Weighted Multilevel Monte Carlo methods

Tony Ware
Joint work with Yu Li

GO20, May 2026, Gozo, Malta

Monte Carlo estimation

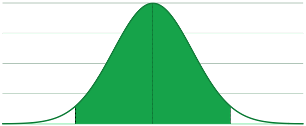
Monte Carlo estimation

Scenario: we have a random variable $P(\omega)$, and we want to compute the quantity $\overline{P} = \mathbb{E}[P]$.

A Monte Carlo estimate of $\overline{P}$ using M samples will take the form

$$_{M}P = \frac{1}{M}\sum_{m=1}^{M} P(\omega_m).$$

This will itself be a random variable, and for sufficiently large M its distribution will resemble $N\left(\overline{P}, \frac{\sigma^2}{M}\right)$, where σ^2 is the variance of P.



We can tighten this distribution by (a) increasing N or (b) finding a way to decrease σ.

Monte Carlo estimation

But: we may have no way to compute samples of P directly.

Instead, we may have access to a sequence of approximate estimators P_l, $l = 0, 1, 2, \ldots$, with $\overline{P}_l = \mathbb{E}[P_l] \to \overline{P}$ as $l \to \infty$, with the cost of computing an estimate P_l being proportional to (say) N_l, and $\left|\mathbb{E}[P_l] - \overline{P}\right| \lesssim N_l^{-q}$.

If we generate a Monte Carlo estimate of $\mathbb{E}[P_l]$ using M samples, the cost will be proportional to MN_l. The total squared error can be expressed using the MSE:

$$\mathbb{V}\left[{}_M P_l\right] + \mathbb{E}\left[{}_M P_l - \overline{P}\right]^2 = \frac{\sigma_l^2}{M} + \frac{1}{N_l^{2q}}.$$

An MSE of ϵ^2 requires a computational cost proportional to $\epsilon^{-\left(2+\frac{1}{q}\right)}$.

Multi-level Monte Carlo (MLMC)

MLMC

Giles (2008 and many others) combines multiple levels of estimates with different numbers of samples to reduce the cost to—under certain conditions—as low as $1/\epsilon^2$.

Suppose that, whenever we generate a sample of P_l, we are able to generate a correlated sample of the variable P^l_{l-1}, with expected value $\overline{P}_{l-1}$ (the same expected value as P_{l-1}).
We collect these together as basic estimators:

$$Y_0 = P_0, \text{ and } Y_l = P_l - P^l_{l-1} \text{ for } l = 1, 2, \ldots,$$

with costs η_l^2 and variances v_l^2. The Y_l are independent of one another, and satisfy $\mathbb{E}[Y_l] = \overline{P}_l - \overline{P}_{l-1}$ for $l > 0$.

A MLMC estimator at level L takes the form $\boxed{\mathcal{P}_L = \sum_{l=0}^{L} {}_{M_l}Y_l}$.

Because of the telescoping property for the means, $\boxed{\mathbb{E}[\mathcal{P}_L] = \overline{P}_L}$.

MLMC

If we have two estimators A and B with the same expected value, but with per-sample costs W_A^2 and W_B^2, and variances V_A^2 and V_B^2, we can compare them by looking at the efforts $E_A = W_A V_A$ and $E_B = W_B V_B$. The estimator with the smallest effort will be preferable.

The computational cost and the variance of a sample of $\mathcal{P}_L$ are

$$\mathcal{W}_L^2 = \sum_{l=0}^{L} M_l \eta_l^2 \quad \text{and} \quad \mathcal{V}_L^2 = \sum_{l=0}^{L} \frac{v_l^2}{M_l}.$$

