Multilevel Monte Carlo for Ergodic SDEs without Contractivity

Wei Fang, Mike Giles

Mathematical Institute, University of Oxford

MCQMC 2018

July 3, 2018

Outline

- Change of measure in MLMC applications
- Existing approaches for ergodic SDEs
- New MLMC scheme with change of measure
- Theoretical results for Lipschitz case
- Numerical results

MLMC estimators using change of measure

This work fits within a growing literature of MLMC applications using change of measure techniques to improve MLMC correction variances:

- Xia, G. (2010)
  Jump-adapted jump-diffusion SDEs – change of measure used to force same jump times on coarse and fine paths
- G (2012, 2015)
  Digital options in finance – change of measure used to force same final state for coarse and fine paths
- Stilger, Poon (2014), Gasparotto (2015), Kebaier, Lelong (2017), Alaya, Hajji, Kebaier (2018)
  Importance sampling for rare events, e.g. deep out-of-the-money digital call options

Ergodic SDEs

The process X_t is *ergodic* if it has a unique invariant distribution π s.t. for each smooth $\varphi \in L^1(\pi)$ and for any fixed initial condition $X_0 = x_0$,

$$\pi(\varphi) := \int \varphi(x)\,\mathrm{d}\pi(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(X_t)\,\mathrm{d}t, \ \ a.s.$$

We'll consider an SDE driven by an m-dimensional Brownian motion:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \tag{1}$$

with a Lipschitz drift $f : \mathbb{R}^m \to \mathbb{R}^m$ satisfying the dissipativity condition: for some $\alpha, \beta > 0$,

$$\langle x, f(x)\rangle \leq -\alpha\|x\|^2 + \beta. \tag{2}$$

Ergodic SDEs

Lemma (Geometric ergodicity)

If the SDE satisfies the previous assumption then it is ergodic with a unique invariant distribution π, and for any φ which grows at worst polynomially there exist positive constants μ^* and λ^* such that

$$\left|\mathbb{E}\left[\varphi(X_t) - \pi(\varphi)\right]\right| \;\leq\; \mu^* \, e^{-\lambda^* t}. \tag{3}$$

We are interested in computing $\pi(\varphi)$, the expectation of some function $\varphi(x)$ with respect to that invariant distribution π.

Several numerical approaches

- Compute the probability density function $\rho(x)$ of π by solving the corresponding stationary Fokker-Planck equation.

  Extremely expensive for high-dimensional problems.

- Compute an ergodic numerical approximation $\widehat{X}_t$ and evaluate

$$\frac{1}{T}\sum_{n=1}^{N} h\,\varphi(\widehat{X}_{nh}) \equiv \frac{1}{N}\sum_{n=1}^{N} \varphi(\widehat{X}_{nh})$$

  Requires the numerical approximation to preserve the ergodicity.

- Estimate $\mathbb{E}\left[\phi(\widehat{X}_T)\right]$ for a sufficiently large T.

  How to apply MLMC to the SDEs without contractivity?

SDEs with contractivity

Assumption (Contractive Lipschitz properties)

There exists a constant $\lambda > 0$ *such that for all* $x, y \in \mathbb{R}^m$, f *satisfies the contractive Lipschitz condition:*

$$\langle x - y, f(x) - f(y) \rangle \ \leq \ -\lambda \, \|x - y\|^2. \tag{4}$$

Lemma (Contractivity)

If the SDE satisfies the contractive assumption, then for any two solutions X_t *and* Y_t*, driven by the same Brownian motion from different initial data,* $\forall \ t > 0, p > 0,$

$$\mathbb{E}\left[\|X_t - Y_t\|^p\right] \leq e^{-\lambda p t} \, \mathbb{E}\left[\|X_0 - Y_0\|^p\right]. \tag{5}$$

Contractivity also ensures that strong convergence is uniform in T.

