Multilevel Function Approximation

Filippo De Angelis, Mike Giles, Christoph Reisinger

Mathematical Institute, University of Oxford

Foundations of Computational Mathematics, Paris

June 20th, 2023

Objective

Want to construct an approximation for a scalar function

$$f : [0,1]^d \to \mathbb{R}$$

with parametric dimension d in the range $1-8$, where $f(\theta)$ is one of the following:

- a functional of the solution $u(\theta; x)$ of a PDE, with θ dependence in the PDE coefficients, the boundary data and/or the functional
- a parametric expectation $\mathbb{E}_\omega[g(\theta;\omega)]$, where $g(\theta;\omega)$ is a functional of the solution of an SDE

Problem: in either case we must approximate $f(\theta)$, and the more accurate the approximation, the greater the computational cost.

Objective: for given ε, lowest cost approximation $\widetilde{f}$ with

$$\|\widetilde{f} - f\| < \varepsilon$$

Outline

- quick recap of key literature:
  - dense and sparse grid linear interpolation
  - MLMC for SDEs, and parametric integration (Heinrich)
- MLFA for PDEs
  - dense grid linear interpolation
  - sparse grid linear interpolation
- MLFA for SDEs – extension of Heinrich's approach
  - randomised MLMC for SDE
  - randomised MLMC and sparse grids
  - MLMC decomposition for SDE
  - MLMC decomposition and sparse grids
- conclusions and references

Dense and sparse grid linear interpolation

If $f \in C^r([0,1]^d)$ for $r \in \{1,2\}$, then piecewise multi-linear interpolation on a standard dense tensor product grid with spacing $2^{-\ell}$ has error

$$\|\widetilde{f} - f\| < c(f)\, 2^{-r\ell}$$

with a cost proportional to $O(2^{d\ell})$.

Smolyak sparse grid interpolation instead has an error bound of the form

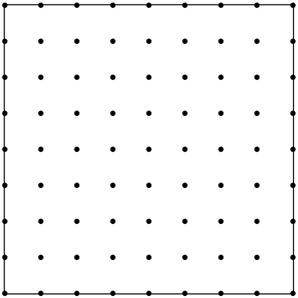
$$\|\widetilde{f} - f\| < c(f)\; 2^{-r\ell} (\ell{+}1)^{d-1}$$

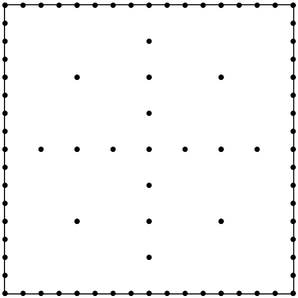
with a cost proportional to $O(2^{\ell}(\ell{+}1)^{d-1})$. However, it needs more regularity in f, including mixed derivatives of degree up to r in each direction:

$$\frac{\partial^{\alpha_1+\alpha_2+\ldots} f}{\partial\theta_1^{\alpha_1}\ \partial\theta_2^{\alpha_2}\ldots}, \quad 0 \leq \alpha_j \leq r \leq 2.$$

Much better than dense grid interpolation for modest values of d, up to 8?

Dense versus sparse grid interpolation





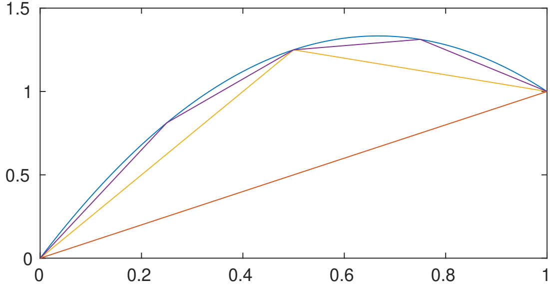
MLMC for parametric integration

Stefan Heinrich's original MLMC research concerned the approximation of $f(\theta) = \mathbb{E}[g(\theta;\omega)]$, given exact sampling of $g(\theta;\omega)$ at unit cost.

In his formulation, the MLMC telescoping sum is

$$f \approx I_L[f] = I_0[f] + \sum_{\ell=1}^{L} I_\ell[f] - I_{\ell-1}[f]$$

where $I_\ell[f]$ represents a level ℓ interpolation.



