Averaging and Sensitivity Analysis for Multiscale Stochastic Reaction Networks

Araz Hashemi[*]
M. Nunez[†], P. Plecháč[‡], D.G. Vlachos[†]

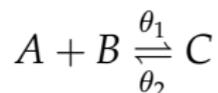[*] Control Science Center, Air Force Research Labs, RQQA
[‡] Dept. of Mathematical Sciences, University of Delaware
[†] Dept. of Chemical Engineering, University of Delaware

araz@udel.edu

January 20, 2016

| Motivation | TTS Rxn Networks | Sensitivity Analysis | Likelihood Ratio | Pathwise Derivatives |
| ●○○○○○○ | ○○○○○○ | ○○○○○ | ○○○○○○○○○○ | ○○○○○○○○ |

Markov Chain Model

- ▶ Example: Chemical species A, B, C

$$A + B \underset{\theta_2}{\overset{\theta_1}{\rightleftharpoons}} C$$

- ▶ $X(t) = [A(t), B(t), C(t)] \in \mathcal{M} \subset \mathbb{Z}^3$ is viewed as a Markov Chain.
  State space $\mathcal{M}$ typically quite large.

- ▶ When reaction r fires, $X \to X + \zeta_r$,
  ζ_r is the **stoichiometric vector** corresponding to rxn r

- ▶ **Propensity Functions:** $a_r(x; \boldsymbol{\theta})$ = instant. prob. of reaction r firing
  Mass Action : $a_r(x; \boldsymbol{\theta}) = \theta_r \cdot b_r(x)$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○●○○○○○	○○○○○○	○○○○○	○○○○○○○○○○	○○○○○○○○

Markov Chain Model

- $X(t)$ has generator $Q = Q(\boldsymbol{\theta})$ determined by propensity functions $a_r(x; \boldsymbol{\theta})$ and stoichiometric vectors ζ_r. Large, but sparse.

$$Q(\boldsymbol{\theta}) = \begin{matrix} & \begin{matrix} \dots & x + \zeta_2 & \dots & x & \dots & x + \zeta_1 & \dots \end{matrix} \\ \begin{matrix} \vdots \\ x \\ \vdots \end{matrix} & \left(\begin{matrix} \dots & a_2(x; \boldsymbol{\theta}) & \dots & -a_0(x; \boldsymbol{\theta}) & \dots & a_1(x; \boldsymbol{\theta}) & \dots \end{matrix} \right) \end{matrix}$$

where $a_0(x; \boldsymbol{\theta}) = \sum_{r=1}^{M} a_r(x; \boldsymbol{\theta})$.

- We are interested in quantities such as $\mathbb{E}_{\boldsymbol{\theta}} \{ f(X(T))) \}$. Most efficient to use Monte Carlo methods to simulate trajectories and average results.

Motivation   TTS Rxn Networks   Sensitivity Analysis   Likelihood Ratio   Pathwise Derivatives
○○●○○○○   ○○○○○○   ○○○○○   ○○○○○○○○○○   ○○○○○○○○

Stiff Systems

ABSORPTION/DESORPTION WITH STIFFNESS ε

$$A \underset{\alpha_2/\varepsilon}{\overset{\alpha_1/\varepsilon}{\rightleftharpoons}} * \qquad A \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B \qquad B \overset{\beta_3}{\rightharpoonup} *$$
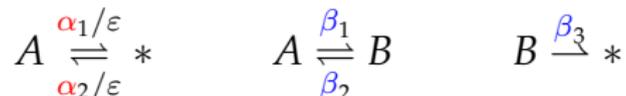
$\varepsilon \ll 1$ is the stiffness parameter

Write $X = [A, *, B]$

Fast Rxns α:

$$\zeta_{\alpha_1} = [-1, 1, 0] \qquad\qquad a_{\alpha_1}^{\varepsilon}(X) = \alpha_1/\varepsilon \cdot X_1$$

$$\zeta_{\alpha_2} = [1, -1, 0] \qquad\qquad a_{\alpha_2}^{\varepsilon}(X) = \alpha_2/\varepsilon \cdot X_2$$

Slow Rxns β:

$$\zeta_{\beta_1} = [-1, 0, 1] \qquad\qquad a_{\beta_1}(X) = \beta_1 \cdot X_1$$

$$\zeta_{\beta_2} = [1, 0, -1] \qquad\qquad a_{\beta_2}(X) = \beta_2 \cdot X_3$$

$$\zeta_{\beta_3} = [0, 1, -1] \qquad\qquad a_{\beta_3}(X) = \beta_3 \cdot X_3$$

| Motivation | TTS Rxn Networks | Sensitivity Analysis | Likelihood Ratio | Pathwise Derivatives |
| 0000000 | 000000 | 00000 | 0000000000 | 00000000 |

Stiff Systems

COMPUTATIONAL PROBLEM

In exact simulation (SSA, Next Reaction, . . .), **every** reaction is computed.

- Prob. of next rxn : $\mathbb{P}\{r^* = r_k\} \propto a_k(x, \theta)$
  As $\varepsilon \to 0$, $\mathbb{P}\{r^* \in \alpha\} \to 1$ and $\mathbb{P}\{r^* \in \beta\} \to 0$.
  Fast rxns dominate computations!

- Each fast rxn advances time clock on micro-scale :
  $\Delta t = O(\varepsilon)$.
  Prohibitive computational load to increment to large time
  horizons (and observe slow dynamics)

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○●○○	○○○○○○	○○○○○	○○○○○○○○○○	○○○○○○○○

Stiff Systems

TAU-LEAPING

[Gillespie 2001, Rathinam et al. 2003, Cao et al 2006]

$$X(t) = X(0) + \sum_{j=1}^{M} R_j(t)\zeta_j$$

$R_j(t) = $ # of reactions of type j which fire by time t

Random Time Change representation:

$$R_j(t) = Y_j \left(\int_0^t a_j\left(X(s); \boldsymbol{\theta}\right) ds \right)$$

Y_j are unit-rate Poisson processes

If the value of $a_j\left(X(s); \boldsymbol{\theta}\right)$ remains essentially constant in $(t, t + \tau]$, then

$$R_j(t + \tau) - R_j(t) \approx \text{Poisson}\left(a_j\left(X(t); \boldsymbol{\theta}\right) \cdot \tau\right)$$

Choosing the appropriate leaping value of τ is a challenge

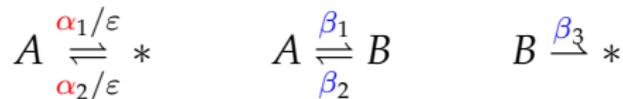| Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○●○	○○○○○○	○○○○○	○○○○○○○○○○	○○○○○○○○

Stiff Systems

STOCHASTIC AVERAGING

[Cao et al. 2005, Samant & Vlachos 2005, E et al. 2007, Kang & Kurtz 2013]

Similar to deterministic quasi-steady state approximations:

$$\dot{X}_t^\varepsilon = f\left(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\right) \qquad X_t^\varepsilon \quad = \text{slow variables}$$

$$\dot{Y}_t^\varepsilon = \frac{1}{\varepsilon} g\left(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\right) \qquad Y_t^\varepsilon \quad = \text{ fast variables}$$

