THE JOURNAL OF CHEMICAL PHYSICS 136, 034115 (2012)

A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems

Patrick W. Sheppard,^{1,a),b)} Muruhan Rathinam,^{2,b),c)} and Mustafa Khammash^{1,d)} ¹Department of Mechanical Engineering, University of California, Santa Barbara Engineering II Bldg., Santa Barbara, California, USA 93106-5070

²Department of Mathematics and Statistics, University of Maryland, Baltimore County 1000 Hilltop Circle, Baltimore, Maryland 21250, USA

(Received 15 July 2011; accepted 27 December 2011; published online 20 January 2012)

Characterizing the sensitivity to infinitesimally small perturbations in parameters is a powerful tool for the analysis, modeling, and design of chemical reaction networks. Sensitivity analysis of networks modeled using stochastic chemical kinetics, in which a probabilistic description is used to characterize the inherent randomness of the system, is commonly performed using Monte Carlo methods. Monte Carlo methods require large numbers of stochastic simulations in order to generate accurate statistics, which is usually computationally demanding or in some cases altogether impractical due to the overwhelming computational cost. In this work, we address this problem by presenting the regularized pathwise derivative method for efficient sensitivity analysis. By considering a regularized sensitivity problem and using the random time change description for Markov processes, we are able to construct a sensitivity estimator based on pathwise differentiation (also known as infinitesimal perturbation analysis) that is valid for many problems in stochastic chemical kinetics. The theoretical justification for the method is discussed, and a numerical algorithm is provided to permit straightforward implementation of the method. We show using numerical examples that the new regularized pathwise derivative method (1) is able to accurately estimate the sensitivities for many realistic problems and path functionals, and (2) in many cases outperforms alternative sensitivity methods, including the Girsanov likelihood ratio estimator and common reaction path finite difference method. In fact, we observe that the variance reduction using the regularized pathwise derivative method can be as large as ten orders of magnitude in certain cases, permitting much more efficient sensitivity analysis than is possible using other methods. © 2012 American Institute of Physics. [doi:10.1063/1.3677230]

I. INTRODUCTION

Stochastic models of chemical reaction networks are critical to capturing the inherent randomness and discrete nature in intracellular networks characterized by low molecule copy numbers. In fact in the presence of nonlinearities, continuous deterministic models described by reaction rate equations may not even capture the average behavior of these systems correctly.¹

Chemical reaction models typically depend on a set of kinetic parameters whose values are often unknown or fluctuate due to an uncertain environment. Even small changes to the parameters may significantly alter the system output, and thus it is critical to quantify the effects of such changes. Parametric sensitivity analysis studies the change of system outputs to variations in kinetic parameters and is an indispensable analysis technique in the study of kinetic models. It enables one to elucidate system robustness properties, to pinpoint critical or rate limiting pathways, and to obtain reduced order models. In the biological context, sensitivity analysis can also guide drug targeting and synthetic parameter design.

Sensitivity analysis methods divide into two different categories, each being suitable for different applications. These categories, respectively, quantify the effects of finite or infinitesimal perturbations of model parameters (Fig. 1). In finite perturbation analyses, the parameter of interest is perturbed by a typically small, but not vanishingly small, amount. The difference between the perturbed and nominal models is then quantified by a finite difference calculation that uses probability densities solved using the finite state projection (FSP) method² or one that uses Monte Carlo simulations.^{3,4} Of particular relevance to this paper is the CRP method⁴ which uses Kurtz's random time change formula with common random numbers to achieve sizable reductions in the estimator variance. The other major sensitivity category is concerned with infinitesimal perturbations. In this analysis, the parameter of interest is perturbed by a vanishingly small amount, i.e., the quantity of interest is the partial derivative of some system functional with respect to a given parameter. The two categories are clearly related in that the infinitesimal sensitivity can be thought of as a limit of the finite sensitivity.

This paper considers the computation of infinitesimal parameter sensitivities for stochastic chemical reaction net-

^{a)}Electronic mail: sheppard@engineering.ucsb.edu.

^{b)}P. W. Sheppard and M. Rathinam contributed equally to this work.

^{c)}Electronic mail: muruhan@umbc.edu.

^{d)}Present address: Department of Biosystems Science and Engineering, ETH Zurich, 1058 7.00 Mattenstrasse 26, 4058 Basel, Switzerland. Electronic mail: mustafa.khammash@bsse.ethz.ch.



FIG. 1. Monte Carlo approaches to computing sensitivity in stochastic models of chemical kinetics.

works. The chemical master equation (CME) is a set of linear ordinary differential equations (ODEs) that describes the probability densities of the state populations of a chemical system.¹ A system of ODEs that exactly describes the evolution of the infinitesimal sensitivity coefficients can be derived directly from the CME by differentiating the probability densities with respect to parameters. Although this system of ODEs for the CME and sensitivities can be very large (or even infinite) making direct integration intractable, the finite state projection (FSP) method can be used to truncate the state space and solve for the sensitivities to within a guaranteed level of accuracy for many systems.⁵ Even for finite systems, however, the effective size of the state space grows exponentially with the number of chemical species and so FSP can be unwieldy for larger systems with many species. One must, therefore, turn to Monte Carlo methods for a computationally tractable way to compute infinitesimal sensitivities for such systems.

There are two primary Monte Carlo approaches to computing sensitivities with respect to infinitesimal perturbations (Fig. 1). The first uses likelihood ratios (LR). Plyasunov and Arkin recently presented a LR algorithm based on the Girsanov measure transformation for the sensitivity analysis of discrete stochastic chemical reaction networks.⁶ The second approach is that of pathwise differentiation (PD), which is commonly called infinitesimal perturbation analysis.^{7,8} This approach uses differentiation of sample path functionals to estimate the infinitesimal sensitivities with variance that can be significantly lower than what is possible using LR methods.⁹ However, the severe drawback of PD is that it is only applicable to a restricted class of problems for which certain regularity conditions are satisfied.

Indeed, the PD method is not directly applicable for many problems that are of interest when studying stochastic chemical and biochemical reaction networks. In this paper, we propose the regularized pathwise derivative (RPD) method, which efficiently estimates parameter sensitivities for discrete stochastic chemical reaction networks using a suitably modified form of PD. The RPD algorithm is derived using the random time change (RTC) representation of Markov processes¹⁰ and is applicable to a large class of sensitivity problems arising in stochastic chemical kinetics. We show via several numerical examples that this method can achieve significantly lower variance estimates compared to the existing methods to compute sensitivities of stochastic chemical kinetics.

The contents of the paper are organized as follows. Section II provides an overview of stochastic chemical kinetics as well as the problem of infinitesimal sensitivity analysis. Section III presents the random time change representation of discrete stochastic chemical systems, which is then used to formulate the regularized pathwise derivative method. A numerical algorithm to implement the method is given in Sec. III A. Several numerical examples in Sec. IV are provided to demonstrate the new algorithm and to compare its performance with alternative methods. Issues regarding the selection of the RPD algorithm parameter, *w*, are discussed in Sec. V. We conclude by summarizing our contributions in Sec. VI.

II. INFINITESIMAL SENSITIVITY ANALYSIS IN STOCHASTIC CHEMICAL KINETICS

In this section, we briefly review stochastic chemical kinetics and introduce the general problem formulation for infinitesimal sensitivity analysis. For general reference on infinitesimal sensitivity analysis of stochastic systems, see Refs. 8 and 9.

Consider a chemical reaction network whose dynamics are modeled using stochastic chemical kinetics.¹ In this framework, the network is described by a continuous-time Markov chain. The populations of n chemical species are discrete values whose elements make up the state vector X(t) $\in \mathbb{Z}^n_+$. There are *M* total reactions in the network, and each time reaction $j \in \{1, ..., M\}$ fires, the populations of the species change by a discrete amount specified by the stoichiometry vector, v_i . Reactions occur randomly in continuous time, and conditioned on X(t) = x, the probability that the *j*th reaction fires during the infinitesimal time interval (t, t + dt]is given by $a_i(x)dt$, where $a_i(x)$ is the propensity function, and dt is an infinitesimally small time increment. The propensities are in general dependent on the state of the system and one or more system parameters, which we will denote by c. For concreteness we shall treat c as a scalar in our discussion bearing in mind that it is easy to extend the results to a vector c of parameters.

