51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Application of Variance Reduction Techniques for Tau-Leaping Systems to Particle Filters

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

Abstract— **We propose several techniques for simulation of the prediction step of particle filters for application to a class of estimation problems. These techniques are based on variance reduction techniques previously proposed for stochastic simulation of tau-leaping processes. Proof of almost sure weak convergence is given for an approach in one dimension, and numerical examples pertaining to building population dynamics are presented in both single and multidimensional tau-leaping cases.**

I. INTRODUCTION

The problem of online estimation of unknown stochastic processes given imperfect observation processes is broadly applicable in most areas of decision and control. Unfortunately, very few special cases like Kalman's linear/Gaussian model [1] allow us to solve this problem directly and exactly. Thus approximate methods to treat these problems were introduced, including particle filtering/sequential Monte Carlo in the 1960s and 70s [10], [12]. While progress in this area was relatively slow in the intervening decades, the 1990s saw an explosion of improvements in this area, as rapidly increasing computing power greatly increased the feasibility of particle methods. An especially big jump was the introduction of the bootstrap filter [9], though it was quickly joined by its own set of improvements, for example work done on the resampling step [14], [15]. There are several excellent works surveying the myriad techniques and improvements [11] as well as convergence results for practitioners [2]. As with all Monte Carlo methods, the drive is ever toward reduced variance. Lower variance in unbiased filters results in lower expected errors per computing cost and leaner estimation systems that are implementable in a rapidly growing class of applications.

One potentially promising class of stochastic systems are tau-leaping systems. Introduced by Gillespie [4] early last decade, tau-leaping is an time-discretizing, approximate sampling method for stochastic systems whose transitions can be classified by a finite number of event channels. Convergence of tau-leaping systems is already quite well understood [5], and several variants of its implementation exist, for example adaptive tau-leaping [6], [8] and implicit tau-leaping [7]. Work has already begun on alternative methods of variance reduction in tau-leaping systems [3] based on stratified and antithetic methods of variance reduction, and here an extension is proposed to the filtering problem applied to tau-leaping systems.

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 introduce necessary preliminary notation on optimal and particle filtering, as well as some background on anticorrelated variance reduction techniques. In section III we construct the anticorrelated particle filter in the one dimensional case and prove almost sure weak convergence of this filter to the optimal Bayesian filter. We also provide numerical support in the form of a room dynamics model. In section IV we outline the extension of the algorithms to the multidimensional setting in the special case of tau-leaping systems. In particular, we present a model for partially observed building dynamics. Finally, section V contains conclusions.

II. PRELIMINARIES

A. Optimal Filtering

We proceed by introducing notation used in Crisan and Doucet [2]. Suppose $X = \{X_t, t \in \mathbb{N}\}$ is a stochastic signal process in $\mathbb{R}^{n_x}$ and $Y = \{Y_t, t \in \mathbb{N} \setminus \{0\}\}$ is a corresponding observation process in $\mathbb{R}^{n_y}$. Let the signal process X be Markov with initial distribution $X_0 \sim \mu(\mathrm{d}x_0)$ and probability transition kernel $K(\mathrm{d}x_t|x_{t-1})$ so that

$$\mathbb{P}(X_t \in A|X_{t-1} = x_{t-1}) = \int_A K(\mathrm{d}x_t|x_{t-1}),\ A \in \mathcal{B}(\mathbb{R}^{n_x})$$
$$\mathbb{P}(Y_t \in B|X_t = x_t) = \int_B g(\mathrm{d}y_t|x_t), \quad B \in \mathcal{B}(\mathbb{R}^{n_y}),$$

where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel σ-algebra on $\mathbb{R}^n$. Consider a probability measure $\pi_{s|t}$ the solution of the optimal filtering problem that denotes the law of X_s conditioned on $Y_1, \ldots, Y_t$. We may obtain $\pi_{t|t}$ via a standard, two-step recursive version of Bayes' Theorem, given by:

Prediction:

$$\pi_{t|t-1}(\mathrm{d}x_t) = \int_{\mathbb{R}^{n_x}} \pi_{t-1|t-1}(\mathrm{d}x_{t-1})K(\mathrm{d}x_t|x_{t-1})$$

Update:

$$\pi_{t|t}(\mathrm{d}x_t) = g(y_t|x_t)\pi_{t|t-1}(\mathrm{d}x_t)\left[\int_{\mathbb{R}^{n_x}} g(y_t|x_t)\pi_{t|t-1}(\mathrm{d}x_t)\right]^{-1}.$$

For φ a function, ν a measure, and Ξ a Markov transition

978-1-4673-2066-5/12/$31.00 ©2012 IEEE

kernel, define the standard notation:

$$(\nu, \varphi) := \int \varphi(x)\nu(\mathrm{d}x)$$
$$\nu\Xi(A) := \int \nu(\mathrm{d}x)\Xi(A|x)$$
$$\Xi\varphi(x) := \int \Xi(\mathrm{d}z|x)\varphi(z).$$

Thus we may more compactly characterize the optimal filtering recursion by:

$$(\pi_{t|t-1}, \varphi) = (\pi_{t-1|t-1}, K\varphi) \qquad \text{Prediction}$$
$$(\pi_{t|t}, \varphi) = (\pi_{t|t-1}, \varphi g)(\pi_{t|t-1}, g)^{-1} \qquad \text{Update}$$

where φ is any continuous, bounded, real-valued function on $\mathbb{R}^{n_x}$.