As $\varepsilon \to 0$, one obtains a singularly perturbed system (X, Y) which can be solved by $\dot{X}_t = f(X_t, Y_t)$ under constraint $g(X, Y) = 0$.

$$A \underset{\alpha_2/\varepsilon}{\overset{\alpha_1/\varepsilon}{\rightleftharpoons}} * \qquad A \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B \qquad B \overset{\beta_3}{\longrightarrow} *$$
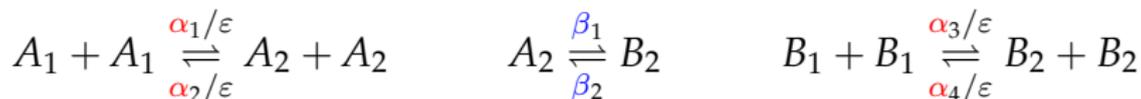
$Y^\varepsilon(t) = (A, *), X^\varepsilon(t) = B$

Can then separate the dynamics of fast-species and slow species in the CME under steady-state constraints.

Motivation   TTS Rxn Networks   Sensitivity Analysis   Likelihood Ratio   Pathwise Derivatives
○○○○○○●   ○○○○○○   ○○○○○   ○○○○○○○○○○   ○○○○○○○○

Stiff Systems

Issues :

1. Rate of convergence in (ε)
2. Should be able to separate dynamics even if one cannot separate species into fast and slow

Example: Fast dimerization, slow isomerization

$$A_1 + A_1 \underset{\alpha_2/\varepsilon}{\overset{\alpha_1/\varepsilon}{\rightleftharpoons}} A_2 + A_2 \qquad A_2 \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B_2 \qquad B_1 + B_1 \underset{\alpha_4/\varepsilon}{\overset{\alpha_3/\varepsilon}{\rightleftharpoons}} B_2 + B_2$$

All species are 'fast', but there are still separable fast and slow dynamics.

▶ For discrete state space, state-space partitioning [Nunez and Vlachos, 2015] allows for a more useful separation of dynamics.

▶ We have gone further to identify the partitioned classes as states of a 'meta' Markov chain which corresponds to a 'macro' rxn network.

▶ This formulation allows for analysis of rate of convergence, as well as implementations of **sensitivity**

$X = [X_1, \ldots, X_{Ns}]$ species

$r_{\alpha_1}, \ldots, r_{\alpha_{Mf}}, r_{\beta_1}, \ldots, r_{\beta_{Ms}}$ Fast and Slow rxns

$\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}] = [\alpha_1, \ldots, \alpha_{Mf}, \beta_1, \ldots, \beta_{Ms}]$ rxn parameters

Propensities:

$$a_{\alpha_j}^{\varepsilon}(x; \boldsymbol{\theta}) = \frac{\alpha_j}{\varepsilon} b_{\alpha_j}(x) \qquad a_{\beta_j}(x; \boldsymbol{\theta}) = \beta_j b_{\beta_j}(x)$$

Take $\mathcal{M}$= State space of $X(t)$, assume $|\mathcal{M}| < \infty$ for simplicity.

As $\varepsilon \to 0$ only fast rxns $\boldsymbol{\alpha}$ fire, so define an equivalence relation on states $s \in \mathcal{M}$ by $s_i \leftrightarrow s_j$ if mutually accessible through only fast rxns. This gives a partition of $\mathcal{M}$ into **fast-classes** $\mathcal{M}_k$:

$$\mathcal{M} = \bigcup_{k=0}^{N_C} \mathcal{M}_k = \left\{ s_1^{(1)}, \ldots, s_{m_1}^{(1)}, s_1^{(2)}, \ldots s_{m_2}^{(2)}, \ldots, \ldots, s_{m_{N_C}}^{(N_C)} \right\}$$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○○○	○●○○○○	○○○○○	○○○○○○○○○○	○○○○○○○○

Formulation

The generator of the exact process can be separated as

$$Q^{\varepsilon}(\boldsymbol{\theta}) = \frac{1}{\varepsilon}\widetilde{Q}(\boldsymbol{\alpha}) + \widehat{Q}(\boldsymbol{\beta})$$

- $\widehat{Q}(\boldsymbol{\beta})$ determined only by slow rxns/propensities $a_{\beta_j}(x, \boldsymbol{\beta})$
- $\widetilde{Q}(\boldsymbol{\alpha}) = \text{diag}\left[Q^{(1)}(\boldsymbol{\alpha}), Q^{(2)}(\boldsymbol{\alpha}), \dots, Q^{(N_C)}(\boldsymbol{\alpha})\right]$
  each $Q^{(k)}$ corresponds to fast-only dynamics of each fast-class $\mathcal{M}_k$ under unscaled propensities $a_{\alpha_j}(x, \boldsymbol{\alpha})$
- Each $\mathcal{M}_k$ is an irreducible class of $\mathcal{M}$ under $\widetilde{Q}(\boldsymbol{\alpha})$, so there is a stationary (steady state) distribution $\pi^{(k)}(\boldsymbol{\alpha})$
- If $X(0) \in \mathcal{M}_k$, then $X(t)$ relaxes to steady-state distn. $\pi^{(k)}$ before slow dynamics take effect
- Slow propensities relax to $\pi^{(k)}$ averages on macro time scale t

$$a_{\beta_j}(X(t); \boldsymbol{\beta}) \approx \bar{a}_{\beta_j}(\mathcal{M}_k, \boldsymbol{\theta}) = \mathbb{E}_{\pi^{(k)}}\left\{a_{\beta_j}(x; \boldsymbol{\beta})\right\}$$

TTS CRN EXAMPLE
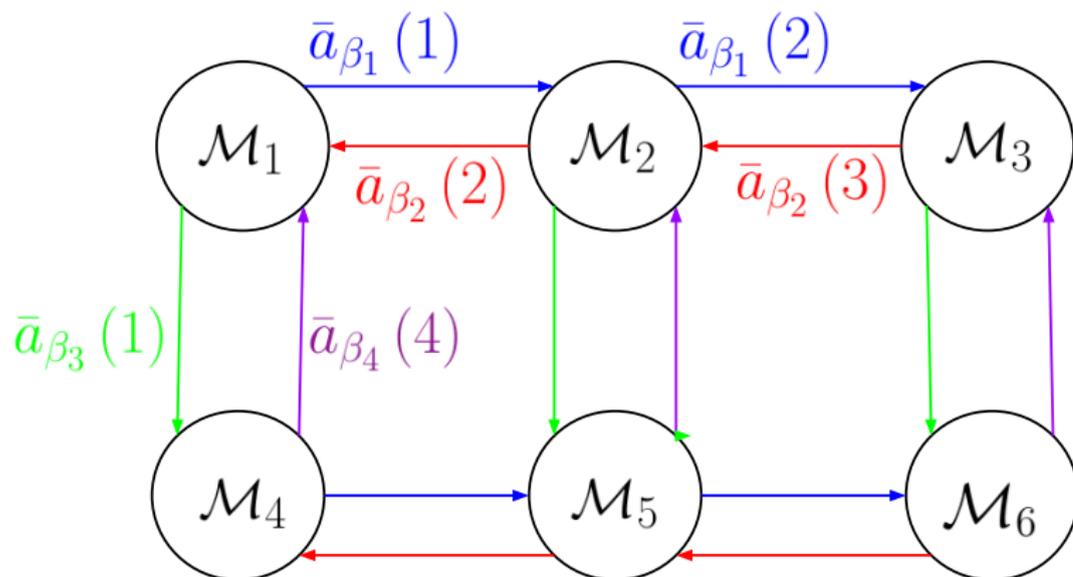
Averaged system $\bar{X}(t)$ forms a meta "macro" Markov chain among fast-classes $\mathcal{M}_k$ with propensities $\bar{a}_\beta\left(\bar{X}(t); \boldsymbol{\beta}, \widetilde{\pi}(\boldsymbol{\alpha})\right)$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
0000000	000●00	00000	0000000000	00000000

