Reduced-Physics Multilevel Monte Carlo Methods for Uncertainty Quantification in Complex Reservoirs

Øystein S. Klemetsdal   Knut-Andreas Lie   Stein Krogstad

Department of Mathematics and Cybernetics, SINTEF Digital, Norway

InterPore 2020 – 12th Annual meeting
31 August – 4 September 2020, Online

Introduction

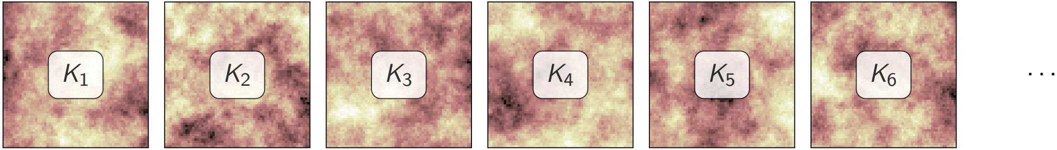
Uncertainty quantification in reservoir simulation

- Quantity of interest u typically derived from simulation results
  Saturation at time t', water production rate, total oil production, etc.

Introduction

Uncertainty quantification in reservoir simulation

- Quantity of interest u typically derived from simulation results
  Saturation at time t', water production rate, total oil production, etc.
  ... which are all random variables due to uncertain subsurface properties



Monte Carlo Method

- Let u be random variable with expected value $\mathbb{E}[u]$ and variance $\mathbb{V}[u]$
- Approximate $\mathbb{E}[u]$ from independent, identically distributed samples $u^1, \ldots, u^N$

$$\mathbb{E}[u] \approx E(u) = \frac{1}{N}\sum_{i=1}^{N} u^i, \quad \mathbb{V}\left[E(u)\right] = \mathbb{E}\left[\left(E(u) - \mathbb{E}\left[E(u)\right]\right)^2\right] = \frac{1}{N}\mathbb{V}[u]$$

Monte Carlo Method

- Let u be random variable with expected value $\mathbb{E}[u]$ and variance $\mathbb{V}[u]$
- Approximate $\mathbb{E}[u]$ from independent, identically distributed samples $u^1, \dots, u^N$

$$\mathbb{E}[u] \approx E(u) = \frac{1}{N}\sum_{i=1}^{N} u^i, \quad \mathbb{V}\left[E(u)\right] = \mathbb{E}\left[\left(E(u) - \mathbb{E}\left[E(u)\right]\right)^2\right] = \frac{1}{N}\mathbb{V}[u]$$

- **Upsides:** Easy to implement and easy to parallelize
- **Downside:** Root mean square error (RMSE) of the estimator is

$$\text{RMSE} = \sqrt{\mathbb{V}\left[E(u)\right]} = \mathcal{O}(N^{-1/2})$$

$\rightarrow$ Accuracy RMSE $< \varepsilon$ requires $N = \mathcal{O}(\varepsilon^{-2})$ samples!

Multilevel Monte Carlo Method (Giles [2015])

- Premise: we can obtain less expensive approximation $u_{\ell-1}$ of u_ℓ for $u_1, \ldots, u_L \equiv u$
- Express expected value as telescopic sum (with $u_0 \equiv 0$)

$$\mathbb{E}[u_L] = \sum_{\ell=1}^{L} \mathbb{E}[u_\ell - u_{\ell-1}]$$

- ... with unbiased estimator

$$\mathbb{E}[u_L] \approx E(u_L) = \sum_{\ell=1}^{L} \left(\frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(u_\ell^{(\ell,i)} - u_{\ell-1}^{(\ell,i)} \right) \right)$$

Multilevel Monte Carlo Method (Giles [2015])

- Premise: we can obtain less expensive approximation $u_{\ell-1}$ of u_ℓ for $u_1, \dots, u_L \equiv u$
- Express expected value as telescopic sum (with $u_0 \equiv 0$)

$$\mathbb{E}[u_L] = \sum_{\ell=1}^{L} \mathbb{E}[u_\ell - u_{\ell-1}]$$

- ... with unbiased estimator

$$\mathbb{E}[u_L] \approx E(u_L) = \sum_{\ell=1}^{L} \left(\frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(u_\ell^{(\ell,i)} - u_{\ell-1}^{(\ell,i)} \right) \right)$$

