

A general-purpose sensitivity method for multiple simultaneous parameter perturbations in stochastic reaction networks

David F. Anderson*

Jingyi Ma[†]

August 23, 2026

Abstract

Stochastic reaction networks are continuous-time Markov chain models for interacting populations, with applications in biochemistry, epidemiology, ecology, and related areas. We study finite-difference sensitivity estimation when a single estimator requires several nearby parameterized paths. Existing variance-reducing couplings are typically pairwise, so that repeated use is either inefficient or requires application-specific choices in multi-path settings. We introduce the multi-path stacked coupling (MSC), a space-time Poisson construction that jointly generates any finite collection of parameterized paths. Each pairwise marginal of MSC has the same law as the corresponding split coupling pair, allowing existing variance bounds to transfer directly; in finite-state settings, we also obtain first-order expansions for the mean and second moment of finite-difference numerators. We apply MSC in three settings of practical importance: estimating many first derivatives simultaneously, estimating a single first derivative using a wider finite-difference stencil, and estimating higher-order derivatives. Numerical experiments on a processive phosphorylation network demonstrate strong performance in each of the three application areas considered, consistent with the theoretical advantages of MSC: across all three applications, MSC achieves the smallest root mean square error (RMSE) among the methods considered over the tested computational budgets.

Keywords: stochastic reaction networks; multi-path stacked coupling; parametric sensitivity analysis; finite-difference estimators; coupling methods; variance reduction; Monte Carlo simulation.

MSC: Primary 65C40; Secondary 60J27, 65C05, 60J22, 92C40.

1 Introduction

Stochastic reaction networks, commonly modeled as continuous-time Markov chains on nonnegative integer lattices, provide a natural framework for systems of interacting populations. They are widely used for the quantitative study of intracellular biochemical systems, with applications including gene regulatory networks [10, 13, 23], viral kinetics [29], signal transduction pathways [2, 20], and metabolic networks [25]. The same continuous-time Markov jump structure also appears in stochastic population, ecological, and epidemic models [1, 9]. Such models depend on a number of parameters, and it is therefore natural to ask how outputs of interest respond to perturbations of those parameters. Parametric sensitivity analysis quantifies this dependence through derivatives of expected model outputs with respect to model parameters, such as the expected abundance of a

*Department of Mathematics, University of Wisconsin-Madison, USA. anderson@math.wisc.edu.

[†]Department of Mathematics, University of Wisconsin-Madison, USA. jma276@wisc.edu

reactant at a fixed time or the probability of a specified model behavior [3, 19]. Such information can help reveal system behavior, identify important parameters and reactions [18], and guide tasks such as parameter estimation and experimental design [18, 26, 30].

The computational study of parametric sensitivities for stochastic reaction networks has developed into a substantial research area, and a number of numerical approaches have been proposed. For example, likelihood-ratio gradient estimators for both discrete-time and continuous-time systems were introduced in a general setting by Glynn [17] and later extended to the stochastic reaction network setting by Arkin and Plyasunov [26]. Pathwise-differentiation methods have also been studied. In particular, Sheppard et al. proposed a regularized pathwise method that is applicable to a subset of stochastic chemical models [28], whereas Anderson and Wolf proposed a hybrid pathwise differentiation method that is unbiased and applicable to a broad class of stochastic chemical models [33]. The hybrid pathwise method is close in spirit to the multi-level Monte Carlo methods developed by Giles [16] and extended to the stochastic reaction network setting by Anderson and Higham [4]. Multiple coupling-based methods built on finite-difference estimators have also been developed to improve efficiency, including the classical common random number method [22], the common reaction path method [27], split coupling (also called coupled finite differences) [3, 5], and the stacked coupling method [8]. Most of these coupling methods are designed for pairwise constructions, though there are important special-purpose extensions to more complicated settings, such as CRP-style shared-path constructions and the double-coupled finite difference method for second derivatives [27, 32]. Despite their differences, these approaches share a common goal: reducing estimator variance in order to lower the overall computational cost of sensitivity estimation, especially in settings where such calculations appear inside optimization routines [31].

Most existing methods are naturally formulated around a single perturbation or a pairwise construction. However, many important computational tasks require several nearby parameter values to be perturbed simultaneously. Perhaps the clearest example is the simultaneous estimation of $K > 1$ first derivatives, where the most straightforward extension of the existing methods is simply to repeat the basic algorithm K times. Similar issues arise when wider finite-difference stencils are used for the estimation of a single first derivative, or when higher-order derivatives are sought. This leads to the central question of the present paper:

Motivating question: Is it possible to exploit the structure of stochastic reaction networks to design efficient couplings for situations in which multiple nearby parameter values are perturbed simultaneously?

In this paper, we address this question by extending the stacked coupling method of Anderson and Yuan [8] to a genuinely multi-path setting. The resulting method, which we term the *multi-path stacked coupling* (MSC), provides a general-purpose coupling framework for finite-difference sensitivity estimation in stochastic reaction networks when multiple nearby parameter values are required simultaneously. Thus, our goal is not merely to handle one additional special case, but rather to provide a clean organizational framework that applies across several natural computational settings.

To keep the discussion concrete, we work throughout with the standard stochastic model for a chemical reaction network: a continuous-time Markov chain $X^\theta = \{X^\theta(t) : t \geq 0\}$ on $\mathbb{Z}_{\geq 0}^d$, parameterized by $\theta \in \mathbb{R}^K$, with R reaction channels, reaction vectors ζ_ℓ , and intensities λ_ℓ^θ , $\ell \in \{1, \dots, R\}$. This process has infinitesimal generator

$$\mathcal{L}^\theta f(x) = \sum_{\ell=1}^R \lambda_\ell^\theta(x) (f(x + \zeta_\ell) - f(x)),$$

and its forward equation, often termed the chemical master equation, is

$$\frac{d}{dt}p^\theta(t, x) = \sum_{\ell=1}^R \lambda_\ell^\theta(x - \zeta_\ell) p^\theta(t, x - \zeta_\ell) - \sum_{\ell=1}^R \lambda_\ell^\theta(x) p^\theta(t, x),$$

where $p^\theta(t, x) = \mathbb{P}(X^\theta(t) = x)$; see, for example, [6, 7].

Under stochastic mass-action kinetics, if reaction ℓ consumes $\nu_{i\ell}$ molecules of species i , then its intensity has the form

$$\lambda_\ell^\theta(x) = \kappa_\ell \prod_{i=1}^d (x_i)^{\nu_{i\ell}},$$

where $(x)_m = x(x-1)\dots(x-m+1)$ and κ_ℓ is the corresponding rate constant. In this case, θ could be the vector of rate constants, in which case $K = R$.

For a quantity of interest of the form

$$g(\theta) = \mathbb{E}[f(X^\theta(T))],$$

where f is a function yielding an observable of interest, we are interested in computational settings where finite-difference approximations require the simultaneous generation of several nearby parameterized processes, such as

$$X^\theta, \quad X^{\theta+\varepsilon_1 e_1}, \quad \dots, \quad X^{\theta+\varepsilon_K e_K},$$

where e_i is the i th canonical basis vector in $\mathbb{R}^K$ and each ε_i is a small perturbation size. The key point is that existing coupling methods involve different tradeoffs in such settings. Repeated application of the split coupling yields favorable pairwise variance behavior, but it requires separate coupled pairs to be generated for each perturbation direction. In contrast, the common reaction path constructions allow randomness to be shared across many perturbed systems simultaneously, thereby reducing the number of simulated trajectories required per Monte Carlo replication, though typically at the cost of weaker variance control. The MSC framework is designed to combine the main advantages of these two viewpoints: it generates all required nearby paths simultaneously, with the same favorable path-count structure as the common reaction path method, while retaining the strong pairwise coupling behavior associated with the split coupling.

We highlight three main application areas, which will be developed further in Section 2 before being numerically studied in Sections 5.2, 5.3, and 5.4.

1. Simultaneous computation of many first derivatives.

This is the clearest and most important application. Suppose one wishes to estimate all K first-order sensitivities simultaneously. In a single Monte Carlo replication, repeated use of split coupling requires the generation of $2K$ sample paths, since one coupled nominal/perturbed pair must be generated for each derivative direction. In contrast, the common reaction path method requires only $K + 1$ sample paths per replication: one nominal path together with the K perturbed paths. The MSC framework has this same $K + 1$ scaling per replication, while retaining the favorable pairwise variance properties associated with the split coupling. Thus, when many derivatives are required simultaneously, MSC combines the main organizational advantage of CRP with the variance reduction of split coupling.

2. Higher-order finite-difference estimation of a single first derivative.

Even when estimating a single first derivative, improved finite-difference formulas often require wider stencils and hence multiple nearby parameter values. MSC provides a systematic way to couple the resulting collection of paths within each Monte Carlo replication. This makes it possible to combine the reduced bias of higher-order finite-difference formulas with favorable variance behavior, and hence to obtain improved overall performance.

3. Higher-order derivatives.

For second-, third-, and higher-order derivatives, as well as mixed partial derivatives, the relevant finite-difference formulas can involve several nearby parameter values, and constructing effective couplings by hand quickly becomes cumbersome. MSC provides a clean, out-of-the-box framework for such settings. While specially tailored methods may perform better in particular problems, MSC offers a broadly applicable general approach that avoids the need for problem-specific coupling constructions.

The remainder of the paper is organized as follows. In Section 2, we make precise the three representative finite-difference settings described above, each of which naturally involves multiple nearby parameterized sample paths. In Section 3, we introduce the multi-path stacked coupling, present the corresponding simulation algorithm, and establish its main theoretical properties. In Section 4, we give a unified bias-variance analysis and derive the optimal RMSE scaling rates for the different path-generation methods. Finally, in Section 5, we introduce the processive phosphorylation network used throughout the numerical study and apply the different methods to the three finite-difference settings considered in Section 2.

2 Finite-difference settings requiring multiple paths

With $g(\theta) = \mathbb{E}[f(X^\theta(T))]$ as in the introduction, we now make precise the three representative finite-difference settings highlighted above. In each case, the estimator involves multiple nearby parameterized sample paths, and the purpose of this brief section is simply to record the corresponding path collections that arise.

We begin with the standard forward finite-difference approximation for a first-order sensitivity. Under the usual smoothness assumptions on g , for $j \in \{1, \dots, K\}$,

$$\frac{\partial}{\partial \theta_j} g(\theta) = \frac{\mathbb{E}[f(X^{\theta+\varepsilon_j e_j}(T))] - \mathbb{E}[f(X^\theta(T))]}{\varepsilon_j} + O(\varepsilon_j), \quad (1)$$

where e_j is the j th unit vector in $\mathbb{R}^K$ and ε_j is the perturbation size. The corresponding Monte Carlo estimator is

$$\hat{D}_j(\theta) = \frac{1}{N} \sum_{i=1}^N d_{[i]}(\varepsilon_j), \quad d_{[i]}(\varepsilon_j) = \frac{f(X_{[i]}^{\theta+\varepsilon_j e_j}(T)) - f(X_{[i]}^\theta(T))}{\varepsilon_j}, \quad (2)$$

where $(X_{[i]}^\theta, X_{[i]}^{\theta+\varepsilon_j e_j})$ is the i th coupled pair of sample paths.

We now record three representative settings in which finite-difference estimators require more than a single coupled pair.

1. Simultaneous computation of many first derivatives.

If θ is multi-dimensional and one wishes to estimate all first-order sensitivities, then for each Monte Carlo replication one is naturally led to the collection of sample paths

$$\left(X_{[i]}^\theta, X_{[i]}^{\theta+\varepsilon_1 e_1}, \dots, X_{[i]}^{\theta+\varepsilon_K e_K}\right). \quad (3)$$

Thus, although each individual derivative is approximated by a two-point finite difference, the overall computation requires a multi-path construction.