B. Particle Filtering

A particle filtering method approximates the optimal filter by maintaining a collection $\{x_t^{(i)}\}_{i=1}^N$ of N samples of the state, known as *particles*, indexed by time. The empirical measure $\pi_{t|t}^N$ of this collection, given by:

$$\pi_{t|t}^N(\mathrm{d}x_t) := \frac{1}{N}\sum_{i=1}^N \delta_{x_t^{(i)}}(\mathrm{d}x_t), \tag{1}$$

is meant to approximate $\pi_{t|t}$, the optimal filtering measure. Note here that δ_x denotes the Dirac delta measure at x. Given this object at time $t-1$, we may draw samples that are approximately distributed according to $\pi_{t|t-1}$ by drawing

$$\tilde{x}^{(i)} \overset{\text{i.i.d.}}{\sim} \pi_{t-1|t-1}^N K(\mathrm{d}x_t) = \frac{1}{N}\sum_{k=1}^N K(\mathrm{d}x_t|x_{t-1}^{(k)}).$$

The update step is performed by computing normalized weights for each particles given the information $Y_t = y_t$, and a resampling is performed to close the loop. One particle filtering algorithm is given by Crisan and Doucet [2]:

At time $t = 0$.
Step 0: *Initialization*
For $i = 1, \ldots, N$, sample $x_0^{(i)} \overset{\text{i.i.d.}}{\sim} \pi_{0|0}$ and set $t = 1$.
At time $t \geq 1$.
Step 1: *Importance Sampling*
For $i = 1, \ldots, N$, sample $\tilde{x}_t^{(i)} \overset{\text{i.i.d.}}{\sim} \pi_{t-1|t-1}^N K$.
For $i = 1, \ldots, N$, calculate the normalized importance weights $w_t^{(i)}$:
$w_t^{(i)} \propto g(y_t|\tilde{x}_t^{(i)})$ such that $\sum_{i=1}^N w_t^{(i)} = 1$.
Step 2: *Resampling*
For $i = 1, \ldots, N$, sample $x_t^{(i)} \overset{\text{i.i.d.}}{\sim} \tilde{\pi}_{t|t}^N := (1/N)\sum_{i=1}^N w_t^{(i)} \delta_{\tilde{x}_t^{(i)}}$

C. Anticorrelated Sampling

Several techniques exist to expedite the convergence of Monte Carlo type methods. Of these, several examples of the larger class of anticorrelated sampling techniques are introduced in [3]. These techniques specifically focus on introducing localized, negative correlation into the simulation of otherwise independent samples in order to reduce variance of Monte Carlo estimates. The three results espoused were antithetic and stratified sampling, as well as a hybridization of the two. They are defined in terms of the simulation of random variables via inversion of their CDFs. In particular, one may simulate a sequence of i.i.d. random variables X_k with distribution function F and law ν by sampling:

$$U_k \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0,1) \tag{2}$$
$$X_k := F^{-1}(U_k), \tag{3}$$

where we formally invert the CDF by defining:

$$F^{-1}(u) := \inf\{x : F(x) \geq u\}. \tag{4}$$

An empirical estimate of ν is a random measure that approximates ν. It can be computed via:

$$\nu^{N,\omega} = \frac{1}{N}\sum_{k=1}^N \delta_{X_k(\omega)} \approx \nu, \tag{5}$$

for N sufficiently large. For convenience we may suppress the explicit dependence on the outcome ω, but all empirical measures constructed in this way are random. We may attempt to accelerate this convergence by introducing localized correlation into the samples, say antithetically, by:

$$U_k^{\mathrm{A}} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0,1) \tag{6}$$
$$X_{k,1}^{\mathrm{A}} := F^{-1}(U_k^{\mathrm{A}}) \tag{7}$$
$$X_{k,2}^{\mathrm{A}} := F^{-1}(1 - U_k^{\mathrm{A}}) \tag{8}$$
$$\nu^{\mathrm{A},N} = \frac{1}{N}\sum_{k=1}^{N/2}\Big(\delta_{X_{k,1}^{\mathrm{A}}} + \delta_{X_{k,2}^{\mathrm{A}}}\Big), \tag{9}$$

or via uniform stratification of the random variate sampled in $[0,1)$:

$$A_j := \left[\tfrac{j-1}{M}, \tfrac{j}{M}\right) \text{ for } j = 1, \ldots, M \tag{10}$$
$$U_{k,j}^{\mathrm{S}} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(A_j) \text{ for } j = 1, \ldots, M \tag{11}$$
$$X_{k,j}^{\mathrm{S}} := F^{-1}(\tilde{U}_{k,j}^{\mathrm{S}}) \text{ for } j = 1, \ldots, M \tag{12}$$
$$\tilde{U}_{k,j}^{\mathrm{S}} := (\Pi^M \vec{U}_k^{\mathrm{S}})_j, \tag{13}$$

where Π^M is a random $M \times M$ permutation matrix. The empirical measure constructed from these samples is given by:

$$\nu_M^{\mathrm{S},Mn} := \frac{1}{Mn}\sum_{k=1}^n \sum_{j=1}^M \delta_{X_{k,j}^{\mathrm{S}}}. \tag{14}$$

