Unbiased estimation of second-order parameter sensitivities for stochastic reaction networks

Quentin BADOLLE
Joint work with Ankit Gupta and Mustafa Khammash

Department of Biosystems Science and Engineering (D-BSSE), ETH Zurich

Online Seminar on Formal Reaction Kinetics and Related Questions (FRK)
March 4, 2025

Intro. part 1 - Introduction to stochastic reaction networks (10 min)

Intro. part 2 - Introduction to parameter sensitivity analysis (5 min)

Introduction to stochastic reaction networks

• Reaction networks are a convenient language to mechanistically describe the evolution of systems in: evolutionary biology, epidemiology, population dynamics, cancerology, neurology and molecular biology, among others.

• Let us consider a system with $N \in \mathbb{N}^*$ types of components which we call species and label $S_1, \ldots, S_N$.

• These species can be types of: individuals in an epidemic, animals in an ecosystem, cells in a body, ion channels in a neuron and molecules in a cell, among others.

• The state of the system is described at any given time by a vector $x = (x_1, \ldots, x_N)$ whose i-th component corresponds to the abundance of species S_i.

• The state can change because of $M \in \mathbb{N}^*$ types of events which we call reactions.

• When a reaction fires, some species interact and are consumed while others are produced.

• This is conveniently expressed as:

$$\nu_{1k} S_1 + \ldots + \nu_{Nk} S_N \longrightarrow \nu'_{1k} S_1 + \ldots + \nu'_{Nk} S_N, \quad k \in [\![1, M]\!]. \tag{1}$$

Stochastic models are often required in molecular biology

• (Observation 1) Cell-to-cell variability within populations of genetically identical cells is omnipresent.

• (Observation 2) Some molecular species are present in low copy numbers within each cell[1].

• We can attribute part of the observed variability in the abundance of molecular species to the random firing of chemical reactions due to the low abundance of some species[2][3].

• (Observation 3) The abundance of molecular species fundamentally takes discrete values.

• Stochastic Reaction Networks (SRNs) account for the random firing of chemical reactions and the discreteness of the abundance of molecular species.

---

[1] Björn Schwanhäusser et al. "Global quantification of mammalian gene expression control". In: *Nature* 473.7347 (2011), pp. 337–342.

[2] Michael B Elowitz et al. "Stochastic gene expression in a single cell". In: *Science* 297.5584 (2002), pp. 1183–1186.

[3] Peter S Swain, Michael B Elowitz, and Eric D Siggia. "Intrinsic and extrinsic contributions to stochasticity in gene expression". In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12795–12800.

- Let us consider an SRN with a finite number $N \in \mathbb{N}^*$ of molecular species.

- The state of the system is described at any given time by a vector $x \in \mathbb{N}^N$ whose i-th component corresponds to the abundance of species S_i.

- Species interact through a finite number $M \in \mathbb{N}^*$ of chemical reactions and every time the k-th reaction fires, the state of the system is displaced by the stoichiometric vector $\zeta_k \in \mathbb{Z}^N$.

- Let us introduce a propensity function $\lambda = (\lambda_k)_{k \in [\![1,M]\!]}$ which depends on the state of the system $x \in \mathbb{N}^N$ and a parameter $\theta \in \mathbb{R}^d$, where $d \in \mathbb{N}^*$.

- The dynamics of the system are expressed as a Continuous-Time Markov Chain (CTMC) $(X_\theta(t))_{t \in \mathbb{R}_+}$ which is fully specified by the stoichiometry vectors ζ_k and propensities λ_k.

The reaction dynamics can be specified in multiple, equivalent ways (1/3)

• Let us introduce the generator $\mathbb{Q}_\theta$ of the CTMC as the (possibly bi-infinite) matrix defined by[4]:

$$[\mathbb{Q}_\theta]_{ij} \coloneqq \begin{cases} \sum_{k \in \mathcal{T}_{ij}} \lambda_k(j, \theta) & \text{if } i \neq j, \\ -\sum_{k=1}^{M} \lambda_k(i, \theta) & \text{if } i = j, \end{cases} \tag{2}$$

where $\mathcal{T}_{ij} \coloneqq \{k \in [\![1, M]\!] \mid i = j + \zeta_k\}$ is the set of reactions which can take the process from state j to state i.

• $\mathbb{Q}_\theta$ can also be seen an operator on probability measures.

• It is used to express the Chemical Master Equation (CME).

• In the general stochastic process literature, $\mathbb{Q}_\theta$ is also know as the transition (or jump rate) matrix and the CME as a Fokker–Planck equation.

[4] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer International Publishing, 2015.

- Let us introduce the generator $\mathbb{A}_\theta$ of the CTMC as the operator defined by[5]:

$$\mathbb{A}_\theta f(x) := \sum_{k=1}^{M} \lambda_k(x,\theta)\Delta_{\zeta_k} f(x) = \sum_{k=1}^{M} \lambda_k(x,\theta)(f(x+\zeta_k) - f(x)), \tag{3}$$

for any bounded, real-valued function f on $\mathbb{N}^N$.

- $\mathbb{A}_\theta$ can also be seen as a (possibly bi-infinite) matrix and it is the adjoint of $\mathbb{Q}_\theta$: $\mathbb{A}_\theta^* = \mathbb{Q}_\theta$.

- It is used to express the Kolmogorov equations and the martingale problem for the process $(X_\theta(t))$[6].

[5] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer International Publishing, 2015.

[6] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

The reaction dynamics can be specified in multiple, equivalent ways (3/3)

• Given a collection of independent, unit-rate Poisson processes $\{(Y_k(t))_{t\in\mathbb{R}_+}\}_{k\in[\![1,M]\!]}$, we associate to each reaction k a counting process $(R_k(t))_{t\in\mathbb{R}_+}$ defined as:

$$R_k(t) := Y_k\left(\int_0^t \lambda_k(X_\theta(s),\theta)ds\right). \tag{4}$$

• Its state increases by 1 every time the k-th reaction fires.

• For an initial state $x \in \mathbb{N}^N$, the dynamics of the CTMC are represented by a stochastic evolution equation[7][8]:

$$X_\theta(t) = x + \sum_{k=1}^{M} \zeta_k R_k(t). \tag{5}$$

• Eq. (4) and (5) specify the so-called Random Time Change (RTC) representation of $(X_\theta(t))$.

[7] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer International Publishing, 2015.

[8] Thomas G Kurtz. "Representations of Markov processes as multiparameter time changes". In: *The Annals of Probability* (1980), pp. 682–715.