MLMC for parametric integration

Heinrich then approximates $(I_\ell - I_{\ell-1})[f]$ through Monte Carlo sampling at required values of θ:

$$(I_\ell - I_{\ell-1})[f] \approx \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} (I_\ell - I_{\ell-1})[g(\,\cdot\,;\omega^{\ell,m})]$$

As $\ell \to \infty$, $(I_\ell - I_{\ell-1})[f] \to 0$ and $\mathbb{V}\left[\,(I_\ell - I_{\ell-1})[g]\,\right] \to 0$, so fewer MC samples needed on finer levels.

Analysis assumes the number of θ points increases exponentially with dimension (as with dense tensor product grid), so the resulting complexity for linear interpolation is of order

$$\begin{array}{ll} \varepsilon^{-2}, & d < 2r \\ \varepsilon^{-2}\,|\log \varepsilon|^2, & d = 2r \\ \varepsilon^{-d/r}, & d > 2r \end{array}$$

assuming $g(\theta;\omega)$ is sufficiently smooth w.r.t. θ

MLMC for parametric integration

Heinrich's work is the key starting point for our current work which extends it in several directions:

- PDEs with appropriate numerical approximation
- sparse grid interpolation to address curse of dimensionality
- weaker assumptions on smoothness of $g(\theta;\omega)$
- numerical approximation of $f(\theta) \equiv \mathbb{E}[g(\theta;\omega)]$ in cases without a finite variance, finite expected cost unbiased estimator

The PDE aspect also follows the outline in the excellent review article

> *"Smolyak's algorithm: a powerful black box for the acceleration of scientific computations", by Tempone & Wolfers*

which presents a unifying framework and meta-analysis which includes multilevel methods.

MLFA for PDEs

The fundamental idea is very simple: building on Stefan Heinrich's approach, if the function f has an interpolation expansion

$$f \;=\; I_0[f] + \sum_{\ell=1}^{\infty} I_\ell[f] - I_{\ell-1}[f] \;=\; \sum_{\ell=0}^{\infty} \Delta I_\ell[f]$$

with $\Delta I_\ell \equiv I_\ell - I_{\ell-1}$, $I_{-1} \equiv 0$, and as $\ell \to \infty$, $\Delta I_\ell[f] \to 0$ and the cost per evaluation increases, then we will use an approximation

$$\widetilde{f} = \sum_{\ell=0}^{L} \Delta I_\ell[f_\ell]$$

where f_ℓ is based on a PDE approximation with grid spacing h_ℓ and

- h_ℓ is small for small ℓ – a few expensive accurate PDE calculations
- h_ℓ is large for large ℓ – lots of cheap PDE calculations

MLFA for PDEs

It follows from the triangle inequality that

$$\|\widetilde{f}-f\| \leq \|(I_L-I)[f]\| + \sum_{\ell=0}^{L} \|(I_\ell-I_{\ell-1})[f_\ell-f]\|.$$

If we assume second order accuracy in the interpolation so that

$$\|(I_L-I)[f]\| < c_1\, 2^{-2L}, \qquad \|\Delta I_\ell[f_\ell-f]\| < c_2\, 2^{-2\ell} h_\ell^q$$

and the cost C_ℓ of constructing $(I_\ell-I_{\ell-1})[f_\ell]$ on level ℓ is bounded by

$$C_\ell < c_3\, 2^{d\ell} h_\ell^{-p}$$

then to achieve an accuracy of ε we can choose L s.t.

$$c_1\, 2^{-2L} \approx \varepsilon/2 \quad \Longrightarrow \quad L = O(|\log \varepsilon|)$$

and . . .

MLFA for PDEs

... choose h_ℓ to minimise

$$c_3 \sum_{\ell=0}^{L} 2^{d\ell} h_\ell^{-p}$$

subject to the requirement that

$$c_2 \sum_{\ell=0}^{L} 2^{-2\ell} h_\ell^{q} \approx \varepsilon/2.$$

Using a Lagrange multiplier gives the optimal h_ℓ as

$$h_\ell = 2^{(d+2)\ell/(p+q)}\, h_0$$

The accuracy requirement then becomes

$$c_2\, h_0^q \sum_{\ell=0}^{L} 2^{-\nu\ell} \approx \varepsilon/2, \quad \nu \equiv (2p-dq)/(d+2)$$

MLFA for PDEs

$\nu > 0$ leads to $h_0 = O(\varepsilon^{1/q})$ and a total cost of $O(\varepsilon^{-p/q})$,

$\nu = 0$ leads to $h_0 = O(\varepsilon^{-1/q} L^{1/q})$ and a cost of $O(\varepsilon^{-p/q} |\log \varepsilon|^{1+p/q})$.