A hybridization of these techniques is easily constructed by

$$U^{\mathrm{H}}_{k,j} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(A_j) \text{ for } j = 1, \ldots, \tfrac{M}{2} \tag{15}$$

$$U^{\mathrm{H}}_{k,j} := 1 - U^{\mathrm{H}}_{k,M-j+1} \text{ for } j = (\tfrac{M}{2}+1), \ldots, M \tag{16}$$

$$X^{\mathrm{H}}_{k,j} := F^{-1}(\tilde{U}^{\mathrm{H}}_{k,j}) \text{ for } j = 1, \ldots, M \tag{17}$$

$$\text{where } \tilde{U}^{\mathrm{H}}_{k,j} := (\Pi^M \vec{U}^{\mathrm{H}}_k)_j \tag{18}$$

$$\nu^{\mathrm{H},Mn}_M := \frac{1}{Mn} \sum_{k=1}^{n} \sum_{j=1}^{M} \delta_{X^{\mathrm{H}}_{k,j}}, \tag{19}$$

for Π^M a random permutation matrix. In order to make precise the concept of convergence of measures, we introduce (weak) convergence of a sequence of measures $\{\mu_N\}_{N=1}^{\infty}$ to another measure μ if, for any $\varphi \in C_b(\mathbb{R}^{n_x})$ the space of continuous, bounded functions:

$$\lim_{N\to\infty} (\mu_N, \varphi) = (\mu, \varphi). \tag{20}$$

Further, one can choose a countable set $\mathcal{A} = \{\varphi_i : i \in \mathbb{N}\}$ such that the above condition holds for every $\varphi \in \mathcal{A}$ if and only if μ_N converges weakly to μ as $N \to \infty$. We may also define a metric d on the set $\mathcal{P}(\mathbb{R}^{n_x})$ of probability measures on $\mathbb{R}^{n_x}$ which induces this weak topology, given by:

$$d(\mu, \nu) = \sum_{i=1}^{\infty} \frac{|(\mu, \varphi_i) - (\nu, \varphi_i)|}{2^i \|\varphi_i\|}, \tag{21}$$

where $\|\varphi\| := \sup_{x\in\mathbb{R}^{n_x}} |\varphi(x)|$, the supremum norm on $C_b(\mathbb{R}^{n_x})$. That is, $\mu_N \to \mu$ weakly as $N \to \infty$ if and only if $\lim_{N\to\infty} d(\mu_N, \mu) = 0$ (see Crisan and Doucet [2]).

III. ANTICORRELATED PARTICLE FILTER IN 1-D

We may express a version of the particle filter in terms of a composition of mappings from $\mathcal{P}(\mathbb{R})$ to itself. First, we construct the two continuous maps used in the optimal filter. Define $b_t : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ to be

$$b_t(\nu)(\mathrm{d}x_t) := \nu K(\mathrm{d}x_t) = \int_{\mathbb{R}} K(\mathrm{d}x_t | x_{t-1}) \nu(\mathrm{d}x_{t-1}), \tag{22}$$

for any $\nu \in \mathcal{P}(\mathbb{R})$. It is shown in [2] that this map is continuous in the sense of weak topology if the Markov transition kernel K is Feller. Observe that for any $\varphi \in \mathcal{C}_b(\mathbb{R})$,

$$(b_t(\nu), \varphi) = (\nu, K\varphi), \tag{23}$$

and also that

$$\pi_{t|t-1} = b_t(\pi_{t-1|t-1}). \tag{24}$$

Similarly, we may define another map on measures, a_t, by its pairing with an arbitrary continuous, bounded function φ as:

$$(a_t(\nu), \varphi) = (\nu, g)^{-1} (\nu, \varphi g), \tag{25}$$

where a sufficient condition for the continuity of a_t is that $g(y_t|\cdot)$ be bounded, continuous and strictly positive. Here, it is also clear that

$$\pi_{t|t} = a_t(\pi_{t|t-1}) = a_t(b_t(\pi_{t-1|t-1})). \tag{26}$$

We may also further define the maps k_t and $k_{1:t}$ to be

$$k_t := a_t \circ b_t \tag{27}$$

$$k_{1:t} := k_t \circ k_{t-1} \circ \cdots \circ k_1, \tag{28}$$

which are continuous if a_t and b_t are for every t. Note then that for initial distribution μ, we may express the optimal filter as

$$\pi_{t|t} = k_t(\pi_{t-1|t-1}) = k_{1:t}(\mu). \tag{29}$$

Now, in order to express a class of particle filters in these terms, define the random mapping $c^{N,\omega}$ from a measure ν to its empirical measure by:

$$c^{N,\omega}(\nu)(\mathrm{d}x_t) = \frac{1}{N} \sum_{j=1}^{N} \delta_{V_j(\omega)}(\mathrm{d}x_t), \tag{30}$$

where $V_j \sim \nu$ are i.i.d.. It is clear from the strong law of large numbers that $c^{N,\omega}(\nu) \to \nu$ weakly as $N \to \infty$ for almost every ω by pairing with an arbitrary $\varphi \in \mathcal{C}_b(\mathbb{R})$. In terms of the above notation, $\nu^{N,\omega} = c^{N,\omega}(\nu)$. The above particle filtering algorithm can thus be expressed as:

$$\begin{aligned} &\text{Prediction:} && \pi^N_{t|t-1} = c^N \circ b_t(\pi^N_{t-1|t-1}) \\ &\text{Update and Resampling:} && \pi^N_{t|t} = c^N \circ a_t(\pi^N_{t|t-1}), \end{aligned}$$

initialized at $\mu^N := c^N(\mu)$. Similarly, define maps k^N_t and $k^N_{1:t}$ to be

$$k^N_t := c^N \circ a_t \circ c^N \circ b_t \tag{31}$$

$$k^N_{1:t} := k^N_t \circ k^N_{t-1} \circ \cdots \circ k^N_1, \tag{32}$$

so that

$$\pi^N_{t|t} = k^N_t(\pi^N_{t-1|t-1}) = k^N_{1:t}(\mu^N). \tag{33}$$

It was shown in [2] that this version of the particle filter converges in the weak sense to the optimal filter (almost surely in the random map $c^{N,\omega}$). We now propose an extension of this particle filter to methods which use negative correlation in the prediction step. Instead of sampling the distributions in an i.i.d. way, sample them according to the above algorithms. For example, construct the antithetic empirical map