Formulation

TTS SSA

(1) Initialize x at macro-time T; $x \in \mathcal{M}_k$ some (unknown) k

(2) Simulate fast-only rxns $Q^{(k)}(\boldsymbol{\alpha})$ until time-averages of observable f and slow propensities a_β relax to steady-state

$$\frac{1}{t} \int_0^t f\left(X^{(k)}(s)\right) ds \to \mathbb{E}_{\pi^{(k)}}\{f(X)\} = \bar{f}(k)$$

$$\frac{1}{t} \int_0^t a_\beta\left(X^{(k)}(s); \boldsymbol{\beta}\right) ds \to \mathbb{E}_{\pi^{(k)}}\{a_\beta(X; \boldsymbol{\beta})\} = \bar{a}_\beta(k; \boldsymbol{\beta})$$

(3) Observe terminal state $\widetilde{x}^{(k)} \sim \pi^{(k)}$. Compute $\bar{a}_{\beta_0} = \sum_{j=1}^{Ms} \bar{a}_{\beta_j}$

(4) Sample time to next slow rxn : $\Delta T \sim \mathcal{E}\mathrm{xp}(\bar{a}_{\beta_0})$

(5) Sample next slow rxn to fire $\beta^* \sim 1/\bar{a}_{\beta_0} [\bar{a}_{\beta_1}, \ldots, \bar{a}_{\beta_{Ms}}]$

(6) Update macro-time $T \leftarrow T + \Delta T$ and move to next fast-class by taking $x = \widetilde{x}^{(k)} + \zeta_{\beta^*}$

(7) Return to **(1)** until macro time horizon is reached

- At the end of a TTS simulation, one obtains terminal state $X(T) = x \sim p_T^0 = p_T^0(X; \boldsymbol{\alpha}, \boldsymbol{\beta})$.
  p_T^0 depends on fast-class stationary distributions $\widetilde{\pi}(\alpha)$ as well as distribution of macro-chain $\bar{p}_T(\bar{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})$.

- What is the error from exact probability distribution $p_T^\varepsilon = p_T^\varepsilon(X^\varepsilon; \boldsymbol{\alpha}/\varepsilon, \boldsymbol{\beta})$?

Theorem 1 (TTS Distribution Error)

Let $\widetilde{\kappa} = \frac{-1}{2} \max\{\mathrm{Re}(\lambda) : \lambda \text{ is a non-zero eigenvalue of a } Q^k\}$. Then

$$||p_T^\varepsilon - p_T^0|| \leq O\left(\varepsilon + \exp\{-\widetilde{\kappa}T/\varepsilon\}\right)$$

- Thus, as long as the macro time horizon T is greater than the relaxation time of the fast dynamics, the induced error of the TTS simulation is $O(\varepsilon)$.

| Motivation | TTS Rxn Networks | Sensitivity Analysis | Likelihood Ratio | Pathwise Derivatives |
| 0000000 | 000000● | 00000 | 0000000000 | 00000000 |

Error Analysis

Often, one is interested in steady state behavior as $T \to \infty$. As the stiffness parameter $\varepsilon \to 0$, this is given by

$$\lim_{\substack{T \to \infty \\ \varepsilon \to 0}} \mathbb{E}_{p_T^\varepsilon} \{f\left(X^\varepsilon(T)\right)\} = \mathbb{E}_{\bar{\pi}} \{\bar{f}\left(\bar{X}\right)\}$$

where $\bar{\pi} = \bar{\pi}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the stationary distribution of the macro-chain associated with generator $\bar{Q} = \widetilde{\pi}(\boldsymbol{\alpha}) \, \widehat{Q}(\boldsymbol{\beta}) \, \mathbf{1}$
Write $\pi = \bar{\pi}\widetilde{\pi}$, where $\widetilde{\pi} = \mathrm{diag}\left[\pi^{(1)}, \ldots, \pi^{(N_C)}\right]$.

Theorem 2 (TTS Expectation Error)

For sufficiently large T, $||p_T^\varepsilon - \pi|| \leq O(\varepsilon)$ and $||\pi^\varepsilon - \pi|| \leq O(\varepsilon)$, where π^ε is the stationary distribution of $\mathcal{M}$ corresponding to Q^ε. Thus for all bounded functions f on $\mathcal{M}$,

$$\left|\mathbb{E}_{\pi^\varepsilon} \{f(X^\varepsilon)\} - \mathbb{E}_{\bar{\pi}} \{\bar{f}(\bar{X})\}\right| \leq ||f||_\infty ||\pi^\varepsilon - \pi|| \leq O(\varepsilon)$$

Motivation	TTS Rxn Networks	**Sensitivity Analysis**	Likelihood Ratio	Pathwise Derivatives
0000000	000000	●0000	0000000000	00000000

Theory

Goal: estimate $\frac{\partial}{\partial \theta_i} \mathbb{E}_{p^\varepsilon(t)} \{f(X^\varepsilon)\}$ for $\theta_i \in \boldsymbol{\alpha}$ or $\theta_i \in \boldsymbol{\beta}$

