Summary of papers for MLMC and IS for stochastic reaction networks

Manuel Holschbach

January 6, 2025

Introduction

- We consider a stochastic reaction network given in the Kurtz random time change representation

$$X(t) = X(0) + \sum_{k=1}^{R} Y_k\left(\int_0^T a_k(X(s))ds\right)\nu_k$$

where Y_k are independent unit-rate Poisson processes, a_k are the propensity functions and ν_k are the stoichiometric coefficients

- We assume that a_j fulfill the non-negativity assumption and that they follow the mass action kinetics, i.e.

$$a_j(x) := \theta_j \prod_{i=1}^{d} \frac{x_i!}{(x_i - \alpha_{j,i})!} \mathbb{1}_{\{x_i \geq \alpha_{j,i}\}}$$

τ-leap method

- We consider the τ-leap method, given by

$$\hat{X}(0) = x_0$$

$$\hat{X}(t_{n+1}) = \hat{X}(t_n) + \sum_{k=1}^{R} \mathcal{P}_{n,j}(a_j(\hat{X}(t_n)\Delta t)\nu_k$$

- We want to apply the Multi-level Monte Carlo method to reduce the variance of the τ-leap estimator

MLMC τ-leap

- Let $\ell \in \{\ell_0, \dots, L\}$, M be positive integers and let $h_\ell = TM^{-\ell}$
- Denoting by $\hat{X}_\ell$ the process generated by the τ-leap method using step size h_ℓ. we obtain

$$\mathbb{E}[g(\hat{X}_L(T))] = \mathbb{E}[g(\hat{X}_{\ell_0}(T))] + \sum_{\ell=\ell_0+1}^{L} \mathbb{E}[g(\hat{X}_\ell(T))] - \mathbb{E}[g(\hat{X}_{\ell-1}(T))]$$

- Denote the estimator of $\mathbb{E}[g(\hat{X}_{\ell_0}(T))]$ by $\hat{Q}_0$ and the estimator of $\mathbb{E}[g(\hat{X}_\ell(T))] - \mathbb{E}[g(\hat{X}_{\ell-1}(T))]$ by $\hat{Q}_\ell$, i.e.

$$\hat{Q}_0 = \frac{1}{n_0} \sum_i g(\hat{X}_0(T)) \qquad \hat{Q}_\ell = \frac{1}{n_\ell} \sum_i g(\hat{X}_\ell(T)) - g(\hat{X}_{\ell-1}(T))$$

to obtain the unbiased estimator for $\mathbb{E}[g(\hat{X}_L(T))]$

$$\hat{Q} = \hat{Q}_0 + \sum_{\ell=\ell_0+1}^{L} \hat{Q}_\ell$$

MLMC τ-leap

The processes $\hat{X}_\ell$ and $\hat{X}_{\ell-1}$ are coupled by

$$\hat{X}_\ell(t) = \hat{X}_\ell(0) + \sum_k \mathcal{P}_{k,1}\Big(\int_0^t a_k(\hat{X}_\ell(s)) \wedge a_k(\hat{X}_{\ell-1}(s))ds\Big)\nu_k$$
$$+ \sum_k \mathcal{P}_{k,2}\Big(\int_0^t a_k(\hat{X}_\ell(s)) - a_k(\hat{X}_\ell(s)) \wedge a_k(\hat{X}_{\ell-1}(s))ds\Big)\nu_k$$

and

$$\hat{X}_{\ell-1}(t) = \hat{X}_{\ell-1}(0) + \sum_k \mathcal{P}_{k,1}\Big(\int_0^t a_k(\hat{X}_\ell(s)) \wedge a_k(\hat{X}_{\ell-1}(s))ds\Big)\nu_k$$
$$+ \sum_k \mathcal{P}_{k,3}\Big(\int_0^t a_k(\hat{X}_{\ell-1}(s)) - a_k(\hat{X}_\ell(s)) \wedge a_k(\hat{X}_{\ell-1}(s))ds\Big)\nu_k$$

where $\mathcal{P}_{k,i}$ are independent unit rate Poisson processes and $a \wedge b := \min\{a, b\}$.