  - Quantities $u_\ell^{(\ell,i)}$ and $u_{\ell-1}^{(\ell,i)}$ from the *same* random sample i, *different* for each ℓ

Multilevel Monte Carlo Method (Giles [2015])

- Total cost C (e.g., CPU time) and total variance V:

$$C = \sum_{\ell=1}^{L} N_\ell C_\ell, \qquad V = \sum_{\ell=1}^{L} \frac{V_\ell}{N_\ell}$$

C_ℓ	Cost of computing single sample of $u_\ell - u_{\ell-1}$	V_ℓ	Variance $\mathbb{V}[u_\ell - u_{\ell-1}]$

- Minimize total cost C for a fixed variance ε^2

$$\min \sum_{\ell=1}^{L} N_\ell C_\ell \quad \text{s.t.} \quad \sum_{\ell=1}^{L} \frac{V_\ell}{N_\ell} = \varepsilon^2 \quad \rightarrow \quad N_\ell = \varepsilon^{-2} \left(\sum_{k=1}^{L} \sqrt{V_k C_k} \right) \sqrt{\frac{V_\ell}{C_\ell}}$$

Multilevel Monte Carlo Method (Giles [2015])

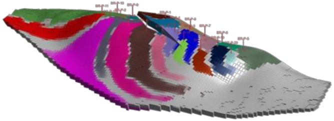
- Coarser levels typically defined by spatial upscaling of the finest level
  - Challenging for complex geomodels (channels, different rock types, etc.)
  - The best methods are computationally expensive and may need hands-on tuning
- ... but *level* does not necessarily refer to spatial upscaling!
  - Temporal discretization
  - Multiscale methods with varying number of basis functions (Efendiev et al. [2013])
  - Solver-based, e.g., fully- and sequential-implicit (Müller et al. [2013, 2014])

Multilevel Monte Carlo Method (Giles [2015])

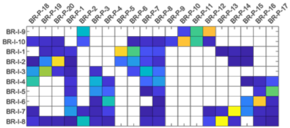
- Coarser levels typically defined by spatial upscaling of the finest level
  - Challenging for complex geomodels (channels, different rock types, etc.)
  - The best methods are computationally expensive and may need hands-on tuning
- … but *level* does not necessarily refer to spatial upscaling!
  - Temporal discretization
  - Multiscale methods with varying number of basis functions (Efendiev et al. [2013])
  - Solver-based, e.g., fully- and sequential-implicit (Müller et al. [2013, 2014])

Herein: Generic MLMC framework with spatial/temporal upscaling, solvers, and tools from *flow diagnostics* to define coarser levels
→ Reduced-physics Multilevel Monte Carlo Method

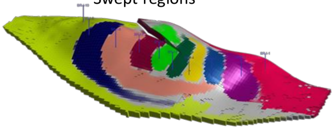
Drainage regions



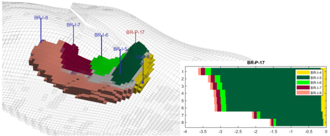
Well-pair communication



Swept regions

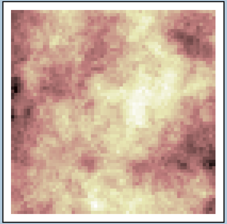


Allocations/volumetrics



Flow diagnostics: a family of *simple and controlled numerical flow experiments* that are run to probe a reservoir model, establish connections and basic volume estimates, and quickly provide a qualitative picture of the flow patterns in the reservoir and quantitative measures of the heterogeneity in dynamic flow paths (Lie [2019]).

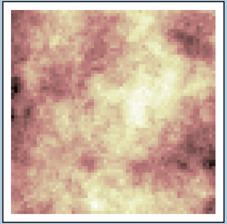
Two examples



Quarter five-spot pattern

- `hello world` of reservoir simulation: 2D quarter five-spot
- Water injection in oil-filled reservoir
- Incompressible flow, quadratic relperms, $\mu_o/\mu_w = 2$
- Model $\log K$ by Gaussian process (Wood and Chan [2017])

Two examples

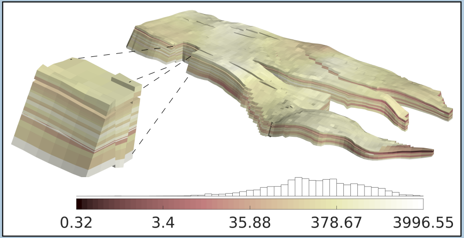


Quarter five-spot pattern

- `hello world` of reservoir simulation: 2D quarter five-spot
- Water injection in oil-filled reservoir
- Incompressible flow, quadratic relperms, $\mu_o/\mu_w = 2$
- Model log K by Gaussian process (Wood and Chan [2017])