If sufficient conditions for existence of invariant distribution $\bar{\pi}$ (e.g. the number of fast-classes $\{\mathcal{M}_k\}_{k=1}^{N_C}$ is finite), then
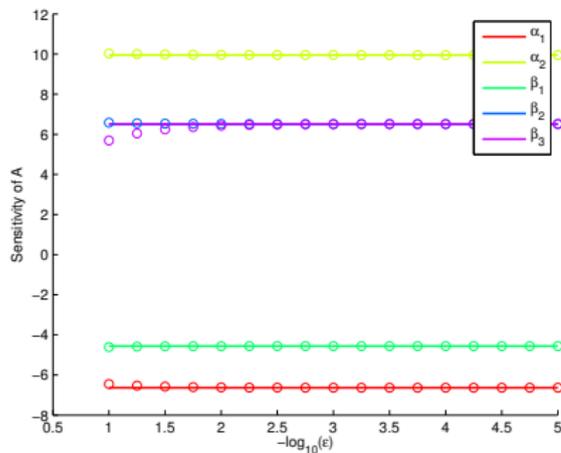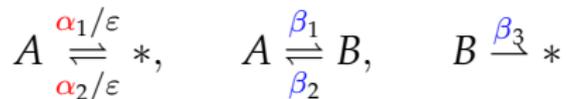
$$\mathbb{E}_{p^\varepsilon(t)} \{f(X^\varepsilon)\} = \mathbb{E}_{\bar{p}(t)} \{\bar{f}(\bar{X})\} + O(\varepsilon)$$
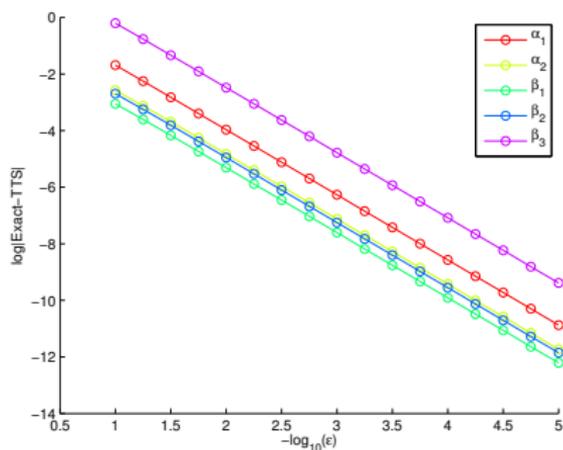
Does it follow that

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{p^\varepsilon(t)} \{f(X^\varepsilon)\} = \frac{\partial}{\partial \theta_i} \mathbb{E}_{\bar{p}(t)} \{\bar{f}(\bar{X})\} + O(\varepsilon)?$$

[Gupta & Khammash 2014] :

$$\lim_{\varepsilon \to 0} \frac{\partial}{\partial \theta_i} \mathbb{E}_{p^\varepsilon(t)} \{f(X^\varepsilon)\} = \frac{\partial}{\partial \theta_i} \mathbb{E}_{\bar{p}(t)} \{\bar{f}(\bar{X})\}$$

Motivation	TTS Rxn Networks	**Sensitivity Analysis**	Likelihood Ratio	Pathwise Derivatives
0000000	000000	0●000	0000000000	00000000

Theory

ANALYTIC SENSITIVITIES OF $|A|$

$$A \underset{\alpha_2/\varepsilon}{\overset{\alpha_1/\varepsilon}{\rightleftharpoons}} *, \qquad A \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B, \qquad B \overset{\beta_3}{\rightarrow} *$$

TTS sensitivity (solid lines) and exact system sensitivity (circles) by ε.

Log-Error of exact sensitivity from TTS sensitivity

Motivation
0000000

TTS Rxn Networks
000000

Sensitivity Analysis
00●00

Likelihood Ratio
0000000000

Pathwise Derivatives
00000000

Stiff vs. Rescaled Sensitivities

STIFF VS. RESCALED SENSITIVITIES

$\boldsymbol{\theta}^{\varepsilon} = [\boldsymbol{\alpha}/\varepsilon, \boldsymbol{\beta}] = [\boldsymbol{\alpha}^{\varepsilon}, \boldsymbol{\beta}].$

What is the difference between $\partial_{\theta_1^{\varepsilon}} \mathbb{E}_{\pi^{\varepsilon}} \{f(X)\} = \partial_{\alpha_1/\varepsilon} \mathbb{E}_{\pi^{\varepsilon}} \{f(X)\}$ vs. $\partial_{\alpha_1} \mathbb{E}_{\pi^{\varepsilon}} \{f(X)\}$?

Proposition 3

Stiff vs. Rescaled Sensitivities

$$\frac{\partial}{\partial \alpha_i^{\varepsilon}} \mathbb{E}_{\pi^{\varepsilon}} \{f(X; \boldsymbol{\theta}^{\varepsilon})\} = \varepsilon \left[\frac{\partial}{\partial \alpha_i} \mathbb{E}_{\pi^{\varepsilon}} \{f(X; \boldsymbol{\theta}^{\varepsilon})\} \right]. \qquad (1)$$

Motivation	TTS Rxn Networks	**Sensitivity Analysis**	Likelihood Ratio	Pathwise Derivatives
0000000	000000	000●0	0000000000	00000000

Stiff vs. Rescaled Sensitivities

Pseudo-inverse representation of steady-state sensitivities :

Lemma 4

Let $Q = Q(\boldsymbol{\theta})$ be the generator of an ergodic continuous-time Markov chain determined by a parameter vector $\boldsymbol{\theta}$. Let $\pi = \pi(\boldsymbol{\theta})$ be the corresponding stationary distribution. Then we can represent the derivatives $\partial_\theta \pi(\boldsymbol{\theta})$ as

$$\frac{\partial \pi}{\partial \theta} = \pi \left(\frac{\partial Q}{\partial \theta} \right) Q^+ \left[\mathbf{1}\pi - I \right]. \tag{2}$$

where Q^+ is the pseudo-inverse of the generator Q.

Possible combination with Finite State Projection algorithm?
[Munsky & Khammash 2006]

Since the macro system $\bar{X}$ is just another reaction network with propensities $\bar{a}_\beta\left(\bar{x}; \beta, \widetilde{\pi}(\boldsymbol{\alpha})\right)$, we can appeal to existing sensitivity analysis estimation methods.

3 categories: **Finite Difference**, **Likelihood Ratio / Girsanov Transform**, and **Pathwise Derivatives**

Finite Differences :

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{\boldsymbol{\theta^0}}\left\{f(X(T))\right\} \approx 1/h \left[\mathbb{E}_{\boldsymbol{\theta^0}+h\cdot e_i}\{f(X(T))\} - \mathbb{E}_{\boldsymbol{\theta^0}}\{f(X(T))\}\right]$$

▶ Low Variance, but each parameter sensitivity must be estimated separately

| Motivation | TTS Rxn Networks | Sensitivity Analysis | Likelihood Ratio | Pathwise Derivatives |
| 0000000 | 000000 | 00000 | ●000000000 | 00000000 |

Theory

Liklihood Ratio / Girsanov Transform :

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{\boldsymbol{\theta^0}} \{f(X(t))\} = \int_{\Omega} f(X(t,\omega)) \frac{\frac{\partial}{\partial \theta_i}\big|_{\boldsymbol{\theta^0}} P(d\omega,t;\boldsymbol{\theta})}{P(d\omega,t;\boldsymbol{\theta})} P(d\omega,t;\boldsymbol{\theta^0})$$
$$= \mathbb{E}_{\boldsymbol{\theta^0}} \{f(X(t)) W_{\theta_i}(t)\}$$

Using the random-time-change representation:

$$X(t) = X(0) + R(t)\boldsymbol{\zeta} = X(0) + \sum_{r=1}^{M} Y_r \left(\int_0^t a_r(X(s),\boldsymbol{\theta}) \, ds \right) \zeta_r$$