MLMC τ-leap

- We can bound the variance of the estimator by

$$Var(\hat{Q}) \leqslant \frac{K_0}{n_0} + \sum_{\ell=\ell_0+1}^{L} \frac{C}{n_\ell}[C_1 h_\ell + C_2 h_\ell^2]$$

  which leads to a computational complexity of

$$n_0 h_{\ell_0}^{-1} + \sum_{\ell=\ell_0+1}^{L} \leqslant \epsilon^{-2}(h_{\ell_0}^{-1} + \ln(\epsilon)^2 + \ln(\epsilon)^{-1}\frac{h_{\ell_0}}{M-1})$$

MLMC τ-leap

- We can also couple the exact process X and $\hat{X}_L$ in the same way to obtain an unbiased estimator for $\mathbb{E}[g(X(T))]$ which results in a computational complexity of

$$\epsilon^{-2}(N(h_L + h_L^2) + h_{\ell_0}^{-1} + \ln(\epsilon)^2 + \ln(\epsilon)^{-1}\frac{h_{\ell_0}}{M-1})$$

  which results in a speed-up of

$$\text{Speed-up factor} \approx \frac{N}{Nh_L^2 + h_{\ell_0^{-1}} + \log(\epsilon)}$$

- Thus

$$\text{Speed-up factor} \gtrapprox \min(h_L^{-2}, Nh_{\ell_0})$$

Importance sampling

- Suppose we want to estimate a rare event, e.g. $\mathbb{E}[g(X(t))] = \mathbb{E}[\mathbb{1}_{\{X_i > x\}}] = \mathbb{P}(X_i > x)$
- Crude Monte Carlo performs poorly when trying to estimate such a quantity, as the variance and as such the required number of samples is very high
- Instead, we aim to sample from a different probability density function $\hat{\rho}_Z$ and shift the measure by

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\rho_X(x)dx = \int_{\mathbb{R}} g(x) \underbrace{\frac{\rho_X(x)}{\hat{\rho}_Z(x)}}_{L(x)} \hat{\rho}_Z(x)dx$$

to reduce the variance

Importance sampling

- In the case of the τ-leap method, the measure change is introduced by changing the rates of the Poisson variables, i.e.

$$\overline{P}_{n,k} = \mathcal{P}_{n,k}(\delta_{n,k}^{\Delta t}(\overline{X}_n^{\Delta t})\Delta t))$$

  where $\delta_{n,k}^{\Delta t}(x) \in \mathcal{A}_{x,k}$ is an admissible control parameter
- The new resulting scheme is then

$$\overline{X}(0) = x_0$$

$$\overline{X}(t_{n+1}) = \overline{X}(t_n) + \sum_{k=1}^{R} \overline{P}_{n,k}\nu_k$$

- Our aim is to determine the optimal parameters $\{\delta_n^{\Delta_t}(x)\}$ that minimize the second moment of the IS estimator

Importance sampling

This is equivalent to solving the following optimization problem

$$u_{\Delta t}(N,x) = g^2(x) \text{ and for } n = N-1,\dots,0, \text{ and } \mathcal{A}_x := \bigtimes_{k=1}^{R} \mathcal{A}_{x,k}$$

$$\begin{aligned} u_{\Delta t}(n,x) = & \inf_{\delta_n^{\Delta t}(x)\in\mathcal{A}_x} \exp\left(\left(-2\sum_{k=1}^{R} a_k(x) + \sum_{k=1}^{R}\delta_{n,k}^{\Delta t}(x)\right)\Delta t\right) \\ & \times \sum_{p\in\mathbb{N}^R}\left(\prod_{k=1}^{R}\frac{(\Delta t.\cdot\delta_{n,k}^{\Delta t}(x))^{p_k}}{p_k!}\left(\frac{a_j(x)}{\delta_{n,k}^{\Delta t}(x)}\right)^{2p_k}\right) \\ & \cdot u_{\Delta t}(n+1,\max(0,x+\nu p)) \end{aligned} \tag{1}$$

Since the sum in the update step is infinite, solving this optimization problem directly is difficult.