In this study, we are interested in determining the infinitesimal sensitivity defined by the partial derivative $\frac{\partial}{\partial c}y(c)$ of the expected value y(c) of some network "output" Y to infinitesimally small perturbations in parameter c. Here, output Y is in general some scalar function of the system trajectory and hence is a random variable and y = E(Y). Examples of y that arise in analyzing chemical reaction networks are the expected population of a particular species at a specified time T, the probability of the state being within a specified subset of the state space at time T, and the expected time of reaching a particular state or subset of states. We note that all of these quantities can be represented by y = E(Y), where Y is a function of the path. In this paper, we focus on outputs of the form $Y(\omega) = f(X(T, \omega))$, where T > 0 is some fixed final time, ω is an underlying random element from the sample space, and f is a scalar function of the state. This does not include

outputs that involve random times such as the time it takes to reach a certain state or the time spent on a certain set of states. Nevertheless it does cover a large class of outputs of interest.

There are two different viewpoints when it comes to modeling the dependence on parameter *c*. One viewpoint regards the sample trajectories of the system to be independent of the parameters *c*, while the probabilities of these sample paths to be dependent on *c*; the other viewpoint regards the trajectories to be dependent on parameters *c* as well as some underlying random elements ω whose probabilities are independent of *c*. Both viewpoints are quite natural and consistent and lead to different approaches to compute the sensitivity $\frac{\partial}{\partial c}y(c)$.

To see the first point of view, we fix an interval [0, T] and write the expected value y(c) of the path function as

$$y(c) = \int_{x} f(x) p(x; c) dx,$$

where x denotes a feasible trajectory (over the interval [0, *T*]) and p(x; c) the corresponding probability density of the trajectory x defined in some suitable sense and the integration is carried over the set of all feasible trajectories. We note here that it is indeed possible to define the density p(x; c) rigorously and derive a simple formula to compute it. If the propensities a_j are differentiable with respect to c, then so would be the density p(x; c). Assuming that differentiation inside the integral is valid we obtain that

$$\frac{\partial}{\partial c}y(c) = \int f(x) \frac{\partial}{\partial c}p(x;c)dx.$$

In order to apply Monte Carlo methods to perform the above integration the following mathematical sleight of hand known as the likelihood ratio method may be used. One writes the above equation as

$$\frac{\partial}{\partial c}y(c) = \int f(x) \frac{\partial}{\partial c} \ln p(x;c) \ p(x;c) \, dx.$$

Thus the required sensitivity may be written as

$$\frac{\partial}{\partial c} y(c) = E\left(f(X) W\right),\tag{1}$$

where the random (weight) variable W is given by

$$W = \frac{\partial}{\partial c} \ln p(X;c),$$

where X denotes the random trajectory over the interval [0, T].

See Appendix A where an analytical formula for the weight W is given. We shall refer to this particular form of the likelihood ratio method as the Girsanov likelihood ratio (GLR) method. Appendix A provides further details as well as the GLR algorithm. The GLR method was introduced to the stochastic chemical kinetic setting in Ref. 6. Further details regarding LR may be found in Refs. 11 and 12.

The second viewpoint can be understood clearly in the context of stochastic simulation where one uses a stochastic simulation algorithm (SSA). Here the sample elements ω can be interpreted as the sequence of random numbers drawn from a random number generator, for instance, from the uniform

distribution $\mathcal{U}(0, 1)$. The probability measure of these random numbers is obviously independent of any system parameters, while the sample paths generated via SSA now clearly depend on *c*. In this case y(c) can be written as

$$y(c) = E[Y] = \int_{\omega \in \Omega} f(X(T, \omega; c)) P(d\omega), \qquad (2)$$

where Ω is the set of all possible random outcomes (for instance, all possible sequences of uniform i.i.d. random numbers) and *P* is the probability measure on these outcomes. If $Y(\omega, c) = f(X(T, \omega; c))$ is differentiable with respect to *c* and $\frac{\partial}{\partial c}Y$ has an analytically tractable form, then one may seek to estimate $\frac{\partial}{\partial c}y(c)$ using the following formula obtained by differentiating inside the expected value (integral) in (2):

$$\frac{\partial}{\partial c}y(c) = E\left[\frac{\partial}{\partial c}Y\right] = E\left[\frac{\partial}{\partial c}f(X(T,\omega;c))\right].$$
 (3)

The quantity $\frac{\partial}{\partial c}Y(\omega, c)$ is the *pathwise derivative* of the output *Y*. If the interchange of derivative and expectation is valid and $\frac{\partial}{\partial c}Y(\omega, c)$ has an analytically tractable form, then one may compute it along with each simulation ω and use the sample average from an i.i.d. sample as an estimate of the sensitivity $\frac{\partial}{\partial c}y(c)$. We shall refer to this method as the *pathwise derivative* method (PD). This approach is also known as infinitesimal perturbation analysis (IPA) in the literature.^{7,8}

We note that the applicability of both LR and PD methods depend on the validity of differentiation inside certain integrals as well as the existence of tractable analytical formulas for the weights (in the case of LR) or the pathwise derivatives (in the case of PD). In the context of stochastic chemical kinetics, analytical formulas for the weights to be used in the GLR were given in Ref. 6. In this paper, we derive the relevant analytical formulas for the pathwise derivatives in Sec. III. Typically, differentiation inside expectation is valid in the context of LR and not always valid in the context of PD. The reason for this is that the probabilities or probability densities are often very well-behaved functions of c, while the random output $Y(\omega, c)$ is not always well behaved as in the example of stochastic chemical kinetics. The PD method has not been applied for sensitivity analysis in the context of stochastic chemical kinetics and may not be applied directly for most outputs of interest. In order to apply the PD method to stochastic chemical kinetics, we regularize the problem in a suitable way and we refer to the resulting method as the RPD. This is presented in Sec. III.

We observe that both methods LR and PD lead to Monte Carlo estimation of sensitivity in the form of the sample mean of a *sensitivity estimator* which we denote by $Z(c, \omega)$. In the case of LR,

$$Z(c, \omega) = Y(\omega)W(\omega, c),$$

where *Y* is the output and *W* is the random weight. In the case of PD,

$$Z(c,\omega) = \frac{\partial}{\partial c} Y(\omega,c)$$

In both cases $E[Z(c)] = \frac{\partial}{\partial c}y(c)$. The Monte Carlo method consists of generating a sample of *N* independent realizations

 $Z_1(c), \ldots, Z_N(c)$ of Z(c), and then computing the sample average, $\frac{1}{N} \sum_{i=1}^{N} Z_i(c, \omega)$.

The efficiency of a sensitivity estimator Z(c) depends on its variance, with smaller variance resulting in greater efficiency. Two different estimators such as those obtained from LR and PD methods will differ in their variances. It has been observed that while PD method is not always applicable, when applicable it has lower variance than the LR method.⁹

III. THE REGULARIZED PATHWISE DERIVATIVE METHOD

The RPD method we introduce in this paper uses the RTC representation of stochastic chemical kinetics. We briefly reproduce relevant aspects of the RTC algorithm from Ref. 4. The RTC description expresses the state X(t) in terms of Poisson processes as follows:

$$X(t, \omega, c) = X(0, \omega, c) + \sum_{j=1}^{M} v_j Y_j(S_j(t, \omega, c), \omega), \quad (4)$$

where

$$S_j(t, \omega, c) = \int_0^t a_j(X(s, \omega, c), c) ds, \quad j = 1, ..., M,$$
 (5)

and $Y_j(., \omega)$, j = 1, ..., M are *independent unit-rate Poissons* associated with each reaction channel. In the RTC equation (4), we refer to S_j as the (dimensionless) *internal time* of the reaction channel *j*. Recall that v_j and $a_j(., c)$ are the stoichiometry vector and propensity function, respectively, for reaction channel *j*. Once we have a realization $Y_1(., \omega), ..., Y_M(., \omega)$ of the noise we can use (4) to solve for $X(., \omega, c)$ as we now describe. We start by denoting the random internal jump times of the Poisson process Y_j by I_i^j where j = 1, ..., M, i = 1, 2, ..., and

$$I_1^j < I_2^j < I_3^j \dots$$

for each *j*. We will also define the *internal time for the next firing of reaction channel j* when regarded at physical time *t* by

$$I_{+}^{j}(t) = \min \left\{ I_{l}^{j} | I_{l}^{j} > S_{j}(t), \ l = 1, 2, \dots \right\},\$$

for j = 1, ..., M. Thus the *i*th firing of the *j*th reaction channel will occur at physical time *t*, if $S_j(t, \omega, c) = I_i^j(\omega)$.