Norne Field Model (OPM [2019])

- Real oil & gas field in the Norwegian sea
- Rock from log-normal distribution (Lorentzen et al. [2019])
- Inject 1 PV water into oil over 2 years
- Compressible flow, quadratic relperms



Two examples

Quantity of interest: **recovery factor (RF)** after injection of 1PV

Two examples

Quantity of interest: **recovery factor (RF)** after injection of 1PV

Strategy 1: Upscaling MLMC

Four levels defined by spatial upscaling (averaging of $\log K$)

- Use fully-implicit solver on each level and compute recovery factor

Two examples

Quantity of interest: **recovery factor (RF)** after injection of 1PV

Strategy 1: Upscaling MLMC

Four levels defined by spatial upscaling (averaging of $\log K$)

- Use fully-implicit solver on each level and compute recovery factor

Strategy 2: Reduced-physics MLMC

Three levels defined by different solution strategies

1. Compute pressure, solve time-of-flight (TOF) and estimate recovery factor
2. Sequential-implicit solve (no outer iterations) and compute recovery factor
3. Fully-implicit solve and compute recovery factor

Two examples

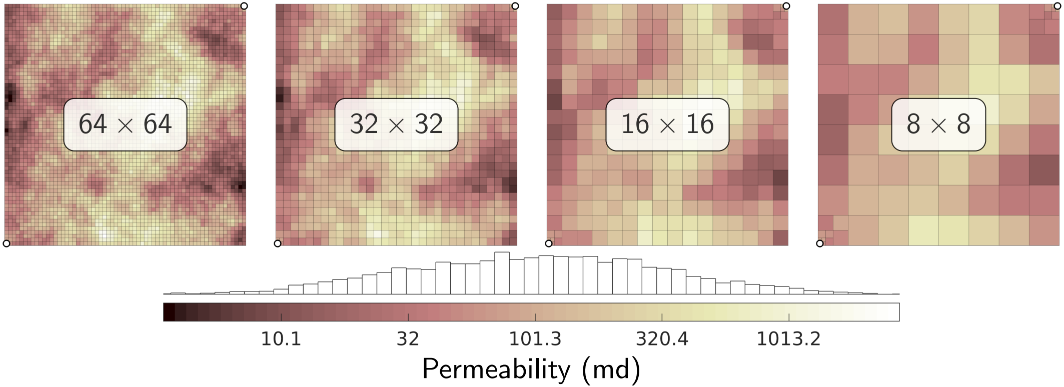
Challenge: TOF-based estimate tends to be biased
Solution: During warm-up, solve the same samples on level 1 and 3 $\rightarrow$ correct level 1

Strategy 2: Reduced-physics MLMC

Three levels defined by different solution strategies

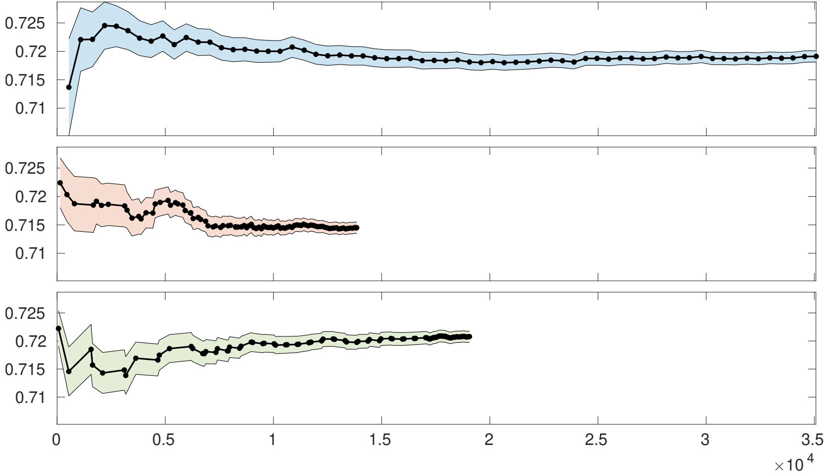
1. Compute pressure, solve time-of-flight (TOF) and estimate recovery factor
2. Sequential-implicit solve (no outer iterations) and compute recovery factor
3. Fully-implicit solve and compute recovery factor