- (Takeaway 1) Reaction networks naturally arise as a descriptor across the life sciences.
- (Takeaway 2) Stochastic models are often required in molecular biology.
- (Takeaway 3) Stochastic reaction networks are continuous-time Markov chains.
- (Takeaway 4) The reaction dynamics can be specified in multiple, equivalent ways.

Introduction to parameter sensitivity analysis

• Given an output function $f : \mathbb{N}^N \to \mathbb{R}$ and a terminal time $T \in \mathbb{R}_+$, we are often interested in predicting the expected network output defined as:

$$\Psi_\theta(x, f, t) := \mathbb{E}[f(X_\theta(t))|X_\theta(0) = x] = \mathbb{E}_x[f(X_\theta(t))]. \tag{6}$$

• The value of the parameter vector θ is frequently uncertain.

• It is therefore crucial to quantify to which degree the predictions made about the system are going to be affected by a change in parameters.

• For any $i \in [\![1, d]\!]$, we introduce the first-order parameter sensitivity of the expected network output with respect to θ_i as:

$$S_\theta^{(i)}(x, f, t) := \frac{\partial \Psi_\theta}{\partial \theta_i}(x, f, t). \tag{7}$$

Sensitivities are the basis for theoretical and numerical approaches

• (Motivation 1) Values for sensitivities are used to assess the robustness of predictions to the value of the parameter vector or the presence/absence of specific reactions.

• (Motivation 2) They can also be used in gradient-based inference procedures.

• (Motivation 3) Expressions for sensitivities can be used as theoretical tools to investigate system properties like Robust Perfect Adaptation (RPA)[9][10].

[9] Stephanie K Aoki et al. "A universal biomolecular integral feedback controller for robust perfect adaptation". In: *Nature* 570.7762 (2019), pp. 533–537.

[10] Ankit Gupta and Mustafa Khammash. "Universal structural requirements for maximal robust perfect adaptation in biomolecular networks". In: *Proceedings of the National Academy of Sciences* 119.43 (2022), e2207802119.

Numerical methods to estimate sensitivities broadly fall into three categories

• Numerical methods to estimate sensitivities are well-established across stochastic models, and broadly fall into three categories[11]:

[11] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*. Vol. 57. Springer, 2007.

The generator gradient estimator has emerged as a fourth class of methods

• In a 2024 paper[12], Wang, Blanchet and Glynn introduced the generator gradient estimator, a novel unbiased gradient estimator for jump-diffusion Stochastic Differential Equations (SDEs).

• In a short note[13], we show that the generator gradient estimator is a close analogue to the exact Integral Path Algorithm (eIPA) estimator previously introduced by Gupta, Rathinam and Khammash[14].

• There, we also demonstrate that generator gradient estimator is an adjoint state method.

• The scaling of adjoint state methods is known to be independent of the number of parameters considered, making it well suited for applications which involve deep neural networks like neural SDEs.

• In the case of the eIPA, the scaling only depends on the number of reactions.

---

[12] Shengbo Wang, Jose Blanchet, and Peter Glynn. "An Efficient High-dimensional Gradient Estimator for Stochastic Differential Equations". In: *arXiv preprint arXiv:2407.10065* (2024).

[13] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "The generator gradient estimator is an adjoint state method for stochastic differential equations". In: *arXiv preprint arXiv:2407.20196* (2024).

[14] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

- (Takeaway 1) Parameter sensitivities are the derivatives of expected network outputs.
- (Takeaway 2) Sensitivities are the basis for theoretical and numerical approaches.
- (Takeaway 3) Numerical methods to estimate sensitivities broadly fall into three categories.
- (Takeaway 4) The generator gradient estimator has emerged as a fourth class of methods.

Part 1 - Estimation of second-order parameter sensitivities (5 min)

Part 2 - An integral formula for second-order parameter sensitivities (5 min)

Part 3 - Construction of an unbiased estimator of second-order sensitivities (10 min)

Part 4 - Numerical examples (5 min)

Part 5 - Summary and outlook (5 min)

Part 1 - Estimation of second-order parameter sensitivities

Higher-order sensitivities provide refined information on network responses (1/2)

- For any $(i, j) \in [\![1, d]\!]^2$, we introduce the second-order parameter sensitivities as:

$$S_\theta^{(i,j)}(x, f, t) := \frac{\partial^2 \Psi_\theta}{\partial \theta_i \partial \theta_j}(x, f, t). \tag{8}$$

- (Motivation 1) They reflect the curvature of the response surface.

- (Motivation 2) Parameter inference procedures commonly aim at optimizing a loss by bringing its gradient close to zero.

- In this context, second-order sensitivities are used to quantify the uncertainty in the estimated parameters which results from the local curvature of the loss landscape.

Higher-order sensitivities provide refined information on network responses (2/2)

• (Motivation 3) In a growing number of experimental setups, the network response can be influenced by cues like chemicals or light and the identifiability of the reaction model varies with the signal[15][16].

• In these settings, second-order sensitivities can be used to select the most informative input as part of experimental design.[17][18].

• (Motivation 4) Finally, second-order sensitivities are also involved in efficient optimisation routines like the Newton-Raphson algorithm.

• The optimisation algorithms can be used for system identification and the design of control mechanisms[19][20].

---

[15] Jakob Ruess et al. "Iterative experiment design guides the characterization of a light-inducible gene expression circuit". In: *Proceedings of the National Academy of Sciences* 112.26 (2015), pp. 8148–8153.

[16] Gianpio Caringella, Lucia Bandiera, and Filippo Menolascina. "Recent advances, opportunities and challenges in cybergenetic identification and control of biomolecular networks". In: *Current Opinion in Biotechnology.* 80 (2023), p. 102893.

[17] Yannis Pantazis and Markos A Katsoulakis. "A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics". In: *The Journal of chemical physics* 138.5 (2013).

[18] Zachary R Fox and Brian Munsky. "The finite state projection based Fisher information matrix approach to estimate information and optimize single-cell experiments". In: *PLoS computational biology* 15.1 (2019), e1006365.

[19] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[20] Maurice Filo and Mustafa Khammash. "Optimal parameter tuning of feedback controllers with application to biomolecular antithetic integral control". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 951–957.

The Girsanov transform method provides an unbiased estimator (1/2)

- The Girsanov Transform (GT) method in currently the only unbiased method to estimate second-order parameter sensitivities of SRNs.