We can uniquely represent the trajectory $X(., \omega, c)$ using the collection (T_i, J_i) for i = 1, 2, ..., where $T_i(\omega, c)$ is the physical time of the *i*th firing of *any* reaction channel, and $J_i(\omega, c) \in \{1, 2, ..., M\}$ is the index of the reaction channel that fires at time T_i . With this notation, we now see that the internal time $S_i(t, \omega, c)$ is piecewise affine in t,

$$S_{j}(t, \omega, c) = S_{j}(T_{i}(\omega, c), \omega, c) + a_{j}(X(T_{i}(\omega, c), c)$$
$$\times (t - T_{i}(\omega, c)), \qquad (6)$$
$$j = 1, \dots, M$$

for $T_i \leq t < T_{i+1}$. This will facilitate easy computation.

Assuming that T_1, \ldots, T_i and J_1, \ldots, J_i are known for some *i*, we also know $X(T_i)$ and $I^j_+(T_i)$ and can compute T_{i+1} and J_{i+1} as follows. First, we note that when the physical time

is equal to T_i , the internal times of the processes are given by $S_j(T_i)$, and that the next firing times of the reactions in their respective internal time frames are given by I_+^j . During $t \in [T_i, T_{i+1})$, the internal times $S_j(t)$ increase at the constant respective rates $a_j(X(T_i))$, and so the physical time $T_{i+1} - T_i$ which elapses before the next firing of a reaction is simply the minimum of $(I_+^j - S_j(T_i))/a_j(X(T_i))$. We then obtain

$$T_{i+1} = T_i + \min\left\{\frac{I_+^j - S_j(T_i)}{a_j(X(T_i))} \middle| j = 1, \dots, M\right\}.$$
 (7)

Here, J_{i+1} is the index of the minimum in Eq. (7) and is unique for almost all ω , and we note that $T_{i+1}(\omega, c) = I_{i+1}^{J_{i+1}(\omega,c)}$.

Assuming $a_i(x; c)$ are smooth functions of c, one may show from Eq. (7) that the times $T_i(\omega, c)$ at which reactions fire are piecewise differentiable in c for each ω . Thus for each ω we may differentiate T_i to compute $\frac{\partial T_i}{\partial c}$. In fact from differentiation of Eq. (7), we obtain the recursive formulas

$$\frac{\partial T_{i+1}}{\partial c} = \frac{\partial T_i}{\partial c} - \frac{I_+^{j^*}(T_i) - S_{j^*}(T_i)}{(a_{j^*}(X(T_i)))^2} \frac{\partial a_{j^*}}{\partial c}(X(T_i)) - \frac{\frac{dS_{j^*}(T_i)}{dc}}{a_{j^*}(X(T_i))},$$
(8)

where $j^* = J_{i+1}$ is the index of the minimum in Eq. (7). The derivatives $\frac{dS_j(T_i)}{dc}$ are found by differentiating Eq. (6),

$$\frac{dS_j(T_{i+1})}{dc} = \frac{dS_j(T_i)}{dc} + \frac{\partial a_j}{\partial c}(X(T_i))(T_{i+1} - T_i) + a_j(X(T_i)) \times \left(\frac{\partial T_{i+1}}{\partial c} - \frac{\partial T_i}{\partial c}\right).$$
(9)

In deriving Eqs. (8) and (9), we have used the fact that for any given nominal value of *c*, with probability one the quantities $X(T_i)$ and $I^j_+(T_i)$ are locally constant in *c*. For each given *c* Eqs. (8) and (9) hold for almost all ω , and from these one can solve recursively for $\frac{\partial T_i}{\partial c}$ for i = 0, 1, 2, ..., using the initial conditions

$$\frac{\partial T_0}{\partial c} = 0, \quad \frac{dS_j(T_0)}{dc} = 0, \quad j = 1, \dots, M.$$
(10)

If we are interested in the sensitivity of $E(T_n)$ with respect to *c* for some fixed *n*, then Eqs. (8) and (9) allow us to compute $\frac{\partial T_n}{\partial c}$ path by path and then take the sample average to estimate $\frac{\partial E(T_n)}{\partial c}$.

But as mentioned earlier, we are interested in the sensitivity of an output *Y* of the form Y = f(X(T)), where *f* is some function of the state and *T* is some final time: i.e., we seek to compute the derivative $\frac{\partial E(f(X(T)))}{\partial c}$. In this case an analogous approach would require us to

In this case an analogous approach would require us to compute $\frac{\partial f(X(T))}{\partial c}$ path by path and then take sample average. But unfortunately this will yield zero as for any given *T* and *c*, for almost all ω , there will be no jump at *T* and hence X(T) will be constant in a neighborhood of that *c* and hence $\frac{\partial f(X(T))}{\partial c}$ = 0. However, $\frac{\partial E(f(X(T)))}{\partial c}$ is typically nonzero, indicating that

$$E\left(\frac{\partial f(X(T))}{\partial c}\right) \neq \frac{\partial E(f(X(T)))}{\partial c},$$

i.e., the expectation and derivative (with respect to c) do not commute. Thus the PD method is not applicable.

To circumvent this issue, we regularize the problem in the following way. Suppose instead of $\frac{\partial}{\partial c}E(f(X(T)))$, we seek to compute the sensitivity

$$\frac{\partial}{\partial c} E\left(\int_{T-w}^{T+w} \frac{1}{2w} f(X(t)) dt\right),\tag{11}$$

where w > 0 is the half width of the regularizing window. In this case, we may compute the pathwise sensitivity

$$\frac{\partial}{\partial c} \left(\int_{T-w}^{T+w} \frac{1}{2w} f(X(t)) dt \right), \tag{12}$$

and then take a sample average. The pathwise sensitivity will not be zero as the probability of a jump in the window [T - w, T + w] is not zero. It turns out that this results in a feasible method. Even though this modifies the sensitivity problem, if w is small enough, the difference between the original and modified sensitivities will be sufficiently small in most practical settings.

Now we shall derive a formula for the pathwise derivative (12) in terms of $\frac{\partial T_n}{\partial c}$ for $n = 1, 2, \dots$ Note that f(X(t, c))is a piecewise constant function of t with jumps and may be written as

$$f(X(t,c)) = f(x_0) + \sum_{i=1}^{\infty} \{ f(X(T_i(c),c)) - f(X(T_{i-1}(c),c)) \} H(t - T_i(c)),$$

where H is the Heavyside function (step function) and we observe that for any (finite) t only finitely many of the terms in the above sum are nonzero. Let us define $\Delta_i(c)$ by

$$\Delta_i(c) = f(X(T_i(c), c)) - f(X(T_{i-1}(c), c)).$$
(13)

Let [T - w, T + w] be a fixed interval. Then for $c = c_0$, with probability 1 none of the $T_i(c_0)$ equal T - w or T + w. Suppose that $i = i_l$ and i_u are lowest and highest indices such that $T_i(c_0) \in (T - w, T + w)$. Then we obtain

$$\frac{\partial}{\partial h}\Big|_{h=0} \int_{T-w}^{T+w} f(X(t,c_0+h))dt$$
$$= \sum_{i=1}^{\infty} \frac{\partial}{\partial h}\Big|_{h=0} \Delta_i(c_0+h) \int_{T-w}^{T+w} H(t-T_i(c_0+h))dt.$$

We make two observations. First with probability 1, $\Delta_i(c_0)$ (+ h) is constant in a neighborhood of h = 0 for any finitely many of indices *i* (we shall only care about $i_l \leq i \leq i_u$). Second, from the definition of the Heavyside function, it can be shown that

$$\frac{\partial}{\partial h}\Big|_{h=0}\int_{T-w}^{T+w}H(t-T_i(c_0+h))dt=-\frac{\partial T_i}{\partial c}(c_0),$$

if $T_i(c_0) \in (T - w, T + w)$ and zero if $T_i(c_0) < T - w$ or $T_i(c_0) > T + w$ (the cases $T_i(c_0) = T - w$ or $T_i(c_0) = T$ + w have zero probability and can be omitted). Thus, we obtain the formula for the regularized pathwise sensitivity

$$\frac{1}{2w} \frac{\partial}{\partial h} \Big|_{h=0} \int_{T-w}^{T+w} f(X(t, c_0 + h)) dt$$
$$= -\frac{1}{2w} \sum_{i=i_l}^{i_u} \Delta_i(c_0) \frac{\partial T_i}{\partial c}(c_0).$$
(14)

Thus the computation of path by path derivative $\frac{\partial}{\partial h}|_{h=0} \int_{T-w}^{T+w} f(X(t, c_0 + h))dt$ is accomplished along with a RTC SSA simulation by using Eqs. (8), (9), (13), and (14).