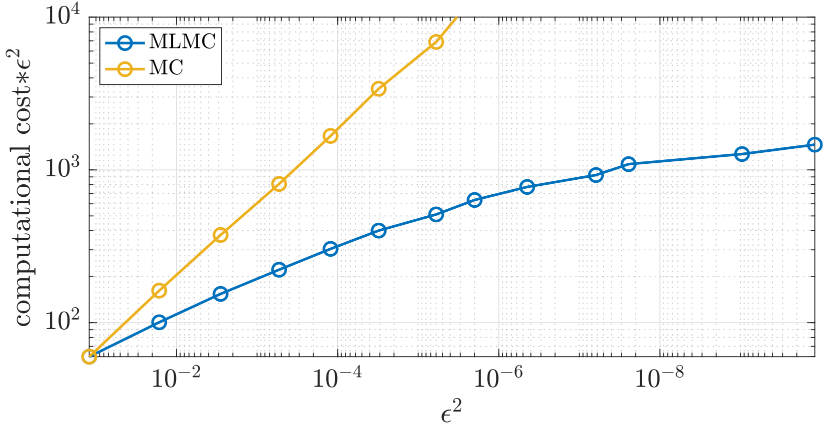
The numbers of samples N_l used to compute a sample of $\mathcal{P}_L$ are chosen to minimise $\mathcal{E}_L = \mathcal{W}_L \mathcal{V}_L$, giving

$$M_l = \frac{\mathcal{W}_L}{\mathcal{V}_L} \frac{v_l}{\eta_l} = \frac{\mathcal{E}_L}{\mathcal{V}_L^2} \frac{v_l}{\eta_l}, \text{ and } \mathcal{E}_L = \sum_{l=0}^{L} \eta_l v_l.$$

$\mathcal{P}_L$ will be preferable to P_L when $\mathcal{E}_L < \eta_L \sigma_L$.
If $q \geq 1$ and $\mathcal{E}_L$ stays bounded as $L \to \infty$, the complexity is $O\big(1/\epsilon^2\big)$;
If $q \geq 1$ and $\mathcal{E}_L$ grows linearly, the complexity is $O\big(\left|\log \epsilon\right| / \epsilon^2\big)$.

MLMC

We show the computational cost of computing an estimate with a MSE of ϵ^2
European call under GBM, Euler Maruyama discretisation, $N_{1+1} = 2N_l$



Weighted Multilevel Monte Carlo (WMLMC)

WMLMC

First idea

The ratios of successive numbers of samples are independent of L, as are $\mathcal{W}_L/\mathcal{V}_L$. so we can write the MLMC estimator recursively:

$$\boxed{\mathcal{P}_0 = {}_{M_0}Y_0,} \text{ and, for } l > 0, \boxed{\mathcal{P}_l = \sum_{k=0}^{l} {}_{M_k}Y_k = {}_{M_l}Y_l + \mathcal{P}_{l-1},}$$

where $M_l = \frac{\mathcal{W}_l}{\mathcal{V}_l}\frac{v_l}{\eta_l}$.

Expanding $\mathcal{P}_l = {}_{M_l}P_l - \left({}_{M_l}P_{l-1}^{l} - \mathcal{P}_{l-1}\right)$, and noticing that $\mathbb{E}[{}_{M_l}P_{l-1}^{l}] = \mathbb{E}[\mathcal{P}_{l-1}]$, leads to …

WMLMC

Second idea

…adding weights (in the spirit of control variates (c.f. Speight 2009, Schaden 2021)):

$$\mathcal{P}_l = {}_{M_l}P_l - \theta_l \left({}_{M_l}P^l_{l-1} - \mathcal{P}_{l-1} \right) = {}_{M_l}Y_l^{\theta_l} + \theta_l \mathcal{P}_{l-1}$$

where $Y_l^\theta = P_l - \theta P^l_{l-1}$.
This results in the weighted MLMC estimate:

$$\boxed{\mathcal{P}_L = \sum_{l=0}^{L} \Theta_l^L \, {}_{M_l}Y_l^{\theta_l},}$$

where $\Theta_l^L = \prod_{k=l+1}^{L} \theta_k$, and M_l (now) depends on $\theta_1, \ldots, \theta_l$.

