(X_t, u_t) & = \rho_{\mathrm{A}}(u_t)\tau & \\
\zeta^2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix} & \lambda^2(X_t, u_t) & = \rho_{\mathrm{R}}(X_t)\tau & \\
\zeta^3 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} & \lambda^3(X_t, u_t) & = \rho_{\mathrm{B}}(X_t)\tau. & (10)
\end{array}$$

An antithetic pair of sample open loop trajectories are shown in Figure 1 for $u \equiv u_{\mathrm{LO}}$, $\kappa_{\mathrm{A}} = \kappa_{\mathrm{R}} = \kappa_{\mathrm{B}} = 0.1$ and $\tau = 1.0\,\mathrm{s}$.

We define the cost function so that closed-loop trajectories try to stabilize the number of molecules of species B:

$$g(x, u) = |x^{\mathrm{B}} - x_{\mathrm{ref}}| \tag{11}$$

where $x_{\mathrm{ref}} = 30$ molecules. Further, we take actions to be 5 second step functions, so that a decision is made every 5 steps of simulation time. We take the length of the finite time horizon $H = 15$ seconds, so that the optimization problem is over 3 actions and thus brute force search over the action space requires only $|\mathcal{U}^3| = 8$ checks. The exhaustive search clearly scales poorly as the size of the admissible control set or window length grow, but is used here for simplicity. Future work would include a more sophisticated optimization technique. An example closed-loop trajectory computed using either 2 iid sample paths or one antithetic pair of sample paths (i.e. N=2) is shown in Figure 2, and its corresponding action sequence is shown in Figure 3.

Because the closed loop trajectories, policies and costs are all stochastic, to compare the performance of iid and antithetic MPC we must take a large ensemble of closed loop realizations for each fixed value of N to compute the expected cost of each algorithm. Figure 4 plots a Monte Carlo estimate of this cost (along with error bars corresponding to the standard error of the mean, an approximation of a single standard deviation of the average cost) using 3.84e3 samples, versus the number of Monte Carlo sample paths to which the model predictive controller has access. While these average cost estimates are somewhat noisy due to high variance in cost incurred by a closed loop trajectory, we can see marked improvement in the antithetic MPC, achieving roughly the same cost using only 2 Monte Carlo samples as the iid MPC achieves using 4 samples. Note that since both the iid and antithetic estimates of the expectation in (7) are consistent [12], the expected cost incurred by each should converge as the ensemble resources N become large.

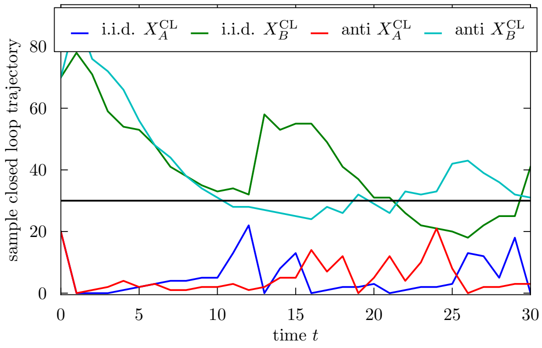

Fig. 2. Two closed loop sample paths of the chemical reaction system with access to only 2 sample paths to estimate the expected value in (7). To estimate the expected cost of a candidate control sequence while running MPC, iid MPC uses two iid sample paths and the antithetic MPC uses two antithetically paired sample paths.

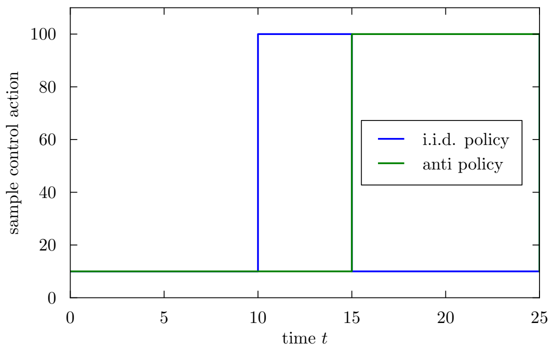

Fig. 3. The implemented policies used by the closed loop paths in Fig. 2. To estimate the expected cost of a candidate control sequence while running MPC, iid MPC uses two iid sample paths and the antithetic MPC uses two antithetically paired sample paths.