2. Higher-order finite-difference estimation of a single first derivative.

Even when only a single first derivative is sought, higher-order finite-difference formulas may require wider stencils.

The standard centered approximation

$$\frac{\partial}{\partial \theta_j} g(\theta) = \frac{\mathbb{E}[f(X^{\theta+\varepsilon_j e_j}(T))] - \mathbb{E}[f(X^{\theta-\varepsilon_j e_j}(T))]}{2\varepsilon_j} + O(\varepsilon_j^2) \quad (4)$$

requires, for each Monte Carlo replication, the two-path collection

$$\left(X_{[i]}^{\theta-\varepsilon_j e_j}, X_{[i]}^{\theta+\varepsilon_j e_j}\right). \quad (5)$$

A higher-order alternative is the fourth-order centered approximation

$$\begin{aligned} \frac{\partial}{\partial \theta_j} g(\theta) = \frac{1}{12\varepsilon_j} \Big[& -\mathbb{E}[f(X^{\theta+2\varepsilon_j e_j}(T))] + 8\mathbb{E}[f(X^{\theta+\varepsilon_j e_j}(T))] \\ & - 8\mathbb{E}[f(X^{\theta-\varepsilon_j e_j}(T))] + \mathbb{E}[f(X^{\theta-2\varepsilon_j e_j}(T))] \Big] + O(\varepsilon_j^4) \end{aligned} \quad (6)$$

which requires, for each Monte Carlo replication, the four-path collection

$$\left(X_{[i]}^{\theta+2\varepsilon_j e_j}, X_{[i]}^{\theta+\varepsilon_j e_j}, X_{[i]}^{\theta-\varepsilon_j e_j}, X_{[i]}^{\theta-2\varepsilon_j e_j}\right). \quad (7)$$

Thus, wider stencils for a single first derivative also naturally lead to a multi-path coupling problem.

3. Higher-order derivatives.

A similar issue arises for higher-order sensitivities. For example, a centered finite-difference approximation for the third derivative is

$$\begin{aligned} \frac{\partial^3}{\partial \theta_j^3} g(\theta) = \frac{1}{2\varepsilon_j^3} \Big[& \mathbb{E}[f(X^{\theta+2\varepsilon_j e_j}(T))] - 2\mathbb{E}[f(X^{\theta+\varepsilon_j e_j}(T))] \\ & + 2\mathbb{E}[f(X^{\theta-\varepsilon_j e_j}(T))] - \mathbb{E}[f(X^{\theta-2\varepsilon_j e_j}(T))] \Big] + O(\varepsilon_j^2). \end{aligned} \quad (8)$$

This approximation again requires, for each Monte Carlo replication, the four-path collection

$$\left(X_{[i]}^{\theta+2\varepsilon_j e_j}, X_{[i]}^{\theta+\varepsilon_j e_j}, X_{[i]}^{\theta-\varepsilon_j e_j}, X_{[i]}^{\theta-2\varepsilon_j e_j}\right). \quad (9)$$

More generally, higher-order derivatives and mixed partial derivatives can require increasingly complicated collections of nearby parameterized sample paths.

The common feature in (3), (7), and (9) is that the relevant finite-difference estimators naturally involve multiple nearby sample paths. This is precisely the setting for which we introduce the multi-path stacked coupling in the next section.

3 The multi-path stacked coupling framework

We now introduce the multi-path stacked coupling framework. The section has two main parts. First, in Section 3.1, we define the MSC construction, give the corresponding simulation algorithm, and provide a simple illustration of the stacked Poisson-space representation. Second, in Section 3.2, we establish the theoretical properties of MSC that will be used in the applications below. In particular, we show that the relevant pairwise marginals under MSC have the same law as the corresponding split-coupled pairs. We then use this pairwise equivalence, together with existing variance bounds for the split coupling, to obtain the variance-scaling results needed for the finite-difference estimators studied later. We also give a finite-state refinement that provides leading-order expansions for the mean and second moment of the corresponding finite-difference numerators.

3.1 Multi-path stacked coupling

To define the multi-path stacked coupling method, we first recall a space-time Poisson representation of a stochastic reaction network, which is an alternative to Kurtz's random-time-change representation [21]. Let $\mathcal{N}$ denote a unit-rate Poisson point process on $[0, \infty) \times [0, \infty)$. For a fixed parameter vector $\theta \in \mathbb{R}^K$, consider a reaction network with reaction channels indexed by $\ell \in \{1, \dots, R\}$, reaction vectors ζ_ℓ , and intensities $\lambda_\ell^\theta(x)$. Define

$$q_\ell^\theta(t) = \sum_{r=1}^{\ell} \lambda_r^\theta(X^\theta(t)), \quad q_0^\theta(t) = 0.$$

Then X^θ may be represented as the solution to

$$X^\theta(t) = X^\theta(0) + \sum_{\ell=1}^R \zeta_\ell \int_{[0,t] \times [0,\infty)} 1_{[q_{\ell-1}^\theta(s-), q_\ell^\theta(s-)+\lambda_\ell^\theta(X^\theta(s-))]}(x) \mathcal{N}(ds \times dx). \quad (10)$$

This representation gives a continuous-time Markov chain with generator

$$\mathcal{L}^\theta f(x) = \sum_{\ell=1}^R \lambda_\ell^\theta(x) (f(x + \zeta_\ell) - f(x)),$$

and hence has the same law as the standard stochastic reaction network with parameter θ .

We now apply this representation to a finite collection of parameter vectors. Let

$$\theta_0, \theta_1, \dots, \theta_M \in \mathbb{R}^K,$$

where K is the dimension of the parameter space and $M+1$ is the number of processes to be constructed simultaneously. In the simultaneous first-derivative application, one could have $M = K$, with $\theta_0 = \theta$ and $\theta_j = \theta + \varepsilon_j e_j$ for $j \in \{1, \dots, K\}$. However, the MSC construction itself applies to any finite collection of parameter vectors.

Equation (10) immediately suggests a coupling of the processes $X^{\theta_0}, X^{\theta_1}, \dots, X^{\theta_M}$: one may construct all processes with the same realization of the Poisson point process $\mathcal{N}$, while allowing each process to use its own cumulative intensity intervals. That is, for each $j \in \{0, \dots, M\}$, we could define

$$X^{\theta_j}(t) = X^{\theta_j}(0) + \sum_{\ell=1}^R \zeta_\ell \int_{[0,t] \times [0,\infty)} 1_{[q_{\ell-1}^{\theta_j}(s-), q_\ell^{\theta_j}(s-)+\lambda_\ell^{\theta_j}(X^{\theta_j}(s-))]}(x) \mathcal{N}(ds \times dx). \quad (11)$$

This common- $\mathcal{N}$ construction is a valid coupling method in its own right. However, it has an important limitation: the same Poisson point may fall in different reaction-channel intervals for different parameter values. Thus, simultaneous jumps need not occur through the same reaction channel [8].

We now develop the multi-path stacked coupling, which fixes this issue by assigning a common region of Poisson space to each reaction channel. For the finite collection of parameter values $\{\theta_j : j = 0, \dots, M\}$, define the channel-wise envelope rates

$$\bar{\lambda}_\ell(s) = \sup_{0 \leq j \leq M} \lambda_\ell^{\theta_j}(X^{\theta_j}(s)), \quad \bar{q}_\ell(s) = \sum_{r=1}^{\ell} \bar{\lambda}_r(s), \quad \bar{q}_0(s) = 0.$$

In the stacked construction, reaction channel ℓ is assigned the vertical strip

$$[\bar{q}_{\ell-1}(s), \bar{q}_{\ell-1}(s) + \bar{\lambda}_\ell(s)],$$

whose height is $\bar{\lambda}_\ell(s)$. Within this strip, the process X^{θ_j} accepts a Poisson point only if that point falls in the subinterval of height $\lambda_\ell^{\theta_j}(X^{\theta_j}(s))$. Consequently, if several processes jump at the same Poisson point, then they necessarily jump through the same reaction channel ℓ .

The multi-path stacked coupling is therefore defined, for $j \in \{0, \dots, M\}$, by

$$X^{\theta_j}(t) = X^{\theta_j}(0) + \sum_{\ell=1}^R \zeta_\ell \int_{[0,t] \times [0,\infty)} \mathbf{1}_{\left[\bar{q}_{\ell-1}(s-), \bar{q}_{\ell-1}(s-) + \lambda_\ell^{\theta_j}(X^{\theta_j}(s-))\right)}(x) \mathcal{N}(ds \times dx), \quad (12)$$

where $\bar{q}_\ell$ and $\bar{\lambda}_\ell$ are as defined above.

We next turn from the stochastic representation (12) to the precise numerical algorithm used to simulate it. For simplicity, and as in the applications below, we state the algorithm for the case in which all parameterized processes have the same fixed initial condition x . At each step, the channel-wise envelope rates $\bar{\lambda}_\ell$ are computed from the current states of all paths, a candidate Poisson point is generated under the total envelope rate, and each path accepts the point according to its own intensity within the selected reaction channel.

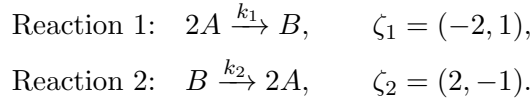
Algorithm 1 Multi-path stacked coupling algorithm

Fix parameter vectors $\theta_j, j \in \{0, \dots, M\}$, a terminal time $T > 0$, and a common initial condition x . Initialize $t = 0$ and $X^{\theta_j} = x$ for all $j \in \{0, \dots, M\}$. Repeat the following steps while $t < T$. All random variables generated below are independent of those generated previously.

- 1: Compute all intensities $\lambda_\ell^{\theta_j}(X^{\theta_j})$ for all $\ell \in \{1, \dots, R\}$ and $j \in \{0, \dots, M\}$. Then compute all $\bar{q}_\ell = \sum_{r=1}^\ell \bar{\lambda}_r$ with $\bar{\lambda}_r = \sup_j \left\{ \lambda_r^{\theta_j}(X^{\theta_j}) \right\}$.
- 2: Generate a unit exponential random variable τ and set $\Delta = \frac{\tau}{\bar{q}_R}$. (If $\bar{q}_R = 0$, set $\Delta = \infty$.)
- 3: **if** $t + \Delta > T$ **then**
- 4: Set $X^{\theta_j}(u) = X^{\theta_j}$ for all $j \in \{0, \dots, M\}$ and $u \in [t, T]$.
- 5: Set $t = T$ and end the loop.
- 6: **else**
- 7: Set $X^{\theta_j}(u) = X^{\theta_j}$ for all $j \in \{0, \dots, M\}$ and $u \in [t, t + \Delta]$.
- 8: Generate a uniform random variable ξ on $[0, 1]$.
- 9: Find $\mu \in \{1, \dots, R\}$ for which $\frac{\bar{q}_{\mu-1}}{\bar{q}_R} \leq \xi < \frac{\bar{q}_\mu}{\bar{q}_R}$.
- 10: **for all** $j \in \{0, \dots, M\}$ **do**
- 11: **if** $\frac{\bar{q}_{\mu-1}}{\bar{q}_R} \leq \xi < \frac{\bar{q}_{\mu-1} + \lambda_\mu^{\theta_j}(X^{\theta_j})}{\bar{q}_R}$ **then**
- 12: $X^{\theta_j} \leftarrow X^{\theta_j} + \zeta_\mu$.
- 13: **end if**
- 14: **end for**
- 15: Set $t \leftarrow t + \Delta$ and return to Step 1.
- 16: **end if**

We close this subsection with a simple example illustrating how the stacked Poisson-space construction determines which processes accept a given Poisson point.