MLMC for infinite time interval

Following an idea of Glynn and Rhee (2014), different levels use an increasing length of time interval T_ℓ, as well as decreasing timesteps, ensuring overall weak convergence:

$$\mathbb{E}\left[\varphi(X_{T_L})\right] = \mathbb{E}\left[\varphi(X_{T_0})\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\varphi(X_{T_\ell}) - \varphi(X_{T_{\ell-1}})\right].$$

Comment: don't need to decide T in advance – MLMC algorithm will automatically terminate at a level L with a sufficiently large T_L.

MLMC for infinite time interval

Since f does not depending explicitly on t, the distribution of the numerical solution simulated on time interval $[-T_\ell, 0]$ is the same as the one simulated on $[0, T_\ell]$ starting from the same initial point.

level 0 $\quad -T_0 \longrightarrow 0$

level 1 $\quad -T_1 \longrightarrow 0$

level 2 $\quad -T_2 \longrightarrow 0$

level 3 $\quad -T_3 \longrightarrow 0$

......

Comment: the fine path and coarse path share the same driving W_t for the time interval $[-T_{\ell-1}, 0]$. The contractivity ensures the exponential decay of the difference at timer $-T_{\ell-1}$.

SDEs without Contractivity

Assumption (One-sided Lipschitz properties)

There exists a constant $\lambda > 0$ such that for all $x, y \in \mathbb{R}^m$, f satisfies:

$$\langle x - y, f(x) - f(y) \rangle \ \leq \ \lambda \, \|x - y\|^2. \tag{6}$$

Examples:

- Double-well potential energy:
  fine path and coarse path may diverge to different wells.
- Stochastic Lorenz equation:
  chaotic as in deterministic case – the multilevel correction variance V_ℓ increases exponentially in T

Comments: moments of the numerical solution are still uniform in T due to the dissipative condition – no impact on standard Monte Carlo method.

New MLMC with change of measure

Key idea: add a "spring" between the fine path and coarse path to prevent divergence.

i.e. instead of:

$$\begin{aligned} \mathbb{Q}^f: \quad \mathrm{d}X_t^f &= f(X_t^f)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{Q}^f}, \\ \mathbb{Q}^c: \quad \mathrm{d}X_t^c &= f(X_t^c)\,\mathrm{d}t + \mathrm{d}W_t^{\mathbb{Q}^c}, \end{aligned}$$

adding a spring term with $2S > \lambda$, we simulate both paths in measure $\mathbb{P}$:

$$\begin{aligned} \mathrm{d}Y_t^f &= S(Y_t^c - Y_t^f)\,\mathrm{d}t + f(Y_t^f)\,\mathrm{d}t + \mathrm{d}W_t, \\ \mathrm{d}Y_t^c &= S(Y_t^f - Y_t^c)\,\mathrm{d}t + f(Y_t^c)\,\mathrm{d}t + \mathrm{d}W_t. \end{aligned}$$

Girsanov's theorem gives

$$\mathbb{E}^{\mathbb{Q}^f}[X_t^f] - \mathbb{E}^{\mathbb{Q}^c}[X_t^c] = \mathbb{E}^{\mathbb{P}}\left[Y_t^f \frac{\mathrm{d}\mathbb{Q}^f}{\mathrm{d}\mathbb{P}} - Y_t^c \frac{\mathrm{d}\mathbb{Q}^c}{\mathrm{d}\mathbb{P}} \right].$$

New MLMC with change of measure

The fine path Y_t^f and coarse path Y_t^c share the same driving Brownian motion W_t in measure $\mathbb{P}$, over the **same** time interval $[0, T]$.

The difference between the new pair of SDE solutions satisfies

$$\mathrm{d}(Y_t^f - Y_t^c) = -2S(Y_t^f - Y_t^c)\,\mathrm{d}t + (f(Y_t^f) - f(Y_t^c))\,\mathrm{d}t,$$

and hence

$$\mathrm{d}\,\|Y_t^f - Y_t^c\|^2 \leq 2(\lambda - 2S)\,\|Y_t^f - Y_t^c\|^2\,\mathrm{d}t,$$

so if $2S > \lambda$ we recover the contractivity between the fine and coarse paths.