WMLMC

- We optimise both the numbers of samples and the weights. It is most natural to do this recursively, minimising the effort associated with $\mathcal{P}_l$ at each stage.
- We start with $\mathcal{P}_0 = Y_0 = P_0$, so that $\mathcal{E}_0 = \eta_0 \sigma_0$.
- Suppose we have an optimised WMLMC estimator $\mathcal{P}_{l-1}$. For fixed θ_l, we can minimise the effort $\mathcal{E}_l$ by setting

$$M_l = \frac{\eta_0}{\sigma_0} \frac{v_l(\theta_l)}{\eta_l},$$

where $v_l^2(\theta) = \sigma_l^2 - 2\theta\rho_l\sigma_l\sigma_{l-1} + \theta^2\sigma_{l-1}^2$, and $\rho_l = \mathrm{Corr}[P_l, P_{l-1}^l]$. Then we have

$$\boxed{\mathcal{E}_l = v_l(\theta_l)\eta_l + |\theta_l|\, \mathcal{E}_{l-1}.}$$

If we are targetting an overall variance of V^2 once we reach level L, we take $M_0 = \frac{\mathcal{E}_L}{V^2}\frac{\sigma_0}{\eta_0}$ and multiply all the other values of M_l by this value.

WMLMC

- If $\rho_l \leq \frac{\mathcal{E}_{l-1}}{\eta_l \sigma_{l-1}}$, we take $\theta_l = 0$, and l becomes our new coarsest level.
- Otherwise, the optimal θ_l is positive, and we have

$$\theta_l = \frac{\rho_l \sigma_l}{\sigma_{l-1}} - \frac{v_l \mathcal{E}_{l-1}}{\eta_l \sigma_{l-1}^2}, \text{ with } v_l = \sigma_l \sqrt{\frac{1-\rho_l^2}{1-\frac{\mathcal{E}_{l-1}^2}{\eta_l^2 \sigma_{l-1}^2}}}$$

- This results in

$$\mathcal{E}_l = \frac{\sigma_l}{\sigma_{l-1}} \left(\rho_l \mathcal{E}_{l-1} + \sigma_l \eta_l \sqrt{(1-\rho_l^2)\left(1 - \frac{\mathcal{E}_{l-1}^2}{\eta_l^2 \sigma_{l-1}^2}\right)} \right).$$

We expect that $\rho_l \to 1$ as $l \to \infty$. If it does so fast enough that $\eta_l \sqrt{1-\rho_l^2} \to 0$, we may be able to show that $\mathcal{E}_l$ does indeed stay bounded. This is the content of the complexity theorem for (W)MLMC proved in Giles (2008).

WMLMC

Comparison with MLMC

- In order to gain insight into the comparative the costs associated with MLMC and WMLMC it is convenient to assume that $\sigma_l \equiv \sigma$ and that $\eta_l^2 = K\eta_{l-1}^2$, and to focus on the two-level case. We have that $\mathcal{E}_0 = \sigma\eta_0 = E_0$ (where $E_l = \sigma_l\eta_l$ is the effort (or cost for unit variance) of a single-level estimator at level l).
- We have, for the optimal weighted version

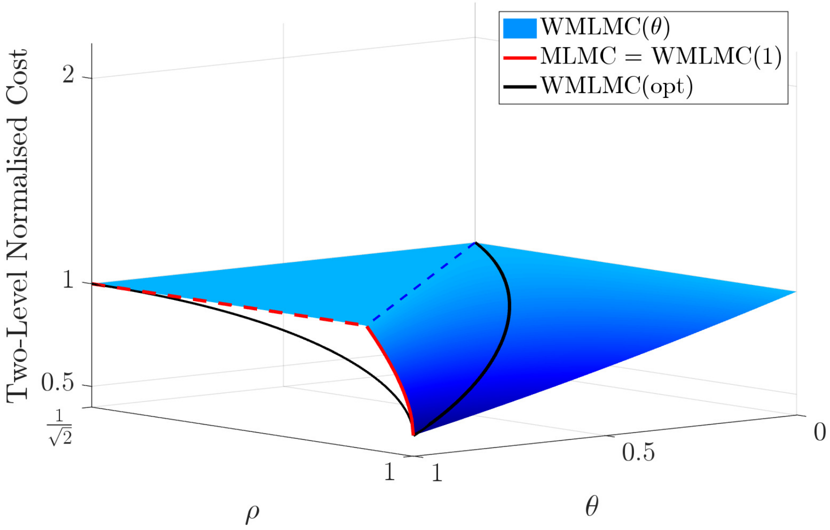
$$\mathcal{E}_1 = \begin{cases} E_0\left(\rho_1 + \sqrt{(1-\rho_1^2)(K-1)}\right) & \text{if } \rho_1 > \frac{1}{\sqrt{K}} \\ E_1\ (= \sigma\eta_1) & \text{otherwise.} \end{cases}$$