Upscaling MLMC hierarchy



Example 1: Quarter five-spot problem

Recovery factor estimate vs. cost (serial simulation time [s])



MC

RF: 0.719

RMSE: 9.99×10^{-4}

MLMC (upscaling)

RF: 0.712

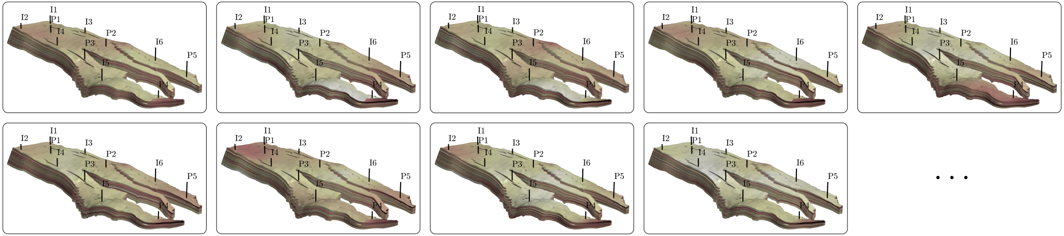
RMSE: 9.96×10^{-4}

MLMC (reduced-phys.)

RF: 0.721

RMSE: 9.97×10^{-4}

Example 2: Norne Field Model

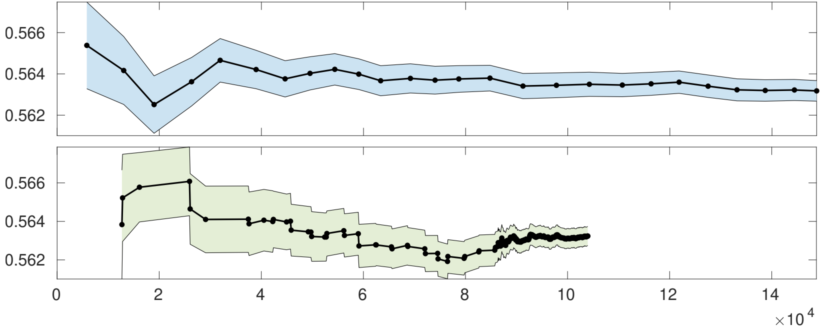


- Corner-point grid: $46 \times 112 \times 22$, 44 915 active cells, complex topology (27 neighbors!)
- Faults, stratigraphic layers, pinch-outs, transmissibility multipliers, etc.

  $\rightarrow$ Significantly more challenging to upscale

Example 2: Norne Field Model

Recovery factor estimate vs. cost (serial simulation time [s])



MC

RF: 0.5632
RMSE: 4.98×10^{-4}

MLMC (reduced-phys.)

RF: 0.5632
RMSE: 4.96×10^{-4}

Concluding Remarks

Conclusions

- Flow diagnostics tools are very well suited as coarse-level solvers in MLMC
- Exemplified here with using time-of-flight to estimate recovery factor
- May be significantly more flexible than upscaling, especially for complex reservoirs

Concluding Remarks

Conclusions

- Flow diagnostics tools are very well suited as coarse-level solvers in MLMC
- Exemplified here with using time-of-flight to estimate recovery factor
- May be significantly more flexible than upscaling, especially for complex reservoirs

Further work

- Experiment with other flow diagnostics tools
  (tracer, drainage regions, data-driven models, etc.)
- Experiment with other types of uncertainty
- Investigate potential for industry-grade fluid physics complexity

Acknowledgements

The research reported in this presentation was funded in part by the Research Council of Norway through grant no. 280950 and in part by Equinor Energy AS, Total E&P Norge AS, and Wintershall DEA Norge AS

mrst.no

Source code coming soon to the open-source MATLAB Reservoir Simulation Toolbox

References

Y. Efendiev, O. Iliev, and C. Kronsbein. Multilevel Monte Carlo methods using ensemble level mixed MsFEM for two-phase flow and transport simulations. *Comput. Geosci.*, 17(5):833–850, 2013. ISSN 14200597. doi: 10.1007/s10596-013-9358-y.

M. B. Giles. Multilevel monte carlo methods. *Acta Numer.*, pages 259–328, 2015. doi: 10.1017/S096249291500001X.

K.-A. Lie. *An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, 2019. doi: 10.1017/9781108591416.