In Appendix **B**, we discuss the validity of differentiation inside expectation in the context of the RPD method.

A. Algorithm

The numerical algorithm follows the conventions of Sec. III and, in addition, uses variables $\tau \triangleq T_{i+1} - T_i$, $\Delta_T \triangleq \frac{\partial T_i}{\partial c} - \frac{\partial T_{i-1}}{\partial c}$, $\Delta_f^{(i)} \triangleq f(X(T_i)) - f(X(T_{i-1}))$, where f is the functional whose sensitivity we are interested in computing, and $dF \triangleq -\sum_{i=i_l}^{i_u} \Delta_f^{(i)} \frac{\partial T_i}{\partial c}$. We assume that each call to the function exprand() returns an independent random number drawn from the exponential distribution with rate one. We also assume that the system depends on a single parameter, c, in order to simplify notation. It is straightforward to modify this algorithm to simultaneously compute sensitivities with respect to a vector of parameters, $c = [c_1, \ldots, c_p]'$. The regularized pathwise derivative method can then be written as the following algorithm.

Regularized pathwise derivative (RPD) sensitivity algorithm

- 1. Specify x_0 , T_f , w, N
- 2. **for** k = 1 to *N* **do**
- 3. Initialization: Set t = 0, $X = x_0$, $f = f(x_0)$, dF = 0, $\frac{\partial T_i}{\partial c} = 0$; $I^j_+ = \text{exprand}()$, $S_j = 0$, $\frac{dS_j}{dc} = 0$ for j = 1, ..., M
- 4. while $t < T_f + w$ do
- 5. Calculate a_j and $\frac{\partial a_j}{\partial c}$, for $j = 1, \dots, M$
- 6. Calculate $\tau = \min_i \{ (I_+^j S_i)/a_i \}$
- Set j^* to the index of the minimum in the above equation 7.
- 8. Calculate $\Delta_T = -\frac{\tau}{a_{j^*}} \frac{\partial a_{j^*}}{\partial c} \frac{dS_{j^*}}{dc} \frac{1}{a_{j^*}}$ 9. Calculate $\frac{dS_j}{dc} = \frac{dS_j}{dc} + \frac{\partial a_j}{\partial c}\tau + a_j\Delta_T$ for j = 1, ..., M10. Update $\frac{\partial T_i}{\partial c} \leftarrow \frac{\partial T_i}{\partial c} + \Delta_T$ 11. Update $S_j \leftarrow S_j + a_j\tau$ for j = 1, ..., M

- Update $t \leftarrow t + \tau$ and $X \leftarrow X + \nu_{i^*}$ 12.
- 13. Update $I_{+}^{j^*} \leftarrow I_{+}^{j^*} + \text{exprand}()$
- 14. Calculate $f_+ = f(X)$ and set $\Delta_f^{(i)} = f_+ f$
- 15. **if** $T_f w < t < T_f + w$ **then**
- 16. Update $dF \leftarrow dF \Delta_f^{(i)} \frac{\partial T_i}{\partial c}$
- 17. end if
- 18. Update $f \leftarrow f_+$
- 19. end while
- 20. return $Z_k = \frac{1}{2w} dF$
- 21. end for
- Compute sensitivity from sample mean $\hat{z} = \frac{1}{N} \sum_{k=1}^{N} Z_k$ 22.

IV. NUMERICAL EXAMPLES

A. Monomolecular birth-death system

We first evaluate the proposed RPD method by estimating the sensitivities for a simple example whose analytical solution is available. The single species, two reaction birth-death process is a basic network in which a chemical species S is

$$\emptyset \xrightarrow{c_1} S \xrightarrow{c_2 x} \emptyset. \tag{15}$$

The propensity functions for the birth and death reactions conditioned on X(t) = x are given by $a_1(x, c) = c_1$ and $a_2(x, c) = c_2x$, respectively. All numerical examples hereafter will consider the system with parameters $c_1 = 10$ and $c_2 = 0.5$.

First, we consider sensitivity analysis of the expected value at a fixed time for this problem. The random variable X(t) can be expressed as the sum of independent Binomial and Poisson random variables, whose expected value at time t is given by

$$E[X(t)] = x_0 e^{-c_2 t} + \frac{c_1}{c_2} (1 - e^{-c_2 t}),$$

where $x_0 = X(0)$ is the initial condition.¹³ The infinitesimal sensitivities with respect to the birth and death rates for this system are readily obtained by taking the partial derivatives

$$\frac{\partial}{\partial c_1} E[X(t)] = \frac{1}{c_2} (1 - e^{-c_2 t}), \tag{16}$$

$$\frac{\partial}{\partial c_2} E[X(t)] = -tx_0 e^{-c_2 t} + \frac{c_1}{c_2} t e^{-c_2 t} - \frac{c_1}{c_2^2} (1 - e^{-c_2 t}).$$
(17)

With initial state $x_0 = 0$, the analytical sensitivities with respect to c_1 and c_2 evolve with time and are shown in Figs. 2(a) and 2(b) (dashed lines).

Whereas the Girsanov method produces an unbiased estimate of the infinitesimal sensitivity, both the RPD and CRP methods solve approximate problems and do not estimate the sensitivity directly. As such, each method introduces a bias and thus there is a tradeoff between bias and variance, which are both functions of the method parameter. In order to compare the methods in a reasonable fashion, we first specify a bias (i.e., the error between the direct sensitivity and the finite difference or regularized sensitivity without considering any numerical errors due to machine precision) and then use w and h that, respectively, achieve this bias.

For the computation of the sensitivity with respect to c_1 to give a constant 0.1% bias, the RPD method requires a window size that changes with time (Fig. 2(e)). The centered finite difference estimator is an unbiased estimator of $\partial/\partial c_1$ for this particular problem, and consequently a very large perturbation size can be used to decrease the variance. In this case h = 9, or 90% of the nominal parameter value of c_1 , was used. The CRP estimator is no longer unbiased for the sensitivity with respect to the death rate, c_2 and so, such as RPD, the algorithm parameter h must be varied with time to give a fixed bias of 0.1%, (Fig. 2(f)). The estimates generated from N $= 10^4$ independent samples at each final time are shown in Figs. 2(a)-2(d). These results show that the Girsanov estimator is unbiased, but that the variance (and hence efficiency) of its estimate is significantly worse than either the RPD or CRP methods at all times. Additionally, its variance increases with

Comparison of sensitivity estimates versus time, birth-death system



FIG. 2. Comparison of sensitivity estimates for $E[X(T)|x_0 = 0]$ for the birth-death process with parameters $c_1 = 10$, $c_2 = 0.5$ as the final time (*T*) is varied, with the method parameters chosen for RPD and CRP to yield a fixed bias of 0.1%. (a) and (b) The sensitivity estimates from each method are compared with the analytical solutions (dashed line). (c) and (d) The variance of the sensitivity estimators at each time. (e) and (f) The algorithm parameter values (blue triangles: RPD, red x: CRP) used in each computation. All estimates were computed independently using $N = 10^4$ sample paths.

t, leading to a variance four orders of magnitude higher than either method at T = 50. Comparing RPD with CRP, with respect to c_1 the CRP method has much lower variance for early *t*, reaching as much as 20-fold lower variance than the RPD estimator at T = 8. However as time increases, the variance of RPD decreases so that by T = 50 the variance of RPD is less than 2-fold higher than that of CRP. For the sensitivity with respect to c_2 , the variance of the CRP estimator grows significantly as *t* increases, whereas the variance of RPD decreases as *t* increases while still producing an identical 0.1% bias. Consequently, CRP quickly loses its initial advantage of 20% lower variance than RPD at T = 1, and by T = 10 CRP variance is already 10-fold higher than RPD, and by T = 50there is a 330-fold higher variance.