- And for MLMLC (which corresponds to taking $\theta_0 = 1$)

$$\mathcal{E}_1 = \begin{cases} E_0\left(1 + \sqrt{2K(1-\rho_1)}\right) & \text{if } \rho_1 > 1 - \frac{1}{2}\left(1 - \frac{1}{\sqrt{K}}\right)^2 \\ E_1 & \text{otherwise.} \end{cases}$$

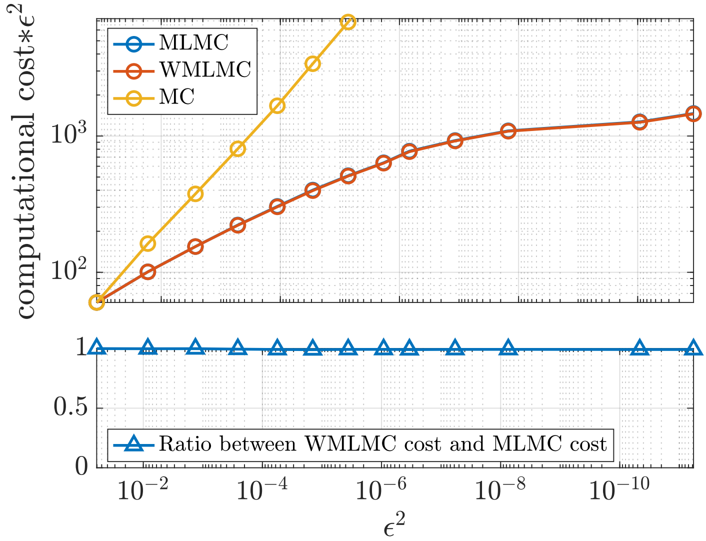
WMLMC

Two-level normalised cost (i.e. $\mathcal{E}_1^2/E_1^2$)