- It relies on an estimator $s_\theta^{(i,j)}(f,T)$ defined as[21][22]:

$$
\begin{aligned}
s_\theta^{(i,j)}(f,T) = f(X_\theta(T)) \Bigg[& \sum_{k=1}^{M} \int_0^T \left(\frac{\partial^2 \log \lambda_k}{\partial\theta_i \partial\theta_j}(X_\theta(s),\theta) dR_k(s) - \frac{\partial^2 \lambda_k}{\partial\theta_i \partial\theta_j}(X_\theta(s),\theta) ds \right) \\
& + \sum_{k=1}^{M} \int_0^T \left(\frac{\partial \log \lambda_k}{\partial\theta_i}(X_\theta(s),\theta) dR_k(s) - \frac{\partial \lambda_k}{\partial\theta_i}(X_\theta(s),\theta) ds \right) \\
& \times \sum_{k=1}^{M} \int_0^T \left(\frac{\partial \log \lambda_k}{\partial\theta_j}(X_\theta(s),\theta) dR_k(s) - \frac{\partial \lambda_k}{\partial\theta_j}(X_\theta(s),\theta) ds \right) \Bigg],
\end{aligned}
\tag{9}
$$

where $X_\theta(0) = x$.

- Depending on the field of application, it is sometimes referred to as the likelihood ratio method or the REINFORCE algorithm[23].

[21] Peter W Glynn. "Likelihood ratio gradient estimation for stochastic systems". In: *Communications of the ACM* 33.10 (1990), pp. 75–84.

[22] Sergey Plyasunov and Adam P Arkin. "Efficient stochastic sensitivity analysis of discrete event systems". In: *Journal of Computational Physics* (2007).

[23] Shakir Mohamed et al. "Monte Carlo Gradient Estimation in Machine Learning.". In: *J. Mach. Learn. Res.* (2020).

The Girsanov transform method provides an unbiased estimator (2/2)

- (Observation 1) The applicability of the GT method depends crucially on the existence of the stochastic exponential and the validity of an interchange of derivation and integration.

- These conditions are challenging to verify and have not yet been derived for second-order sensitivities[24].

- (Observation 2) Recall the expression from the previous slide:

$$s_\theta^{(i,j)}(f,T) = f(X_\theta(T))\left[\sum_{k=1}^{M}\int_0^T\left(\frac{\partial^2 \log\lambda_k}{\partial\theta_i\partial\theta_j}(X_\theta(s),\theta)dR_k(s) - \frac{\partial^2\lambda_k}{\partial\theta_i\partial\theta_j}(X_\theta(s),\theta)ds\right)\right.$$
$$\left. + \sum_{k=1}^{M}\int_0^T\left(\frac{\partial \log\lambda_k}{\partial\theta_i}(X_\theta(s),\theta)dR_k(s) - \frac{\partial\lambda_k}{\partial\theta_i}(X_\theta(s),\theta)ds\right)\sum_{k=1}^{M}\int_0^T\left(\frac{\partial \log\lambda_k}{\partial\theta_j}(X_\theta(s),\theta)dR_k(s) - \frac{\partial\lambda_k}{\partial\theta_j}(X_\theta(s),\theta)ds\right)\right].$$

- From this, it is clear that the GT estimator is not defined whenever some of the propensities are proportional to θ_i and/or θ_j with the parameter being evaluated at zero.

- This precludes the investigation of the sensitivity of an output to the presence/absence of a reaction in the important case of mass action kinetics.

- (Observation 3) When applicable, the GT method has been shown to suffer from large variance not only for SRNs but also for a range of other stochastic models[25].

[24] Ting Wang and Muruhan Rathinam. "On the validity of the Girsanov transformation method for sensitivity analysis of stochastic chemical reaction networks". In: *Stochastics* 93.8 (2021), pp. 1227–1248.

[25] Shakir Mohamed et al. "Monte Carlo Gradient Estimation in Machine Learning.". In: *J. Mach. Learn. Res.* (2020).

Finite difference methods provide biased estimators (1/2)

• Pick an $\epsilon \in \mathbb{R}_+^*$ and introduce $\hat{\theta} := (\theta_1^\epsilon, \theta_2^\epsilon, \theta_3^\epsilon, \theta_4^\epsilon)$ specified by:

$$\theta_1^\epsilon := \theta + (e_i + e_j)\epsilon, \quad \theta_2^\epsilon := \theta + e_i\epsilon, \quad \theta_3^\epsilon := \theta + e_j\epsilon, \quad \theta_4^\epsilon := \theta. \tag{10}$$

• Finite difference methods provide a biased estimation of the second-order sensitivity by approximating it as[26]:

$$S_\theta^{(i,j)}(x, f, T) \approx \mathbb{E}_x \left[\frac{f\left(X^{(1)}_{\theta^\epsilon_1}(T)\right) - f\left(X^{(2)}_{\theta^\epsilon_2}(T)\right) - f\left(X^{(3)}_{\theta^\epsilon_3}(T)\right) + f\left(X^{(4)}_{\theta^\epsilon_4}(T)\right)}{\epsilon^2} \right], \tag{11}$$

where each process indexed by ℓ has generator $\mathbb{A}_{\theta^\epsilon_\ell}$.

• Eq. (11) is usually replaced by its centred equivalent, leading to a so-called centred finite difference estimator with bias $O(\epsilon^2)$.

• In addition, the four processes are often coupled to reduce the variance of the estimator. One of these coupling schemes is the split coupling with variance $O(\epsilon^{-3})$[27].