We next examine sensitivity analysis with respect to a non-smooth functional for which the Girsanov method was previously shown to be particularly efficient relative to FD methods.⁶ For this case, the functional f(X(T))= $\mathbf{1}_{\{a \le X(T) \le b\}}$ is considered, where $\mathbf{1}_{\{\}}$ is the indicator function that takes value 1 when X(T) is in the interval [a, b], a < b, and zero otherwise. Note that $E[f(X(T))] = \mathbb{P}(a \le X(T) \le b)$. This type of functional arises frequently when one is interested in estimating the sensitivity of the probability that a chemical species is within a given range, for instance, a range that specifies some threshold levels. Again, the RPD, Girsanov, and CRP methods were used to compute the sensitivity of this functional with respect to c_2 for three different ranges of values at fixed T = 10 (Fig. 3).



FIG. 3. Sensitivity analysis for $E(\mathbf{1}_{\{a \le X(T) \le b\}})$ for the birth-death process with parameters T = 10, $c_1 = 10$, $c_2 = 0.5$, and $x_0 = 0$. (a) The sensitivity estimates are shown for the methods indicated by color and compared with the analytical result (black circle). Error bars indicate 95% confidence intervals of the estimates. (b) The variance of the sensitivity estimates is shown, with the vertical axis in log scale. All estimates were obtained using $N = 10^4$ sample paths. For RPD and CRP, w and h were selected to achieve a 0.1% bias relative to the exact sensitivities (w = 0.978, h = 0.0069 for $15 \le X(T) \le 16$; w = 0.876, h = 0.0094 for $10 \le X(T) \le 20$). For $1 \le X(T) \le 100$ w = 0.876 and h = 0.0024 were used.

When estimating sensitivities for the narrow interval 15 $\leq X(T) \leq 16$, all three methods produce estimates from 10^4 simulations whose confidence intervals contain the exact sensitivity. However, the RPD estimate is most efficient with a variance (12.8) three-fold lower than the variance of the Girsanov estimate (37.3) and over 22-fold lower than CRP (288.5). When considering a wider interval of $10 \le X(T)$ \leq 20, the RPD method still produces low variance estimates (variance 16.8), while the variance of the Girsanov estimator increases nearly five-fold to 155.1. The CRP estimate has a lower variance (176.3) but is still over 10-fold higher than RPD. Finally, when considering the interval $1 \le X(T) \le 100$ for which the functional is nearly constant and its sensitivity is very small (-9×10^{-8}), both RPD and CRP compute a sensitivity of zero with no variance, while the Girsanov estimate is nonzero (-0.292) and has high variance (318.0).

Revisiting the sensitivity of E[X(T)], the analytical formulas show that the sensitivity with respect to c_1 is independent of the initial state x_0 (16), while the sensitivity with respect to c_2 is an affine function of x_0 (17). We, therefore, examined whether the performance of the methods are dependent on the initial condition by considering the conditional expectation $E[X(T)|x_0]$ for this problem. RPD, Giranov, and CRP were again used to estimate the sensitivity as x_0 was varied between 0 and 5000 (Fig. 4).

As in the previous examples, w and h were selected, respectively, so that the RPD and CRP estimates would have a fixed bias of 0.1% relative to the exact sensitivity. Compared with the CRP method, RPD produces estimates with at least 20% lower variance for the sensitivity with respect to c_2 (Fig. 4(d)). The RPD is again less efficient than CRP with respect to c_1 sensitivity estimates (10-fold higher variance) due to the unbiasedness of the CRP estimator for this particular problem (Fig. 4(c)).

The results clearly demonstrate that for the sensitivity estimates with respect to c_1 , the variance of the estimates is independent of x_0 when using RPD and CRP, but that the variance rapidly increases with the Girsanov estimator as x_0 is changed (Fig. 4(c)). The variance of the estimates with respect to c_2 increases as x_0 increases for all three methods; however, the variance of the Girsanov estimator is several orders of magnitude higher than the other two methods (Fig. 4(d)). Thus, while the Girsanov method still produces unbiased estimates for this functional, it has the undesirable feature that its variance may be adversely affected by changes in initial conditions in certain problems.

B. Reversible isomerization system

The reversible isomerization system consists of two species, S_1 and S_2 , which switch between states via the following two reactions:

$$\mathcal{S}_1 \xrightarrow{c_1 \mathcal{X}_1} \mathcal{S}_2,$$
 (18)

$$\mathcal{S}_2 \xrightarrow{c_2 X_2} \mathcal{S}_1. \tag{19}$$

In this network, stochastic trajectories fluctuate about the expected value, the evolution of which undergoes an initial



FIG. 4. Comparison of sensitivity estimates for $E[X(T)|x_0]$ for the birth-death process with parameters T = 3, $c_1 = 10$, $c_2 = 0.5$ as the initial state x_0 is varied. (a) and (b) The sensitivity estimates from each method are compared with the analytical solution (dashed line). (c) and (d) Variance of the sensitivity estimates from each method. The window lengths ($w_1 = 0.289$, $w_2 = 0.309$) and perturbation sizes ($h_1 = 9.0$, $h_2 = 0.041$) were chosen for RPD and CRP, respectively, to yield a fixed bias of 0.1% of the exact sensitivity with $x_0 = 0$. All estimates were computed using $N = 10^4$ sample paths.

transient before reaching a stationary value (Fig. 5(a)). The expected value of the population of both species can be computed analytically for this simple example at a fixed time

$$E[X_1(t)] = X_1(0) + \frac{(1 - e^{-(c_1 + c_2)t})}{c_1 + c_2} (c_2 X_2(0) - c_1 X_1(0)),$$
(20)

$$E[X_2(t)] = X_2(0) - \frac{(1 - e^{-(c_1 + c_2)t})}{c_1 + c_2} \left(c_2 X_2(0) - c_1 X_1(0)\right),$$
(21)

where $X_1(0)$ and $X_2(0)$ are the initial populations of species S_1 and S_2 . In the numerical examples considered here, we assume initial conditions of $X_1(0) = 50$ and $X_2(0) = 0$ and parameters $c_1 = 0.3, c_2 = 0.5$. The exact sensitivities are obtained by differentiating the result with respect to each of the parameters.

We examine the performance of the RPD method relative to Girsanov and CRP methods by estimating the sensitivity of $E[X_1(T)]$ using 10⁴ independent trajectories at several choices of w for RPD and finite difference parameters h for CRP. While the Girsanov estimator is again unbiased, both RPD and CRP estimates produce a bias which depends on the algorithm parameters w and h. The algorithm parameters w and h can be chosen to achieve a specified bias for RPD and CRP, respectively, for this simple example, and so this permits comparison between the efficiency (variance) between RPD and CRP when the accuracy (bias) is fixed. The results examining the variance for several biases between 10⁻⁶ and 10⁻¹ demonstrate that during the transient regime (T = 3), both RPD and CRP achieve estimates with 5-fold lower variance or less compared to the Girsanov estimator with bias even as low as 10^{-6} (Fig. 5(b), solid lines). When the bias is permitted to increase, larger *w* and *h* can be used to reduce the variance further.

While at T = 3, the RPD estimator and CRP estimators had similar variance (RPD was 15%-25% lower variance than CRP for small bias values), the situation changes significantly when the estimators are used to evaluate the sensitivity in the later time regime at T = 40 (Fig. 5(b), dashed lines). At later times, the expected value has nearly reached stationarity allowing larger regularization windows to be used for RPD, whereas for CRP smaller perturbations must be used to maintain a low variance. At the smallest bias considered (10^{-6}) , CRP has only 1.73 fold-lower variance compared to the Girsanov estimator, whereas the RPD estimator has a variance over 6 500-fold lower than Girsanov. As h and w are increased to give a larger relative bias of 10^{-2} , the CRP estimate has over 500-fold lower variance compared with Girsanov but is still an order of magnitude higher than the RPD estimate, whose variance is 42 000-fold lower than Girsanov.