Example 3.1. We consider a reaction network with two species and two reactions,



For a parameter vector $\theta = (k_1, k_2)$, the intensity functions are

$$\lambda_1^\theta(x_A, x_B) = k_1 x_A (x_A - 1), \quad \lambda_2^\theta(x_A, x_B) = k_2 x_B.$$

We compare three parameter vectors,

$$\theta_0 = (20, 20), \quad \theta_1 = (25, 25), \quad \theta_2 = (30, 30),$$

so that, for example, both rate constants are equal to 20 for X^{θ_0} , both are equal to 25 for X^{θ_1} , and both are equal to 30 for X^{θ_2} . We use the common initial condition

$$X^{\theta_0}(0) = X^{\theta_1}(0) = X^{\theta_2}(0) = (20, 200).$$

To make the construction explicit, suppose that the underlying space-time Poisson point process has two relevant points, one at time $t = 1$ with vertical coordinate 3000, and one at time $t = 2$ with vertical coordinate 14800. We now determine which processes accept these points under the MSC construction.

At the initial state $(20, 200)$, the reaction-1 intensities are

$$\lambda_1^{\theta_0} = 20 \cdot 20 \cdot 19 = 7600, \quad \lambda_1^{\theta_1} = 25 \cdot 20 \cdot 19 = 9500, \quad \lambda_1^{\theta_2} = 30 \cdot 20 \cdot 19 = 11400,$$

and the reaction-2 intensities are

$$\lambda_2^{\theta_0} = 20 \cdot 200 = 4000, \quad \lambda_2^{\theta_1} = 25 \cdot 200 = 5000, \quad \lambda_2^{\theta_2} = 30 \cdot 200 = 6000.$$

Taking the maximum over the three processes in each reaction channel yields

$$\bar{\lambda}_1 = 11400, \quad \bar{\lambda}_2 = 6000.$$

Thus, before the first jump, reaction channel 1 occupies the vertical interval $[0, 11400)$ and reaction channel 2 occupies the vertical interval $[11400, 17400)$.

The first Poisson point, which we recall is assumed to be at $(1, 3000)$, has vertical coordinate 3000. Since $3000 < 11400$, the point lies in the reaction-1 region. Moreover,

$$3000 < 7600, \quad 3000 < 9500, \quad 3000 < 11400,$$

so all three processes accept this point as a reaction-1 event. Therefore all three processes are updated by $\zeta_1 = (-2, 1)$, giving

$$X^{\theta_0}(1) = X^{\theta_1}(1) = X^{\theta_2}(1) = (18, 201).$$

Before the second Poisson point, all three processes are therefore at state $(18, 201)$. The reaction-1 intensities are

$$\lambda_1^{\theta_0} = 20 \cdot 18 \cdot 17 = 6120, \quad \lambda_1^{\theta_1} = 25 \cdot 18 \cdot 17 = 7650, \quad \lambda_1^{\theta_2} = 30 \cdot 18 \cdot 17 = 9180,$$

and the reaction-2 intensities are

$$\lambda_2^{\theta_0} = 20 \cdot 201 = 4020, \quad \lambda_2^{\theta_1} = 25 \cdot 201 = 5025, \quad \lambda_2^{\theta_2} = 30 \cdot 201 = 6030.$$

Hence

$$\bar{\lambda}_1 = 9180, \quad \bar{\lambda}_2 = 6030.$$

Thus reaction channel 1 occupies the vertical interval $[0, 9180)$, while reaction channel 2 occupies the vertical interval $[9180, 15210)$.

The second Poisson point, which we recall is assumed to be at $(2, 14800)$, has vertical coordinate 14800. Since

$$9180 < 14800 < 15210,$$

this point lies in the reaction-2 region. Within that region, the acceptance intervals for the three processes are

$$[9180, 13200) \text{ for } X^{\theta_0}, \quad [9180, 14205) \text{ for } X^{\theta_1}, \quad [9180, 15210) \text{ for } X^{\theta_2}.$$

Since $14205 < 14800 < 15210$, only X^{θ_2} accepts the point. Consequently,

$$X^{\theta_2}(2) = (18, 201) + (2, -1) = (20, 200),$$

while X^{θ_0} and X^{θ_1} remain unchanged at $(18, 201)$.

Figure 1 gives a visual representation of this construction for the two Poisson points considered above. $\triangle$

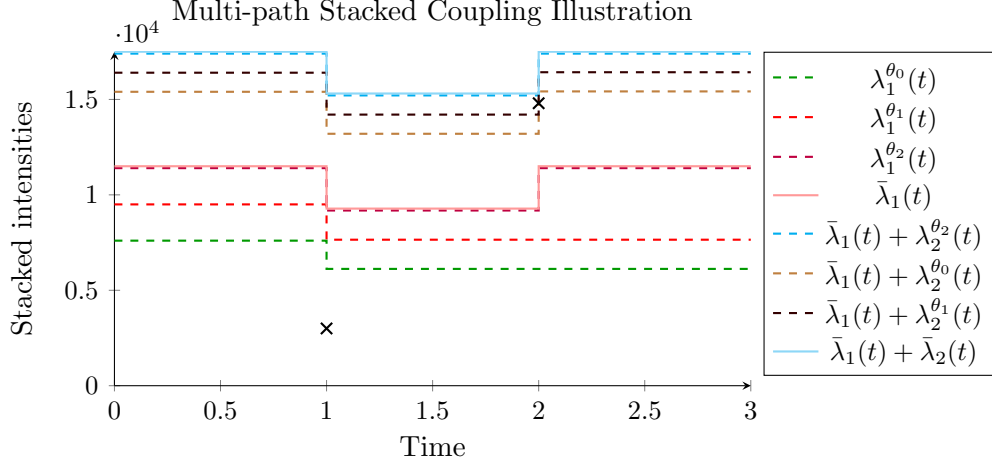


Figure 1: Illustration of the multi-path stacked coupling for the three processes X^{θ_0} , X^{θ_1} , and X^{θ_2} . The lower region corresponds to reaction channel 1, with upper boundary $\bar{\lambda}_1(t)$. The upper region corresponds to reaction channel 2, with upper boundary $\bar{\lambda}_1(t) + \bar{\lambda}_2(t)$. Dashed lines within each region show the acceptance boundaries for the individual processes. The Poisson point at (1, 3000) is accepted by all three processes as a reaction-1 event, while the Poisson point at (2, 14800) is accepted only by X^{θ_2} as a reaction-2 event. The two solid envelope boundaries are shifted upward by 100 units for visibility.

3.2 Theoretical properties of MSC

In this subsection, we establish the main theoretical properties of MSC. There are two main points. First, each relevant pairwise marginal of the MSC construction has the same law as the corresponding split-coupled pair. This allows us to transfer existing variance bounds for split coupling directly to MSC. Second, we give a finite-state refinement of this variance theory: using a perturbation expansion for the Q -matrix of a coupled process, we obtain leading-order expansions for the mean and second moment of finite-difference numerators. This second result is new and provides sharper asymptotic information than the order bounds alone. Such information can be useful for understanding leading variance constants, choosing perturbation sizes, comparing coupling methods beyond order estimates, and guiding more refined Monte Carlo implementations.

Throughout, the subscript MSC denotes processes generated by the multi-path stacked coupling, while the subscript SC denotes processes generated by the split coupling. We also write $a \wedge b = \min\{a, b\}$.

The split coupling, also known as the coupled finite-difference method, is a standard coupling for a pair of processes with parameter vectors θ_0 and θ_j . The resulting pair is itself a continuous-time Markov chain on the product state space: for each reaction channel, the two coordinates share jumps at the minimum of their two intensities, while the remaining excess intensity drives jumps only in the coordinate with the larger intensity. For fixed $j \in \{1, \dots, M\}$, define

$$m_\ell(s) = \lambda_\ell^{\theta_0} \left(X_{\text{SC}}^{\theta_0}(s) \right) \wedge \lambda_\ell^{\theta_j} \left(X_{\text{SC}}^{\theta_j}(s) \right).$$

Then the split-coupled pair can be represented as

$$\begin{aligned}
X_{\text{SC}}^{\theta_j}(t) &= X_{\text{SC}}^{\theta_j}(0) + \sum_{\ell=1}^R Y_{\ell,1} \left(\int_0^t m_\ell(s) ds \right) \zeta_\ell \\
&\quad + \sum_{\ell=1}^R Y_{\ell,2} \left(\int_0^t (\lambda_\ell^{\theta_j} (X_{\text{SC}}^{\theta_j}(s)) - m_\ell(s)) ds \right) \zeta_\ell, \\
X_{\text{SC}}^{\theta_0}(t) &= X_{\text{SC}}^{\theta_0}(0) + \sum_{\ell=1}^R Y_{\ell,1} \left(\int_0^t m_\ell(s) ds \right) \zeta_\ell \\
&\quad + \sum_{\ell=1}^R Y_{\ell,3} \left(\int_0^t (\lambda_\ell^{\theta_0} (X_{\text{SC}}^{\theta_0}(s)) - m_\ell(s)) ds \right) \zeta_\ell,
\end{aligned} \tag{13}$$

where the $Y_{\ell,i}$ are independent unit-rate Poisson processes; see [3].

The following theorem shows that each nominal/perturbed marginal pair under MSC has exactly the same distribution as the corresponding split-coupled pair. Although we state the result for the nominal process and one perturbed process, the proof only uses the two projected coordinates. Thus, after relabeling, the same conclusion applies to any two coordinates of an MSC construction provided the corresponding split-coupled pair and MSC marginal pair are nonexplosive.

Theorem 3.2. *Fix $j \in \{1, \dots, M\}$. Let $(X_{\text{SC}}^{\theta_0}, X_{\text{SC}}^{\theta_j})$ be the split-coupled pair generated by (13), and let $(X_{\text{MSC}}^{\theta_0}, X_{\text{MSC}}^{\theta_1}, \dots, X_{\text{MSC}}^{\theta_M})$ be the multi-path process generated by the MSC representation (12). Suppose that the split-coupled pair and the MSC marginal pair $(X_{\text{MSC}}^{\theta_0}, X_{\text{MSC}}^{\theta_j})$ are nonexplosive and have the same initial distribution. Then*

$$(X_{\text{SC}}^{\theta_0}, X_{\text{SC}}^{\theta_j}) \stackrel{d}{=} (X_{\text{MSC}}^{\theta_0}, X_{\text{MSC}}^{\theta_j}).$$

Proof. It suffices to show that the MSC marginal pair is a continuous-time Markov chain with the same generator as the split-coupled pair. Indeed, since the two pair processes have the same initial distribution and are nonexplosive, equality of their generators implies equality in law by the standard uniqueness theorem for nonexplosive continuous-time Markov chains on countable state spaces; see, for example, Theorem 2.8.4 of [24].