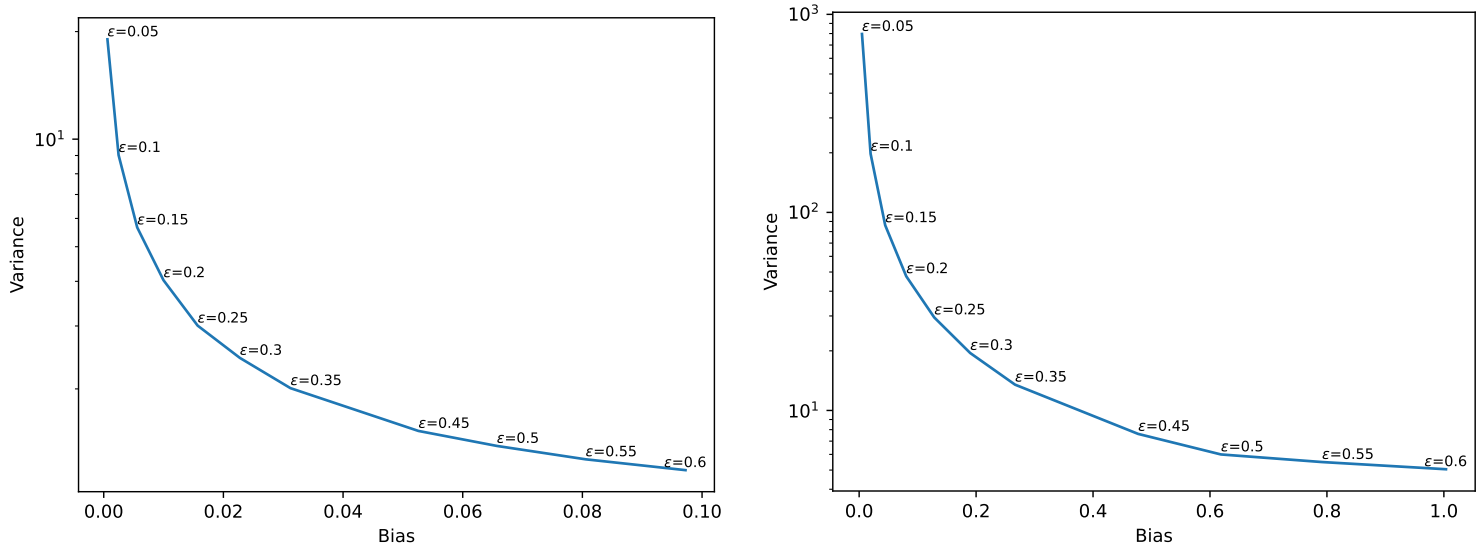
• (Observation 1) Any effort in reducing the bias will translate in a sharp increase in variance.

[26] Elizabeth Skubak Wolf and David F Anderson. "A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks". In: *The Journal of chemical physics* 137.22 (2012).

[27] David F Anderson. "An efficient finite difference method for parameter sensitivities of continuous time Markov chains". In: *SIAM Journal on Numerical Analysis* 50.5 (2012), pp. 2237–2258.

Finite difference methods provide biased estimators (2/2)

• This effect is more pronounced for second-order than for first-order sensitivity for which the corresponding estimator again has bias $O(\epsilon^2)$ but a variance growing like $O(\epsilon^{-1})$ only.



(a) First-order split coupling

(b) Second-order split coupling

Figure: Bias-variance tradeoff for the first- and second-order split coupling methods in the case of a constitutive gene expression network. The y-axis is in log scale.

• (Observation 2) The GT method can generate one sample of the whole Hessian for multiple output functions from a single trajectory of the process.

• For non-centred finite differences, $1 + d + d(d+1)/2$ trajectories are needed to get one sample of the Hessian while $1 + 2d^2$ trajectories are needed for centred finite differences.

Part 2 - An integral formula for second-order parameter sensitivity

- Recall that:

$$\Psi_\theta(x, f, t) = \mathbb{E}[f(X_\theta(t))|X_\theta(0) = x],$$
$$S_\theta^{(i)}(x, f, t) = \frac{\partial \Psi_\theta}{\partial \theta_i}(x, f, t),$$
$$S_\theta^{(i,j)}(x, f, t) = \frac{\partial^2 \Psi_\theta}{\partial \theta_i \partial \theta_j}(x, f, t).$$

Theorem[28]

Let $(X_\theta(t))$ be the CTMC with generator $\mathbb{A}_\theta$ defined in eq. (3). Under appropriate assumptions, $S_\theta^{(i,j)}(x_0, f, T)$ exists and is given by $S_\theta^{(i,j)}(x_0, f, T) = \mathbb{E}_{x_0}[s_\theta^{(i,j)}(f, T)]$ where:

$$\begin{aligned} s_\theta^{(i,j)}(f, T) = \sum_{k=1}^{M} \Bigg[& \int_0^T \frac{\partial^2 \lambda_k}{\partial \theta_i \partial \theta_j}(X_\theta(s), \theta) \Delta_{\zeta_k} \Psi_\theta(X_\theta(s), f, T - s) ds \\ & + \int_0^T \frac{\partial \lambda_k}{\partial \theta_i}(X_\theta(s), \theta) \Delta_{\zeta_k} S_\theta^{(j)}(X_\theta(s), f, T - s) ds \\ & + \int_0^T \frac{\partial \lambda_k}{\partial \theta_j}(X_\theta(s), \theta) \Delta_{\zeta_k} S_\theta^{(i)}(X_\theta(s), f, T - s) ds \Bigg]. \end{aligned} \tag{12}$$

[28] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "Unbiased estimation of second-order parameter sensitivities for stochastic reaction networks". In: *arXiv preprint arXiv:2410.11471* (2024).

- The proof builds on ideas introduced in ref.[29][30][31].

Outline of the proof

Start by introducing the four coupled processes from ref.[32].

- Step 1: Define the second-order finite difference involving the coupled processes and take its limit to zero.
  - Step 1.1: Define a partition of the sample space based on the splitting time between the processes.
  - Step 1.2: Enumerate the reaction counting processes which can have caused the splitting.
  - Step 1.3: Take the limit to zero of the quantities thereby created.
- Step 2: Simplify the expression obtained in step 1.

[29] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer International Publishing, 2015.

[30] Ankit Gupta and Mustafa Khammash. "Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks". In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2598–A2620.

[31] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

[32] Elizabeth Skubak Wolf and David F Anderson. "A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks". In: *The Journal of chemical physics* 137.22 (2012).

Part 3 - Construction of an unbiased estimator of second-order sensitivity

Estimation of the first term in eq. (12) (1/2)

• The Double Bernoulli Path Algorithm (Double BPA or DBPA) follows an approach analogous to that introduced in ref.[33].

• We use the notation $(X_\theta(t,x))$ to explicitly indicate that a process with generator $\mathbb{A}_\theta$ started from state x at time 0, whenever needed.

• Introduce what we will call the first-order auxiliary processes:

$$\begin{cases} X_\theta^{(p,k,1)}(t) & := X_\theta(t, X_\theta(\sigma_p)), \\ X_\theta^{(p,k,2)}(t) & := X_\theta(t, X_\theta(\sigma_p) + \zeta_k), \end{cases} \tag{13}$$

where σ_p is the p-th jump time of the process $(X_\theta(t))$ which we call the main process, taking $\sigma_0 = 0$ for convenience.

• Observe that $(X_\theta^{(p,k,2)}(t))$ starts from the state $X_\theta(\sigma_p)$ perturbed by the stoichiometry vector ζ_k of reaction k.

