Sensitivity Analysis of Multiscale Reaction Networks with Stochastic Averaging

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

* Dept. of Mathematical Sciences, University of Delaware
† Dept. of Chemical Engineering, University of Delaware

araz@udel.edu

July 31, 2015
Example: Chemical species $A, B, C$

$$A + B \xrightleftharpoons{\theta_1 \theta_2} C$$

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

State space $\mathcal{M}$ typically quite large.

When reaction $r$ fires, $X \rightarrow X + \zeta_r$,

$\zeta_r$ is the stoichiometric vector corresponding to rxn $r$

Propensity Functions: $a_r(x; \theta) = \text{instant. prob. of reaction } r$ firing

Mass Action: $a_r(x; \theta) = \theta_r \cdot b_r(x)$
**Absorption/Desorption with stiffness $\varepsilon$**

$$A \xrightleftharpoons[\alpha_2/\varepsilon]{\alpha_1/\varepsilon} * \quad A \xrightarrow[\beta_2]{\beta_1} B \quad B \xrightarrow[\beta_3]{\beta} *$$

$\varepsilon \ll 1$ is the stiffness parameter

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

**Fast Rxns $\alpha$:**

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

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

**Slow Rxns $\beta$:**

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

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

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

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

- Prob. of next rxn: \( P\{r^* = r_k\} \propto a_k(x, \theta) \)
  As \( \varepsilon \to 0 \), \( P\{r^* \in \alpha\} \to 1 \) and \( P\{r^* \in \beta\} \to 0 \).
  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)
\[ X = [X_1, \ldots, X_{Ns}] \text{ species} \]

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

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

Propensities:

\[ a_{\alpha_j}^\varepsilon (x; \theta) = \frac{\alpha_j}{\varepsilon} b_{\alpha_j} (x) \quad a_{\beta_j} (x; \theta) = \beta_j b_{\beta_j} (x) \]

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

As \( \varepsilon \to 0 \) only fast rxns \( \alpha \) 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 \textbf{fast-classes} \( \mathcal{M}_k \):

\[ \mathcal{M} = \bigcup_{k=0}^{N_C} \mathcal{M}_k = \left\{ s_1^{(1)}, \ldots, s_{m_1}^{(1)}, s_1^{(2)}, \ldots, s_{m_2}^{(2)}, \ldots, \ldots, s_{m_{NC}}^{(N_C)} \right\} \]
The generator of the exact process can be separated as

$$Q^\varepsilon(\theta) = \frac{1}{\varepsilon} \tilde{Q}(\alpha) + \hat{Q}(\beta)$$

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

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

Averaged system $\bar{X}(t)$ forms a meta “macro” Markov chain among fast-classes $\mathcal{M}_k$ with propensities $\bar{a}_\beta (\bar{X}(t); \beta, \bar{\pi}(\alpha))$
At the end of a TTS simulation, one obtains terminal state $X(T) = x \sim p_T = p_T(X; \alpha, \beta)$. $p_T$ depends on fast-class stationary distributions $\tilde{\pi}(\alpha)$ as well as distribution of macro-chain $\bar{p}_T(\bar{X}; \alpha, \beta)$.

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

**Theorem 1 (TTS Distribution Error)**

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

$$||p_T^\varepsilon - p_T|| \leq O(\varepsilon + \exp\{-\tilde{\kappa}T/\varepsilon\})$$

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
- Ongoing Work

**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

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

where $\bar{\pi} = \bar{\pi}(\alpha, \beta)$ is the stationary distribution of the macro-chain associated with generator $\bar{Q}$.

Write $\pi = \bar{\pi} \tilde{\pi}$, where $\tilde{\pi} = \text{diag} \left[ \pi^{(1)}, \ldots, \pi^{(N_C)} \right]$.

**Theorem 2 (TTS Expectation Error)**

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

$$|\mathbb{E}_{\pi^\varepsilon} \{f(X^\varepsilon)\} - \mathbb{E}_{\bar{\pi}} \{\bar{f}(\bar{X})\}| \leq ||f||_\infty ||\pi^\varepsilon - \pi|| \leq O(\varepsilon)$$
Our goal is to estimate \( \frac{\partial}{\partial \theta_i} \mathbb{E}_{\pi^\varepsilon} \{ f(X^\varepsilon) \} \) for \( \theta_i \in \alpha \) or \( \theta_i \in \beta \).