$W_\theta(t)$ can be represented as

$$W_\theta(t) = \sum_{r=1}^{M} \int_0^t \frac{\frac{\partial a_r}{\partial \theta}(X(s^-),\boldsymbol{\theta^0})}{a_r(X(s^-),\boldsymbol{\theta^0})} dR_r(s) - \sum_{r=1}^{M} \int_0^t \frac{\partial a_r}{\partial \theta}(X(s^-),\boldsymbol{\theta^0}) \, ds$$

Motivation TTS Rxn Networks Sensitivity Analysis **Likelihood Ratio** Pathwise Derivatives
0000000   000000   00000   ○●○○○○○○○○   00000000

Theory

N_S = number of samples
Likelihood Ratio Estimator:

$$\widehat{\mathbf{LR}}(N_S, \theta_i) \triangleq \frac{1}{N_S} \sum_{n=1}^{N_S} [\widehat{f(x(T))}]_n [\widehat{W_{\theta_i}(T)}]_n.$$

$W_\theta(T)$ is a martingale with zero mean but variance which grows $O(T)$
Makes LR inefficient for estimation at large time horizons

Can reduce variance by *centering*

Centered Likelihood Ratio Estimator:

$$\widehat{\mathbf{CLR}}(N_S, \theta_i) \triangleq \widehat{\mathbf{LR}}(N_S, \theta_i) - \frac{1}{N_S^2} \left\{ \sum_{n=1}^{N_S} [\widehat{f(x(T))}]_n \right\} \left\{ \sum_{n=1}^{N_S} [\widehat{W_{\theta_i}(T)}]_n \right\}$$

STEADY-STATE SENSITIVITIES

How to estimate steady-state sensitivities $\partial_{\theta_i} \mathbb{E}_{\pi(\boldsymbol{\theta})} \{f(X)\}$?

Terminal Distribution : $\mathbb{E}_{p_T(\boldsymbol{\theta})} \{f(X(T))\} = \mathbb{E}_{\pi(\boldsymbol{\theta})} \{f(X)\} + O\left(e^{-\kappa T}\right)$

[Gupta & Khammash 2014] : $\lim_{T \to \infty} \partial_{\theta_i} \mathbb{E}_{p(T)(\boldsymbol{\theta})} \{f(X)\} = \partial_{\theta_i} \mathbb{E}_{\pi(\boldsymbol{\theta})} \{f(X)\}$, so can estimate via LR/CLR with large time horizon $f(X(T))W_{\theta_i}(T)$

Variance grows proportional to $\mathbb{V}\mathrm{ar}\{f(X(T))\}\mathbb{V}\mathrm{ar}\{W_{\theta_i}(T)\} = O(T)$
Bias/Variance tradeoff in choice of T

Instead of using the terminal distribution $f(X(T))$ in LR estimator, could use ergodic average : $1/T \int_0^T f(X(s))ds$.
Larger bias $O(1/T)$, but diminishing variance $O(1/T)$

Then $\mathbb{V}\mathrm{ar}\left\{1/T \int_0^T f(X(s))ds W_\theta(T)\right\} \approx O(1)$, and should have $O(1/T)$ bias

Motivation
0000000

TTS Rxn Networks
000000

Sensitivity Analysis
00000

Likelihood Ratio
0000●00000

Pathwise Derivatives
00000000

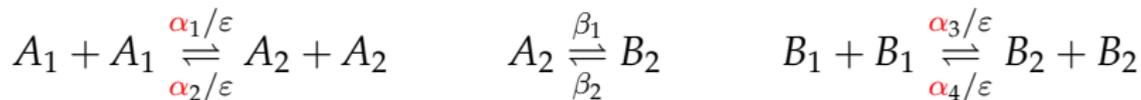CELR Estimator

CELR FOR STEADY-STATE SENSITIVITIES

Thus we propose the **Ergodic Likelihood Ratio**

$$\widehat{\mathbf{ELR}}(N_S, \theta_i) \triangleq \frac{1}{N_S} \sum_{n=1}^{N_S} \frac{1}{T} \left[\widehat{\int_0^T f(x(s))ds} \right]_n [\widehat{W_{\theta_i}(T)}]_n.$$

and again center it for improved efficiency to obtain the
Centered Ergodic Likelihood Ratio

$$\widehat{\mathbf{CELR}}(N_S, \theta_i) \triangleq \widehat{\mathbf{ELR}}(N_S, \theta_i)$$
$$- \frac{1}{N_S^2} \left\{ \sum_{n=1}^{N_S} \frac{1}{T} \left[\widehat{\int_0^T f(x(s))ds} \right]_n \right\} \left\{ \sum_{n=1}^{N_S} [\widehat{W_{\theta_i}(T)}]_n \right\}$$

$$A_1 + A_1 \overset{\alpha_1/\varepsilon}{\underset{\alpha_2/\varepsilon}{\rightleftharpoons}} A_2 + A_2 \qquad A_2 \overset{\beta_1}{\underset{\beta_2}{\rightleftharpoons}} B_2 \qquad B_1 + B_1 \overset{\alpha_3/\varepsilon}{\underset{\alpha_4/\varepsilon}{\rightleftharpoons}} B_2 + B_2$$

$x(0) = [10, 15, 15, 10]$, $|\mathcal{M}| = 5850$, $T = 100$

$\boldsymbol{\alpha} = [1, 2, 3, 4]$, $\boldsymbol{\beta} = [1.5, 2.5]$, $\varepsilon = 10^{-3}$, $\boldsymbol{\theta} = [\boldsymbol{\alpha}/\varepsilon, \boldsymbol{\beta}]$
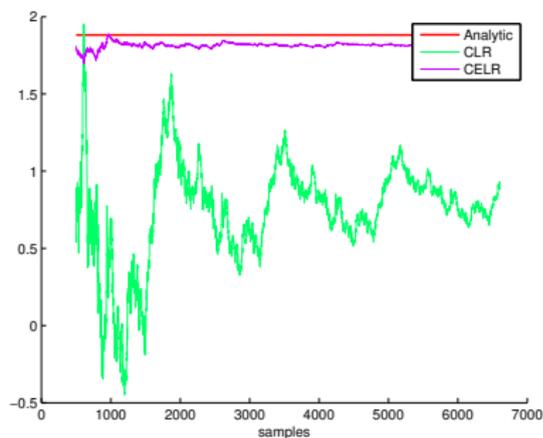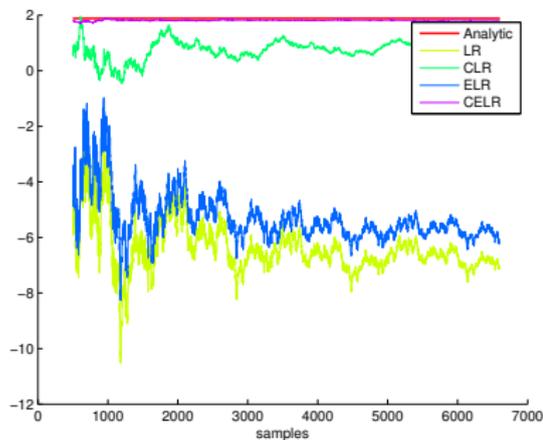
$f(X) = |A_2|$