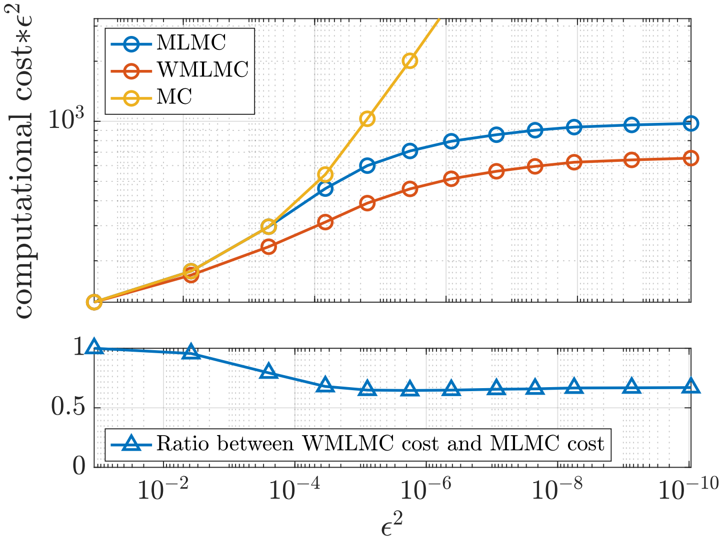
WMLMC

European call under GBM, Euler Maruyama discretisation



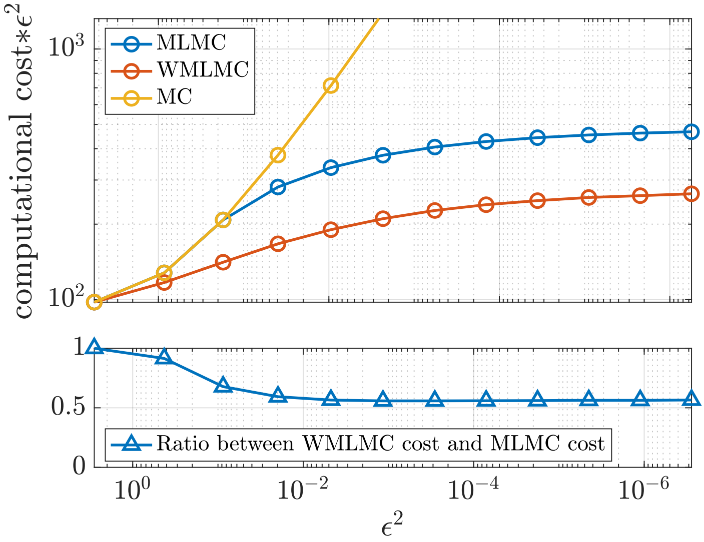
WMLMC

Asian call under GBM, Milstein discretisation



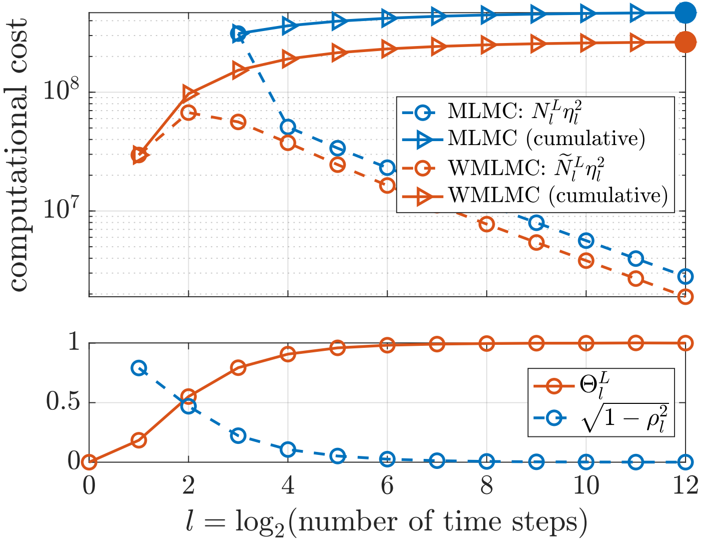
WMLMC

European call under IGBM, Milstein discretisation



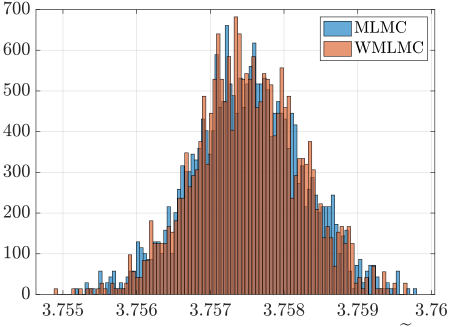
WMLMC

European call under IGBM, Milstein discretisation

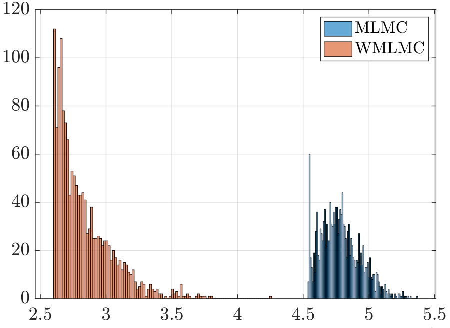


WMLMC

European call under IGBM, Milstein discretisation, 1500 simulations of the WMLMC estimator $\mathcal{P}_L$ and the MLMC estimator $\widetilde{\mathcal{P}}_L$ with $K = 2$. The optimal values of θ_l and M_l were estimated in each case (leading to the observed variation in computational costs). The left hand graph shows histograms of the two sets of estimates. The computed MSE values are 1.11×10^6 for MLMC, and 1.06×10^6 for WMLMC. The right hand graph shows histograms of the computational costs $\sum_{l=0}^{L} M_l \eta_l^2$. The value of L determined by the algorithm was either 11 or 12, depending on the run.



Histograms of 1500 samples of $\mathcal{P}_L$ and $\widetilde{\mathcal{P}}_L$

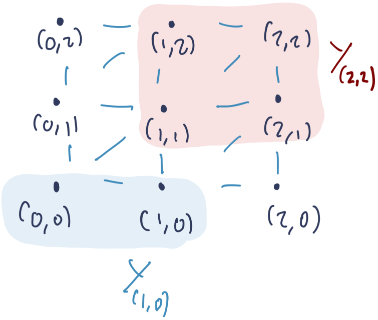


1500 samples of computational costs

Weighted Multi-Index MLMC

Weighted Multi-Index Multilevel Monte Carlo

- Our estimates are now indexed by multi-indices $\lambda \in \mathbb{N}^2$.
- The MIMLMC method uses alternating signs to define Y_λ.
- It can (again) be written in a recursive form, and weights can be added.
- There is no longer an explicit expression for the optimal weights, but they can be determined numerically at each node.