If the number of fast-classes \( \{ M_k \}_{k=1}^{N_C} \) is finite, (or sufficient conditions for existence of invariant distribution \( \bar{\pi} \)), then

\[
\mathbb{E}_{\pi^\varepsilon} \{ f(X^\varepsilon) \} = \mathbb{E}_{\bar{\pi}} \{ \bar{f}(\bar{X}) \} + O(\varepsilon)
\]

\[
= \sum_{k=1}^{N_C} \bar{f}(k; \alpha) \cdot \bar{\pi}(\alpha, \beta) + O(\varepsilon)
\]

Thus

\[
\frac{\partial}{\partial \theta_i} \mathbb{E}_{\pi^\varepsilon} \{ f(X^\varepsilon) \} = \frac{\partial}{\partial \theta_i} \mathbb{E}_{\bar{\pi}} \{ \bar{f}(\bar{X}) \} + O(\varepsilon)
\]

So, if we can compute the sensitivity of the macro system with \( \bar{f}(\bar{X}) \), then this provides an \( O(\varepsilon) \) estimate of the exact system.
Since the macro system $\bar{X}$ is just another reaction network with propensities $\bar{a}_\beta (\bar{x}; \beta, \bar{\pi}(\alpha))$, we can appeal to existing sensitivity analysis estimation methods.

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

**Finite Differences:**

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

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

\(W_{\theta_i}(t)\) is computable from a sample trajectory in terms of:

\(a_r(X(s); \theta)\) and \(\partial \theta_i a_r(X(s), \theta)\).
Pathwise Derivatives: Application of I.P.A. of discrete-event systems to rxn networks in the path space [Sheppard et al. 2012]

Viewing parameter-dependence on trajectory $X_\theta(t, \omega)$, can compute derivatives of jump (rxn) times $T_n = T_n(\theta, \omega)$

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

$a_r(X_n; \theta)$ and $\partial \theta_i a_r(X_n; \theta)$
In the TTS SSA, we have observable and propensities

\[ \tilde{f}(\bar{X}; \alpha) = \sum_{x \in \bar{X}} f(x) \pi(\bar{X})(x; \alpha) \]

\[ \tilde{a}_{\beta_r}(\bar{X}; \alpha, \beta) = \sum_{x \in \bar{X}} \beta_r b_{\beta_r}(x) \pi(\bar{X})(x; \alpha) = \beta_r \bar{b}_{\beta_r}(\bar{X}; \alpha) \]

Computing sensitivity for \( \theta_i = \beta_i \) using above methods is straightforward, since

\[ \frac{\partial}{\partial \beta_i} \tilde{f}(\bar{X}; \alpha) = 0 \quad \frac{\partial}{\partial \beta_i} a_{\beta_r}(\bar{X}; \alpha) = \bar{b}_{\beta_r}(\bar{X}; \alpha) \cdot \delta_{i,r} \]

Thus, we may compute : \( \bar{W}_{\beta_i} \) or \( \partial \beta_i \bar{T}_n \) for the macro process and use to estimate \( \partial \beta_i \mathbb{E}_{\pi(\alpha, \beta)} \{ f(\bar{X}) \} \)
For $\theta_i = \alpha_i$, the situation is more complicated.

- Induces a product rule:

\[
\frac{\partial}{\partial \alpha_i} \mathbb{E}_{p_T(\theta)} \left\{ \bar{f}(\bar{X}(T); \alpha) \right\} = \mathbb{E}_{p_T(\theta)} \left\{ \frac{\partial \bar{f}}{\partial \alpha_i}(\bar{X}; \alpha) + \bar{f}(\bar{X}; \alpha)\bar{W}_{\alpha_i}(T) \right\}
\]

\[
\frac{\partial}{\partial \alpha_i} \mathbb{E}_{p_T(\theta)} \left\{ \frac{1}{T} \int_0^T \bar{f}(\bar{X}(s); \alpha)ds \right\}
\]

\[
= \tilde{N}(T, \theta) \sum_{n=0} \left[ \left( \frac{\partial \bar{f}}{\partial \alpha_i}(\bar{X}; \alpha) \right) (\Delta \bar{T}_n(\alpha, \beta)) + (\bar{f}(\bar{X}_n; \alpha)) \left( \frac{\partial}{\partial \alpha_i} \Delta \bar{T}_n(\alpha, \beta) \right) \right]
\]

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

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

Cumulative errors from $\bar{f}(\bar{X}_n; \theta), \bar{a}_\beta(\bar{X}_n; \theta)$ could grow large in variance.

$$\frac{d}{dt} \bar{p}(t) = \bar{p}(t)\bar{Q}(I + \delta)$$

$\delta$ = local estimation error (Gaussian).

How does this error propagate?
Pseudo-inverse representation of sensitivities

For M.C. with generator $Q(\theta)$ and stationary distribution $\pi(\theta)$, we can compute sensitivities against the stationary measure by

$$\frac{\partial \pi}{\partial \theta_i} = \pi \left( \frac{\partial Q}{\partial \theta_i} \right) Q^+ [\mathbf{1}_\pi - I]$$

If system is small enough to make solving $\pi Q = 0$ feasible, then additional computation of the pseudo-inverse $Q^+$ provides the steady-state sensitivity.

Possible combination with Finite-State Projection algorithm? [Munsky and Khammash, 2006]
Thanks!