Let

$$z = (x_0, x_j) \in \mathbb{Z}_{\geq 0}^d \times \mathbb{Z}_{\geq 0}^d,$$

and define

$$\lambda_{\ell,j}^{\min}(z) = \lambda_\ell^{\theta_0}(x_0) \wedge \lambda_\ell^{\theta_j}(x_j).$$

For any bounded function $h : \mathbb{Z}_{\geq 0}^d \times \mathbb{Z}_{\geq 0}^d \rightarrow \mathbb{R}$, the split-coupled pair has generator

$$\begin{aligned}
\mathcal{L}_j h(z) &= \sum_{\ell=1}^R \left\{ \lambda_{\ell,j}^{\min}(z) (h(z + (\zeta_\ell, \zeta_\ell)) - h(z)) \right. \\
&\quad + \left(\lambda_\ell^{\theta_0}(x_0) - \lambda_{\ell,j}^{\min}(z) \right) (h(z + (\zeta_\ell, 0)) - h(z)) \\
&\quad \left. + \left(\lambda_\ell^{\theta_j}(x_j) - \lambda_{\ell,j}^{\min}(z) \right) (h(z + (0, \zeta_\ell)) - h(z)) \right\}.
\end{aligned}$$

We now consider the MSC process and project onto the coordinates $(0, j)$. Fix a full MSC state

$$z^* = (x_0, x_1, \dots, x_M),$$

and write $z = (x_0, x_j)$ for its projection onto coordinates 0 and j . In the reaction- ℓ block of the stacked Poisson representation, the two coordinates $X_{\text{MSC}}^{\theta_0}$ and $X_{\text{MSC}}^{\theta_j}$ both jump through reaction ℓ on a subinterval of height

$$\lambda_\ell^{\theta_0}(x_0) \wedge \lambda_\ell^{\theta_j}(x_j) = \lambda_{\ell,j}^{\min}(z).$$

Only $X_{\text{MSC}}^{\theta_0}$ jumps on a subinterval of height

$$\lambda_\ell^{\theta_0}(x_0) - \lambda_{\ell,j}^{\min}(z),$$

and only $X_{\text{MSC}}^{\theta_j}$ jumps on a subinterval of height

$$\lambda_\ell^{\theta_j}(x_j) - \lambda_{\ell,j}^{\min}(z).$$

All remaining points in the reaction- ℓ block affect neither coordinate of the projected pair and therefore contribute no transition to the projected process.

Thus, after projection onto the coordinates $(0, j)$, every nontrivial transition rate depends only on the projected state $z = (x_0, x_j)$, not on the remaining coordinates of z^* . Events that change only the other coordinates are invisible to the projection. Hence the projected pair $(X_{\text{MSC}}^{\theta_0}, X_{\text{MSC}}^{\theta_j})$ is itself a continuous-time Markov chain, with infinitesimal generator $\mathcal{L}_j$, the same generator as that of the split-coupled pair.

Therefore the MSC marginal pair and the split-coupled pair have the same generator, and the equality in law follows from the uniqueness result cited at the start of the proof. $\square$

We next record the variance estimate that will be used repeatedly below. This is essentially Theorem 3.1 of [3], applied to the pairwise marginals of MSC. In particular, we assume that the hypotheses of that theorem hold for the parameter perturbations considered here: the intensity functions and drift are uniformly Lipschitz on the relevant state space, the change in the intensity functions and drift is $O(\varepsilon)$ under an $O(\varepsilon)$ parameter perturbation, and f is Lipschitz on the relevant state space.

Corollary 3.3. *Suppose there is a constant $C_0 > 0$ such that, for each sufficiently small $\varepsilon > 0$, the parameter vectors $\alpha_\varepsilon, \beta_\varepsilon \in \mathbb{R}^K$ satisfy*

$$\|\alpha_\varepsilon - \beta_\varepsilon\| \leq C_0 \varepsilon.$$

Suppose that the hypotheses of Theorem 3.2 hold for the corresponding split-coupled pair and MSC marginal pair, and that the hypotheses of Theorem 3.1 of [3] hold for the split-coupled pair, with constants that can be chosen independently of ε . If $X_{\text{MSC}}^{\alpha_\varepsilon}$ and $X_{\text{MSC}}^{\beta_\varepsilon}$ are two coordinates of an MSC construction, then for each fixed $T > 0$ there is a constant $C_T > 0$ such that

$$\mathbb{E} \left[\sup_{t \leq T} \left(f(X_{\text{MSC}}^{\alpha_\varepsilon}(t)) - f(X_{\text{MSC}}^{\beta_\varepsilon}(t)) \right)^2 \right] \leq C_T \varepsilon.$$

In particular,

$$\text{Var} \left(f(X_{\text{MSC}}^{\alpha_\varepsilon}(T)) - f(X_{\text{MSC}}^{\beta_\varepsilon}(T)) \right) = O(\varepsilon).$$

Proof. By Theorem 3.2, after relabeling the two coordinates if necessary, the pair $(X_{\text{MSC}}^{\alpha_\varepsilon}, X_{\text{MSC}}^{\beta_\varepsilon})$ has the same law as the corresponding split-coupled pair. The result then follows directly from Theorem 3.1 of [3], applied with perturbation size of order ε . $\square$

We next state the version needed for wider stencils and higher-order derivative formulas.

Corollary 3.4. *Fix an integer $J \geq 2$ and coefficients $c_1, \dots, c_J$ satisfying*

$$\sum_{r=1}^J c_r = 0.$$

For each sufficiently small $\varepsilon > 0$, let $\theta_1^\varepsilon, \dots, \theta_J^\varepsilon \in \mathbb{R}^K$ satisfy

$$\max_{1 \leq r, s \leq J} \|\theta_r^\varepsilon - \theta_s^\varepsilon\| \leq C_0 \varepsilon$$

for some constant $C_0 > 0$, and suppose the hypotheses of Corollary 3.3 hold uniformly for all pairs of parameter vectors in this finite collection. Let the processes $X_{\text{MSC}}^{\theta_1^\varepsilon}, \dots, X_{\text{MSC}}^{\theta_J^\varepsilon}$ be generated by MSC, and define the finite-difference numerator

$$N_\varepsilon(T) = \sum_{r=1}^J c_r f\left(X_{\text{MSC}}^{\theta_r^\varepsilon}(T)\right).$$

Then, for each fixed $T > 0$,

$$\mathbb{E}[N_\varepsilon(T)^2] = O(\varepsilon), \quad \text{Var}(N_\varepsilon(T)) = O(\varepsilon).$$

Proof. Since $\sum_{r=1}^J c_r = 0$, we may rewrite $N_\varepsilon(T)$ using the first process as a reference:

$$N_\varepsilon(T) = \sum_{r=2}^J c_r \left(f\left(X_{\text{MSC}}^{\theta_r^\varepsilon}(T)\right) - f\left(X_{\text{MSC}}^{\theta_1^\varepsilon}(T)\right) \right).$$

For each r , the parameter vectors θ_r^ε and θ_1^ε differ by $O(\varepsilon)$. Corollary 3.3 therefore gives

$$\mathbb{E}\left[\left(f\left(X_{\text{MSC}}^{\theta_r^\varepsilon}(T)\right) - f\left(X_{\text{MSC}}^{\theta_1^\varepsilon}(T)\right)\right)^2\right] = O(\varepsilon).$$

Since there are only finitely many terms in the sum, the bound

$$\left(\sum_{r=2}^J u_r\right)^2 \leq (J-1) \sum_{r=2}^J u_r^2$$

implies $\mathbb{E}[N_\varepsilon(T)^2] = O(\varepsilon)$. The variance bound follows immediately from

$$\text{Var}(N_\varepsilon(T)) \leq \mathbb{E}[N_\varepsilon(T)^2].$$

□

The preceding corollaries provide the variance-order estimates needed for the later bias-variance analysis. In finite-state settings, a Q -matrix perturbation argument gives sharper information: the mean and second moment of a coupled finite-difference numerator admit first-order asymptotic expansions. We record this as a finite-state refinement, which can be useful for understanding the leading constants underlying the order bounds, choosing perturbation sizes, and comparing coupling methods beyond order estimates.

Theorem 3.5. Let $\{Z^\varepsilon\}_{\varepsilon \geq 0}$ be a family of continuous-time Markov chains on a finite state space E , with common initial distribution μ_0 and respective Q -matrices Q^ε . Suppose there exists a matrix G such that

$$Q^\varepsilon = Q^0 + \varepsilon G + R^\varepsilon, \quad \text{where} \quad |R^\varepsilon| = o(\varepsilon),$$

as $\varepsilon \downarrow 0$, where $|\cdot|$ denotes any matrix norm. Suppose also that there is a subset $\Delta \subset E$ such that μ_0 is supported on Δ and Δ is invariant for the chain with Q -matrix Q^0 . Let $H : E \rightarrow \mathbb{R}$ satisfy $H(x) = 0$ for all $x \in \Delta$. Then, for each fixed $T > 0$,

$$\begin{aligned} \mathbb{E}[H(Z^\varepsilon(T))] &= a_T \varepsilon + o(\varepsilon), \\ \mathbb{E}[H(Z^\varepsilon(T))^2] &= b_T \varepsilon + o(\varepsilon), \end{aligned}$$

where $a_T \in \mathbb{R}$ and $b_T \geq 0$ are given by

$$\begin{aligned} a_T &= \mu_0 \left(\int_0^T e^{(T-s)Q^0} G e^{sQ^0} ds \right) H, \\ b_T &= \mu_0 \left(\int_0^T e^{(T-s)Q^0} G e^{sQ^0} ds \right) H^2, \end{aligned}$$

and H^2 denotes the pointwise square of H . Consequently,

$$\text{Var}(H(Z^\varepsilon(T))) = b_T \varepsilon + o(\varepsilon).$$

Proof. Since E is finite, the transition matrix of Z^ε at time T is

$$P_T^\varepsilon = e^{TQ^\varepsilon}.$$

The standard first-order expansion of the matrix exponential gives

$$e^{TQ^\varepsilon} = e^{TQ^0} + \varepsilon \int_0^T e^{(T-s)Q^0} G e^{sQ^0} ds + o(\varepsilon).$$

Identifying functions on E with column vectors and μ_0 with a row vector, for any function $\varphi : E \rightarrow \mathbb{R}$,

$$\mathbb{E}[\varphi(Z^\varepsilon(T))] = \mu_0 e^{TQ^\varepsilon} \varphi = \mu_0 e^{TQ^0} \varphi + \varepsilon \mu_0 \left(\int_0^T e^{(T-s)Q^0} G e^{sQ^0} ds \right) \varphi + o(\varepsilon).$$

Because μ_0 is supported on Δ , Δ is invariant for the chain with Q -matrix Q^0 , and both H and H^2 vanish on Δ , we have

$$\mu_0 e^{TQ^0} H = 0, \quad \text{and} \quad \mu_0 e^{TQ^0} H^2 = 0.$$

Applying the preceding expansion with $\varphi = H$ and then with $\varphi = H^2$ gives the stated formulas for a_T and b_T . Finally,

$$(\mathbb{E}[H(Z^\varepsilon(T))])^2 = O(\varepsilon^2),$$

and hence

$$\begin{aligned} \text{Var}(H(Z^\varepsilon(T))) &= \mathbb{E}[H(Z^\varepsilon(T))^2] - (\mathbb{E}[H(Z^\varepsilon(T))])^2 \\ &= b_T \varepsilon + o(\varepsilon). \end{aligned}$$

□

Theorem 3.5 applies directly to finite-difference estimators generated by MSC. For example, for some $j \in \{1, \dots, K\}$, let

$$Z^\varepsilon = \left(X^{\theta + a_1 \varepsilon e_j}, \dots, X^{\theta + a_J \varepsilon e_j} \right), \quad H(x_1, \dots, x_J) = \sum_{r=1}^J c_r f(x_r),$$

where $\sum_{r=1}^J c_r = 0$, and suppose each coordinate process has finite state space S . If the coordinate processes share the same initial condition, then the initial law of Z^ε is supported on the diagonal

$$\Delta = \{(x, \dots, x) : x \in S\}.$$

Moreover, H vanishes on Δ , and when $\varepsilon = 0$, all parameter vectors equal θ , so the MSC construction leaves Δ invariant. Thus, whenever the Q -matrix of Z^ε has the first-order expansion assumed in Theorem 3.5, the theorem yields leading-order expansions for the mean and second moment of the finite-difference numerator.