$$c^{\mathrm{A},N}(\nu)(\mathrm{d}x_t) := \nu^{\mathrm{A,N}} = \frac{1}{N} \sum_{k=1}^{N/2} \Big(\delta_{V^{\mathrm{A}}_{k,1}} + \delta_{V^{\mathrm{A}}_{k,2}} \Big), \tag{34}$$

where the samples $(V^{\mathrm{A}}_{k,1}, V^{\mathrm{A}}_{k,2})$, $k = 1, \ldots, N/2$ used are i.i.d. in k and $V^{\mathrm{A}}_{k,1}$ and $V^{\mathrm{A}}_{k,2}$ are pairwise antithetically sampled as above. We may then define the antithetic particle filter $\pi^{\mathrm{A},N}_{t|t}$

$$\begin{aligned} &\text{Initialization:} && \pi^{\mathrm{A},N}_{0|0} = \mu^{\mathrm{A},N} := c^{\mathrm{A},N}(\mu) \\ &\text{Prediction:} && \pi^{\mathrm{A},N}_{t|t-1} = c^{\mathrm{A},N} \circ b_t(\pi^{\mathrm{A},N}_{t-1|t-1}) \\ &\text{Update and Resampling:} && \pi^{\mathrm{A},N}_{t|t} = c^N \circ a_t(\pi^{\mathrm{A},N}_{t|t-1}). \end{aligned}$$

Similarly, we may define stratified and hybrid particle filters using M uniform strata of $[0, 1)$ in terms of the empirical map

$$c^{\alpha,N}(\nu)(\mathrm{d}x_t) := \nu^{\alpha,N}_M = \frac{1}{N} \sum_{k=1}^{N/M} \sum_{j=1}^{M} \delta_{V^{\alpha}_{k,j}}, \tag{35}$$

for $\alpha \in \{\mathrm{S}, \mathrm{H}\}$, that is, where $V_{k,j}^{\mathrm{S}} \sim \nu$ and $V_{k,j}^{\mathrm{H}} \sim \nu$ are sampled via stratification and the hybrid technique, respectively. We construct the stratified and hybrid particle filters exactly as above with A replaced with S and H, respectively. We now show convergence of these techniques in the weak sense almost surely, and we expect these techniques to be MSE non-increasing by results proven in [3].

A. Almost Sure Convergence

We proceed by proving a strong property of the convergence of $c^{\alpha,N}$ to identity, and then the desired result follows immediately via a Lemma in [2].

Lemma 1: For $c^{\alpha,N,\cdot}$ as above and any sequence of measures $\{\nu_N\}_{N=1}^{\infty}$ such that $\nu_N \to \nu$ as $N \to \infty$, then $c^{\alpha,N,\cdot}(\nu_N) \to \nu$ almost surely for $\alpha \in \{\mathrm{A}, \mathrm{S}, \mathrm{H}\}$

Proof: Suppose $\{\nu_N\}_{N=1}^{\infty}, \nu \in \mathcal{P}(\mathbb{R})$ are any such measures. For any $\varphi_i \in \mathcal{A}$, any empirical map $c^{\alpha,N}, \alpha \in \{\mathrm{A}, \mathrm{S}, \mathrm{H}\}$, and any number M of correlated variables per i.i.d. sample (i.e. $M = 2$ for antithetic sampling, and M is the number of strata for stratified or hybrid),

$$\mathbb{E}\left[\left((c^{\alpha,N}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4\right] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{k=1}^{N/M}\sum_{j=1}^{M}[\varphi_i(V_{k,j}^{\alpha}) - (\nu_N, \varphi_i)]\right)^4\right]. \quad (36)$$

For compactness, define $a_{k,j}^i := \varphi_i(V_{k,j}^{\alpha}) - (\nu_N, \varphi_i)$ and observe that $a_{k,j}^i,\ k \in \{1, \dots, n/M\}, j \in \{1, \dots, M\}$ are random, independent in k, correlated in j and $\mathbb{E}[a_{k,j}^i] = 0$ since $V_{k,j}^{\alpha} \sim \nu_N$. Then we have, using these facts

$$\mathbb{E}\left[\left((c^{\alpha,N}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4\right] \quad (37)$$

$$= \mathbb{E}\left[\left(\frac{1}{N}\sum_{k=1}^{N/M}\sum_{j=1}^{M}a_{k,j}^i\right)^4\right] \quad (38)$$

$$= \frac{1}{N^4}\mathbb{E}\left[\sum_{k=1}^{N/M}\left(\sum_{j=1}^{M}a_{k,j}^i\right)^4\right] + \frac{6}{N^4}\mathbb{E}\left[\sum_{\substack{k_1=1\\k_2=k_1+1}}^{N/M}\left(\sum_{j_1=1}^{M}a_{k,j}^i\right)^2\left(\sum_{j_2=1}^{M}a_{k,j}^i\right)^2\right] \quad (39)$$

$$= 2^4\|\varphi_i\|^4\frac{M^3 + 3NM^2 - 3M^3}{N^3} \quad (40)$$

$$\leq \frac{48M^2\|\varphi_i\|^4}{N^2}, \quad (41)$$

for sufficiently large N. Thus

$$\mathbb{E}\left[\sum_{N=1}^{\infty}\left((c^{\alpha,N}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4\right] \leq 48M^2\|\varphi_i\|^4\sum_{N=1}^{\infty}\frac{1}{N^2} < \infty. \quad (42)$$