R. Lorentzen, X. Luo, T. Bhakta, and R. Valestrand. History matching the full norne field model using seismic and production data. *SPE J.*, 2019. doi: 10.2118/194205-PA.

F. Müller, P. Jenny, and D. W. Meyer. Multilevel Monte Carlo for two phase flow and Buckley-Leverett transport in random heterogeneous porous media. *J. Comput. Phys.*, 250:685–702, 2013. ISSN 10902716. doi: 10.1016/j.jcp.2013.03.023.

F. Müller, D. W. Meyer, and P. Jenny. Solver-based vs. grid-based multilevel Monte Carlo for two phase flow and transport in random heterogeneous porous media. *J. Comput. Phys.*, 268:39–50, 2014. ISSN 10902716. doi: 10.1016/j.jcp.2014.02.047. URL `http://dx.doi.org/10.1016/j.jcp.2014.02.047`.

OPM. The Open Porous Media (OPM) Initative, 2019. URL `https://opm-project.org/`.

A. T. A. Wood and G. Chan. Simulation of stationary gaussian processes in $[0,1]^d$. 3(4):409–432, 2017. doi: 10.1080/10618600.1994.10474655.

Developed by nuclear physicist Stanislaw Ulam during the Manhattan Project in the late 1940's

It was at that time that I suggested an obvious name for the statistical method – a suggestion not unrelated to the fact that Stan had an uncle who would borrow money from relatives because he "just had to go to Monte Carlo"

— Nicholas Metropolis, *The Beginning of the Monte Carlo Method* (1987)

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$\mathcal{E}^{\mathrm{ML}}(u_L)^2 = \mathbb{E}\left[\left(E(u_L) - \mathbb{E}[u]\right)^2\right]$$

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$
\begin{aligned}
\mathcal{E}^{\mathrm{ML}}(u_L)^2 &= \mathbb{E}\left[(E(u_L) - \mathbb{E}[u])^2\right] \\
&= \mathbb{E}\left[(E(u_L) - \mathbb{E}\left[E(u_L)\right] + \mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u])^2\right]
\end{aligned}
$$

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$
\begin{aligned}
\mathcal{E}^{\mathrm{ML}}(u_L)^2 &= \mathbb{E}\left[\left(E(u_L) - \mathbb{E}[u]\right)^2\right] \\
&= \mathbb{E}\left[\left(E(u_L) - \mathbb{E}\left[E(u_L)\right] + \mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u]\right)^2\right] \\
&= \underbrace{\mathbb{E}\left[\left(E(u_L) - \mathbb{E}\left[E(u_L)\right]\right)^2\right]}_{\text{Sampling error}} + \underbrace{\left(\mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u]\right)^2}_{\text{Approximation error}}
\end{aligned}
$$

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$\mathcal{E}^{\mathrm{ML}}(u_L)^2 = \underbrace{\mathbb{V}\left[E(u_L)\right]}_{\text{Sampling error}} + \underbrace{\left(\mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u]\right)^2}_{\text{Approximation error}}$$

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$\mathcal{E}^{\mathrm{ML}}(u_L)^2 = \underbrace{\mathbb{V}\left[E(u_L)\right]}_{\text{Sampling error}} + \underbrace{\left(\mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u]\right)^2}_{\text{Approximation error}}$$

- Sampling error $< \varepsilon^2/2$ and approximation error $< \varepsilon^2/2$ ensures $\mathcal{E}^{\mathrm{ML}}(u_L) < \varepsilon$

Extra: Multilevel Monte Carlo

- Generally, u_ℓ is obtained by simulation, so that u_L is approximation of u
- Mean square error is now

$$\mathcal{E}^{\mathrm{ML}}(u_L)^2 = \underbrace{\mathbb{V}\left[E(u_L)\right]}_{\text{Sampling error}} + \underbrace{\left(\mathbb{E}\left[E(u_L)\right] - \mathbb{E}[u]\right)^2}_{\text{Approximation error}}$$

- Sampling error $< \varepsilon^2/2$ and approximation error $< \varepsilon^2/2$ ensures $\mathcal{E}^{\mathrm{ML}}(u_L) < \varepsilon$
- Simplify notation: let E_ℓ be MC estimator of $u_\ell - u_{\ell-1}$

$$E_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(u_\ell^{(\ell,i)} - u_{\ell-1}^{(\ell,i)} \right), \quad E(u_L) = \sum_{\ell=0}^{L} E_\ell$$