C. Genetic oscillator

Sensitivity analysis can be a particularly valuable tool when trying to identify which system parameters are the key regulators of a response in a very complex and high-dimensional network. Here, we consider a higher dimensional example to demonstrate such a case and compare the two infinitesimal sensitivity methods. The genetic oscillator, adapted from Vilar *et al.*,¹⁴ consists of nine genes, mRNAs, and proteins which interact in a network of 16 biochemical



FIG. 5. Reversible isomerization process with parameters $c_1 = 0.3$, $c_2 = 0.2$, $X_1(0) = 50$, $X_2(0) = 0$. (a) The evolution of the expected value for each species (dashed line) is shown and compared with a representative sample path (solid lines) generated using SSA. (b) Variance and bias of sensitivity estimates for $E[X_1(T)]$ with respect to c_1 at T = 3 (solid) and T = 40 (dashed) obtained from $N = 10^4$ independent samples for all methods. Note the Girsanov estimator (green) is unbiased, and its variance from one ensemble of $N = 10^4$ samples is shown for reference only. The RPD estimates (triangles) and CRP estimates (x) were computed using parameters w and h, respectively, which achieved the identical relative bias indicated on the horizontal axis. Representative w and h values are labeled.

reactions to produce oscillations. The system reactions and parameters of the network are listed in Table I. The initial populations for the gene promoters P_r and P_a were assumed to be 1, with all other species initialized to 0.

The RPD and Girsanov methods were used to evaluate the sensitivity of the expected value of the activator protein, A, at a fixed final time to all system parameters to evaluate the relative effects of each of the parameters. Because the reaction rates can vary by several orders of magnitude, we consider the sensitivities normalized by their respective parameters, i.e., $\frac{\partial E(X(T))}{\partial c_i}c_i$, to allow direct comparison. The results show that the RPD estimator significantly outperforms the Girsanov estimator for evaluating the sensitivity with respect to each parameter in the system (Fig. 6). The RPD estimator converges to the infinitesimal sensitivity with far lower variance than the Girsanov estimator, as indicated by the very tight confidence intervals. In fact, for this problem the RPD estimator achieves variance reductions between 1.4×10^4 (with respect to α_A) and 7 × 10¹¹ (with respect to γ_C) fold lower compared with the Girsanov method. These results are consistent when even larger samples sizes ($N = 10^6$) are considered (data not shown).

TABLE I. Model reactions, propensity functions, and parameters for the genetic oscillator example.

Reactions	Parameter	Value
$P_a \xrightarrow{\alpha_A P_a} P_a + mRNA_a$	α_A	50.0
$P_{a}A \xrightarrow{\alpha_{a} \alpha_{A}} \xrightarrow{P_{a}A} P_{a}A + mRNA_{a}$	α_R	0.01
$P_r \xrightarrow{\alpha_R P_r} P_r + mRNA_r$	eta_A	50.0
$P_{r_A} \xrightarrow{\alpha_r \ \alpha_R \ P_{r_A}} P_{r_A} + mRNA_r$	β_R	5.0
$mRNA_a \xrightarrow{\beta_A \ mRNA_a} mRNA_a + A$	γc	20.0
$mRNA_r \xrightarrow{\beta_R mRNA_r} mRNA_r + R$	γ_A	1.0
$A + R \xrightarrow{\gamma_C A R} C$	θ_A	50.0
$P_a + A \xrightarrow{\gamma_A P_a A} P_{a_A} P_{a_A}$	Ϋ́R	1.0
$P_r + A \xrightarrow{\frac{\gamma_R P_r A}{\gamma_R P_r A}} P_r A$	θ_R	100.0
$A \xrightarrow{\delta_A A} \varnothing$	δ_A	1.0
$R \xrightarrow{\delta_R R} \varnothing$	δ_R	0.2
$mRNA_a \xrightarrow{\delta_{MA} mRNA_a} \varnothing$	δ_{MA}	10.0
$mRNA_r \xrightarrow{\delta_{MR} mRNA_r} \varnothing$	δ_{MR}	0.5
$C \xrightarrow{\delta'_A C} R$	δ'_A	1.0
	α_a	10.0
	α_r	5000

An important point to note is that the sensitivities with respect to all 16 reaction parameters are computed from only N sample paths using either infinitesimal sensitivity method (RPD or Girsanov). In contrast, 16 times as many simulations would be needed for the CRP method with central difference scheme. More generally in high-dimensional problems with p



FIG. 6. Sensitivity analysis for E[A(T)] for the genetic oscillator at T = 5 using $N = 10^4$ samples at parameter values listed in Table I. (a) The sensitivity estimates are shown for the RPD method (w = 0.1) and the Girsanov method. Error bars indicate the 95% confidence intervals of the sensitivity estimates. (b) The estimator variance is shown for the respective methods, with the same coloring as in (a).

parameters, the computational expense of sensitivity analysis can be particularly expensive when using finite difference methods. To state this more precisely, in order to compute the sensitivities with respect to p parameters using FD from Nsamples, one must compute 2Np simulations with a centered FD scheme, or alternatively N(p + 1) simulations with a (higher variance) forward FD scheme. In contrast, both RPD and Girsanov estimators can produce sensitivity estimates with respect to all p parameters using only N simulations. Although each simulation becomes more expensive as pgrows, the computational expense scales linearly with p for both RPD and Girsanov. This additional expense of RPD and Girsanov is typically much less than FD methods, thus making either method preferred over FD when computing sensitivities for large number of parameters.

For this particular example and implementation of the algorithms (code written in C and simulations performed on a 2 GHz Intel Core i7 laptop with 4GB memory), we observed that a single sensitivity sample for one parameter requires 0.0146 s of computation time using CRP, 0.0112 s using Girsanov, and 0.0174 s using RPD. However, it takes an additional 0.0146, 0.0008, and 0.0018 s for each additional parameter sensitivity sample generated using CRP, Girsanov, and RPD, respectively. Thus it is evident that when N samples of only two parameter sensitivities are desired, CRP is already more expensive than the either two methods, and that it becomes even more expensive as p increases. From our results, the Girsanov estimator is the least computationally expensive algorithm, requiring only 75% - 45% of the simulation time required for the same number of samples generated using RPD. However in this example (and indeed, all other examples considered above), this additional computational expense per sample using RPD is more than compensated by the significant variance reductions. In other words, although each simulation takes less time using Girsanov, many more simulations will be required to generate an estimate of the same variance as RPD. The comparison between CRP and RPD is more nuanced due to their dependence on algorithmic parameters that affect variance and bias, but it is clear from the previous examples that the combination of reduced computational expense (for p > 1) and lower variance using RPD can be significant for many problems.

D. Toggle switch

The stochastic genetic toggle switch originally proposed by Gardner *et al.*¹⁵ models reaction kinetics using rational propensity functions rather than the affine (in *c*) propensities that are obtained from networks consisting entirely of elementary chemical reactions. The toggle switch consists of two mutually repressing proteins, \mathcal{U} and \mathcal{V} , which switches between two states in which one species has a large population, while the other has only a few copies present. The populations of species, \mathcal{U} and \mathcal{V} are described by the state vector, X(t, c)= $[x_1 x_2]'$. The reaction network consists of the following four reactions: where the propensity functions a_j and stoichiometry vectors v_i associated with each reaction channel *j* are given by

$$a_{1}(x, c) = \frac{\alpha_{1}}{1 + x_{2}^{\beta}},$$

$$a_{2}(x, c) = x_{1},$$

$$a_{3}(x, c) = \frac{\alpha_{2}}{1 + x_{1}^{\gamma}},$$

$$a_{4}(x, c) = x_{2},$$

$$\nu_{1} = [1 \ 0]', \ \nu_{2} = [-1 \ 0]', \ \nu_{3} = [0 \ 1]', \ \nu_{4} = [0 \ -1]'.$$

For this example, we analyze the system at the nominal parameter set $c = c_0$ with values,

$$\alpha_1 = 50, \quad \alpha_2 = 16, \quad \beta = 2.5, \quad \gamma = 1.$$

All simulations start at t = 0 with initial conditions $x_1 = x_2 = 0$.

We consider sensitivity analysis of the expected value of species \mathcal{U} at a fixed time *T*. We first bring attention to a limitation of both infinitesimal sensitivity methods that arises in this problem. Namely, due to the particular form of the propensities, the partial derivatives $\frac{\partial a_1}{\partial \beta}$ and $\frac{\partial a_3}{\partial \gamma}$ may be singular whenever $x_1 = 0$ or $x_2 = 0$, respectively. As these partial derivatives are used in the numerical RPD algorithm (Sec. III A), the RPD method fails to accurately estimate the sensitivity with respect to these parameters for this problem. The Girsanov algorithm cannot be used either in this case for the same reason (see Appendix A).