Thus, with probability 1

$$\sum_{N=1}^{\infty}\left((c^{\alpha,N}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4 < \infty \quad (43)$$

$$\implies \lim_{N\to\infty}|(c^{\alpha,N}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)| = 0 \quad (44)$$

for any $i \in \mathbb{N}$. Thus $\lim_{N\to\infty} d(c^{\alpha,N}(\nu_N), \nu_n) = 0$, almost surely, and by the triangle inequality

$$\lim_{N\to\infty} d(c^{N,\omega}(\nu_N), \nu) \leq \lim_{N\to\infty} d(c^{N,\omega}(\nu_N), \nu_N) + \lim_{N\to\infty} d(\nu_N, \nu) = 0.$$

■

Theorem 2: For transition kernel K Feller and likelihood function g bounded, continuous and strictly positive, then $\lim_{N\to\infty}\pi_{t|t}^{\alpha,N} = \pi_{t|t}$ almost surely for $\alpha \in \{\mathrm{A}, \mathrm{S}, \mathrm{H}\}$

Proof: As was proven in [2], for a_t, b_t continuous, $c^N, c^{\alpha,N}$ endowed with the property proven in Lemma 1, for $\lim_{N\to\infty}\mu^{\alpha,N} = \lim_{N\to\infty} c^{\alpha,N}(\mu) = \mu$,

$$\lim_{N\to\infty}\pi_{t|t}^{\alpha,N} = \lim_{N\to\infty} k_{1:t}^{\alpha,N}(\mu^{\alpha,N}) = k_{1:t}(\mu) = \pi_{t|t}. \quad (45)$$

■

B. Room Population Dynamics

As a concrete illustration of the proof above, consider X_t to be the population of a single, initially empty room into which people enter at exponentially spaced times, and their arrivals are independent of each other. This can be modeled as a Poisson process with, say, unit rate. Suppose that, at every time t, an observer attempts to count the number of occupants but has some probability of over or undercounting. Let this count be denoted Y_t and suppose that we can model the accuracy of the observer by

$$Y_t = X_t + N_t \quad (46)$$

where N_t is an i.i.d. noise process with distribution with pmf

$$\mathbb{P}(N_t = z) = \begin{cases}\frac{\kappa}{z^4} & \text{if } z \in \mathbb{Z}\setminus\{0\}\\ \kappa & \text{if } z = 0,\end{cases} \quad (47)$$

where $\kappa = (\frac{\pi^4}{45} + 1)^{-1}$. We will henceforward refer to this distribution as a quartic power law.

To accurately guess the state X_t, given all of the past measurements and knowledge of the system, we must solve the optimal Bayesian filtering problem for the measure $\pi_{t|t}$. However, this is in general a difficult quantity to calculate, so we seek some computationally efficient method of approximation, the particle filtering distribution $\pi^N t|t$.

We initialize a population of particles $\{X_t^{(i)}\}_{i=1}^N$ at the point $X_0^{(i)} = 0$. Now, given a particle population at time $t-1$, we use our knowledge of the process X_t to predict the next transition. This problem is simple enough that we can efficiently determine the distribution function of the measure $\pi_{t|t}^N K \in \mathcal{P}(\mathbb{R})$, which is just a set of points which advance

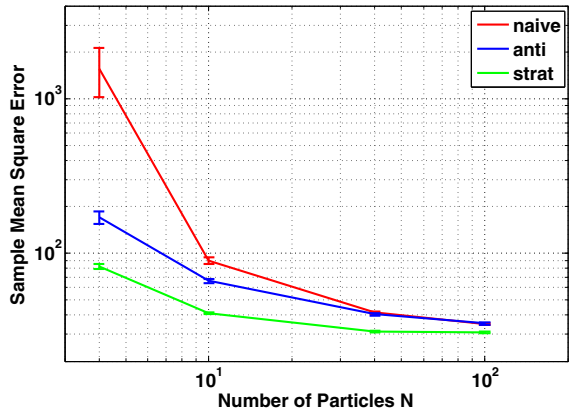

Fig. 1. Plot of sample mean square error of the mean of the empirical particle distribution from an ensemble of random sample paths of the Poisson process. Error bars shown are standard error of the mean. Here, ensemble size is 100. Note the apparent convergence to the Bayesian limit.

with independent Poisson distributions. In this case, it is given by

$$F(z) = \mathbb{P}(\tilde{X}_t^{(i)} \leq z) = \frac{1}{N}\sum_{k=1}^{N} F_\lambda(\max\{z - X_{t-1}^{(k)}, 0\}), \tag{48}$$

where $F_\lambda(z)$ is the CDF of a Poisson distribution with rate λ. Given access to a CDF, implementation of the antithetic or stratified filters is trivial, so we may correlate our predictions as above. Finally, we compute likelihood weights for each of our samples, and resample to get a new population.

C. Numerical Results

Consider a single sample of this process, comprised of a sample trajectory X_t^e, a sample measurement process Y_t^e, and a particle filter of N particles, with corresponding mean estimator

$$\hat{X}_t^{e,N} = \frac{1}{N}\sum_{j=1}^{N} X_t^{e,(j)}. \tag{49}$$

If, for every e we compute the mean square error (MSE) $|X^e - \hat{X}^e|^2$ and average over an ensemble $\{e\}_{e=1}^{100}$, we will approximate the expected MSE of the particle filter estimator, $\mathbb{E}[|X - (\pi^N, x)|^2]$. Further, as the number of particles N becomes sufficiently large, this quantity should approach the optimal Bayesian limit $\mathbb{E}[|X - (\pi, x)|^2] = \mathbb{E}[|X - \mathbb{E}[X|Y]|^2] > 0$ for a non-fully observed system. Fig. 1 collects the results of these numerical experiments. Note that, relative to the Bayesian limit, the anticorrelated samplers can produce more than an order of magnitude reduction MSE of the naive sampler.