Numerical implementation

We use standard Euler-Maruyama (Milstein) method for path simulation.

For level 0, it is the same as the standard MLMC.

For level $\ell > 1$, we simulate the SDE with the spring terms using timestep $h = 2^{-\ell} h_0$ for the fine path and $2h$ for the coarse path.

- At time $t_0 = 0$ we set $\widehat{Y}^f_{t_0} = \widehat{Y}^c_{t_0} = x_0$.
- At odd timesteps $t_{2n+1} = t_{2n} + h$ we update both paths:

$$\begin{aligned}
\widehat{Y}^c_{t_{2n+1}} &= \widehat{Y}^c_{t_{2n}} + S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}})\, h + f(\widehat{Y}^c_{t_{2n}})\, h + \Delta W_{2n}, \\
\widehat{Y}^f_{t_{2n+1}} &= \widehat{Y}^f_{t_{2n}} + S(\widehat{Y}^c_{t_{2n}} - \widehat{Y}^f_{t_{2n}})\, h + f(\widehat{Y}^f_{t_{2n}})\, h + \Delta W_{2n}.
\end{aligned}$$

- At even timesteps $t_{2n+2} = t_{2n+1} + h$ for $n \geq 0$, we update the spring and drift terms for the fine path, and update both paths:

$$\begin{aligned}
\widehat{Y}^c_{t_{2n+2}} &= \widehat{Y}^c_{t_{2n+1}} + S(\widehat{Y}^f_{t_{2n}} - \widehat{Y}^c_{t_{2n}})\, h + f(\widehat{Y}^c_{t_{2n}})\, h + \Delta W_{2n+1}, \\
\widehat{Y}^f_{t_{2n+2}} &= \widehat{Y}^f_{t_{2n+1}} + S(\widehat{Y}^c_{t_{2n+1}} - \widehat{Y}^f_{t_{2n+1}})\, h + f(\widehat{Y}^f_{t_{2n+1}})\, h + \Delta W_{2n+1}.
\end{aligned}$$

Radon-Nikodym Derivatives

We calculate the exact Radon-Nikodym derivatives $\mathrm{R}^f = \mathrm{d}\widehat{\mathbb{Q}}^f/\mathrm{d}\mathbb{P}$ for the fine path and $\mathrm{R}^c = \mathrm{d}\widehat{\mathbb{Q}}^c/\mathrm{d}\mathbb{P}$ for coarse path, step by step:

- At $t_0 = 0$, we set $\mathrm{R}^f_{t_0} = \mathrm{R}^c_{t_0} = 1$.
- At odd timesteps $t_{2n+1} = t_{2n} + h$ we only update R^f:

$$\mathrm{R}^f_{t_{2n+1}} \;=\; \mathrm{R}^f_{t_{2n}}\,\mathrm{R}\left(\,\widehat{Y}^f_{t_{2n+1}}\,|\,\widehat{Y}^f_{t_{2n}}\right).$$

- At even timesteps $t_{2n+2} = t_{2n+1} + h$ we update both R^f and R^c:

$$\begin{aligned}
\mathrm{R}^f_{t_{2n+2}} &= \mathrm{R}^f_{t_{2n+1}}\mathrm{R}(\widehat{Y}^f_{t_{2n+2}}\,|\,\widehat{Y}^f_{t_{2n+1}}),\\
\mathrm{R}^c_{t_{2n+2}} &= \mathrm{R}^c_{t_{2n}}\;\mathrm{R}(\widehat{Y}^c_{t_{2n+2}}\,|\,\widehat{Y}^c_{t_{2n}}).
\end{aligned}$$