Importance sampling

Instead, we apply a Taylor expansion around $\Delta t = 0$ and truncate the expression to arrive at the expression

$$\begin{aligned}
&\overline{u}_{\Delta t}(n, x) \\
&= \Delta t \cdot \sum_{k=1}^{R} \inf_{\delta_k \in \mathcal{A}_{x,k}} \left(\frac{a_k(x)^2}{\delta_k} \overline{u}_{\Delta t}(n+1, \max(0, x+\nu_k) + \delta_k \overline{u}_{\Delta t}(n+1, x) \right) \\
&+ (1 - 2\Delta t \sum_{k=1}^{R} a_k(x)) \overline{u}_{\Delta t}(n+1, x)
\end{aligned}$$

The computational cost for solving this equation for the state space $\times_{i=1}^{d}[0, S_i]$ is given by

$$W_{dp}(\overline{S}, \Delta t) \approx \left(\max_{i=1,\ldots,d} S_i \right)^d \frac{T}{\Delta t} R$$

i.e. the cost scales exponentially with the number of chemical species d.

Importance sampling

- Instead of solving this equation directly, we can use a learning based approach by using the ansatz function

$$\hat{u}(t,x;\beta) = \frac{1}{1+e^{-(1-t)(\langle \beta_{space},x\rangle+\beta_{time})-b_0-\beta_0 x_i}}$$

  where b_0 and β_0 are determined by imposing the final condition $\hat{u}(1,x;\beta) \approx g^2(x)$
- The parameters $\beta_{space}, \beta_{time}$ are then chosen to minimize the second moment of the IS estimator using a stochastic optimizer
- Then, the new IS measure is used to generate M sample paths to obtain the estimator

$$\mu^{IS}_{M,\Delta t} = \frac{1}{M}\sum_{i=1}^{M} L_i \cdot g(\overline{X}^{\Delta t,\beta}_{[i],N})$$

Importance sampling

- The computational cost for this approach is given by

$$W_{IS-TL}(TOL) = W_{pl}(I, M_0, \Delta t_{pl}) + const \frac{Var[g(\overline{X}_N^{\Delta t, \beta})L]}{TOL^3}$$

where $W_{pl}(I, M_0, \Delta t_{pl})$ is the offline cost for learning the parameters $\beta_{space}, \beta_{time}$

- The computational cost for standard MC-TL is given by

$$W_{MC-TL}(TOL) = const \frac{Var[g(\hat{X}_N^{\Delta t})]}{TOL^3}$$

- Since we can achieve a low offline learning cost, the resulting variance reduction is given by

$$Var[g(\overline{X}_N^{\Delta t, \beta})L] \ll Var[g(\hat{X}_N^{\Delta t})]$$

HJB equations

- Another way to deal with the equations in (1) is by considering the continuous time analogue, i.e.

$$\tilde{u}(T,x) = g^2(x)$$

$$-\frac{d\tilde{u}}{dt}(t,x) = \inf_{\delta(t,x)\in\mathcal{A}_x}\left(-2\sum_{k=1}^{R} a_k(x) + \sum_{k=1}^{R}\delta_k(t,x)\right)\tilde{u}(t,x)$$
$$+\sum_{k=1}^{R}\frac{a_k(x)}{\delta_k(t,x)}\tilde{u}(t,\max(0,x+\nu_k))$$

- Assuming that $\tilde{u}(t,x) > 0$, we obtain the optimal parameters

$$\tilde{\delta}_k(t,x) = a_k(x)\sqrt{\frac{\tilde{u}(t,\max(0,x+\nu_k))}{\tilde{u}(t,x)}}$$

- To ensure this is the case, we replace the indicator function $g(x) = \mathbb{1}_{\{x_i>x\}}$ by the sigmoid

$$\tilde{g}(x) = \frac{1}{\quad} > 0$$

Markovian projection

- However, the cost of solving the HBJ equations also sclaes exponentially with the number of chemical species d
- To address this problem, we aim to project the system to a lower dimensional space

$$P : \mathbb{R}^d \to \mathbb{R}^{\overline{d}}$$
$$x \mapsto Px$$

  where $\overline{d} \ll d$
- In the case of rare event estimation, i.e. $g(x) = \mathbb{1}_{\{x_i > x\}}$, the projection is given by

$$P(x) = \left(0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0\right) x$$

Markovian projection

- Assuming that a_k are bounded, the projection $S(t) = P(X(t))$ has the same distribution as the process

$$\overline{S}(t) = P(x_0) + \sum_{k=1}^{R} \overline{\mathcal{P}}_k\left(\int_0^t \overline{a}_k(\tau, \overline{S}(\tau))d\tau\right)P(\nu_k)$$

  where $\overline{\mathcal{P}}_k$ are independent unit rate Poisson processes and

$$\overline{a}_k(t,s) = \mathbb{E}[a_k(X(t))|P(X(t)) = s, X(0) = x_0]$$

- We can approximate the Markovian propensities $\overline{a}_k$ via L^2-regression by

$$\overline{a}_k(\cdot,\cdot) \approx \arg\min_h \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{n=0}^{N-1}(a_k(\tilde{X}^{\Delta t}_{[m],n}) - h(t_n, P(\tilde{X}^{\Delta t}_{[m],n})))^2$$