[33] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

- In the previous slide, the first-order auxiliary processes were introduced as:

$$\begin{cases} X_\theta^{(p,k,1)}(t) & = X_\theta(t, X_\theta(\sigma_p)), \\ X_\theta^{(p,k,2)}(t) & = X_\theta(t, X_\theta(\sigma_p) + \zeta_k). \end{cases}$$

- Using the fact that trajectories of the main process are constant between jump times and the tower property of conditional expectations, observe in eq. (12) that:

$$\begin{aligned} &\sum_{k=1}^{M} \mathbb{E}\left[\int_0^T \frac{\partial^2 \lambda_k}{\partial\theta_i \partial\theta_j}(X_\theta(s),\theta)\Delta_{\zeta_k}\Psi_\theta(X_\theta(s), f, T-s)ds\right] \\ &= \mathbb{E}\left[\sum_{\substack{p=0 \\ \sigma_p<T}}^{\infty}\sum_{k=1}^{M} \frac{\partial^2 \lambda_k}{\partial\theta_i \partial\theta_j}(X_\theta(\sigma_p),\theta)\int_{\sigma_p}^{\sigma_{p+1}\wedge T}\Delta_{\zeta_k}\Psi_\theta(X_\theta(\sigma_p), f, T-s)ds\right] \qquad (14) \\ &= \mathbb{E}\left[\sum_{\substack{p=0 \\ \sigma_p<T}}^{\infty}\sum_{k=1}^{M} \frac{\partial^2 \lambda_k}{\partial\theta_i \partial\theta_j}(X_\theta(\sigma_p),\theta)\int_{T-(\sigma_{p+1}\wedge T)}^{T-\sigma_p}(f(X_\theta^{(p,k,2)}(s)) - f(X_\theta^{(p,k,1)}(s)))ds\right]. \end{aligned}$$

The double Bernoulli path algorithm (1/3)

• The second and third term in eq. (12) are estimated by introducing so-called second-order auxiliary processes similar to the first-order auxiliary processes in eq. (13).

• This approach allows to generate one sample of the whole Hessian for multiple output functions from a single trajectory of the main process.

• Estimates for the average output $\Psi_\theta(x, f, t)$ and its Jacobian can be obtained from the same trajectory of the main process.

• (Refinement 1) The split coupling used in ref.[34] for finite differences is used to couple auxiliary processes and reduce the variance of the estimator.

• (Refinement 2) Bernoulli random variables whose rate depend on the derivatives of the propensities at the current state of the auxiliary processes are introduced to control the number of such processes.

• This provides an efficient way to modulate the computational cost per trajectory of the estimator.

[34] David F Anderson. "An efficient finite difference method for parameter sensitivities of continuous time Markov chains". In: *SIAM Journal on Numerical Analysis* 50.5 (2012), pp. 2237–2258.

The double Bernoulli path algorithm (2/3)

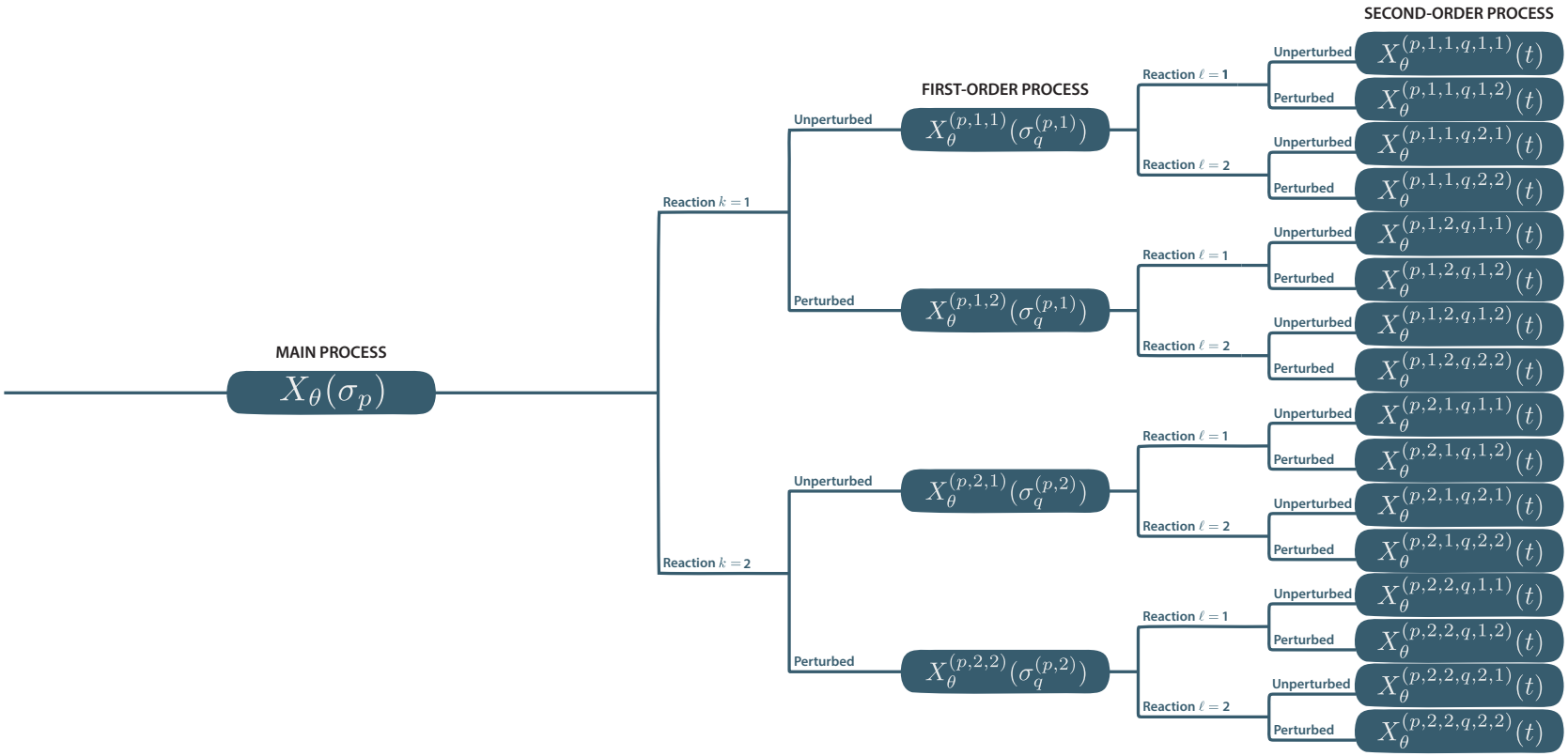


Figure: Representation of the Double BPA as a balanced binary tree where siblings are coupled. The full tree has $4M^2$ leaves where M is the number of reactions.