$\partial_{\theta_1} \mathbb{E}_{\pi^\varepsilon} \{f(X)\} = \partial_{\alpha_1/\varepsilon} \mathbb{E}_{\pi^\varepsilon} \{f(X)\}$:

$$\partial_{\theta_6} \mathbb{E}_{\pi^\varepsilon} \{f(X)\} = \partial_{\beta_2} \mathbb{E}_{\pi^\varepsilon} \{f(X)\} :$$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
0000000	000000	00000	0000000●000	00000000

Multiscale LR
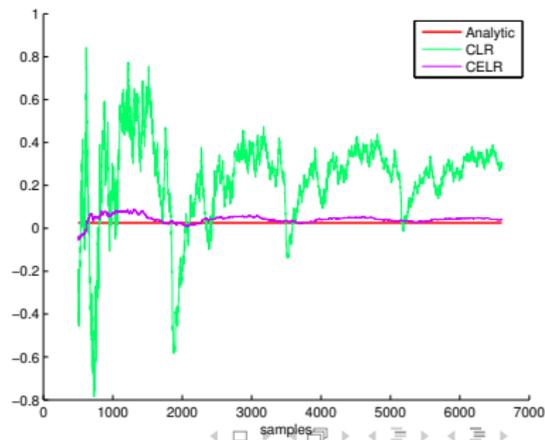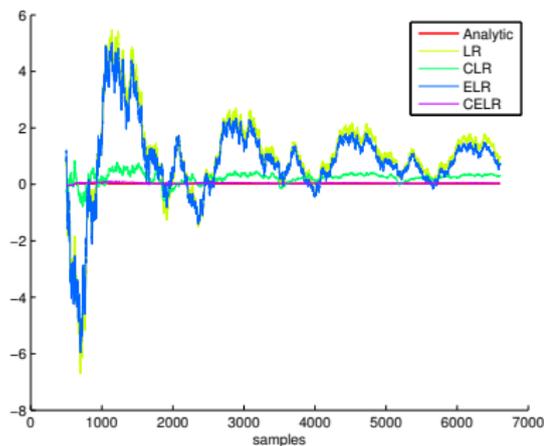
- $\lim_{\varepsilon \to 0} \frac{\partial}{\partial \theta_i} \mathbb{E}_{p^\varepsilon(t)} \{f(X^\varepsilon)\} = \frac{\partial}{\partial \theta_i} \mathbb{E}_{\overline{p}(t)} \left\{ \overline{f}(\overline{X}; \boldsymbol{\theta}) \right\}$

- $\overline{X}(t)$ is just another reaction network with propensities $\overline{a}_{\beta_j}(\overline{x}; \boldsymbol{\theta})$ and observable $\overline{f}(\overline{x}; \boldsymbol{\theta})$

- Thus the macro-sensitivities can be represented by

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_{\overline{p}_T(\boldsymbol{\theta})} \left\{ \overline{f}\left(\overline{X}(T); \boldsymbol{\theta}\right) \right\} = \mathbb{E}_{\overline{p}_T(\boldsymbol{\theta})} \left\{ \frac{\partial}{\partial \theta_i} \overline{f}\left(\overline{X}(T); \boldsymbol{\theta}\right) + \overline{f}\left(\overline{X}; \boldsymbol{\theta}\right) \overline{W}_{\theta_i}(T) \right\}$$

$$\overline{W}_{\theta_i} = \sum_{r=1}^{M_s} \int_0^T \frac{\frac{\partial}{\partial \theta_i} \overline{a}_{\beta_r}\left(\overline{X}(s); \boldsymbol{\theta}\right)}{\overline{a}_{\beta_r}\left(\overline{X}(s); \boldsymbol{\theta}\right)} dR_{\beta_r}(s) - \sum_{r=1}^{M_s} \int_0^T \frac{\partial}{\partial \theta_i} \overline{a}_{\beta_r}\left(\overline{X}(s); \boldsymbol{\theta}\right) ds.$$

- Sensitivity w.r.t. slow parameters $\theta_i = \beta_i$ straightforward, since

$$\frac{\partial}{\partial \beta_i} \overline{f}(\overline{X}; \boldsymbol{\alpha}) = 0 \qquad \frac{\partial}{\partial \beta_i} a_{\beta_r}(\overline{X}; \boldsymbol{\alpha}) = \overline{b}_{\beta_r}(\overline{X}; \boldsymbol{\alpha}) \cdot \delta_{i,r}$$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○○○	○○○○○○	○○○○○	○○○○○○○●○○	○○○○○○○○

Multiscale LR

For $\theta_i = \alpha_i$, the situation is more complicated.

$$\frac{\partial}{\partial \alpha_i} \mathbb{E}_{\bar{p}_T(\boldsymbol{\theta})} \left\{ \bar{f}(\bar{X}(T); \boldsymbol{\alpha}) \right\} = \mathbb{E}_{\bar{p}_T(\boldsymbol{\theta})} \left\{ \frac{\partial \bar{f}}{\partial \alpha_i}(\bar{X}; \boldsymbol{\alpha}) + \bar{f}(\bar{X}; \boldsymbol{\alpha}) \overline{W}_{\alpha_i}(T) \right\}$$

▶ $\bar{f}(\bar{X}; \boldsymbol{\alpha})$ and $\bar{a}_{\beta_r}(\bar{X}; \boldsymbol{\alpha})$ depend on α_i only indirectly through underlying fast-class steady-state measure $\pi^{(\bar{X})}(\boldsymbol{\alpha})$.

▶ However, we may estimate $\partial_{\alpha_i} \bar{f}(\bar{X}; \boldsymbol{\alpha})$ and $\partial_{\alpha_i} \bar{a}_{\beta_r}(\bar{X}; \boldsymbol{\alpha})$ using micro-sensitivity estimates inside each fast-class $\bar{X}$.

$$\frac{\partial}{\partial \alpha_i} \bar{f}(\overline{X}; \boldsymbol{\theta}) \approx \mathbb{E}_{\widetilde{p}_t^{(\overline{X})}} \left\{ \frac{1}{t} \left[\int_0^t f(\widetilde{X}(s)) ds \right] \widetilde{W}_{\alpha_i}(t) \right\}$$

$$\frac{\partial}{\partial \alpha_i} \bar{a}_{\beta_r}(\overline{X}; \boldsymbol{\theta}) \approx \mathbb{E}_{\widetilde{p}_t^{(\overline{X})}} \left\{ \frac{1}{t} \left[\int_0^t a_{\beta_r}(\widetilde{X}(s)) ds \right] \widetilde{W}_{\alpha_i}(t) \right\},$$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○○○	○○○○○○	○○○○○	○○○○○○○○●○	○○○○○○○○

Multiscale LR

$$A \underset{\alpha_2/\varepsilon}{\overset{\alpha_1/\varepsilon}{\rightleftharpoons}} *, \qquad A \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} B, \qquad B \overset{\beta_3}{\rightarrow} *$$

$(A_0, B_0, *_0) = (30, 60, 10)$

$(\alpha_1^\varepsilon, \alpha_2^\varepsilon, \beta_1, \beta_2, \beta_3) = (1/\varepsilon, 1.5/\varepsilon, 2, 1, 0.4)$