Markovian projection

- We can use the MP approach to address the curse of dimensionality when solving the HJB equations
- To that end, consider for a suitable projection P and final condition $\tilde{g}(P(x)) = g(x)$

$$\tilde{u}_{\overline{d}}(T, s) = \tilde{g}^2(s)$$

$$\frac{d\tilde{u}_{\overline{d}}}{dt}(t, s) = -2 \sum_{k=1}^{R} \overline{a}_k(t, s)(\sqrt{\tilde{u}_{\overline{d}}(t, s)(t, \max(0, s + \overline{\nu}_j))} - \tilde{u}_{\overline{d}}(t, s)$$

for $t \in [0, T], s \in \mathbb{N}^{\overline{d}}$

- Using the solution to these equations, we can obtain the optimal parameters

$$\overline{\delta_k}(t, x) = a_k(x)\sqrt{\frac{\tilde{u}_{\overline{d}}(t, \max(0, P(x + \nu_k)))}{\tilde{u}_{\overline{d}}(t, P(x))}}$$

Combining MLMC and IS

- Our goal is to combine the MLMC approach and the IS approach
- To that end, consider

$$\mathbb{E}[g_L] = \sum_{\ell=1}^{L} \mathbb{E}[g_\ell - g_{\ell-1}] + \mathbb{E}[g_0]$$

where

$$\mathbb{E}[g_0] = \int g_0 d\mathbb{P}_0, \quad \mathbb{E}[g_\ell - g_{\ell-1}] = \int (g_\ell - g_{\ell-1}) d\mathbb{P}_\ell$$

where $\mathbb{P}_\ell$ is the coupling measure of level ℓ and $\mathbb{P}_0$ is the single level measure

Combining MLMC and IS

- Consider now the simple case of one chemical species and one reaction
- In that case, the coupling on level ℓ is done in the following way

$$\begin{aligned}
&\overline{X}_\ell(t_{n+1}) - \overline{X}_{\ell-1}(t_{n+1}) = \overline{X}_\ell(t_n) - \overline{X}_{\ell-1}(t_n) \\
&+ \nu_1(\mathcal{P}'''_n(\Delta a^1_{\ell-1,n}\Delta t_\ell)\mathbb{1}_{\{\Delta a^1_{\ell-1,n}>0\}} - \mathcal{P}''_n(-\Delta a^1_{\ell-1,n}\Delta t_\ell)\mathbb{1}_{\{\Delta a^1_{\ell-1,n}<0\}}) \\
&+ \nu_1(\mathcal{Q}'''_n(\Delta a^2_{\ell-1,n}\Delta t_\ell)\mathbb{1}_{\{\Delta a^2_{\ell-1,n}>0\}} - \mathcal{Q}''_n(-\Delta a^2_{\ell-1,n}\Delta t_\ell)\mathbb{1}_{\{\Delta a^2_{\ell-1,n}<0\}})
\end{aligned}$$

where $\mathcal{P}'''_n, \mathcal{P}''_n, \mathcal{Q}'''_n, \mathcal{Q}''_n$ are conditionally independent Poisson processes and $\Delta a^1_{\ell-1,n} = a(\overline{X}_\ell(t_n)) - a(\overline{X}_{\ell-1}(t_n))$ and $\Delta a^2_{\ell-1,n} = a(\overline{X}_\ell(t_n + \Delta t_\ell)) - a(\overline{X}_{\ell-1}(t_n))$

Combining MLMC and IS

- Note that $\Delta a^1_{\ell-1,n}$ and $\Delta a^2_{\ell-1,n}$ are random variables
- As such , we consider the sigma algebra

$$\mathcal{F}_{n_\ell} = \sigma((\Delta a_{\ell,k})_{k=0,\dots,n_\ell}), \quad n_\ell = 0, \dots, N_{\ell-1}$$

to ensure that the $\Delta a_{\ell-1,n}$ are measurable with respect to $\mathcal{F}_{n_\ell}$