Remark 3.6. *In the subsequent application sections, the processive phosphorylation network (19) has a finite reachable state space by conservation of the total amounts of substrate, kinase, and phosphatase. On this state space, the mass-action intensity functions are Lipschitz, and because the perturbed parameters enter as rate constants, the intensities and drift change by $O(\varepsilon)$ under the finite-difference perturbations used below. Thus the hypotheses needed for the variance-order estimates are satisfied, and the network also lies in the finite-state setting of Theorem 3.5.*

For comparison, we also record the corresponding finite-state variance limit for independently generated paths, which will be used in the RMSE analysis of the next section.

Theorem 3.7. *Let $\{X^\theta : \theta \in \mathbb{R}^K\}$ be a family of continuous-time Markov chains on a finite state space E , with common initial distribution μ_0 and respective Q -matrices Q^θ . Fix $\theta \in \mathbb{R}^K$, an integer $J \geq 2$, coefficients $c_1, \dots, c_J \in \mathbb{R}$, and a function $f : E \rightarrow \mathbb{R}$. For each $r \in \{1, \dots, J\}$, let θ_r^ε satisfy*

$$Q^{\theta_r^\varepsilon} \longrightarrow Q^\theta \quad \text{as } \varepsilon \downarrow 0.$$

Let $X_{\text{Indep}}^{\theta_1^\varepsilon}, \dots, X_{\text{Indep}}^{\theta_J^\varepsilon}$ be generated independently, and define

$$N_\varepsilon(T) = \sum_{r=1}^J c_r f\left(X_{\text{Indep}}^{\theta_r^\varepsilon}(T)\right).$$

Then, for each fixed $T > 0$,

$$\text{Var}(N_\varepsilon(T)) = \left(\sum_{r=1}^J c_r^2 \right) \text{Var}\left(f\left(X^\theta(T)\right)\right) + o(1).$$

Proof. By independence,

$$\text{Var}(N_\varepsilon(T)) = \sum_{r=1}^J c_r^2 \text{Var}\left(f\left(X_{\text{Indep}}^{\theta_r^\varepsilon}(T)\right)\right).$$

Since E is finite and $Q^{\theta_r^\varepsilon} \rightarrow Q^\theta$, continuity of the matrix exponential gives

$$e^{TQ^{\theta_r^\varepsilon}} \longrightarrow e^{TQ^\theta}.$$

Hence the first and second moments of $f(X_{\text{Indep}}^{\theta_r^\varepsilon}(T))$ converge to those of $f(X^\theta(T))$, and therefore

$$\text{Var}\left(f\left(X_{\text{Indep}}^{\theta_r^\varepsilon}(T)\right)\right) = \text{Var}\left(f\left(X^\theta(T)\right)\right) + o(1).$$

Summing over r gives the result. $\square$

4 General optimal RMSE scaling for finite-difference estimators

All finite-difference estimators considered below exhibit the same basic bias-variance tradeoff. They differ only in (i) the order of the finite-difference bias and (ii) the variance scaling of the finite-difference numerator under the chosen coupling method. We therefore carry out the general RMSE optimization once here and summarize the resulting optimal scaling rates for the different coupling methods.

Fix $T > 0$, let $f : \mathbb{Z}_{\geq 0}^d \rightarrow \mathbb{R}$ be an observable of interest, and, for $\theta \in \mathbb{R}^K$, define

$$g(\theta) \stackrel{\text{def}}{=} \mathbb{E} \left[f(X^\theta(T)) \right].$$

Let $D(\theta)$ denote a parameter derivative of g of total order $k \geq 1$, fix an integer $J \geq 2$, and let $e_1, \dots, e_K$ denote the canonical unit vectors in $\mathbb{R}^K$. For fixed coefficients a_{rj} , define

$$\theta_r^\varepsilon = \theta + \varepsilon \sum_{j=1}^K a_{rj} e_j, \quad r = 1, \dots, J.$$

We consider finite-difference approximations of the form

$$D_\varepsilon(\theta) = \frac{\sum_{r=1}^J c_r g(\theta_r^\varepsilon)}{\varepsilon^k}, \quad (14)$$

where $c_1, \dots, c_J$ are the finite-difference coefficients and the coefficients a_{rj} specify the stencil points. In particular, we assume

$$\sum_{r=1}^J c_r = 0.$$

The corresponding Monte Carlo estimator is

$$\widehat{D}_\varepsilon(\theta) = \frac{1}{N} \sum_{i=1}^N d_{[i]}(\varepsilon), \quad d_{[i]}(\varepsilon) = \frac{\sum_{r=1}^J c_r f \left(X_{[i]}^{\theta_r^\varepsilon}(T) \right)}{\varepsilon^k}. \quad (15)$$

For each i , the processes $\{X_{[i]}^{\theta_r^\varepsilon}\}_{r=1}^J$ form one collection of coupled sample paths, and the N collections are independent and identically distributed.

We measure the accuracy of the estimator using its root mean square error (RMSE),

$$\begin{aligned} \text{RMSE} \left(\widehat{D}_\varepsilon(\theta) \right) &= \left(\mathbb{E} \left[\left(\widehat{D}_\varepsilon(\theta) - D(\theta) \right)^2 \right] \right)^{1/2} \\ &= \left(\text{Var} \left(\widehat{D}_\varepsilon(\theta) \right) + \left(\mathbb{E} \left[\widehat{D}_\varepsilon(\theta) \right] - D(\theta) \right)^2 \right)^{1/2}. \end{aligned}$$

This criterion captures both sources of error relevant here: the bias inherent to the finite-difference estimator and the statistical error due to the variance of the Monte Carlo estimator. In particular, an RMSE below η implies that both the absolute bias is below η and the variance is below η^2 .

We assume that the finite-difference approximation has bias order $b > 0$ with a nonzero leading coefficient; that is,

$$\begin{aligned} D_\varepsilon(\theta) - D(\theta) &= B\varepsilon^b + o(\varepsilon^b), \\ \mathbb{E}[\widehat{D}_\varepsilon(\theta)] - D(\theta) &= B\varepsilon^b + o(\varepsilon^b), \end{aligned} \tag{16}$$

for some constant $B \neq 0$, where the second equality follows from $\mathbb{E}[\widehat{D}_\varepsilon(\theta)] = D_\varepsilon(\theta)$.

To treat the different coupling methods within a common framework, define the finite-difference numerator

$$N_{[i]}^\varepsilon = \sum_{r=1}^J c_r f\left(X_{[i]}^{\theta_r^\varepsilon}(T)\right).$$

Suppose that, under the coupling method being considered,

$$\text{Var}\left(N_{[i]}^\varepsilon\right) = C_V \varepsilon^q + o(\varepsilon^q) \tag{17}$$

for some $C_V > 0$ and $q < 2k$. Since the N collections of paths are independent and identically distributed,

$$\text{Var}\left(\widehat{D}_\varepsilon(\theta)\right) = C_V N^{-1} \varepsilon^{q-2k} + o\left(N^{-1} \varepsilon^{q-2k}\right).$$

Moreover, the bias expansion (16) gives

$$\left(\mathbb{E}\left[\widehat{D}_\varepsilon(\theta)\right] - D(\theta)\right)^2 = B^2 \varepsilon^{2b} + o(\varepsilon^{2b}).$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{D}_\varepsilon(\theta) - D(\theta)\right)^2\right] &= \text{Var}\left(\widehat{D}_\varepsilon(\theta)\right) + \left(\mathbb{E}\left[\widehat{D}_\varepsilon(\theta)\right] - D(\theta)\right)^2 \\ &= C_V N^{-1} \varepsilon^{q-2k} + B^2 \varepsilon^{2b} + o\left(N^{-1} \varepsilon^{q-2k}\right) + o(\varepsilon^{2b}). \end{aligned}$$

For fixed N , the leading-order terms are minimized at

$$\varepsilon^* = \left(\frac{(2k-q)C_V}{2bB^2N}\right)^{\frac{1}{2b+2k-q}}.$$

It follows that the corresponding optimal mean square error and root mean square error satisfy

$$\begin{aligned} \text{MSE}_{\text{opt}} &= O\left(N^{-\frac{2b}{2b+2k-q}}\right), \\ \text{RMSE}_{\text{opt}} &= O\left(N^{-\frac{b}{2b+2k-q}}\right). \end{aligned} \tag{18}$$

It remains to identify the numerator-variance exponent q for each coupling method. In the finite-state setting, under the hypotheses of Theorem 3.7, independently generated paths satisfy

$$\text{Var}\left(N_{[i]}^\varepsilon\right) = C_{\text{ind}} + o(1),$$

where

$$C_{\text{ind}} = \left(\sum_{r=1}^J c_r^2\right) \text{Var}\left(f\left(X^\theta(T)\right)\right).$$

Thus $q = 0$, and (18) yields

$$\text{RMSE}_{\text{opt}} = O\left(N^{-\frac{b}{2(b+k)}}\right).$$

For MSC, Corollary 3.4 gives the general variance bound $O(\varepsilon)$, while Theorem 3.5 gives

$$\text{Var}\left(N_{[i]}^{\varepsilon}\right) = C_{\text{cpl}}\varepsilon + o(\varepsilon),$$

provided the hypotheses of that theorem hold. Hence $q = 1$, and

$$\text{RMSE}_{\text{opt}} = O\left(N^{-\frac{b}{2b+2k-1}}\right).$$

The resulting optimal RMSE scaling rates are summarized in Table 1.

Table 1: Optimal RMSE scaling for the estimator (15) under the different coupling methods

	Independent	SC	MSC
Numerator-variance exponent q	0	1	1
Optimal RMSE	$O\left(N^{-\frac{b}{2(b+k)}}\right)$	$O\left(N^{-\frac{b}{2b+2k-1}}\right)$	$O\left(N^{-\frac{b}{2b+2k-1}}\right)$

We do not include CRP in the theoretical scaling comparison because, to our knowledge, there is no corresponding small-perturbation variance theory for CRP. Previous work has shown that split coupling can be obtained, in a precise limiting sense, from CRP by allowing increasingly frequent recoupling [5], while numerical studies have also shown that CRP can lose effectiveness over time and, over intermediate time scales, may perform little better than independent simulation [3]. We therefore treat CRP empirically in the numerical comparisons below rather than assigning it a theoretical RMSE scaling rate.

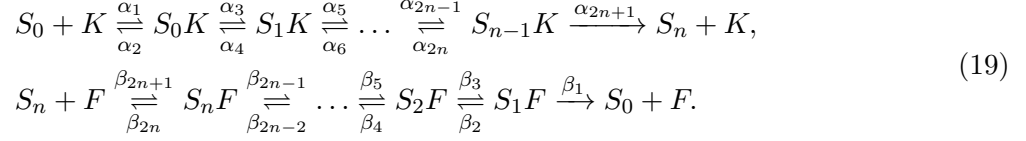
5 Numerical applications of MSC to a processive phosphorylation network

We begin by introducing the processive phosphorylation/dephosphorylation model that we will use for all of our numerical experiments and then present three applications of the multi-path stacked coupling method: simultaneous computation of many first derivatives, higher-order finite-difference estimation of a first derivative, and higher-order derivative estimation. Appendix A gives the details of the numerical experiments, together with application-specific implementation details in Appendices B, C, and D.