1000 TTS samples computed to time horizon $T = 100$

Motivation TTS Rxn Networks Sensitivity Analysis **Likelihood Ratio** Pathwise Derivatives
0000000   000000   00000   000000000●   00000000

Multiscale LR
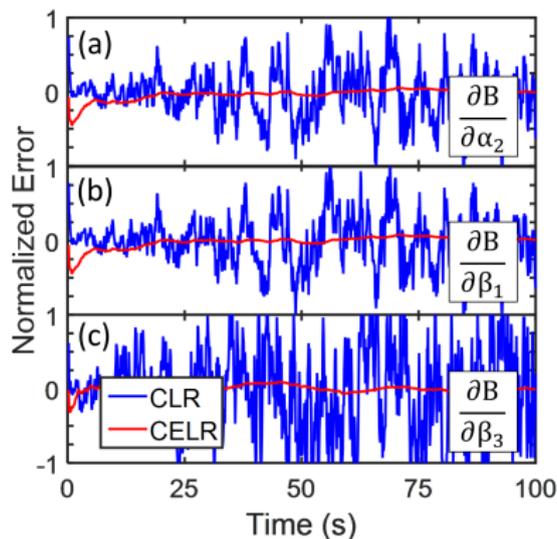
	Percent Error			
	CLR	CELR	CLR	CELR
	$t = 1.3$s		$t = 100$s	
α_1	-1.4	-40.0	-83.5	0.0
α_2	-3.0	-40.3	-63.0	-1.1
β_1	0.2	-39.1	-64.1	1.2
β_2	-1.1	-16.7	-22.6	-2.8
β_3	-1.4	-23.3	20.6	-0.1
	Half-length of 95%			
	Confidence Interval			
α_1	12	8	99	11
α_2	12	7	94	11
β_1	12	7	91	11
β_2	13	10	108	12
β_3	21	14	171	18

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
0000000	000000	00000	0000000000	●0000000

Theory

Pathwise Derivatives :

Application of I.P.A. of discrete-event systems to rxn networks in the **path space** [Sheppard et al. 2012]

$$X(t, \boldsymbol{\theta}, \omega) = X(0, \boldsymbol{\theta}, \omega) + \sum_{r=1}^{M} \zeta_r Y_r \left(S_r(t, \boldsymbol{\theta}, \omega), \omega \right)$$

Internal times : $S_r(t, ; \omega, \boldsymbol{\theta}) = \int_0^t a_r \left(X(s; \omega, \boldsymbol{\theta}); \boldsymbol{\theta} \right)$;

- ► For a fixed ω, the event-times of the Poisson processes $Y_r(\cdot, \omega)$ are fixed

- ► When the (increasing, piecewise-linear) internal times $S_r(t, \boldsymbol{\theta}, \omega)$ hit the event-times of $Y_r(\cdot, \omega)$, reaction channel r will fire

- ► $I_+^r(t) = \inf \{u \geq S_r(t, \boldsymbol{\theta}) : Y_r(u) = Y_r(S_r(t, \boldsymbol{\theta}) + 1)\}$ = the internal time for the next event of $Y_r(\cdot)$ after the current internal time $S_r(t, \boldsymbol{\theta})$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○○○	○○○○○○	○○○○○	○○○○○○○○○○	○●○○○○○○

Theory

Viewing parameter-dependence on trajectory $X_{\boldsymbol{\theta}}(t, \omega)$, can compute derivatives of jump (rxn) times $T_n = T_n(\boldsymbol{\theta}, \omega)$ in terms of derivatives of internal times for each rxn

$$T_{n+1} = T_n + \min_{r=1,\dots,M} \left\{ \frac{I_+^r(T_n) - S_r(T_n)}{a_r(X(T_n))} \right\}$$

$\partial\theta_i T_n(\boldsymbol{\theta}; \omega)$ and $\partial\theta_i S_r\left(T_n(\boldsymbol{\theta}; \omega), \boldsymbol{\theta}; \omega\right)$ can be solved recursively from $T_0 = 0$.

$\partial\theta_i\, T_n(\boldsymbol{\theta}, \omega)$ can be computed in terms of :

$$a_r(X_n; \boldsymbol{\theta}) \text{ and } \partial\theta_i\, a_r(X_n; \boldsymbol{\theta})$$

For path-space differentiation,
$\partial\theta_i\mathbb{E}\left\{f(X(T))\right\} \neq \mathbb{E}\left\{\partial\theta_i f(X(T))\right\}$, since the RHS $\equiv 0$

Instead, must **smooth** the RHS by instead using an integral of the form $L(\boldsymbol{\theta}) = \int_A^B F\left(\boldsymbol{\theta}, X_{\boldsymbol{\theta}}(s)\right) ds$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
0000000	000000	00000	0000000000	00●00000

Theory

Several choices can be considered

- **RPD** : Regularized Pathwise Derivative [Sheppard et al. 2012]

$$L_{\mathrm{RPD}}(\boldsymbol{\theta}) = \frac{1}{2w} \int_{T-w}^{T+w} f\left(X_{\boldsymbol{\theta}}(s)\right) ds \approx f(X_{\boldsymbol{\theta}}(T))$$
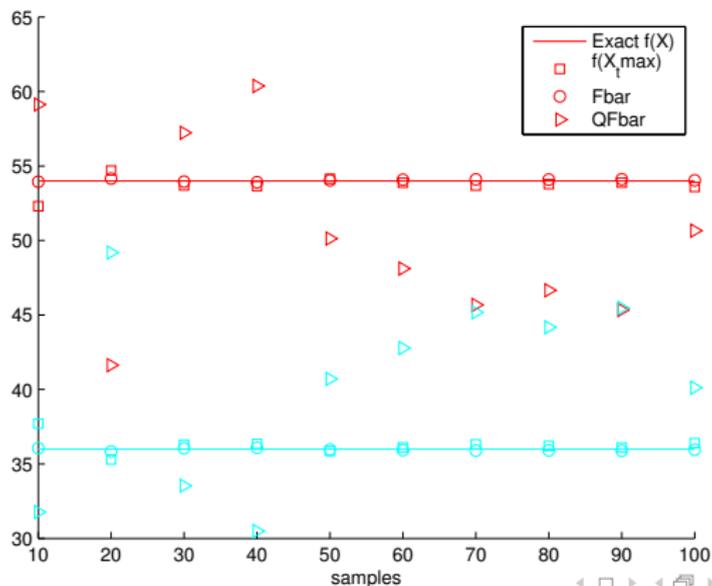
Bias-Variance trade-off in w

- **GS** : Generator Smoothing (Dynkin Formula) [Wolf and Anderson 2015]

$$L_{\mathrm{GS}}(\boldsymbol{\theta}) = f(X_{\boldsymbol{\theta}}(0)) + \int_0^T (Q(\boldsymbol{\theta})f)(X_{\boldsymbol{\theta}}(s))ds$$