$$
\begin{aligned}
\mathcal{P}_\lambda &= \mathrm{M}_\lambda Y_\lambda(t_\lambda) + \sum_{\mu \in \square_\lambda^-} \theta_\mu^\lambda \mathcal{P}_\mu \\
&= \sum_{\mu \le \lambda} \Theta_\mu^\lambda \ \mathrm{M}_\mu Y_\mu .
\end{aligned}
$$

Weighted Multi-Index Multilevel Monte Carlo

Some more details

We write

$$\begin{aligned}\mathcal{P}_\lambda &= {}_{M_\lambda}Y_\lambda(t_\lambda) + \sum_{\mu \in \Box_\lambda^-} \theta_\mu^\lambda \mathcal{P}_\mu \\ &= \sum_{\mu \le \lambda} \Theta_\mu^\lambda \;\; {}_{M_\mu}Y_\mu.\end{aligned}$$

- The index set $\Box_\lambda^-$ is the set of nodes directly connected to λ, at which we will produce samples correlated to P_λ.
- Θ_μ^λ is the sum over all paths from μ to λ of the product of the values of $\theta_{\mu'}^{\lambda'}$ along each path.
- The effort M_μ is independent of the path from μ to λ by virtue of the recursive construction.

Weighted Multi-Index Multilevel Monte Carlo

Some more details

- Given the recursive relation

$$\mathcal{P}_\lambda = {}_{M_\lambda}Y_\lambda(t_\lambda) + \sum_{\mu \in \square_\lambda^-} \theta_\mu^\lambda \mathcal{P}_\mu,$$

  we can, given $t_\lambda = [\theta_\mu^\lambda]_{\mu \in \square_\lambda^-}$, determine M_λ so as to minimise the effort $\mathcal{E}_\lambda$.
- It remains to optimise over t_λ, and this involves minimising the sum of the square roots of two quadratic forms defined by t_λ. This minimisation is performed numerically at each node.
- The unweighted MIMLMC method consists in setting $\theta_\mu^\lambda = (-1)^{1+|\lambda-\mu|}$.

Weighted Multi-Index Multilevel Monte Carlo

Zakai SPDE, finite difference schemes

We consider the Zakai SPDE (Reisinger/Wang 2018):

$$\mathrm{d}v = -\mu \frac{\partial v}{\partial x}\mathrm{d}t + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}\mathrm{d}t - \sqrt{\rho}\frac{\partial v}{\partial x}\mathrm{d}W_t.$$

The quantity of interest is $L_t = 1 - \int_0^\infty v(t,x)\mathrm{d}x$. We compute with a timestep $k = 4^{-m+1}$ and a grid spacing in x of $h = 2^{-n}$, and compare the effort required to achieve a fixed variance using the unweighted and weighted MIMLMC:

$n \backslash m$	0	1	2	3
0	1.0	1.0	1.0	1.0
1	1.6	1.6	1.6	1.7
2	1.8	1.8	1.8	1.9
3	1.8	1.8	1.8	1.9

Weighted Multi-Index Multilevel Monte Carlo

A hybrid Fourier-Monte-Carlo basket option valuation

- Here we express a $(d+2)$-asset basket option payoff as a function $\Lambda(Z_1, Z_2)$ of independent normal random variables $Z_1 \in \mathbb{R}^2$ and $Z_2 \in R^d$. The option value can be written as a nested expectation:

$$\mathbb{E}[\Lambda_2(Z_2)], \quad \text{with } \Lambda_2(z_2) = \mathbb{E}[\Lambda(Z_1, z_2)].$$

- For each sample of Z_2, the inner expectation is computed using a two-dimensional cosine series, with $(2^{1+L_i} - 1)$ modes used in dimension i (for $i = 1, 2$).
- We again compare MIMC with WMIMC. For some low-resolution combinations of L_1 and L_2, MIMC provides no advantage over a single-level estimate, so we do not record the ratio.