Extra: Multilevel Monte Carlo

Theorem

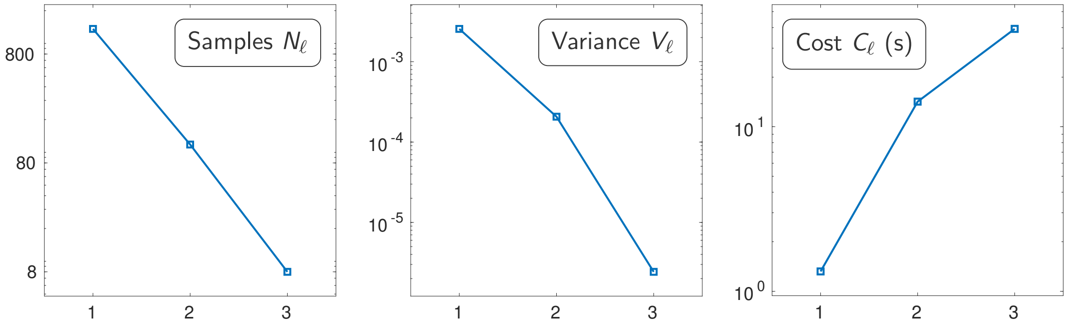
If there exists independent estimators E_ℓ based on N_ℓ MC samples, with expected cost C_ℓ and variance V_ℓ, and $\alpha, \beta, \gamma, c_1, c_2, c_3 > 0$ such that $\alpha \geq \min(\beta, \gamma)/2$, and

1. $|\mathbb{E}[u_\ell - u]| \leq c_1 2^{-\alpha\ell}$ *(Increase in accuracy)*
2. $V_\ell \leq c_2 2^{-\beta\ell}$ *(Decrease in variance)*
3. $C_\ell \leq c_3 2^{\gamma\ell}$ *(Increase in cost)*

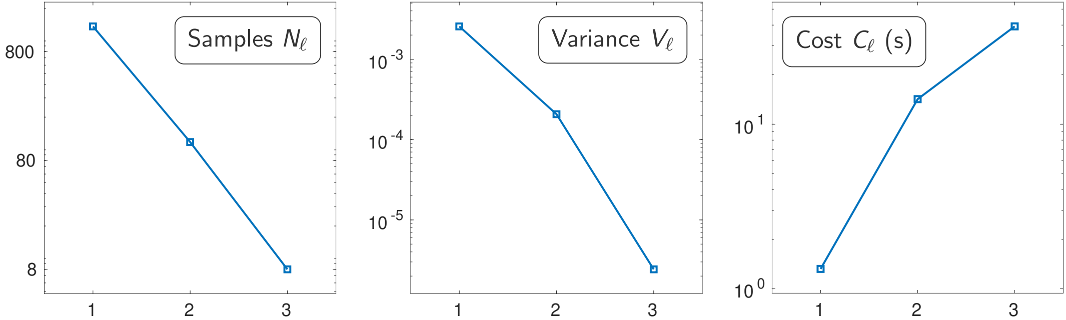
Then, there exists $c_4 > 0$ such that for $\varepsilon < e^{-1}$, there are L, N_ℓ for which the estimator $E(u_L) = \sum_\ell E_\ell$ has $\mathcal{E}^{\mathrm{ML}}(u_L) < \varepsilon$, and

$$\mathbb{E}[C] \leq \begin{cases} c_4 \varepsilon^{-2} & \beta > \gamma \\ c_4 \varepsilon^{-2} \log(\varepsilon)^2 & \beta = \gamma \\ c_4 \varepsilon^{-2-(\gamma-\beta)/\alpha} & \beta < \gamma \end{cases}$$

Extra: Example 1



Extra: Example 1



- Estimated RMSE $\approx 2.0 \times 10^{-3}$, Total cost of MLMC: $\sum_\ell N_\ell C_\ell \approx 3.79 \times 10^3$ s
- Assuming $\mathbb{V}[u_L] \approx \mathbb{V}[u_1]$ and cost of computing one sample of $u_L \approx C_L$
  $\rightarrow$ Cost of MC simulation with the same accuracy $= 2.46 \times 10^6$ s