5.1 The processive phosphorylation/dephosphorylation model

Phosphorylation and dephosphorylation are ubiquitous mechanisms in cellular signaling and regulation, and phosphorylation cycles have also been studied as mechanisms for integral regulation and adaptation [14, 15]. For our numerical examples, we use the sequential, processive n -site phosphorylation/dephosphorylation network studied by Conradi and Shiu [11]. Broader classes of processive multisite phosphorylation networks, including variants with irreversible reactions and

product inhibition, were subsequently studied by Eithun and Shiu [12]. The network used here is



The parameter vector is

$$\theta = (\theta_1, \dots, \theta_{4n+2}) = (\alpha_1, \dots, \alpha_{2n+1}, \beta_1, \dots, \beta_{2n+1}).$$

Let $X^\theta(t) = (S_0^\theta(t), S_n^\theta(t), K^\theta(t), F^\theta(t), (S_0 K)^\theta(t), \dots, (S_{n-1} K)^\theta(t), (S_1 F)^\theta(t), \dots, (S_n F)^\theta(t))$ denote the state vector of the chemical reaction network at time t .

For all parameter vectors θ , we use the initial condition

$$\begin{aligned} S_0^\theta(0) &= 100, \quad S_n^\theta(0) = 0, \quad K^\theta(0) = 10, \quad F^\theta(0) = 10, \\ (S_i K)^\theta(0) &= 0, \quad i \in \{0, \dots, n-1\}, \quad (S_j F)^\theta(0) = 0, \quad j \in \{1, \dots, n\}. \end{aligned}$$

This initial condition is fixed throughout the numerical experiments. As a representative quantity of interest, we consider the expected number of S_0 molecules at a fixed terminal time $T = 30$:

$$g(\theta) = \mathbb{E} \left[f \left(X^\theta(T) \right) \right] = \mathbb{E} \left[S_0^\theta(T) \right]. \quad (20)$$

In the numerical experiments below, we compare the performance of all three coupling methods (CRP, SC, and MSC) with the baseline independent simulation (Indep) for estimating parameter sensitivities of $g(\theta)$ using the following rate constants

$$\begin{aligned} \alpha_1 &= 0.02, \quad \alpha_2 = 0.2, \quad \alpha_{2n+1} = 2.0, \\ \alpha_{2k+1} &= 1, \quad \alpha_{2k+2} = 0.5, \quad k \in \{1, \dots, n-1\}; \\ \beta_1 &= 1.5, \quad \beta_{2n} = 0.15, \quad \beta_{2n+1} = 0.015, \\ \beta_{2k-2} &= 0.4, \quad \beta_{2k-1} = 0.8, \quad k \in \{2, \dots, n\}. \end{aligned}$$

To keep the subsequent application subsections concise, we note here that the processive n -site network (19) has three conserved quantities:

$$\begin{aligned} S_{\text{tot}}^\theta &= S_0^\theta + S_n^\theta + \sum_{i=0}^{n-1} (S_i K)^\theta + \sum_{i=1}^n (S_i F)^\theta, \\ K_{\text{tot}}^\theta &= K^\theta + \sum_{i=0}^{n-1} (S_i K)^\theta, \\ F_{\text{tot}}^\theta &= F^\theta + \sum_{i=1}^n (S_i F)^\theta. \end{aligned}$$

These three conserved quantities together involve every species in the network. Consequently, each species count is bounded by the corresponding conserved total, and the reachable state space is finite. Together with the common initial condition and the form of the mass-action intensities, this verifies the hypotheses needed to apply Theorems 3.5 and 3.7; see also Remark 3.6. We therefore apply the conclusions of those results directly in the subsections below.

5.2 Simultaneous computation of many first derivatives

We first consider the simultaneous estimation of all K first-order parametric sensitivities $\frac{\partial}{\partial \theta_j} g(\theta)$, $j \in \{1, \dots, K\}$, where g is defined in (20) with $T = 30$. For each parameter direction, we use the forward finite-difference approximation introduced in Section 2, which has bias $O(\varepsilon_j)$ and requires simulation of the pair $(X^\theta, X^{\theta + \varepsilon_j e_j})$.

For this application, the key distinction among the methods is the combination of variance reduction and the number of simulated component paths required to estimate all K sensitivities. For each parameter direction, Theorem 3.2 implies that the corresponding MSC and SC finite-difference estimators have the same variance, while previous numerical studies have found that SC often provides greater variance reduction than CRP [3, 30]. The relevant notion of efficiency is the total computational work required to achieve a given RMSE. Wall-clock time, however, depends strongly on implementation details and hardware. We therefore use the total number of simulated component paths as a simple proxy for computational effort that is less sensitive to implementation details than wall-clock time. Since all methods simulate the same reaction network over the same time interval and at nearby parameter values, this metric provides a natural way to quantify the computational savings obtained by sharing paths across parameter directions.

Under this path-count metric, suppose that N Monte Carlo replications are used for each parameter direction. SC simulates a separate nominal–perturbed pair for each of the K directions and therefore requires $2KN$ component paths. By contrast, each MSC replication jointly simulates the nominal path X^θ and all K perturbed paths $X^{\theta + \varepsilon_j e_j}$, $j \in \{1, \dots, K\}$, and therefore requires only $(K+1)N$ component paths. CRP and Indep likewise require $(K+1)N$ component paths under this accounting. Thus, MSC has the pairwise variance-reduction properties of SC while requiring the same number of simulated component paths as CRP and Indep.

To assess overall accuracy across all parameter directions, we use the sum of the empirically optimized RMSEs over the K directions. Following the numerical procedure described in Appendix A, together with the application-specific details in Appendix B, we plot this quantity against the total number of simulated component paths on a base-2 log-log scale. Each plotted value is estimated using 500 independent repetitions of the entire estimation procedure to reduce Monte Carlo variability.

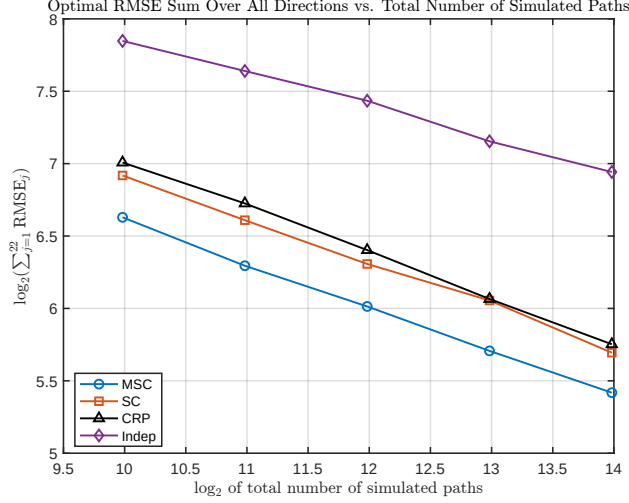


Figure 2: Base-2 log-log plot of the sum of the empirically optimized RMSEs over all parameter directions versus the total number of simulated component paths for the $O(\varepsilon_j)$ -bias forward finite-difference estimators.

Table 2: Theoretical and empirical slopes under different coupling methods.

	Indep	CRP	SC	MSC
Theoretical slope	−0.2500	—	−0.3333	−0.3333
Empirical slope	−0.2263	−0.3137	−0.3060	−0.3028

The empirical slopes for Indep, SC, and MSC closely match the theoretical slopes derived in Section 4. The theoretical and empirical slopes are reported in Table 2. Specifically, Indep exhibits a slope of approximately $-\frac{1}{4}$, whereas MSC and SC exhibit slopes of approximately $-\frac{1}{3}$. At present, there is no corresponding theoretical slope available for CRP. Among the methods considered in this example, MSC achieves the smallest sum of empirically optimized RMSEs over the range of simulated-path budgets considered.

5.3 Higher-order finite-difference estimation of a first derivative

We next examine whether MSC improves efficiency when higher-order centered finite-difference approximations are used to reduce the bias in computing sensitivities. For g as defined in (20) with $T = 30$, and for a fixed parameter θ_j , we consider the centered finite-difference approximations (4) and (6), which have bias of order $O(\varepsilon_j^2)$ and $O(\varepsilon_j^4)$, respectively. The $O(\varepsilon_j^2)$ -bias estimator uses the two-path collection $(X_{[i]}^{\theta-\varepsilon_j e_j}, X_{[i]}^{\theta+\varepsilon_j e_j})$ and the $O(\varepsilon_j^4)$ -bias estimator uses the four-path collection

$$(X_{[i]}^{\theta+2\varepsilon_j e_j}, X_{[i]}^{\theta+\varepsilon_j e_j}, X_{[i]}^{\theta-\varepsilon_j e_j}, X_{[i]}^{\theta-2\varepsilon_j e_j}).$$

For the numerical comparison, we estimate the sensitivity $\frac{\partial}{\partial \theta_1} g(\theta)$ using both centered finite-difference estimators and compare MSC, CRP, and Indep. Section 4 predicts more favorable optimal RMSE scaling for MSC than for Indep for both estimators. We omit SC from this comparison for two reasons. For the two-path stencil, Theorem 3.2 implies that the SC and MSC constructions have

the same joint distribution, and both require two component paths per Monte Carlo replication. Thus, there is no distinction between the two methods for this estimator. For the four-path stencil, there is no unique pairwise split-coupling construction, since the four paths can be grouped into coupled pairs in several different ways, and hence there is no single canonical “SC” estimator to use as a benchmark.

We evaluate the methods using the empirically optimized RMSE. Following the numerical procedure described in Appendix A, together with the application-specific modification for Application 2 given in Appendix C, we plot this quantity against the total number of simulated component paths on a base-2 log-log scale. Each plotted value is estimated using 500 independent repetitions of the complete estimation procedure.

Figure 3 presents the results for both estimators, while the theoretical and empirical slopes are reported in Table 3.

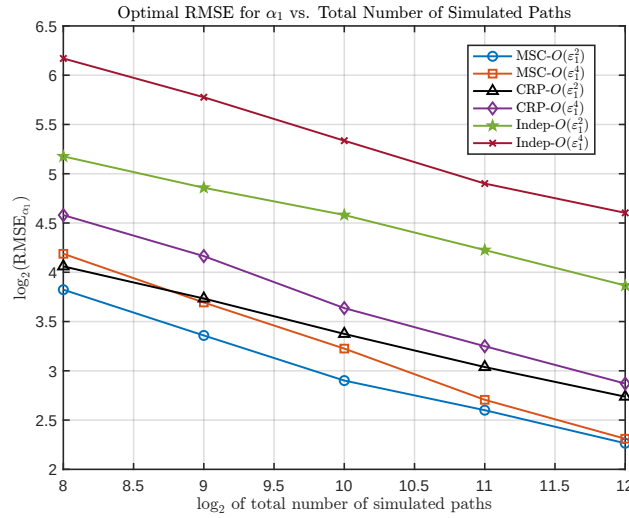


Figure 3: Base-2 log-log plot of the RMSEs for the $O(\varepsilon_1^2)$ -bias and $O(\varepsilon_1^4)$ -bias estimators versus the total number of simulated component paths.

Table 3: Theoretical and empirical convergence rates.

Method	Indep	CRP	MSC
Centered difference with bias $O(\varepsilon_j^2)$			
Theoretical rate	−0.3333	—	−0.4000
Empirical rate	−0.3275	−0.3305	−0.3896
Centered difference with bias $O(\varepsilon_j^4)$			
Theoretical rate	−0.4000	—	−0.4444
Empirical rate	−0.3917	−0.4274	−0.4692

The empirical slopes for Indep and MSC are close to the theoretical predictions for both finite-difference estimators. As predicted by the theory, the $O(\varepsilon_1^4)$ -bias estimator exhibits a more favorable asymptotic RMSE slope than the $O(\varepsilon_1^2)$ -bias estimator. Over the range of path-count budgets

considered here, however, this asymptotic advantage has not yet overcome the additional computational cost of the four-point stencil, and the two-point MSC estimator retains a slightly smaller RMSE. For each stencil, MSC achieves the smallest RMSE among the methods considered.