IV. Multidimensional Anticorrelated Particle Filtering

One way to extend the variance reduced algorithms presented above to a multi-dimensional setting is via the tau-leaping method of Gillespie [4] for simulating Markov processes with a finite number of event channels. Consider the random time-change representation of a Markov process $X(t) \in \mathbb{R}^D$, $t \in [0, T]$, with I event channels, each with propensity function $\rho^i(t, X(t))$, defined by

$$X(t) = X(0) + \sum_{i=1}^{I} \Upsilon^i\left(\int_0^t \rho^i(s, X(s))\,\mathrm{d}s\right)\zeta^i, \tag{50}$$

where Υ^i is a unit-rate Poisson process and $\zeta^i \in \mathbb{R}^D$ are the state jump vectors, so that $\zeta^i = X(t^+) - X(t^-)$, if the ith event channel experiences a transition at time t. The evolution of such a process can be approximated in discrete time using the tau-leaping method. For time-step increment τ, let $t_\ell = \ell\tau$ and $\widetilde{X}_\ell \approx X(t_\ell)$ for $\ell \in \{0, \ldots, L\}$, where $L := \max\{\ell : t_\ell \leq T\}$. Then $\widetilde{X}_\ell$ evolves via

$$\widetilde{X}_{\ell+1} = \widetilde{X}_\ell + \sum_{i=1}^{I} S_\ell^i\left(\rho^i\big(t_\ell, \widetilde{X}_\ell\big)\tau\right)\zeta^i, \tag{51}$$

where $S_\ell^i(\lambda) \sim \mathrm{Pois}(\lambda)$. For compactness, define $\lambda_\ell^i = \rho^i\big(t_\ell, \widetilde{X}_\ell\big)\tau$ and denote $S_\ell^i\left(\lambda_\ell^i\right)$ by S_ℓ^i. Thus (51) becomes

$$\widetilde{X}_{\ell+1} = \widetilde{X}_\ell + \sum_{i=1}^{I} S_\ell^i\zeta^i. \tag{52}$$

If we couple this discrete time stochastic process with an observation process Y_t, say

$$\widetilde{X}_{t+1} = \widetilde{X}_t + \sum_{i=1}^{I} S_t^i\zeta^i \tag{53}$$

$$Y_t = f(X_t, N_t), \tag{54}$$

where N_t is some independent noise process, and $X_0 \sim \mu$, then we have produced a whole class of mulitdimensional filtering problems for finding $\pi_{t|t}$, the law of X_t given the observations $Y_1, Y_2, \ldots, Y_t$. The particle filtering approximation to the solution of this problem can be obtained exactly as above, though often practitioners will implement a slightly modified algorithm known as the bootstrap particle filter [9]. The primary difference between a bootstrap filter and the particle filter presented above is that, in the prediction step, instead of drawing

$$\tilde{x}_t^{(j)} \overset{\text{i.i.d.}}{\sim} \pi_{t-1|t-1}^N K(\mathrm{d}x_t) = \frac{1}{N}\sum_{k=1}^{N} K(\mathrm{d}x_t|x_{t-1}^{(k)}).$$

we merely simulate the particle $x_t^{(j)}$ moving forward in time in isolation, that is

$$\tilde{x}_t^{(j)} \sim K(\mathrm{d}x_{t-1}|x_{t-1}^{(j)}).$$

So in this case, given a particle population $\{x_t^{(j)}\}_{j=1}^N$ at time t, to perform the prediction step, for each particle $x_t^{(j)}$ we sample S_t^i for each $i \in \{1, \ldots, I\}$ and compute

$$\tilde{x}_{t+1}^{(j)} = x_t^{(j)} + \sum_{i=1}^{I} S_t^i\zeta^i, \tag{55}$$

and the rest of the algorithm follows as before. We can see now that the special structure of this model alows us

to reduce the stochastic simulation of a multidimensional process to the simulation of I random, real-valued variables, corresponding with each reaction process. Thus the technique we propose for implementing anticorrelated particle filtering is the application of the one dimensional techniques to each of these random variables. Since the techniques as constructed produced fully realized (yet correlated) samples from their respective distributions, transitions simulated this way are fair sample transitions of the process. To proceed we first discuss the application of the stratified technique for M strata A_j constructed as before. For each $j \in \{1, \ldots, M\}$ and for any $i \in \{1, \ldots, I\}$ and any time step t, take $U_{j,t}^{\mathrm{S},i} \overset{\text{ind.}}{\sim} \mathrm{Unif}(A_j)$. For a uniformly distributed random permutation $\Pi_t^{\mathrm{S},M}$, set $\vec{V}_t^{\mathrm{S},i} = \Pi_t^{\mathrm{S},M} \vec{U}_t^{\mathrm{S},i}$. Now, for each $r \in \{1, \ldots, M\}$ define

$$\widetilde{X}_{r,t+1}^{\mathrm{S}} = \widetilde{X}_{r,t}^{\mathrm{S}} + \sum_{i=1}^{I} S_{r,t}^{\mathrm{S},i} \zeta^i, \tag{56}$$

for given deterministic $\widetilde{X}_{r,0}^{\mathrm{S}} = X_r^{\mathrm{S}}(0)$, where we sample $S_{r,t}^{\mathrm{S},i}$ via

$$S_{r,t}^{\mathrm{S},i} := F^{-1}_{\lambda_{r,t}^{\alpha,i}} \left(V_{r,t}^{\mathrm{S},i} \right), \tag{57}$$

and where $\lambda_{r,t}^{\alpha,i} := \rho^i\big(t_t, \widetilde{X}_{r,t}^{\mathrm{S}}\big)\tau$ and F_λ is the Poisson CDF with parameter λ.