The double Bernoulli path algorithm (3/3)

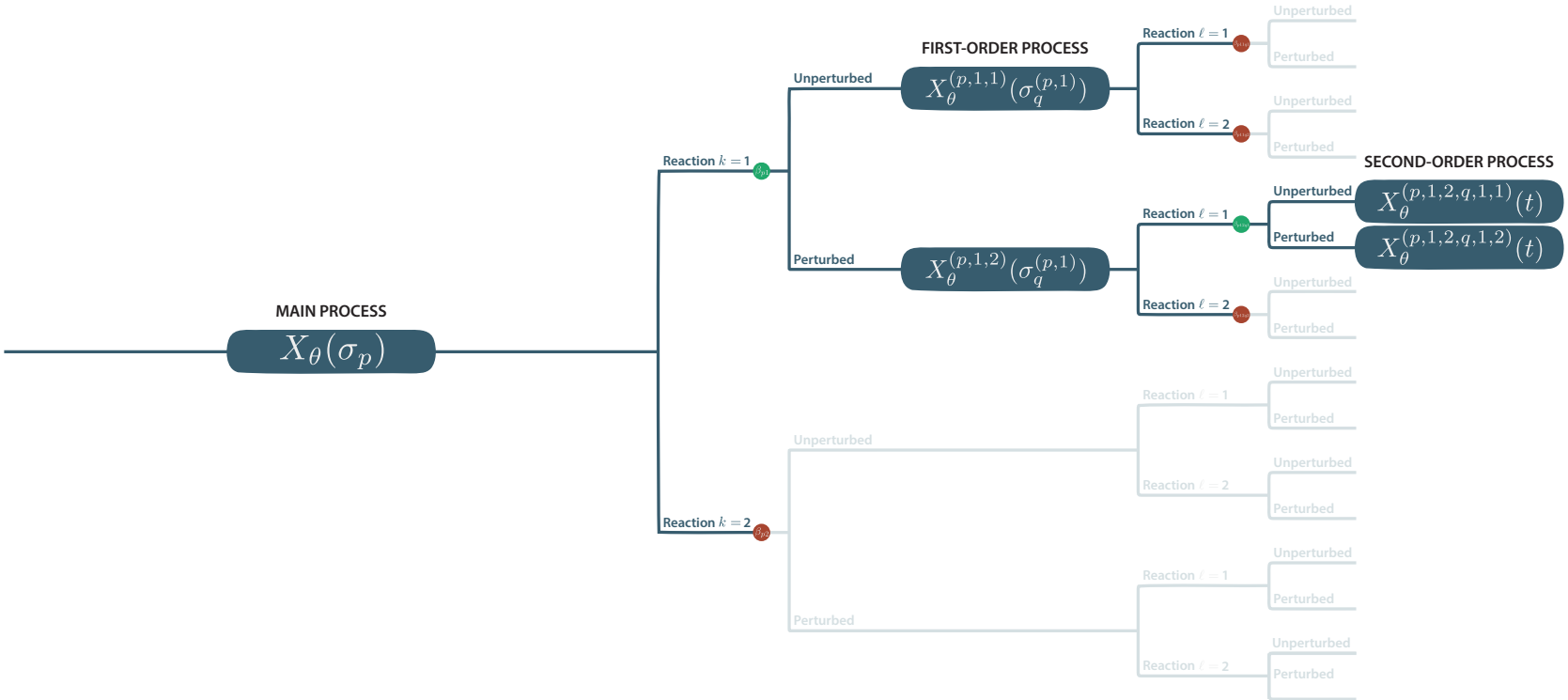


Figure: Representation of the Double BPA as a balanced binary tree where siblings are coupled. Bernoulli random variables are symbolised by dots. They are green when they equal 1 (meaning two coupled auxiliary processes get simulated) and red otherwise. After the introduction of these variables, only parts of the tree get generated while still leaving the estimator unbiased.

Part 4 - Numerical examples

Numerical examples (1/3)

• We now compare the performance of the estimator associated to the DBPA with the only unbiased alternative, which uses the GT method.

• While both are guaranteed to converge to the exact value as the number of samples increases, we nonetheless expect them to lead to estimates with different variances or mean-squared errors for a finite-number of samples.

• Given that the cost per sample, as measured by the average simulation time per sample, *a priori* differs between both methods, we assess their performance in terms of the mean-squared error $\mathrm{MSE}(t_{\mathrm{comp}})$ defined as:

$$\mathrm{MSE}(t_{\mathrm{comp}}) := \sigma_Y^2 / \bar{n}(t_{\mathrm{comp}}), \tag{15}$$

where σ_Y^2 is the variance of one sample from the DBPA or the GT method and $\bar{n}(t_{\mathrm{comp}})$ is the average number of samples generated within a given computational time budget t_{comp}.

• Introducing $\mathcal{M}_0 := t_{\mathrm{comp}} / \bar{n}(t_{\mathrm{comp}})$ the average simulation time per sample and $\mathcal{M} := \mathcal{M}_0 \sigma_Y^2$ the variance-adjusted cost per sample, we have:

$$\mathrm{MSE}(t_{\mathrm{comp}}) = \mathcal{M} / t_{\mathrm{comp}}. \tag{16}$$

• This means that it is sufficient to report $\mathcal{M}$ to fully characterise the relative performance of the two unbiased methods for any computational time t_{comp}.

• We consider a gene expression network under the control of the antithetic integral controller from ref.[35][36]:

$$\begin{aligned} &S_3 \xrightarrow{\lambda_1} S_1 + S_3, \quad S_1 \xrightarrow{\lambda_2} S_1 + S_2, \quad S_1 \xrightarrow{\lambda_3} \emptyset, \\ &S_2 \xrightarrow{\lambda_4} \emptyset, \quad S_2 \xrightarrow{\lambda_5} S_2 + S_4, \quad S_3 + S_4 \xrightarrow{\lambda_6} \emptyset, \quad \emptyset \xrightarrow{\lambda_7} S_3, \end{aligned} \tag{17}$$