5.4 Higher-order derivatives

We use the final application to illustrate a central advantage of MSC: it provides a clean framework for estimators that require several coupled paths simultaneously. This is particularly useful for finite-difference estimation of higher-order sensitivities, where the required stencil may involve several nearby parameter values.

For example, the centered finite-difference approximation (8) for a third derivative requires the four-path collection

$$\left(X^{\theta+2\varepsilon_j e_j}, X^{\theta+\varepsilon_j e_j}, X^{\theta-\varepsilon_j e_j}, X^{\theta-2\varepsilon_j e_j} \right).$$

MSC couples these four processes within a single joint construction. In contrast, pairwise SC does not specify a unique four-path coupling: the four paths can be grouped into coupled pairs in several different ways, leading to different joint constructions. Thus, MSC provides a general-purpose framework that avoids the need to design an application-specific arrangement of pairwise couplings, an advantage that becomes more pronounced as the stencil becomes wider. Specialized methods can provide stronger variance reduction for particular estimators; for example, the double-coupled finite-difference method of Wolf and Anderson [32] is tailored to second-order sensitivities. Our goal here is different: to provide a single coupling framework that applies directly to arbitrary collections of nearby parameter values.

For g as defined in (20) with $T = 30$, we estimate the third-order sensitivity $\frac{\partial^3}{\partial \theta_3^3} g(\theta)$ using the centered finite-difference approximation (8), which has bias $O(\varepsilon_3^2)$. We compare the empirically optimized RMSEs obtained using MSC, CRP, and Indep under the same total simulated-path budget. Following the numerical procedure described in Appendix A, together with the application-specific details in Appendix D, we plot the RMSE against the total number of simulated component paths on a base-2 log-log scale. Each plotted value is estimated using 500 independent repetitions of the complete estimation procedure.

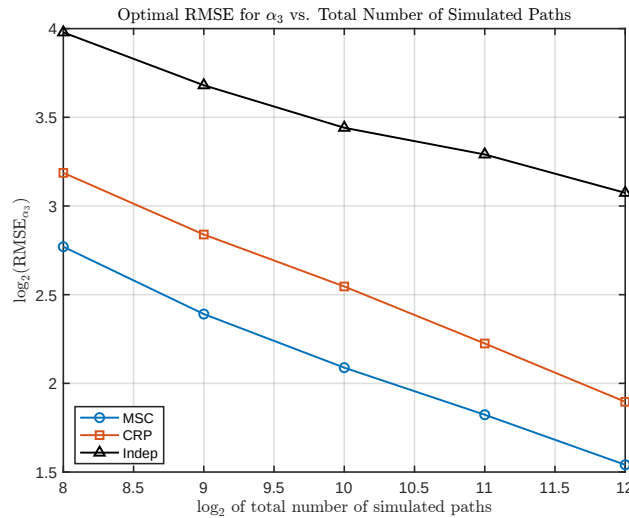


Figure 4: Base-2 log-log plot of the empirically optimized RMSE for the $O(\varepsilon_3^2)$ -bias third-order sensitivity estimator versus the total number of simulated component paths.

Table 4: Theoretical and empirical convergence rates for the empirically optimized RMSE

Method	Indep	CRP	MSC
Theoretical rate	−0.2000	—	−0.2222
Empirical rate	−0.2261	−0.3227	−0.3074

Figure 4 presents the numerical results, and the theoretical and empirical slopes are reported in Table 4. Over the range of path-count budgets considered, the empirical slopes for Indep and MSC are steeper than the corresponding asymptotic theoretical predictions; no theoretical slope is available for CRP. We believe this discrepancy reflects the fact that the simulations have not yet entered the small- ε regime underlying the asymptotic theory. Nevertheless, MSC achieves the smallest empirically optimized RMSE among the methods considered throughout the displayed range.

6 Discussion

The motivation for MSC is simple. There are numerous finite-difference settings in which one wants to generate several nearby parameterized paths simultaneously, including the estimation of many first derivatives, the use of wider finite-difference stencils for a single first derivative, and the estimation of higher-order derivatives. In such settings, the common reaction path method has an attractive path-count structure because all required processes can be generated jointly, while split coupling provides stronger variance reduction but is naturally pairwise. We wanted a construction that combined these two advantages. MSC does exactly that: it generates the full collection of nearby parameterized paths simultaneously, while retaining the pairwise behavior of split coupling.

The theory supports this picture. By Theorem 3.2, each pairwise marginal of an MSC construction has the same law as the corresponding split-coupled pair, and Corollary 3.4 extends the resulting variance control to general finite-difference contrasts involving multiple nearby paths. The finite-state theory further gives the leading-order variance behavior used in the RMSE analysis. The numerical experiments are consistent with these results: MSC achieved the smallest empirically optimized RMSE among the methods considered in each of the three applications.

Acknowledgments

DFA gratefully acknowledges financial support from the Office of the Vice Chancellor for Research at UW–Madison, with funding from the Wisconsin Alumni Research Foundation; from the Trustees of the William F. Vilas Estate; and from NSF grant DMS-2051498.

ChatGPT was used for sentence-level editing and to provide initial coding templates for the numerical simulations.

References

- [1] Linda JS Allen. *An introduction to stochastic processes with applications to biology*. CRC press, 2010.
- [2] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2019.

- [3] David F. Anderson. An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM Journal on Numerical Analysis*, 50(5):2237–2258, 2012.
- [4] David F. Anderson and Desmond J. Higham. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Modeling & Simulation*, 10(1):146–179, 2012.
- [5] David F. Anderson and Masanori Koyama. An asymptotic relationship between coupling methods for stochastically modeled population processes. *IMA Journal of Numerical Analysis*, 35(4):1757–1778, 2015.
- [6] David F. Anderson and Thomas G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits: engineering approaches to systems and synthetic biology*, pages 3–42. Springer, 2011.
- [7] David F. Anderson and Thomas G. Kurtz. *Stochastic analysis of biochemical systems*, volume 674. Springer, 2015.
- [8] David F. Anderson and Chaojie Yuan. Low variance couplings for stochastic models of intracellular processes with time-dependent rate functions. *Bulletin of mathematical biology*, 81(8):2902–2930, 2019.
- [9] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*. Springer Science & Business Media, 2012.
- [10] Adam Arkin, John Ross, and Harley H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics*, 149(4):1633–1648, 1998.
- [11] Carsten Conradi and Anne Shiu. A global convergence result for processive multisite phosphorylation systems. *Bulletin of mathematical biology*, 77(1):126–155, 2015.
- [12] Mitchell Eithun and Anne Shiu. An all-encompassing global convergence result for processive multisite phosphorylation systems. *Mathematical biosciences*, 291:1–9, 2017.
- [13] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.
- [14] Zhou Fang, Bayu Jayawardhana, Abraham van der Schaft, and Chuanhou Gao. Integral regulation mechanism in phosphorylation cycles. In *Proc. of the 56th IEEE Conference on Decision and Control*. IEEE, 2017.
- [15] Zhou Fang, Bayu Jayawardhana, Arjan van der Schaft, and Chuanhou Gao. Adaptation mechanisms in phosphorylation cycles by allosteric binding and gene autoregulation. *IEEE Transactions on Automatic Control*, 65(8):3457–3470, 2019.
- [16] Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations research*, 56(3):607–617, 2008.
- [17] Peter W. Glynn. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84, 1990.
- [18] Rudyanto Gunawan, Yang Cao, Linda Petzold, and Francis J Doyle. Sensitivity analysis of discrete stochastic systems. *Biophysical journal*, 88(4):2530–2540, 2005.

- [19] Ankit Gupta and Mustafa Khammash. Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks. *SIAM Journal on Scientific Computing*, 35(6):A2598–A2620, 2013.
- [20] Boris N. Kholodenko. Cell-signalling dynamics in time and space. *Nature reviews Molecular cell biology*, 7(3):165–176, 2006.
- [21] Thomas G. Kurtz. Representations of Markov processes as multiparameter time changes. *The Annals of Probability*, pages 682–715, 1980.
- [22] Averil M. Law and W. David Kelton. *Simulation modeling and analysis*, volume 3. Mcgraw-Hill New York, 2007.
- [23] Harley H. McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.
- [24] James R. Norris. *Markov chains*. Number 2. Cambridge university press, 1998.
- [25] Bernhard Ø. Palsson. *Systems biology: properties of reconstructed networks*. Cambridge university press, 2006.
- [26] Sergey Plyasunov and Adam P. Arkin. Efficient stochastic sensitivity analysis of discrete event systems. *Journal of Computational Physics*, 221(2):724–738, 2007.
- [27] Muruhan Rathinam, Patrick W. Sheppard, and Mustafa Khammash. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of chemical physics*, 132(3), 2010.
- [28] Patrick W. Sheppard, Muruhan Rathinam, and Mustafa Khammash. A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems. *The Journal of chemical physics*, 136(3), 2012.
- [29] R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *J. Theor. Biol*, 218:309–321, 2002.
- [30] Rishi Srivastava, David F. Anderson, and James B. Rawlings. Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models. *The Journal of chemical physics*, 138(7), 2013.
- [31] Rishi Srivastava and James B. Rawlings. Parameter estimation in stochastic chemical kinetic models using derivative free optimization and bootstrapping. *Computers & chemical engineering*, 63:152–158, 2014.
- [32] Elizabeth Skubak Wolf and David F. Anderson. A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks. *The Journal of chemical physics*, 137(22), 2012.
- [33] Elizabeth Skubak Wolf and David F. Anderson. Hybrid pathwise sensitivity methods for discrete stochastic models of chemical reaction systems. *The Journal of chemical physics*, 142(3), 2015.

A Numerical methodology

We fix $n = 5$ in the model (19) throughout the numerical experiments. For each method, let N denote the number of independent Monte Carlo replications averaged to form the sensitivity estimator. If one replication requires p_m component paths under method m , then the total number of simulated paths is Np_m .

For example, in Application 1, MSC, CRP, and Indep require $K + 1$ component paths per replication, while SC requires $2K$ component paths per replication. Thus, N replications correspond to $(K + 1)N$ simulated paths for MSC, CRP, and Indep, and $2KN$ simulated paths for SC. In Application 2, the $O(\varepsilon_1^2)$ -bias estimator requires 2 component paths per replication, while the $O(\varepsilon_1^4)$ -bias estimator requires 4 component paths per replication. Thus, N replications correspond to $2N$ and $4N$ simulated paths, respectively. In Application 3, the centered finite-difference estimator for the third derivative requires 4 component paths per replication, so N replications correspond to $4N$ simulated paths.

1. Numerical reference values.

For each target sensitivity $D_j(\theta)$, we compute a numerical reference value $D_{j,\text{ref}}$ using MSC with a large Monte Carlo replication count, $N_{\text{ref}} = 2^{17}$, and a small application-specific perturbation size. The finite-difference estimator and perturbation size used to construct the reference value are specified in the corresponding application-specific appendix.

2. Empirical calibration of the perturbation size.

For each coupling method m , target sensitivity $D_j(\theta)$, and Monte Carlo replication count $N_0 = 2^8$, we consider perturbations of the form

$$\varepsilon_{m,j}(C) = C|\theta_j|N_0^{-\gamma_m}, \quad (21)$$

where γ_m denotes the exponent appropriate to the coupling method and finite-difference estimator under consideration, as determined by Section 4 when such a theoretical result is available. For CRP, for which we have no corresponding theory, the exponent used in the numerical experiments is treated only as an empirical tuning convention.