However, there are no apparent limitations in estimating the sensitivity $\frac{\partial}{\partial \alpha_1} E[\mathcal{U}(T)]$. The sensitivity at fixed final time T = 5 for this example was computed accurately using the Finite State Projection method.² The exact result (1.19)was compared with the results obtained from 10⁵ samples using the CRP method, the Girsanov method, and the RPD method at various sizes of the regularizing window (Fig. 7). The results show that both the Girsanov and CRP approaches give excellent estimates of the sensitivity with low variances for this large number of samples. In particular, the Girsanov method yields an estimate of 1.25 ± 0.09 , while the CRP method yields an estimate of 1.19 ± 0.04 . The RPD estimates, in contrast, have a variance several orders of magnitude larger than the competing methods even for large w. With the largest regularization window considered (w = 1.0), the RPD gave a rather poor estimate of 2.35 ± 1.96 , which is over 400-fold higher than the Girsanov estimator and 2000-fold higher than the CRP estimator. The results are similar when also considering the sensitivity with respect to α_2 . Thus for this problem both the CRP and Girsanov estimators are better able to provide accurate sensitivity estimates than the RPD estimator.

V. ON THE SELECTION OF THE PARAMETER w

The performance of the RPD algorithm depends on the choice of the window-size algorithmic parameter, w. On one hand, a large w is desirable because it smooths the problem and leads to reduced variance estimates and hence higher efficiency. On the other hand, choosing larger w solves a problem further away from the unregularized sensitivity problem,



FIG. 7. Sensitivity analysis for the toggle switch example, $\frac{\partial}{\partial \alpha_1} E[\mathcal{U}(T)]$ at T = 10 using $N = 10^5$ samples and parameter values listed in the text. (a) Sensitivity estimates computed from the RPD, Girsanov, and CRP methods are compared with the exact result (1.1897), calculated using the finite state projection. Colored bands indicate the 95% confidence intervals of the point estimates for each method. RPD was computed using windows of varying widths between $w = 10^{-4}$ and 10^0 . A perturbation of size h = 0.01 was used for the CRP estimates. (b) Variance of the estimates.

potentially introducing a significant error in the estimate. Therefore, one is faced with the challenge of choosing w that properly compromises between accuracy and efficiency. A similar issue exists with the finite difference methods as well where the perturbation size h plays a role analogous to that of w. The authors are unaware of any standard practice for effectively choosing h for stochastic finite difference sensitivity estimators. Although there are some reasonable guidelines for choosing the parameter to balance Taylor series truncation error with round-off error in the case of deterministic systems,¹⁶ this choice is obfuscated in stochastic sensitivity analysis as smaller perturbations lead to increased variance and hence larger sampling error. Results deriving the perturbation size that optimizes asymptotic convergence in sample size N exist (see, e.g., Chap. VII in Ref. 9, and references therein), but are of limited value in practice as they involve constants that are either unknown or require information about higher order derivatives, which can be exceedingly expensive computationally.

The numerical results suggest heuristic strategies that may be used to select w. For example, it was seen that very large w could be used while maintaining low bias once the system is near stationarity (Figs. 2 and 5). This is not surprising since near stationarity, time averages equal ensemble averages. Smaller w must be used, in contrast, when the system is in highly transient regimes. This suggests that a small number of pilot trajectories could be generated to evaluate how f(X(t)) changes within a time interval around the fixed time T, and to check whether a particular choice of w seems reasonable. Adaptive methods could also be implemented to adjust the window size if the sample variances exceed some specified tolerance. Under mild regularity assumptions it is easy to show via Taylor expansions that the bias of the RPD method vanishes as $O(w^2)$ as $w \to 0$. This is analogous to the $O(h^2)$ bias when using central finite differences. Under the assumption that the bias is $O(w^2)$ (or $O(h^2)$ for finite differences), one may use multiple values of w (or h) to estimate and remove the bias using regression. The authors intend to explore the numerical efficiency of this idea in a future work.

In some situations the regularized problem (11) may be of equal or more interest than the unregularized one. For instance, instead of considering the sensitivity of the probability that a gene is on or off at a single instant T in time, it may be more natural to consider the sensitivity of the fraction of time it is turned on during an interval of time [T - w, T + w] that is of interest biologically. Additionally, experimental constraints in how precisely time course data can be obtained may actually make it more useful to consider small intervals of time that are consistent with the experimental system which is being modeled than only considering a single time instance.

VI. CONCLUSIONS

In this work, we developed the RPD method as an efficient Monte Carlo estimator for the infinitesimal sensitivity analysis of discrete stochastic chemical networks. The RPD sensitivity estimator was constructed by using the random time change description of continuous-time Markov processes and by considering a regularized function to make pathwise differentiation possible for many problems in which a naive implementation of PD would otherwise not be valid. We discussed the conditions for commuting differentiation and expectation and hence assuring validity of PD methods and showed that the RPD method is expected to satisfy these conditions for many typical problems.

We demonstrated the applicability of RPD using several numerical examples, and found that RPD accurately and efficiently estimated the sensitivities of a variety of path functionals of interest.

Comparison of RPD with alternative sensitivity methods using likelihood ratios (Girsanov) and finite difference quotients (common reaction path) showed that RPD estimates typically have lower variance, with RPD variances often orders of magnitude lower than alternative methods.

In fact for the particular example of a genetic oscillator, variance reductions of up to 11 orders of magnitude were observed, indicating that computational expense can be reduced dramatically by using RPD. We also illustrated some limitations of the RPD estimator by presenting a problem for which the RPD estimator is not well suited, demonstrating that for certain problems the RPD estimator is no longer valid and an alternative method must be used.

Though the RPD method is applicable to a large class of sensitivity problems, it may not be applicable to typical problems involving random times, such as the expected time to reach a certain state. Future work includes developing proper regularization techniques to deal with random times, and developing strategies for automated selection of the algorithm parameter.

ACKNOWLEDGMENTS

The authors wish to acknowledge financial support from the National Science Foundation under Grant Nos. NSF-ECCS-0835847 and NSF-ECCS-0802008 and from the National Institute of Health under Grant No. R01-GM04983.

APPENDIX A: LIKELIHOOD RATIO METHOD

In this Appendix, we describe the GLR method in the context of chemical kinetics. This method was introduced in the context of chemical kinetics in Ref. 6 for systems, where propensity functions are affine in parameters. A more general related formula for the Girsanov transformation that considers the ratios of the probability densities p(x; c) and p(x; c') for two different parameters c and c' may be found in Ref. 17; our presentation here does not assume any particular form for the propensities except that they are continuously differentiable with respect to the parameter c. An alternative derivation of the Girsanov method without restrictions on the form of propensity functions may also be found in Ref. 18.

Recall that the required sensitivity z may be written as

$$z = E\left(f(X)W\right),\tag{A1}$$

where the random (weight) variable W is given by

$$W = \frac{\partial}{\partial c} \ln p(X; c).$$

It can be shown that the variable W can be written as a sum $W = \sum_{j=1}^{M} W_j$, where W_j is the random weight corresponding to reaction channel j and W_i is given by

$$W_j = \int_{[0,T]} \frac{\frac{\partial a_j}{\partial c}(X(s-))}{a_j(X(s-))} dR_j(s) - \int_0^T \frac{\partial a_j}{\partial c}(X(s)) ds.$$
(A2)

In (A2) $R_i(s)$ is the process that counts the number of firings of the *j*th channel during [0, s]. We note that the first integral is a Lebesgue-Stieltjes integral which simply sums the values of $\frac{\partial a_j}{\partial c}(X(s-))/a_j(X(s-))$ corresponding to the finitely many values of s at which a firing of the *j*th reaction occurs. Thus if $t_i^1, t_i^2, \ldots, t_i^{m_j}$ are the times at which *j*th reaction fires, then

$$\int_{[0,T]} \frac{\frac{\partial a_j}{\partial c}(X(s-))}{a_j(X(s-))} dR_j(s) = \sum_{l=1}^{m_j} \frac{\frac{\partial a_j}{\partial c}(X(t_j^l-))}{a_j(X(t_j^l-))}.$$

In other words, each time the reaction channel *j* fires, this integral is increased by the partial derivative of the natural logarithm of the propensity function of the reaction channel *j* evaluated just prior to the firing.