$\nu < 0$ leads to $h_0 = O(\varepsilon^{-1/q} 2^{\nu L/q})$ and a cost of $O(\varepsilon^{-d/2})$.

Thus the total cost is

$$\begin{array}{cl} \varepsilon^{-p/q}, & p/q > d/2 \\ \varepsilon^{-p/q}\,|\log \varepsilon|^{1+p/q}, & p/q = d/2 \\ \varepsilon^{-d/2}, & p/q < d/2 \end{array}$$

Note:

- $O(\varepsilon^{-p/q})$ is the cost of a single ε-accurate PDE calculation
- $O(\varepsilon^{-d/2})$ is the cost of an ε-accurate interpolation of unit cost data

In this sense the method has near-optimal asymptotic efficiency

MLFA for PDEs with sparse interpolation

With sparse interpolation the accuracy requirement becomes

$$c_2 \sum_{\ell=0}^{L} 2^{-2\ell} (\ell+1)^{d-1} h_\ell^q \;\approx\; \varepsilon/2.$$

and the cost becomes

$$C = c_3 \sum_{\ell=0}^{L} 2^{\ell} (\ell+1)^{d-1} h_\ell^{-p}$$

Optimising this results in the total cost being

$$\begin{array}{rl} \varepsilon^{-p/q}, & p/q > 1/2 \\ \varepsilon^{-p/q}\, |\log \varepsilon|^{3d/2}, & p/q = 1/2 \\ \varepsilon^{-1/2}\, |\log \varepsilon|^{3(d-1)/2}, & p/q < 1/2 \end{array}$$

Note:

- $O(\varepsilon^{-p/q})$ is again the cost of a single ε-accurate PDE calculation
- $O(\varepsilon^{-1/2}\, |\log \varepsilon|^{3(d-1)/2})$ is the cost of an ε-accurate sparse interpolation of unit cost data

MLFA for SDEs

If $\beta > \gamma$ in the standard SDE discretisation sense, then randomised MLMC (see [RG15]) can be used to give an unbiased estimator Y, with $\mathbb{E}[Y(\theta;\omega)] = f(\theta)$ and finite variance and expected cost. If

$$\begin{aligned} \|(I_\ell - I)\,[f]\,\| &< c_1\, 2^{-r\ell} \\ \mathbb{V}\,[(I_\ell - I_{\ell-1})\,[Y]\,] &< c_2\, 2^{-s\ell} \end{aligned}$$

and the total expected cost is bounded by $c_3 \sum_0^L 2^{d\ell} M_\ell$, for M_ℓ samples per level, then ε r.m.s. accuracy can be achieved with cost of order

$$\begin{cases} \varepsilon^{-2}, & s > d \\ \varepsilon^{-2} |\log \varepsilon|^2, & s = d \\ \varepsilon^{-2-(d-s)/r}, & s < d \end{cases}$$

MLFA for SDEs

The previous result is a slight generalisation of Heinrich's analysis which assumed $s = 2r$.

With sparse interpolation, the cost is reduced to order

$$
\begin{cases}
\varepsilon^{-2}, & s > 1 \\
\varepsilon^{-2} |\log \varepsilon|^{2+3(d-1)}, & s = 1 \\
\varepsilon^{-2-(1-s)/r} |\log \varepsilon|^{(3+(1-s)/r)(d-1)}, & s < 1
\end{cases}
$$

MLFA for SDEs

If $\beta \leq \gamma$, then we can use a MIMC combination of path-based MLMC and Heinrich's MLMC. The starting point is the interpolation decomposition:

$$f \approx \sum_{\ell=0}^{L} (I_\ell - I_{\ell-1})[f], \qquad I_{-1}[f] \equiv 0,$$

where I_ℓ uses a dense interpolation with spacing proprtional to $2^{-\ell}$.

We then replace f with a timestep approximation expansion

$$f \approx \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L'_\ell} \Delta I_\ell[\Delta f_{\ell'}], \qquad \Delta I_\ell[\Delta f_{\ell'}] \equiv (I_\ell - I_{\ell-1})[f_{\ell'} - f_{\ell'-1}]$$

in which L'_ℓ is a decreasing function of ℓ, since $\Delta I_\ell[f]$ becomes smaller as ℓ increases and so less relative accuracy is required in its approximation.