For each candidate value of C , we independently repeat the complete estimation procedure based on N_0 replications $L_{\text{cal}} = 20$ times. If $\widehat{D}_{m,j}^{(\ell)}(C; N_0)$ denotes the estimate from repetition ℓ , we compute

$$\widehat{\text{RMSE}}_{m,j}(C; N_0) = \left[\frac{1}{L_{\text{cal}}} \sum_{\ell=1}^{L_{\text{cal}}} \left(\widehat{D}_{m,j}^{(\ell)}(C; N_0) - D_{j,\text{ref}} \right)^2 \right]^{1/2}.$$

We denote by $\widehat{C}_{m,j}$ the candidate value that minimizes this empirical RMSE. For other Monte Carlo replication counts N , the perturbation is then chosen as

$$\varepsilon_{m,j}(N) = \widehat{C}_{m,j}|\theta_j|N^{-\gamma_m}. \quad (22)$$

3. Evaluation at fixed path-count budgets.

Let p_m denote the number of component paths required to generate one Monte Carlo replication for the method and finite-difference estimator under consideration. For a fixed total path-count budget B , the Monte Carlo replication count is

$$N_m(B) = \frac{B}{p_m},$$

where the budgets are selected so that $N_m(B)$ is an integer for every method being compared. Each method is then evaluated using $N_m(B)$ replications and the perturbation size in (22).

4. Estimation of the RMSE.

For each method and path-count budget, we repeat the complete estimation procedure $L_{\text{eval}} = 500$ times. The reported RMSE for sensitivity $D_j(\theta)$ is

$$\widehat{\text{RMSE}}_{m,j}(B) = \left[\frac{1}{L_{\text{eval}}} \sum_{\ell=1}^{L_{\text{eval}}} \left(\widehat{D}_{m,j}^{(\ell)}(B) - D_{j,\text{ref}} \right)^2 \right]^{1/2}.$$

The resulting RMSEs are plotted against the relevant path-count budget on a base-2 log-log scale.

B Application 1: implementation details

In Application 1, with g defined in (20), the target quantities are all $K = 22$ first-order sensitivities

$$D_j(\theta) = \frac{\partial}{\partial \theta_j} g(\theta), \quad j \in \{1, \dots, K\}.$$

The estimators being compared use the forward finite-difference approximation introduced in Section 2.

For the numerical reference values, we use the centered finite-difference estimator

$$\widehat{D}_{j,\text{ref}} = \frac{1}{N_{\text{ref}}} \sum_{i=1}^{N_{\text{ref}}} \frac{f\left(X_{[i]}^{\theta+h_j^{\text{ref}} e_j}(T)\right) - f\left(X_{[i]}^{\theta-h_j^{\text{ref}} e_j}(T)\right)}{2h_j^{\text{ref}}},$$

where

$$N_{\text{ref}} = 2^{17}, \quad h_j^{\text{ref}} = 0.005|\theta_j|.$$

The required paths for all parameter directions are generated jointly using MSC. The resulting reference values are checked for stability by repeating the calculation with $h_j^{\text{ref}}/2$. The sum of the absolute changes in the 22 reference sensitivities was 3.53, approximately 8.3% of the smallest aggregate RMSE reported in Application 1. We therefore regard the reference values as sufficiently stable for the numerical comparison.

The perturbation exponents and the number of component paths required for one Monte Carlo replication are summarized below.

Table 5: Perturbation exponents and component paths per replication for Application 1.

Method	γ_m	p_m
MSC	$\frac{1}{3}$	$K + 1 = 23$
SC	$\frac{1}{3}$	$2K = 44$
CRP	$\frac{1}{4}$	$K + 1 = 23$
Indep	$\frac{1}{4}$	$K + 1 = 23$

For CRP, for which Section 4 provides no corresponding theoretical scaling result, we use $\gamma_{\text{CRP}} = 1/4$ as an empirical tuning convention.

For each method m and parameter direction j , we calibrate $\widehat{C}_{m,j}$ at $N_0 = 2^8$ using the candidate values

$$C \in \{0.1, 0.2, \dots, 15.0\}$$

and the procedure in Appendix A. Thus, for the final simulations,

$$\varepsilon_{m,j}(N) = \widehat{C}_{m,j} |\theta_j| N^{-\gamma_m}.$$

To compare the methods at identical total path-count budgets, we use

$$B_k = (K+1)(2K)2^k, \quad k \in \{0, \dots, 4\}.$$

This choice is a common multiple of the $K+1$ paths required by MSC, CRP, and Indep and the $2K$ paths required by SC. Consequently,

$$\begin{aligned} N_{\text{MSC}}(B_k) &= N_{\text{CRP}}(B_k) = N_{\text{Indep}}(B_k) = (2K)2^k, \\ N_{\text{SC}}(B_k) &= (K+1)2^k. \end{aligned}$$

For each method and budget, we estimate the componentwise RMSEs using $L_{\text{eval}} = 500$ independent repetitions and report the aggregate quantity

$$\mathcal{R}_m(B_k) = \sum_{j=1}^K \widehat{\text{RMSE}}_{m,j}(B_k),$$

which is the quantity plotted in Figure 2.

C Application 2: implementation details

In Application 2, with g defined in (20), the target quantity is the first-order sensitivity

$$D(\theta) = \frac{\partial}{\partial \theta_1} g(\theta).$$

The estimators being compared are the centered finite difference approximations with bias of order $O(\varepsilon_1^2)$ and $O(\varepsilon_1^4)$.

For the numerical reference value, we use the centered finite-difference estimator

$$\widehat{D}_{\text{ref}} = \frac{1}{N_{\text{ref}}} \sum_{i=1}^{N_{\text{ref}}} \frac{-f\left(X_{[i]}^{\theta+2h^{\text{ref}}e_1}(T)\right) + 8f\left(X_{[i]}^{\theta+h^{\text{ref}}e_1}(T)\right) - 8f\left(X_{[i]}^{\theta-h^{\text{ref}}e_1}(T)\right) + f\left(X_{[i]}^{\theta-2h^{\text{ref}}e_1}(T)\right)}{12h^{\text{ref}}},$$

where

$$N_{\text{ref}} = 2^{17}, \quad h^{\text{ref}} = 0.1|\theta_1|.$$

The four required paths are generated jointly using MSC. The resulting reference value is checked for stability by repeating the calculation with $h^{\text{ref}}/2$. The resulting change in the reference value was 0.046, approximately 0.95% of the smallest RMSE reported in Application 2. We therefore regard the reference values as sufficiently stable for the numerical comparison.

The perturbation exponents $\gamma_{m,b}$ and the numbers of component paths per Monte Carlo replication $p_{m,b}$ are summarized below.

Table 6: Perturbation exponents and component paths per replication for Application 2.

Method	Bias order	$\gamma_{m,b}$	$p_{m,b}$
MSC	$O(\varepsilon_1^2)$	$\frac{1}{5}$	2
MSC	$O(\varepsilon_1^4)$	$\frac{1}{9}$	4
CRP	$O(\varepsilon_1^2)$	$\frac{1}{6}$	2
CRP	$O(\varepsilon_1^4)$	$\frac{1}{10}$	4
Indep	$O(\varepsilon_1^2)$	$\frac{1}{6}$	2
Indep	$O(\varepsilon_1^4)$	$\frac{1}{10}$	4

For CRP, for which Section 4 provides no corresponding theoretical scaling result, we use $\gamma_{\text{CRP},2} = 1/6$ for $O(\varepsilon_1^2)$ and $\gamma_{\text{CRP},4} = 1/10$ for $O(\varepsilon_1^4)$ as empirical tuning conventions.

For each method m and bias order $b \in \{2, 4\}$, we calibrate the corresponding constant $\hat{C}_{m,b}$ at $N_0 = 2^8$ using the candidate values

$$C \in \{0.05, 0.1, \dots, 3.0\}$$

and the procedure in Appendix A. Thus, for the final simulations,

$$\varepsilon_{m,b}(N) = \hat{C}_{m,b} |\theta_1| N^{-\gamma_{m,b}}.$$

To compare the methods at identical total path-count budgets, we use

$$B_k = 2^k, \quad k \in \{8, \dots, 12\}.$$

Consequently, for each method $m \in \{\text{MSC}, \text{CRP}, \text{Indep}\}$,

$$N_{m,2}(B_k) = \frac{B_k}{2} = 2^{k-1},$$

$$N_{m,4}(B_k) = \frac{B_k}{4} = 2^{k-2},$$

where the second subscript denotes the bias order of the finite-difference estimator. For each method, estimator and budget, we estimate the RMSE using $L_{\text{eval}} = 500$ independent repetitions and report the quantity

$$\widehat{\text{RMSE}}_{m,b}(B_k),$$

which is the quantity plotted in Figure 3.

D Application 3: implementation details

In Application 3, with g defined in (20), the target quantity is the third-order sensitivity

$$D(\theta) = \frac{\partial^3}{\partial \theta_3^3} g(\theta).$$

We use the centered finite-difference approximation (8), which has bias of order $O(\varepsilon_3^2)$.

For the numerical reference value, we use the centered finite-difference estimator

$$\widehat{D}_{\text{ref}} = \frac{1}{N_{\text{ref}}} \sum_{i=1}^{N_{\text{ref}}} \frac{f\left(X_{[i]}^{\theta+2h^{\text{ref}}e_3}(T)\right) - 2f\left(X_{[i]}^{\theta+h^{\text{ref}}e_3}(T)\right) + 2f\left(X_{[i]}^{\theta-h^{\text{ref}}e_3}(T)\right) - f\left(X_{[i]}^{\theta-2h^{\text{ref}}e_3}(T)\right)}{2(h^{\text{ref}})^3},$$

where

$$N_{\text{ref}} = 2^{17}, \quad h^{\text{ref}} = 0.2|\theta_3|.$$

The four required paths are generated jointly using MSC. The resulting reference value is checked for stability by repeating the calculation with $h^{\text{ref}}/2$. The resulting change in the reference value was 0.069, approximately 2.36% of the smallest RMSE reported in Application 3. We therefore regard the reference value as sufficiently stable for the numerical comparison.

The perturbation exponents and the number of component paths required for one Monte Carlo replication are summarized below.

Table 7: Perturbation exponents and component paths per replication for Application 3.

Method	γ_m	p_m
MSC	$\frac{1}{9}$	4
CRP	$\frac{1}{10}$	4
Indep	$\frac{1}{10}$	4

For CRP, for which Section 4 provides no corresponding theoretical scaling result, we use $\gamma_{\text{CRP}} = 1/10$ as an empirical tuning convention.

For each method m , we calibrate $\widehat{C}_m$ at $N_0 = 2^8$ using the candidate values

$$C \in \{0.01, 0.02, \dots, 2.00\}$$

and the procedure in Appendix A. Thus, for the final simulations,

$$\varepsilon_m(N) = \widehat{C}_m |\theta_3| N^{-\gamma_m}.$$

To compare the methods at identical total path-count budgets, we use

$$B_k = 2^k, \quad k \in \{8, \dots, 12\}.$$

Consequently,

$$N_{\text{MSC}}(B_k) = N_{\text{CRP}}(B_k) = N_{\text{Indep}}(B_k) = \frac{2^k}{4} = 2^{k-2}, \quad k \in \{8, \dots, 12\}.$$

For each method and budget, we estimate the RMSE using $L_{\text{eval}} = 500$ independent repetitions and report the quantity

$$\widehat{\text{RMSE}}_m(B_k),$$

which is the quantity plotted in Figure 4.