Finally, the multilevel correction estimator on level ℓ is

$$\varphi(\widehat{Y}^f_T)\,\mathrm{R}^f_T - \varphi(\widehat{Y}^c_T)\,\mathrm{R}^c_T. \tag{7}$$

Stability and strong error results

Theorem

Under the Lipschitz and dissipativity assumptions, there exists $h_0 > 0$ such that for all $h < h_0$, and any $p \geq 1$, $T > 0$, there exists a constant $C_1 > 0$ which is independent of p and T such that

$$\sup_{0\leq n\leq N} \mathbb{E}\left[\|\widehat{Y}^f_{t_n}\|^p\right]^{1/p} \leq C_1\, p^{1/2}, \quad \sup_{0\leq n\leq N} \mathbb{E}\left[\|\widehat{Y}^c_{t_n}\|^p\right]^{1/p} \leq C_1\, p^{1/2}. \qquad (8)$$

Furthermore, if $2S > \lambda$, then for there exists a constant $C_2 > 0$, independent of p and T, such that

$$\sup_{0\leq n\leq N} \mathbb{E}\left[\|\widehat{Y}^f_{t_n} - \widehat{Y}^c_{t_n}\|^p\right]^{1/p} \leq C_2\ \min\left(p^{1/2}h^{1/2}, ph\right). \qquad (9)$$

Variance Estimation

Theorem (Moment of level estimator)

If $\varphi: \mathbb{R}^m \to \mathbb{R}$ is locally Lipschitz, with at worst polynomial growth, and the original SDE satisfies the Lipschitz and dissipativity conditions, then using $2S > \lambda$ and a sufficiently small h, for any $p \geq 1$, $T > 0$, there exists a constant $C_3 > 0$, independent of p and T such that

$$\mathbb{E}\left[\left|\varphi(\widehat{Y}_T^f)R^f - \varphi(\widehat{Y}_T^c)R^c\right|^p\right]^{1/p} \leq C_3\, p^2 \sqrt{T}\, h.$$

Note that this implies that the variance of the MLMC estimator (7) is bounded by $C_3^2 p^4 T h^2$ which increases linearly in T.

Idea of Proof

By Jensen's inequality and Holder inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\widehat{Y}_T^f R^f - \widehat{Y}_T^c R^c\right\|^p\right] \\
&= \mathbb{E}\left[\left\|\widehat{Y}_T^f(R^f - 1) + (\widehat{Y}_T^f - \widehat{Y}_T^c) - \widehat{Y}_T^c(R^c - 1)\right\|\right] \\
&\le 3^{p-1}\mathbb{E}\left[\left\|\widehat{Y}_T^f - \widehat{Y}_T^c\right\|^p\right] + 3^{p-1}\mathbb{E}\left[\left\|\widehat{Y}_T^f\right\|^{2p}\right]^{1/2}\mathbb{E}\left[\left|R^f - 1\right|^{2p}\right]^{1/2} \\
&\quad + 3^{p-1}\mathbb{E}\left[\left\|\widehat{Y}_T^c\right\|^{2p}\right]^{1/2}\mathbb{E}\left[\left|R^c - 1\right|^{2p}\right]^{1/2}.
\end{aligned}$$

The blue terms have the order:

$$R^f - 1 \sim \exp\left(\int_0^T S(\widehat{Y}_s^f - \widehat{Y}_s^c)\,\mathrm{d}W_s\right) - 1 \sim \int_0^T S(\widehat{Y}_s^f - \widehat{Y}_s^c)\mathrm{d}W_s,$$

and

$$\mathbb{E}\left[\left\|\int_0^T S(\widehat{Y}_s^f - \widehat{Y}_s^c)\mathrm{d}W_s\right\|^p\right] \sim O(T^{p/2}h^p).$$