V. Conclusions

In this paper, we proposed a new algorithm for model predictive control of Markov jump processes using variance-reduced trajectory sampling. We showed how this anticorrelated stochastic simulation algorithm could be useful to reduce the Monte Carlo budget of estimating the expected cost of a candidate control sequence. Further, we demonstrated a factor of 2 reduction in the number of Monte Carlo paths necessary to achieve the same closed loop cost in the resource-poor limit for a simple, non-linear, non-Gaussian chemical reaction system, though only modest improvement in the average cost of the closed loop controller for fixed Monte Carlo resources. More computation is necessary to numerically explore if the antithetic MPC scheme will outperform iid simulation for a greater range of online ensemble

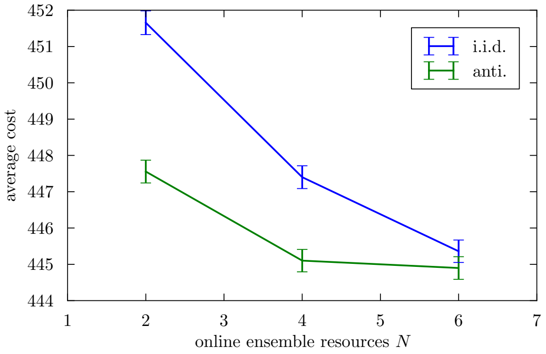

Fig. 4. The estimated expected closed loop cost incurred by iid MPC and antithetic MPC versus the number of Monte Carlo samples to which they have access for online estimation of expected cost. Average costs are computed using 38,400 sample closed loop paths. Note the antithetic technique requires approximately half the ensemble resources to achieve the same average cost. The error bars show +/- standard error of the mean, which is approximately one standard deviation of the sample average cost.

resources N, and analytical study of their properties will be sought to prove this improvement result for all N. In future work, we hope to characterize the properties of the system that govern how much improvement antithetic simulation will provide, and we also hope to incorporate probabilistic constraints.

Acknowledgment

The authors acknowledge AFOSR support under grant FA9550-15-1-0059.

References

[1] David F. Anderson. Incorporating postleap checks in tau-leaping. *The Journal of Chemical Physics*, 128(5):054103, 2008.

[2] D. Bernardini and A. Bemporad. Scenario-based model predictive control of stochastic constrained linear systems. *Joint 48th IEEE CDC and 28th IEEE CCC, Shanghai, P.R. China*, 2009.

[3] 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*, 2012.

[4] G. C. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51:742–753, 2006.

[5] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient stepsize selection for the tau-leaping simulation method. *Journal of Chemical Physics*, 124:044109, 2006.

[6] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[7] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[8] N. Kantas, J. M. Maciejowski, and A. Lecchini-Visintini. Sequential Monte Carlo for model predictive control. In L. Magni, D. M. Raimondo, and F. Allgöwer, editors, *Nonlinear model predictive control*, pages 263–273. Springer, Berlin, 2009.

[9] B. Kouvaritakis and M. Cannon. Stochastic model predictive control. In T. Samad and J. Baillieul, editors, *Encyclopedia of systems and control*. Springer, 2014.

[10] T. G. Kurtz. Approximation of population processes. *CBMS-NSF Regional Conf. Ser. in Appl. Math*, 1981.

[11] A. Lecchini-Visintini, W. Glover, J. Lygeros, and J. M. Maciejowski. Monte Carlo optimization for conflict resolution in air traffic control. *IEEE Transactions on Intelligent Transportation Systems*, 7(4):470–482, 2006.

[12] P. A. Maginnis. Variance reduction for Poisson and Markov jump processes. Master's thesis, University of Illinois at Urbana-Champaign, 2011.

[13] P. A. Maginnis, M. West, and G. E. Dullerud. Anticorrelated discrete-time stochastic simulation. *IEEE Conference on Decision and Control*, 2013.

[14] D. Q. Mayne. Model predictive control: recent developments and future promise. *Automatica*, 50:2967–2986, 2014.

[15] R. R. Negenbord, B. De Schutter, M. A. Wiering, and H. Hellendoorn. Learning-based model predictive control for markov decision processes. *Proceedings of the 16th IFAC World Congress, Prague, Czech Republic*, 2005.

[16] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *Journal of Chemical Physics*, 119:12784, 2003.

[17] 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*, 4(3):867–895, 2005.

[18] M. Rathinam, P. Sheppard, and M. Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *Journal of Chemical Physics*, 132(3):034103, 2010.

[19] 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*, 114, 2009.

[20] M. Vidyasagar. Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, 37:1515–1528, 2001.

[21] W. Whitt. Bivariate distributions with given marginals. *Annals of Statistics*, 4(6):1280–1289, 1976.