MLFA for SDEs

The final step is to replace $\Delta I_\ell[\Delta f_{\ell'}]$ by a Monte Carlo estimate, giving the MIMC-style estimator

$$\widetilde{f} = \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L'_\ell} \left(\frac{1}{M_{\ell,\ell'}} \sum_{m=1}^{M_{\ell,\ell'}} \Delta I_\ell[\Delta g_{\ell'}(\cdot;\omega^{\ell,\ell',m})] \right)$$

We now need to choose $L, L'_\ell, M_{\ell,\ell'}$ to achieve the desired accuracy at the minimum cost.

$$\begin{aligned} \mathbb{E}[\widetilde{f}-f] &= (I_L - I)[f] + \sum_{\ell=0}^{L} (I_\ell - I_{\ell-1})[f_{L'(\ell)} - f] \\ \Longrightarrow \quad \left\| \mathbb{E}[\widetilde{f}-f] \right\| &\leq \|(I_L - I)[f]\| + \sum_{\ell=0}^{L} \|(I_\ell - I_{\ell-1})[f_{L'(\ell)} - f]\| \end{aligned}$$

and

$$\mathbb{V}[\widetilde{f}] = \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L'_\ell} \left(\frac{1}{M_{\ell,\ell'}} \mathbb{V}\left[\Delta I_\ell[\Delta g_{\ell'}(\cdot;\omega)]\right] \right)$$

MLFA for SDEs

If we have

$$\begin{aligned} \| \Delta I_{\ell'}[\Delta f_\ell] \, \| &< c_1 \, 2^{-\alpha\ell - r\ell'} \\ \mathbb{V}\left[\Delta I_{\ell'}[\Delta g_\ell]\right] &< c_2 \, 2^{-\beta\ell - s\ell'} \end{aligned}$$

and the total cost is bounded by

$$c_3 \sum_{\ell=0}^{L} \sum_{\ell'=0}^{L'_\ell} 2^{\gamma\ell + d\ell'} M_{\ell,\ell},$$

then ε RMS accuracy can be achieved at a computational cost of order

$$\begin{aligned} &\varepsilon^{-2}, && \eta < 0 \\ &\varepsilon^{-2-\eta} \, |\log \varepsilon|^p, && \eta \geq 0 \end{aligned}$$

for some p (see MIMC analysis by [HNT16]), where

$$\eta = \max\left(\frac{\gamma - \beta}{\alpha}, \frac{d-s}{r} \right).$$

MLFA for SDEs

Note: in the best case when $\eta < 0$, the dominant contribution to the total cost comes from the base level $\ell = \ell' = 0$, which is why there are no log terms in its complexity.

With sparse interpolation the corresponding cost is of order

$$\begin{array}{ll} \varepsilon^{-2}, & \eta < 0 \\ \varepsilon^{-2-\eta} \, |\log \varepsilon|^q, & \eta \geq 0 \end{array}$$

for some q, where now

$$\eta = \max\left(\frac{\gamma - \beta}{\alpha}, \frac{1-s}{r} \right).$$

Conclusions and future work

Conclusions:

- excellent asymptotic efficiency in approximating parametric functions arising from PDEs and SDEs – nearly optimal in some cases
- meta-theorems make various assumptions which need to be verified, especially for mixed derivatives when using sparse grid interpolation

On-going work:

- numerical results
- numerical analysis of PDEs to prove validity of mixed derivative assumptions in specific cases (building on prior research within the sparse grid community)
- numerical analysis of SDEs to prove validity of mixed derivative assumptions in specific cases (building on prior analysis by Giles and Sheridan-Methven)

References

H.-J. Bungartz, M. Griebel. 'Sparse grids'. pp.147-269, Acta Numerica, 2004.

F. De Angelis. PhD thesis, in preparation, Oxford University, 2024.

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

M.B. Giles, O. Sheridan-Methven, 'Analysis of nested multilevel Monte Carlo using approximate Normal random variables'. SIAM/ASA Journal on Uncertainty Quantification 10(1):200-226, 2022.

A.-L. Haji-Ali, F. Nobile, R. Tempone. 'Multi-index Monte Carlo: when sparsity meets sampling'. Numerische Mathematik, 132:767-806, 2016.

S. Heinrich. 'Multilevel Monte Carlo methods'. Lecture Notes in Computer Science, 2179:58-67, 2001.

C.-H. Rhee, P.W. Glynn. 'Unbiased estimation with square root convergence for SDE models'. Operations Research, 63(5):1026-1043, 2015.

R. Tempone, S. Wolfers. 'Smolyak's algorithm: a powerful black box for the acceleration of scientific computations'. pp.201-228 in Sparse Grids and Applications, Springer, 2018.