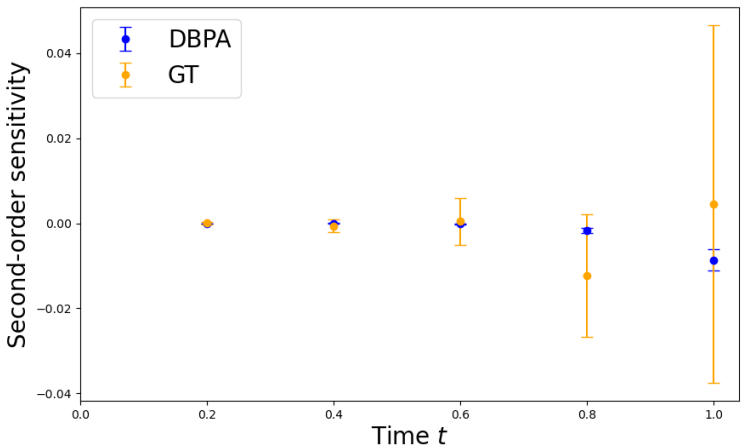
where:

$$\begin{aligned} \lambda_1(x,\theta) &= \theta_1 x_3, & \lambda_2(x,\theta) &= \theta_2 x_1, \\ \lambda_3(x,\theta) &= \theta_3 x_1, & \lambda_4(x,\theta) &= \theta_4 x_2, \\ \lambda_5(x,\theta) &= \theta_5 x_2, & \lambda_6(x,\theta) &= \theta_6 x_3 x_4, \\ \lambda_7(x,\theta) &= \theta_7. \end{aligned}$$

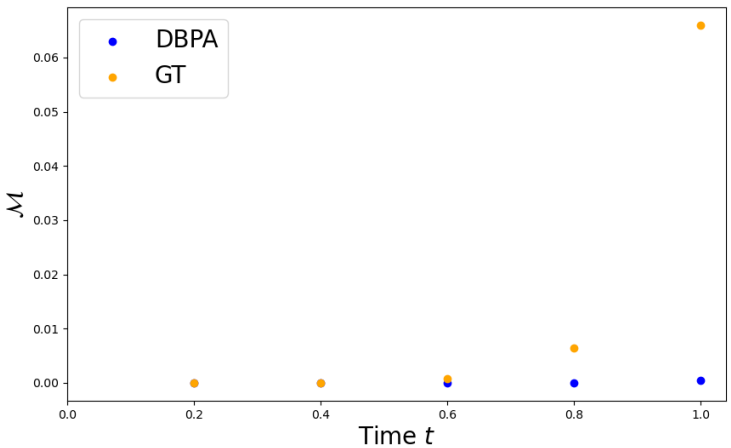
[35] Corentin Briat, Ankit Gupta, and Mustafa Khammash. "Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks". In: *Cell systems* 2.1 (2016), pp. 15–26.

[36] Stephanie K Aoki et al. "A universal biomolecular integral feedback controller for robust perfect adaptation". In: *Nature* 570.7762 (2019), pp. 533–537.

Numerical results (3/3)



(a) $S_\theta^{(5,7)}(x, f, t)$ with $f(x) = (x_2)^2$



(b) $\mathcal{M}$ for $S_\theta^{(5,7)}(x, f, t)$ with $f(x) = (x_2)^2$

Figure: Antithetic integral controller. The sensitivity $S_\theta^{(i,j)}(x, f, t)$ is computed using 10^4 DBPA simulations and 5×10^5 GT simulations. The parameters of the network are set to $\theta_1 = 1.0$, $\theta_2 = 1.0$, $\theta_3 = 2.5$, $\theta_4 = 0.5$, $\theta_5 = 0.0023$, $\theta_6 = 1.0$, $\theta_7 = 1.0$, $\theta_8 = 1.0$, $\theta_9 = 0.5$, $\theta_{10} = 0.0023$. The initial state is chosen to be $x = (0, 0, 0, 0)$. In the panels on left-hand side, error bars correspond to two standard deviations. While both methods are unbiased and will lead to the same sensitivity estimate asymptotically, the DBPA can offer large performance improvements over the GT method by having a lower variance-adjusted cost per sample $\mathcal{M}$, as showcased in the panels on the right-hand side.

Summary and outlook

Summary

• (Summary 1 - part 2) We provided conditions on the existence of second-order sensitivities of an expected output of interest (see manuscript[37]).

• (Summary 2 - part 2) We also derived a new integral representation for these second-order derivatives.

• (Summary 3 - part 3) Based on this formula, we introduced the Double BPA which can generate unbiased samples of the Hessian of an average output.

• (Summary 4 - part 4) We illustrate on numerical examples that the Double BPA can provide a substantial performance improvement over the only unbiased alternative previously available which is based on the GT method.

[37] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "Unbiased estimation of second-order parameter sensitivities for stochastic reaction networks". In: *arXiv preprint arXiv:2410.11471* (2024).

Outlook (1/3)

• (Outlook 1) The Double BPA relies on exact simulations of the reaction dynamics.

• These simulations can be computationally demanding, especially in the presence of time-scale separation.

• In the spirit of ref.[38], we could extend the approach developed here to rely instead on approximate simulations like those from the tau-leap algorithms, using again eq. (12) as the starting point.

• (Outlook 2) In ref.[39], the equivalent of eq. (12) for first-order sensitivities was used in the DeepCME framework.

• In this hybrid Deep Learning-Monte Carlo approach, the expected output $\Psi_\theta(x, f, t)$ was first approximated using a neural network and this surrogate was used instead of the auxiliary paths when evaluating the integral over paths.

• Similarly, we envision leveraging eq. (12) together with a neural approximation of $\Psi_\theta(x, f, t)$ and its first-order sensitivities to avoid the simulation of first- and second-order auxiliary paths altogether.

---

[38] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

[39] Ankit Gupta, Christoph Schwab, and Mustafa Khammash. "DeepCME: A deep learning framework for computing solution statistics of the chemical master equation". In: *PLoS computational biology* 17.12 (2021), e1009623.

• (Outlook 3) Let $\alpha := (\alpha_1, \ldots, \alpha_p)$ be a multi-index. We introduce the α-th order parameter sensitivity as:

$$S_\theta^{(\alpha)}(x, f, t) := \partial_\theta^{(\alpha)} \Psi_\theta(x, f, t). \tag{18}$$

• By inspection of the proof for the first-order sensitivity formula in ref.[40][41] and that for second-order sensitivity[42], it becomes clear that under appropriate assumptions, it holds that $S_\theta^{(\alpha)}(x_0, f, T)$ is the expectation of:

$$s_\theta^{(\alpha)}(f, T) = \sum_{k=1}^{M} \sum_{\substack{\beta \text{ s.t.} \\ \mathbf{0} \prec \beta \preccurlyeq \alpha}} \binom{\alpha}{\beta} \left[\int_0^T \partial_\theta^{(\beta)} \lambda_k(X_\theta(s), \theta) \Delta_{\zeta_k} \partial_\theta^{(\alpha-\beta)} \Psi_\theta(X_\theta(s), f, T-s) ds \right]. \tag{19}$$

[40] Ankit Gupta and Mustafa Khammash. "Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks". In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2598–A2620.