To estimate z via Monte Carlo simulations, one generates N number of independent trajectories $X^{(i)}$, i = 1, ..., N and along with each trajectory computes the corresponding $W^{(i)}$. Then the sensitivity z is estimated by

$$\hat{z} = \frac{1}{N} \sum_{i=1}^{N} f(X^{(i)}) W^{(i)}.$$
 (A3)

An algorithm for computing sensitivities via the GLR method is given below. This algorithm uses the Gillespie direct method SSA (Ref. 1) and is nearly identical to that presented previously by Plyasunov and Arkin.⁶ However, the following is modified to accommodate propensities which are not necessarily affine in the parameters. In the following, we assume that calls to the function rand() return independent uniform random numbers in the interval (0, 1).

Girsanov likelihood ratio sensitivity algorithm

- 1. Specify x_0, T_f, N
- 2. for k = 1 to N do
- 3. Initialization: Set t = 0, $X = x_0$, $W_i = 0$ for $j = 1, \dots, M$
- while $t < T_f$ do 4.

- 5. Calculate a_j and $\frac{\partial a_j}{\partial c}$ for j = 1, ..., M6. Set $a_0 = \sum_{j=0}^{M} a_j$ 7. Generate $u_1 = \text{rand}()$ and $u_2 = \text{rand}()$
- 8. Set $\tau = -\ln(u_1)/a_0$
- 9. Set $j^* \in [1, ..., M]$ to the index for which 10. Set $\Delta W_{j^*} = \frac{\partial a_{j^*}}{\partial c} \frac{1}{a_{j^*}} - \tau \frac{\partial a_{j^*}}{\partial c}$
- For $j \neq j^*$, set $\Delta W_j = -\tau \frac{\partial a_j}{\partial c}$ 11.
- Update $W_j \leftarrow W_j + \Delta W_j$ for $j = 1, \dots, M$ 12.
- Update $t \leftarrow t + \tau$, $X \leftarrow X + \nu_{i^*}$ 13.
- 14. end while
- Set $W = \sum_{j=1}^{M} W_j$ 15.
- 16. **return** weighted sensitivity estimate $Z_k = f(X)W$
- 17. end for
- Compute sensitivity from sample mean $\hat{z} = \frac{1}{N} \sum_{k=1}^{N} Z_k$ 18.

APPENDIX B: VALIDITY OF DIFFERENTIATION INSIDE EXPECTATION

In this Appendix, we discuss the validity of differentiation inside expectation. Let $Y(c, \omega)$ be a function of the random element ω and the parameter c. We state an important lemma.9

Lemma 1: Let c_0 be a specified value of c. Suppose the following hold:

- 1. For a set of ω with probability one, $Y(c, \omega)$ is differentiable with respect to c at $c = c_0$.
- 2. There exists an interval (c_l, c_u) containing $c = c_0$ (independent of ω) on which $Y(c, \omega)$ is Lipschitz (in c) for a set of ω with probability one, with constant K which may depend on ω . In other words, for any c_1 , c_2 in the interval (c_l, c_u) , the following holds:

$$|Y(c_1, \omega) - Y(c_2, \omega)| \le K(\omega)|c_1 - c_2|.$$

- 3. E(K) is finite.
- 4. $E(|Y(c, \omega)|)$ is finite for all c in (c_l, c_u) .

Then the following commutativity holds:

$$\frac{d}{dc}\Big|_{c=c_0} E(Y(c)) = E\left(\frac{d}{dc}\Big|_{c=c_0} Y(c)\right).$$

Suppose we are interested in E(f(X(T))), where f is some function of the state. When using the RPD method, one obtains the function $Y(c, \omega)$ given by

$$Y(c) = \frac{1}{2w} \int_{T-w}^{T+w} f(X(t,c))dt = \frac{1}{2w} \sum_{i=1}^{\infty} \Delta_i(c) \\ \times \int_{T-w}^{T+w} H(t-T_i(c))dt.$$
(B1)

We observe that condition 1 of Lemma 1 holds as explained earlier and the derivative of *F* at $c = c_0$ is given by Eq. (14). We note that condition 1 holds in the nonregularized case *Y*(*c*) = *f*(*X*(*T*; *c*)) as well.

Now we shall show that the condition 2 of Lemma 1 holds under the following important assumption on the chemical system.

Assumption 1: At every possible state x, if $a_j(x) > 0$ and $a_k(x) > 0$ for $j \neq k$, then $a_k(x + v_j) > 0$ and $a_j(x + v_k) > 0$.

Under assumption 1 it follows that the jump times $T_i(c)$ are continuous in c.⁴

First, under assumption 1 we claim that Y(c) given by (B1) is a continuous function of c for each realization ω . To see this, first we observe that for each i, the integral $\int_{T-w}^{T+w} H(t-T_i(c)) dt$ is continuous in c because $T_i(c)$ is continuous in c. The terms $\Delta_i(c)$ are only piecewise continuous in c for each fixed i. However, a jump discontinuity in $\Delta_i(c)$ occurs when $T_i(c) = T_{i+1}(c)$ or $T_i(c) = T_{i-1}(c)$, thus two or more jump times coincide. In this case, two or more jumps occur at the same time and the sum of Δ_i for these jumps remains continuous in c. As a result of this crossover phenomenon the sum in the above equation remains continuous in c. Second, we observe that for a given realization ω , in any bounded interval of c values, the derivative $\frac{\partial}{\partial c}Y(c)$ exists for all c except at finitely many points where two or more $T_i(c)$ crossover (these are the points where $\frac{\partial T_i}{\partial c}$ fail to exist). Third from the form of Eqs. (8) and (9), we see that for a given re-alization ω , the derivatives $\frac{\partial T_i}{\partial c}$ and hence $\frac{\partial}{\partial c}Y(c)$ are bounded functions on any bounded interval of c values bounded away from c = 0. Thus for any given bounded interval of c values bounded away from c = 0 and for any realization, with probability 1, the function Y(c) is a Lipschitz continuous function whose Lipschitz constant may depend on the realization and on the interval.

Thus we have established that condition 2 of Lemma 1 holds under assumption 1. In the nonregularized case, we note that condition 2 does not hold. This is because, given any interval (c_l, c_u) containing c_0 , the set of ω (i.e., the set of trajectories) for which $Y(c, \omega) = f(X(T, \omega; c))$ does not have a discontinuity in (c_l, c_u) is nonzero.

It is harder to prove that conditions 3 and 4 of Lemma 1 hold (especially condition 3 is harder). However, these are typically expected to hold.

- ¹D. Gillespie, J. Phys. Chem. 81, 2340 (1977).
- ²B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
- ³R. Gunawan, Y. Cao, L. Petzold, and F. Doyle, Biophys. J. 88, 2530 (2005).
- ⁴M. Rathinam, P. Sheppard, and M. Khammash, J. Chem. Phys. **132**, 034103 (2010).
- ⁵S. Dandach and M. Khammash, PLOS Comput. Biol. 6 (2010).
- ⁶S. Plyasunov and A. Arkin, J. Comput. Phys. 221, 724 (2007).
- ⁷Y. Ho and X. Cao, J. Optim. Theory Appl. 40, 559 (1983).
- ⁸P. Glasserman, Gradient Estimation Via Perturbation Analysis (Springer, New York, 1990).
- ⁹S. Asmussen and P. Glynn, *Stochastic Simulation: Algorithms and Analysis* (Springer, New York, 2007).
- ¹⁰S. Ethier and T. Kurtz, *Markov Processes: Characterization and Conver*gence (Wiley, New York, 1986).
- ¹¹P. Glynn, Commun. ACM 33, 75 (1990).
- ¹²R. Rubinstein and A. Shapiro, Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method (Wiley, New York, 1993).
- ¹³M. Rathinam and H. E. Samad, J. Comput. Phys. 224, 897 (2007).
- ¹⁴J. Vilar, H. Kueh, N. Barkai, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. 99, 5988 (2002).
- ¹⁵T. Gardner, C. Cantor, and J. Collins, Nature (London) 403, 339 (2000).
- ¹⁶B. P. Flannery, W. H. Press, S. A. Teukolsky, and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1992).
- ¹⁷P. Bremaud, Point Processes and Queues: Martingale Dynamics (Springer Verlag, New York, 1981).
- ¹⁸P. B. Warren and R. J. Allen, "Trajectory reweighting schemes in stochastic modeling: Determination of steady-state parameter sensitivity coefficients" (unpublished).