MC / std MLMC complexity

Weak error $O(\varepsilon)$ requires $T = O(\frac{1}{\lambda^*}|\log\varepsilon|)$, and $h = O(\varepsilon)$, so the standard MC cost is

$$\mathrm{C}_{MC} = O\left(\varepsilon^{-3}|\log\varepsilon|\right)$$

For standard MLMC we get

$$V_\ell = O\left((h_0\, 2^{-\ell})^2\, 2^{\kappa T}\right),$$

and the requirement that $V_1 < V_0$ means that (provided $\kappa > 2\lambda^*$)

$$h_0 = O(2^{-\kappa T/2}) = O(\varepsilon^{-\kappa/2\lambda^*})$$

and hence the standard MLMC cost is

$$\mathrm{C}_{std} = O\left(\varepsilon^{-2-\kappa/2\lambda^*}|\log\varepsilon|\right).$$

New MLMC complexity

For the new MLMC with the change of measure:

$$V_\ell = O\left((h_0\, 2^{-\ell})^2\, T\right)$$

In order to get a good coupling and stability of Radon-Nikodym derivatives, we need

$$h_0 = O(T^{-1}),$$

and therefore

$$\mathrm{C}_{new} = O(\varepsilon^{-2} |\log \varepsilon|^2).$$

In comparison, for SDEs with contractivity, MLMC for the infinite time interval achieves the optimal computational cost $O(\varepsilon^{-2})$.

Numerical Results

We first present the numerical results for Lipschitz version of stochastic Lorenz equation. We run 10000 sample paths from $T = 0$ to 20 to get the following results. Our interest is to compute $\pi(\varphi)$.
For standard MLMC:

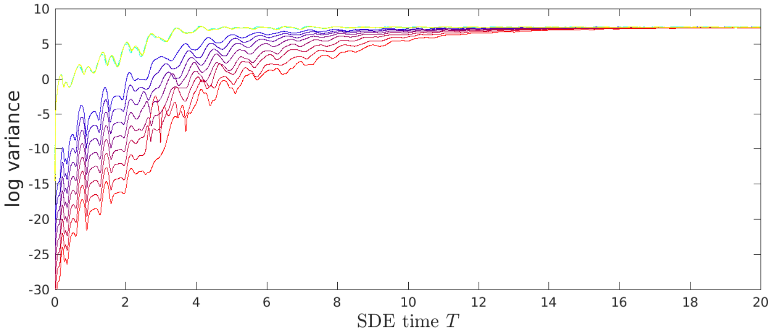
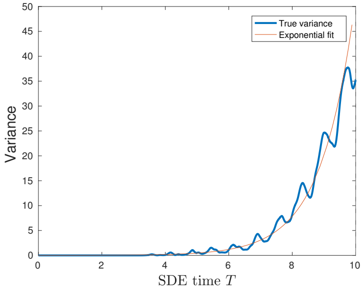


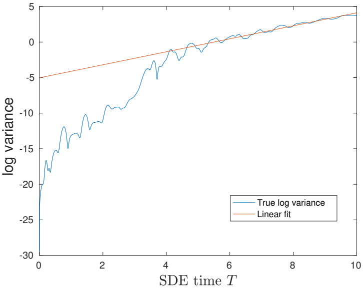
Figure: Variance for each level without change of measure

Numerical Results

Without the new change of measure, the κ we fit for the exponential growth of the variance is 1.3601.



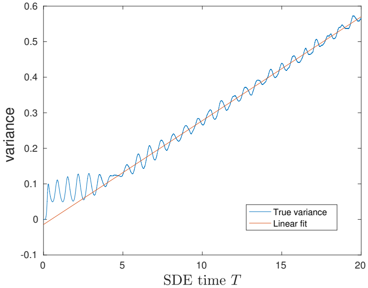
(a) Exponential increase of variance



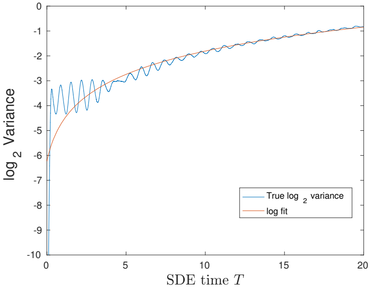
(b) Linear increase of log variance

Numerical Results

With the change of measure, we see a linear growth in the variance of the MLMC estimator