The construction of the samples used for the other variance-reduced pathwise mean estimators follow a similar development. Hybrid paths $\widetilde{X}_{r,t}^{\mathrm{H}}$ are simulated almost identically to the stratified case, save that

$$U_{j,t}^{\mathrm{H},i} \sim \mathrm{Unif}(A_j) \qquad \text{if } 1 \le j \le \frac{M}{2} \tag{58}$$

$$U_{j,t}^{\mathrm{H},i} = 1 - U_{M+1-j,t}^{\mathrm{H},i} \qquad \text{otherwise.} \tag{59}$$

Then, for a uniformly distributed random permutation $\Pi_t^{\mathrm{H},M}$, $\vec{V}_t^{\mathrm{H},i} := \Pi_t^{\mathrm{H},M} \vec{U}_t^{\mathrm{H},i}$, and $\widetilde{X}_{r,t}^{\mathrm{H}}$ is constructed as above.

Finally, in the case of the antithetic estimator, to generate an even number M of paths $\widetilde{X}_r^{\mathrm{A}}$, simulate $V_{r,t}^{\mathrm{A},i} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(0,1)$ for each $r \in \{1, \ldots, \frac{M}{2}\}$, and for each event channel i and timestep t. We then simulate $\widetilde{X}_{r,t}^{\mathrm{A}}$ as above, save that we define

$$S_{r,t}^{\mathrm{A},i} := \begin{cases} F^{-1}_{\lambda_{r,t}^{\alpha,i}} \left(V_{r,t}^{\mathrm{A},i} \right) & \text{if } 1 \le r \le \frac{M}{2} \\ F^{-1}_{\lambda_{r,t}^{\alpha,i}} \left(1 - V_{M+1-r,t}^{\mathrm{A},i} \right) & \text{if } \frac{M}{2} + 1 \le r \le M. \end{cases} \tag{60}$$

While the multidimensional extension of the technique presented here appears natural and well motivated, extension of the proof of Theorem 2 to the multidimensional case may require a different mathematical approach, and proof of such a theorem is reserved for future work.

A. Building Population Dynamics

Take as an example a simple model of the movement of people through departures side of the three domestic terminals of O'Hare International airport. The airport is represented by a six node graph, with each terminal comprised of a pre- and post-security node, connected as shown in Figure 2. The state $X_t \in \mathbb{R}^6$ represents the population at each

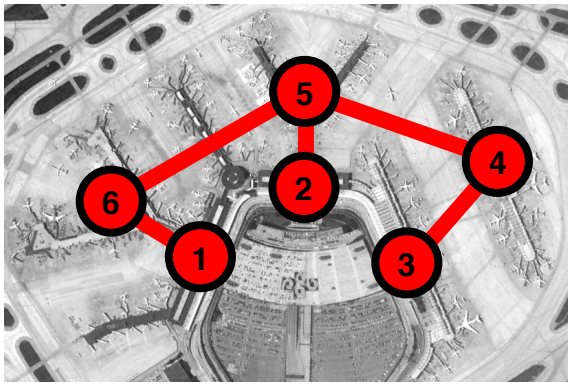

Fig. 2. Six node graph of O'Hare International Airport's domestic terminals [16]. State $X_t \in \mathbb{R}^6$ is the population of each node. Measurements Y_t are taken at nodes 1, 2, 3, and 5.

node at time t. There are source type event channels at nodes 1, 2 and 3 which correspond to jump vectors $\zeta^i = u_i$ the ith standard unit vector in $\mathbb{R}^6$ for $i \in \{1, 2, 3\}$. The corresponding rates at which these transitions occur are simple time dependent, piecewise linear functions (though they needn't be, in general), $\rho_i(t, X_t) = \rho_i(t)\eta$, where $\eta = 100$ is a model scaling factor. There are sinks at nodes 4, 5 and 6, with corresponding jump vectors $\zeta^i = -u_i$ for $i \in \{4, 5, 6\}$, and time dependent, linear rate functions $\rho_i(t, X_t) = \rho_i(t)(X_t)_i$, where $(x)_i$ denotes the ith component of $x \in \mathbb{R}^6$, and $\rho_i(t)$ is also piecewise linear. A jump transition i where a person moves from node ℓ to node m along an edge of the graph has corresponding jump vector $\zeta^i = u_m - u_\ell$. Such transitions from node $\ell \in \{1, 2, 3\}$ to node $m \in \{4, 5, 6\}$ along an edge of the graph are taken to be linear in the state, with no explicit time dependence, of the form $\rho_i(t, X_t) = \frac{3\gamma}{2}(X_t)_\ell$, where $\gamma = 0.12$ is another scaling factor. Transitions in the opposite direction, that is from $m \in \{4, 5, 6\}$ to $\ell \in \{1, 2, 3\}$ are taken to have rate 0. Finally people are free to move in any direction along a single edge from node ℓ to node m, where $\ell, m \in \{4, 5, 6\}$. Such a transition i is taken to have nonlinear rate function