Higher variance, but unbiased $\mathbb{E}\{L_{\mathrm{GS}}(\boldsymbol{\theta})\} = \mathbb{E}\{f(X_{\boldsymbol{\theta}}(T))\}$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
○○○○○○○	○○○○○○	○○○○○	○○○○○○○○○○	○○○●○○○○

Theory

For steady-state estimation, the ergodic average will have lower variance than the distribution at terminal time T. Thus we use the **Ergodic Pathwise Derivative**:

$$L_{\text{EPD}}(\boldsymbol{\theta}) = \frac{1}{T} \int_0^T f(X_{\boldsymbol{\theta}}(s))ds = \mathbb{E}_{\pi(\boldsymbol{\theta})}\{f(X))\} + O(1/T)$$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	**Pathwise Derivatives**
0000000	000000	00000	0000000000	0000●000

Theory

For any of the above functionals $L(\boldsymbol{\theta})$, one can compute the pathwise derivative via

$$
\frac{\partial}{\partial \theta_i} \int_0^t f(X^{\boldsymbol{\theta}}(s); \boldsymbol{\theta}) ds = \sum_{l=0}^{N(t)-1} \left[f(\hat{X}_l^{\boldsymbol{\theta}}; \boldsymbol{\theta}) \left(\frac{\partial}{\partial \theta_i} \Delta_l^{\boldsymbol{\theta}} \right) + \left(\frac{\partial}{\partial \theta_i} f(\hat{X}_l^{\boldsymbol{\theta}}; \boldsymbol{\theta}) \right) \Delta_l^{\boldsymbol{\theta}} \right]
$$
$$
+ \left[\frac{\partial}{\partial \theta_i} f(\hat{X}_{N(t)}^{\boldsymbol{\theta}}; \boldsymbol{\theta}) \right] \left(t - T_{N(t)}^{\boldsymbol{\theta}} \right) - f(\hat{X}_{N(t)}^{\boldsymbol{\theta}}, \boldsymbol{\theta}) \left[\frac{\partial}{\partial \theta_i} T_{N(t)}^{\boldsymbol{\theta}} \right],
$$

where $\Delta_l = T_{l+1} - T_l$.

$$
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\boldsymbol{\theta}} \left\{ f(X^{\boldsymbol{\theta}}(T)) \right\} = \mathbb{E}_{\boldsymbol{\theta}} \left\{ \frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta}) \right\} + bias_{smoothing} + bias_{PD}
$$

$bias_{smoothing} = \mathbb{E}_{\boldsymbol{\theta}} \left\{ L(\boldsymbol{\theta}) \right\} - \mathbb{E}_{\boldsymbol{\theta}} \left\{ f(X(T)) \right\}$, can be controlled

The pathwise derivative bias, $bias_{PD} = \partial_{\theta_i} \mathbb{E}_{\boldsymbol{\theta}} \left\{ L(\boldsymbol{\theta}) \right\} - \mathbb{E}_{\boldsymbol{\theta}} \left\{ \partial_{\theta_i} L(\boldsymbol{\theta}) \right\}$, can be more problematic.

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	Pathwise Derivatives
0000000	000000	00000	0000000000	00000●00

Theory

Example: Single-Scale with only fast reactions $A \xrightarrow{\alpha_1} * \xrightarrow{\alpha_2} A$, $f(X(t)) = |A(t)|$, time horizon $t = 37.5$ seconds, small sample size: 100 samples
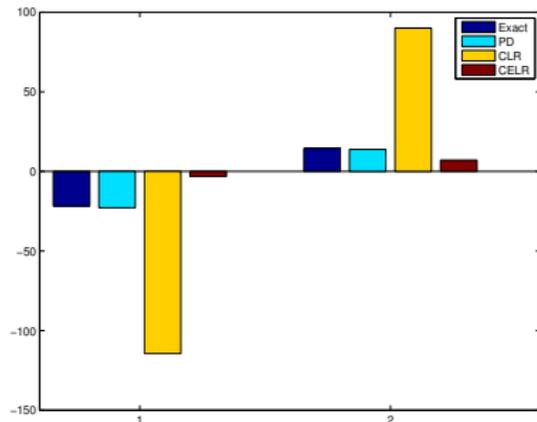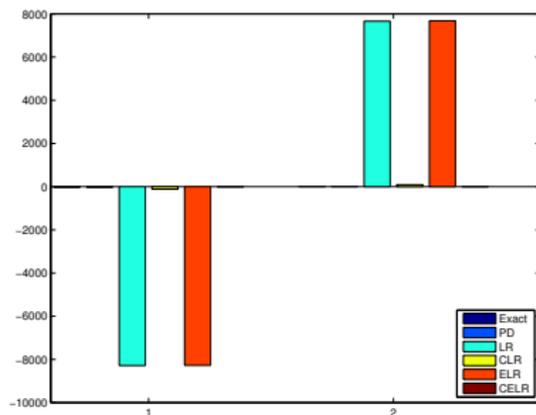
Comparison of (ergodic) Pathwise Derivative against Girsanov Transform estimates.

When applicable, achieves very accurate estimates with a small number of samples; can be used for estimation of micro-sensitivities $\partial_{\alpha_i} \overline{a}_\beta(\overline{X})$

Motivation	TTS Rxn Networks	Sensitivity Analysis	Likelihood Ratio	**Pathwise Derivatives**
○○○○○○○	○○○○○○	○○○○○	○○○○○○○○○○	○○○○○○●○

Multiscale PD

$$\lim_{\varepsilon \to 0} \frac{\partial}{\partial \theta_i} \mathbb{E}_{p_T^\varepsilon(\boldsymbol{\theta})} \left\{ f\left(X^\varepsilon(T) \right) \right\} \frac{\partial}{\partial \theta_i} = \mathbb{E}_{\overline{p}_T(\boldsymbol{\theta})} \left\{ \overline{f}\left(\overline{X}(T) \right) \right\}$$

The process $\overline{X}(T)$ is simply another reaction network with reactions

- propensities $\overline{a}_{\beta_r}(\overline{X}; \boldsymbol{\theta})$
- jump-times $\overline{T}_l^{\boldsymbol{\theta}}$
- internal times $\overline{S}_{\beta_r}^{\boldsymbol{\theta}}(\overline{t}) = \int_0^{\overline{t}} \overline{a}_{\beta_r}\left(\overline{X}(s) \right) ds$

Again, if one can compute $\overline{a}(\overline{X}, \boldsymbol{\theta})$ and $\partial_{\theta_i} \overline{a}(\overline{X}, \boldsymbol{\theta})$, then one can compute a multiscale pathwise derivative

Can combine Multiscale PD with Multiscale LR for a Multiscale Hybrid PD similar to [Wolf & Anderson 2015].

SUMMARY

- TTS framework for multiscale stochastic reaction networks

- Rate of convergence of probability measures

- Pseudo-inverse representation of stationary sensitivities

- CELR estimator for stationary sensitivities

- EPD estimator for stationary sensitivities

- Multiscale LR and PD

A. Hashemi, M. Nunez, P. Plechac, and D. G. Vlachos, " Stochastic Averaging and Sensitivity Analysis for Two Scale Reaction Networks, " arXiv:1509.03802, Sep. 2015.