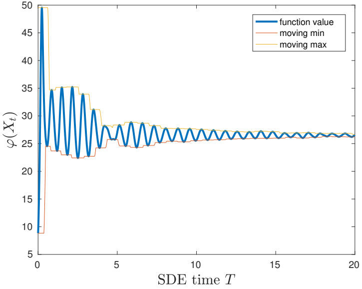
(c) Linear increase of variance



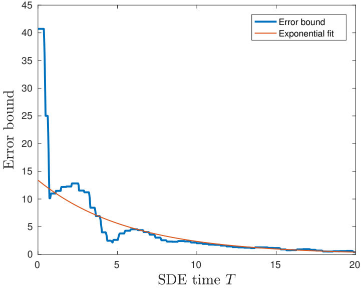
(d) Log increase of log variance

Numerical Results

We can also estimate the convergence rate to the invariant measure $\lambda^* = 0.1741$.



(e) $\mathbb{E}\ \varphi(\widehat{X}_t)$



(f) Error bound

Stochastic Lorenz equation

We simulate the standard stochastic Lipschitz SDE (no far-field adjustment to make it globally Lipschitz) with initial value $x_0 = [0, 0, 0]$ to time $T = 10$, and use the adaptive function:

$$h^{\delta}(x) = \frac{\max(100, \|x\|^2)}{2^{11} \max(100, \|f(x)\|^2)} \delta$$

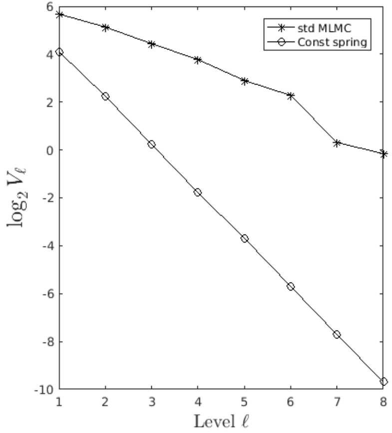
with $\delta = 2^{-\ell}$ for each level ℓ.

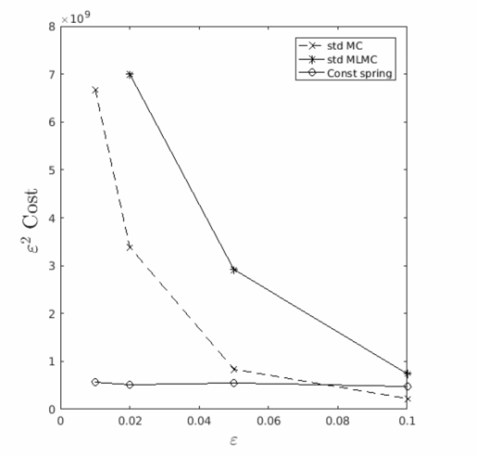
We compare two different schemes:

- standard MLMC with adaptive timestep.
- MLMC with adaptive timestep and change of measure with constant spring coefficient $S = 10$.

A possible third scheme is the scheme with adaptive spring which requires to calculate the largest positive eigenvalue of the Jacobian matrix $\frac{\partial f}{\partial x}$.

Stochastic Lorenz equation





Conclusion

- The previous MLMC works well for SDEs with contractivity at all times, or on average (negative Lyapunov exponent)
- The new MLMC with change of measure works well for SDEs with positive Lyapunov exponent – can greatly reduce the MLMC variance.
- The new MLMC improves the computational cost for the stochastic Lorenz equation by a huge amount – the benefits are more limited for a double-well potential energy.
- The numerical analysis is only valid for SDEs with globally Lipschitz drift using uniform timestepping, but in practice it works fine for locally Lipschitz drift using adaptive timestepping