Weighted Multi-Index Multilevel Monte Carlo

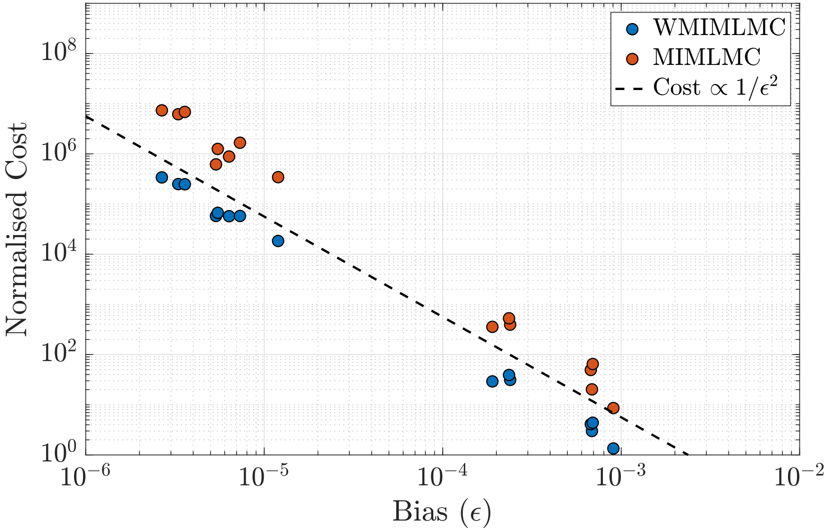
A hybrid Fourier-Monte-Carlo basket option valuation

The ratio between the unweighted and weighted MIMLMC costs to achieve a given variance

$L_1 \backslash L_2$	0	1	2	3	4	5	6
0							
1	1.49	1.59	1.57	1.56	1.75	1.71	1.70
2	1.46	1.61	1.86	2.85	3.51	3.17	4.26
3	1.59	1.81	1.93	6.18	6.57	11.78	14.58
4	1.58	2.17	4.25	11.93	10.44	15.22	18.47
5	1.61	2.36	5.81	12.39	18.51	24.46	27.37
6	1.64	2.35	5.82	13.25	21.41	28.56	31.48

Weighted Multi-Index Multilevel Monte Carlo

A hybrid Fourier-Monte-Carlo basket option valuation



Conclusions

- MLMC and MIMLMC can be formulated recursively, and in that way are naturally viewed as nested control variate variance reduction techniques.
- As such, weights can be added, and the optimal weights computed at each node.
- The addition of weights does not change the asymptotic rate of convergence, but it does allow for more efficient use of estimates at coarser resolutions (and lower correlations) than with unweighted MLMC, resulting in potentially significant gains in performance.
- The gains are relatively insensitive to the optimal choice of weights.
- The multi-index version offers potentially even greater relative improvement.

References

References I

Amri, M. R. E., Mycek, P., Ricci, S., and De Lozzo, M. (2023). Multilevel surrogate-based control variates. arXiv preprint arXiv:2306.10800.

Giles, M. B. (2008). Multilevel Monte Carlo path simulation. Operations Research-Baltimore, 56(3):607–617.

Giles, M. B. (2015). Multilevel Monte Carlo methods. Acta Numerica, 24:259–328.

Giorgi, D., Lemaire, V., and Pagès, G. (2020). Weak error for nested multilevel Monte Carlo. Methodology and Computing in Applied Probability, 22(3):1325–1348.

Lemaire, V. and Pagès, G. (2017). Multilevel Richardson-Romberg extrapolation. Bernoulli, 23(4A):2643–2692.

Pagès, G. (2007). Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity.

Reisinger, C. and Wang, Z. (2018). Analysis of Multi-Index Monte Carlo estimators for a Zakai SPDE. Journal of Computational Mathematics, 36(2):202–236.

Ruijter, M. and Oosterlee, C. (2012). Two-dimensional Fourier Cosine series expansion method for pricing financial options. SIAM Journal on Scientific Computing, 34(5):B642–B671.

Schaden, D. (2021). Variance Reduction with Multilevel Estimators. PhD thesis, Technische Universität München.

Speight, A. (2009). A multilevel approach to control variates. Journal of Computational Finance, 12(1–25).