$$\rho_i(t, X_t) = \frac{\gamma}{2} \left((X_t)_5 + \frac{(X_t)_5^2}{2\eta} \right) \tag{61}$$

if the transition is leaving node 5 and

$$\rho_i(t, X_t) = \frac{\gamma}{3} \left((X_t)_\ell + \frac{(X_t)_\ell^2}{2\eta} \right), \ \ell \in \{4, 6\} \tag{62}$$

if the transition is entering node 5. Noisy, low resolution measurements of part of the state are available at nodes 1, 2, 3, and 5, and are given by

$$(Y_t)_\ell = \chi \left(\left\lceil \frac{(X_t)_\ell}{\chi} \right\rceil + N_t \right), \ \ell \in \{1, 2, 3, 5\} \tag{63}$$

where χ is a scaling factor determining resolution as well as variance of the noise, and N_t is an i.i.d. noise sequence drawn from a quartic power law.

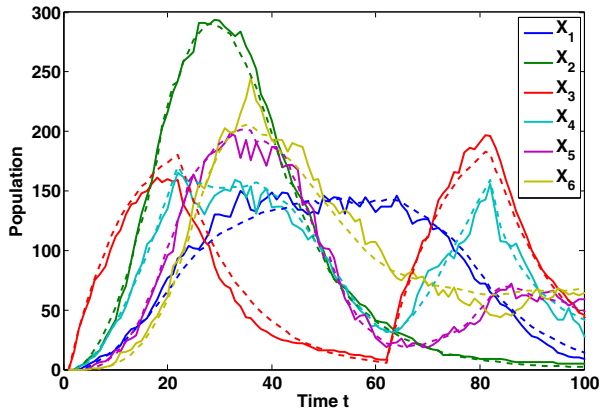

Fig. 3. Illustrative sample path X_t, shown with mean path estimator (dashed lines) of a corresponding particle filter with 100 particles. Note that here the measurements have a resolution of only $\chi = 50$ and nodes 4 and 6 are not even observed directly.

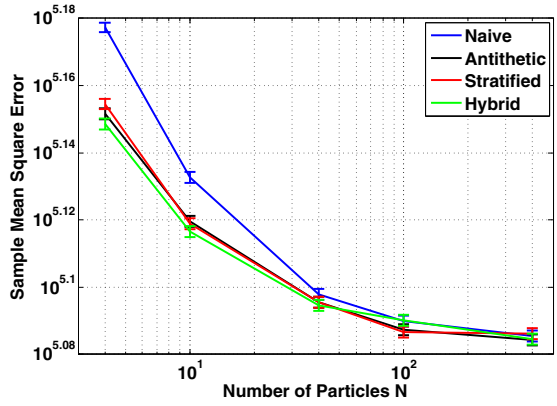

Fig. 4. Plot of sample mean square error of the mean of the empirical particle distribution from an ensemble of random sample paths of the airport model. Error bars shown are standard error of the mean. Here, ensemble size is 16000. Note the apparent convergence to the Bayesian limit.

B. Numerical Results

We see in Fig. 3 a single sample path of this process, along with the mean estimator produced by a 100 particle filter. We can repeat the experiment in the scalar case, save that we now compute MSE in terms of norms squared $\|X^e - \hat{X}^e\|^2$ and approximate the expected MSE of the multidimensional particle filter estimator, $\mathbb{E}[\|X-(\pi^N, x)\|^2]$. Again, as the number of particles N becomes sufficiently large, this quantity should converge to the optimal Bayesian limit $\mathbb{E}[\|X - (\pi, x)\|^2] = \mathbb{E}[\|X - \mathbb{E}[X|Y]\|^2] > 0$. Fig. 4 plots these estimated MSEs versus number of particles for each of the techniques presented here. Observe that the estimators appear to closely approximate the limit in a relatively small number of particles, and that the MSEs of the anticorrelated particle filter mean estimators appear to be upper bounded by the naive particle filter. Also note that relative MSE reductions do not appear to be as drastic as in the one dimensional case.

V. CONCLUSION

In this work, we proposed the idea of introducing anti-correlation into the Monte Carlo prediction step of particle filters. In the one dimensional case, we proved that almost every of these (random) measures converge to the Bayesian limit in the weak sense, and then supported this with numerical results for a Poisson process model that showed reductions of greater than an order of magnitude in MSE (relative to Bayesian optimal) from the naive technique. We also showed that these techniques can be extended to higher dimensions in the case of bootstrap filter estimation of tau-leaping models. We showed more modest reductions in MSE for the anticorrelated techniques in this case.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal filtering*. Prentice-Hall, 1979.

[2] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners", *IEEE Trans. Sign. Proc.*, vol. 50, 2002, pp. 736-746.

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

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

[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] 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.

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

[9] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *Proc. Inst. Elect. Eng. F*, vol. 140, 1993, pp. 107-113.

[10] H. Akashi and H. Kumamoto, "Construction of discrete-time nonlinear filter by Monte Carlo methods with variance-reducing techniques", *Systems and Control*, vol. 19, 1975, pp. 211221.

[11] S. Maskell and N. Gordon, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174188, 2001.

[12] J. E. Handschin and D. Q. Mayne, "Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering", *International Journal of Control*, vol. 9, 1969, pp. 547559.

[13] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, 2nd edition, 2004.

[14] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 125, 1996.

[15] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems", *Journal of the American Statistical Association*, vol. 93, pp. 10321044, 1998.

[16] http://www.allairports.net/chicago/chicago-airport-terminal-map.htm