[41] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

[42] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "Unbiased estimation of second-order parameter sensitivities for stochastic reaction networks". In: *arXiv preprint arXiv:2410.11471* (2024).

- (Outlook 4) Assume the process $(X_\theta(t))$ is exponentially ergodic with stationary distribution π_θ.

- We define the steady-state first- and second-order sensitivities as:

$$S_\theta^{(i)}(f) := \lim_{t\to\infty} S_\theta^{(i)}(x, f, t), \tag{20}$$

$$S_\theta^{(i,j)}(f) := \lim_{t\to\infty} S_\theta^{(i,j)}(x, f, t). \tag{21}$$

- Introduce F_θ and G_θ as solution of so-called Poisson equations[43].

Theorem

$$\begin{aligned} S_\theta^{(i,j)}(f) = \sum_{k=1}^{M} \sum_{x\in\mathbb{N}^N} \Bigg[& \frac{\partial^2 \lambda_k}{\partial\theta_i \partial\theta_j}(x,\theta)\Delta_{\zeta_k} F_\theta(x) \\ & + \frac{\partial \lambda_k}{\partial\theta_i}(x,\theta)\Delta_{\zeta_k} G_\theta^{(j)}(x) + \frac{\partial \lambda_k}{\partial\theta_j}(x,\theta)\Delta_{\zeta_k} G_\theta^{(i)}(x) \Bigg] \pi_\theta(x). \end{aligned} \tag{22}$$

[43] Patrik Dürrenberger, Ankit Gupta, and Mustafa Khammash. "A finite state projection method for steady-state sensitivity analysis of stochastic reaction networks". In: *The Journal of chemical physics* 150.13 (2019).

Thank you for your attention!

References I

[1] Björn Schwanhäusser et al. "Global quantification of mammalian gene expression control". In: *Nature* 473.7347 (2011), pp. 337–342.

[2] Michael B Elowitz et al. "Stochastic gene expression in a single cell". In: *Science* 297.5584 (2002), pp. 1183–1186.

[3] Peter S Swain, Michael B Elowitz, and Eric D Siggia. "Intrinsic and extrinsic contributions to stochasticity in gene expression". In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12795–12800.

[4] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer International Publishing, 2015.

[5] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[6] Thomas G Kurtz. "Representations of Markov processes as multiparameter time changes". In: *The Annals of Probability* (1980), pp. 682–715.

[7] Stephanie K Aoki et al. "A universal biomolecular integral feedback controller for robust perfect adaptation". In: *Nature* 570.7762 (2019), pp. 533–537.

[8] Ankit Gupta and Mustafa Khammash. "Universal structural requirements for maximal robust perfect adaptation in biomolecular networks". In: *Proceedings of the National Academy of Sciences* 119.43 (2022), e2207802119.

[9] Søren Asmussen and Peter W Glynn. *Stochastic simulation: algorithms and analysis*. Vol. 57. Springer, 2007.

[10] Shengbo Wang, Jose Blanchet, and Peter Glynn. "An Efficient High-dimensional Gradient Estimator for Stochastic Differential Equations". In: *arXiv preprint arXiv:2407.10065* (2024).

[11] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "The generator gradient estimator is an adjoint state method for stochastic differential equations". In: *arXiv preprint arXiv:2407.20196* (2024).

[12] Ankit Gupta, Muruhan Rathinam, and Mustafa Khammash. "Estimation of parameter sensitivities for stochastic reaction networks using tau-leap simulations". In: *SIAM Journal on Numerical Analysis* 56.2 (2018), pp. 1134–1167.

[13] Jakob Ruess et al. "Iterative experiment design guides the characterization of a light-inducible gene expression circuit". In: *Proceedings of the National Academy of Sciences* 112.26 (2015), pp. 8148–8153.

[14] Gianpio Caringella, Lucia Bandiera, and Filippo Menolascina. "Recent advances, opportunities and challenges in cybergenetic identification and control of biomolecular networks". In: *Current Opinion in Biotechnology.* 80 (2023), p. 102893.

[15] Yannis Pantazis and Markos A Katsoulakis. "A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics". In: *The Journal of chemical physics* 138.5 (2013).

[16] Zachary R Fox and Brian Munsky. "The finite state projection based Fisher information matrix approach to estimate information and optimize single-cell experiments". In: *PLoS computational biology* 15.1 (2019), e1006365.

[17] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[18] Maurice Filo and Mustafa Khammash. "Optimal parameter tuning of feedback controllers with application to biomolecular antithetic integral control". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 951–957.

[19] Peter W Glynn. "Likelihood ratio gradient estimation for stochastic systems". In: *Communications of the ACM* 33.10 (1990), pp. 75–84.

[20] Sergey Plyasunov and Adam P Arkin. "Efficient stochastic sensitivity analysis of discrete event systems". In: *Journal of Computational Physics* (2007).

[21] Shakir Mohamed et al. "Monte Carlo Gradient Estimation in Machine Learning.". In: *J. Mach. Learn. Res.* (2020).

[22] Ting Wang and Muruhan Rathinam. "On the validity of the Girsanov transformation method for sensitivity analysis of stochastic chemical reaction networks". In: *Stochastics* 93.8 (2021), pp. 1227–1248.

[23] Elizabeth Skubak Wolf and David F Anderson. "A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks". In: *The Journal of chemical physics* 137.22 (2012).

[24] David F Anderson. "An efficient finite difference method for parameter sensitivities of continuous time Markov chains". In: *SIAM Journal on Numerical Analysis* 50.5 (2012), pp. 2237–2258.

[25] Quentin Badolle, Ankit Gupta, and Mustafa Khammash. "Unbiased estimation of second-order parameter sensitivities for stochastic reaction networks". In: *arXiv preprint arXiv:2410.11471* (2024).

[26] Ankit Gupta and Mustafa Khammash. "Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks". In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2598–A2620.

[27] Corentin Briat, Ankit Gupta, and Mustafa Khammash. "Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks". In: *Cell systems* 2.1 (2016), pp. 15–26.

[28] Ankit Gupta, Christoph Schwab, and Mustafa Khammash. "DeepCME: A deep learning framework for computing solution statistics of the chemical master equation". In: *PLoS computational biology* 17.12 (2021), e1009623.

[29] Patrik Dürrenberger, Ankit Gupta, and Mustafa Khammash. "A finite state projection method for steady-state sensitivity analysis of stochastic reaction networks". In: *The Journal of chemical physics* 150.13 (2019).