On Multilevel Best Linear Unbiased Estimators

Daniel Schaden, Elisabeth Ullmann

Workshop “Optimization and Inversion under Uncertainty”

RICAM Linz

Outline

- Linear models, best linear unbiased estimators (BLUEs)
- Variance bound, monomial example
- Sample allocation, PDE example
- Asymptotic analysis of multilevel BLUEs

Motivation (1/3)

- **Goal:** Estimate expected value $\mathbb{E}[Z]$ of scalar valued random variable Z
- Can generate samples of approximations $Z_\ell \approx Z$, $\ell = 1, \dots, L$
- Expectations $\mu = (\mu_1, \dots, \mu_L)^\top \in \mathbb{R}^L$
- Covariance Matrix $C = (\mathrm{Cov}(Z_k, Z_\ell))_{k,\ell=1}^L \in \mathbb{R}^{L \times L}$

Motivation (2/3)

- Many linear, unbiased estimators for $\mu_L = \mathbb{E}[Z_L] \approx \mathbb{E}[Z]$
  - Monte Carlo
  $$\widehat{\mu}_L^{MC} = \frac{1}{m_L} \sum_{i=1}^{m_L} Z_L(\omega_i)$$
  - Multilevel Monte Carlo (Giles 2008)
  $$\widehat{\mu}_L^{MLMC} = \sum_{\ell=1}^{L} \frac{1}{m_\ell} \sum_{i=1}^{m_\ell} (Z_\ell(\omega_i^\ell) - Z_{\ell-1}(\omega_i^\ell))$$
  - Multifidelity Monte Carlo (Peherstorfer et al. 2016)
  - Approximate Control Variates (Gorodetsky et al. 2018)
- **Are these estimators optimal?** $\longrightarrow$ minimal variance

Motivation (3/3)

- **Linear combination** of Monte Carlo estimators:

$$\widehat{\mu}_L = \sum_{k=1}^{K} \sum_{\ell \in S^k} \beta_\ell^k \frac{1}{m_k} \sum_{i=1}^{m_k} Z_\ell(\omega_i^k)$$

- Unbiasedness (row sum constraint): $\sum_{k=1}^{K} \beta_\ell^k = \delta_{\ell,L}$

MC

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5						
Z_4						
Z_3						
Z_2						
Z_1						

MLMC

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5					1.00	-1.00
Z_4				1.00	-1.00	
Z_3			1.00	-1.00		
Z_2		1.00	-1.00			
Z_1	1.00	-1.00				

Example

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5				-1.00	2.00	-1.00
Z_4		0.25			-0.50	0.25
Z_3	-1.00	1.00	-1.00	1.00		
Z_2		-3.00	3.00			
Z_1	2.00	-2.00				

Example (1/3)

- Multilevel Monte Carlo with *independent* inputs $\omega_1^2, \omega_2^2, \omega_1^1$

$$\widehat{\mu}_2^{MLMC} = \frac{1}{2}\sum_{i=1}^{2}(Z_2(\omega_i^2) - Z_1(\omega_i^2)) + Z_1(\omega_1^1)$$
$$\approx (\mu_2 - \mu_1) + \mu_1$$

- Two samples to estimate $\mu_2 - \mu_1$ and one sample to estimate μ_1

- **Rearrange:**

$$\widehat{\mu}_2^{MLMC} = \frac{1}{2}\left((Z_2(\omega_1^2) - Z_1(\omega_1^2)) + (Z_2(\omega_2^2) - Z_1(\omega_2^2)\right) + Z_1(\omega_1^1)$$
$$= \frac{1}{2}(\mathbf{Z}_2(\omega_1^2) + \mathbf{Z}_2(\omega_2^2)) + \mathbf{Z}_1(\omega_1^1) - \frac{1}{2}(\mathbf{Z}_1(\omega_1^2) + \mathbf{Z}_1(\omega_2^2))$$

- **Three samples of $\mathbf{Z}_1$** and **two samples of $\mathbf{Z}_2$**

Example (2/3)

- **Key idea:** treat model evaluations as observations of unknown "truth" $\mathbb{E}[Z]$
- Example: five (correlated) observations of Z_1 and Z_2
- Reformulation as **linear model**:

$$\begin{bmatrix} Z_1(\omega_1^2) \\ Z_2(\omega_1^2) \\ Z_1(\omega_2^2) \\ Z_2(\omega_2^2) \\ Z_1(\omega_1^1) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \end{bmatrix} + \begin{bmatrix} Z_1(\omega_1^2) - \mu_1 \\ Z_2(\omega_1^2) - \mu_2 \\ Z_1(\omega_2^2) - \mu_1 \\ Z_2(\omega_2^2) - \mu_2 \\ Z_1(\omega_1^1) - \mu_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \eta^2(\omega_1^2) \\ \eta^2(\omega_2^2) \\ \eta^1(\omega_1^1) \end{bmatrix}$$

- Observation vector Y, Design matrix H, parameter μ, mean zero noise ε

$$Y = H\mu + \varepsilon$$

Example (3/3)

- Model groups $S^1 = \{1\}$ and $S^2 = \{1,2\}$ with $S^1, S^2 \subseteq \{1,2\}$, $L = 2$
- $m_1 = 1,\ m_2 = 2$

$$Z^1 = \begin{bmatrix} Z_1 \end{bmatrix} = \begin{bmatrix} \mu_1 \end{bmatrix} + \begin{bmatrix} Z_1 - \mu_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mu + \eta^1, \qquad R^1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$Z^2 = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} Z_1 - \mu_1 \\ Z_2 - \mu_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mu + \eta^2, \qquad R^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} Z_1(\omega_1^2) \\ Z_2(\omega_1^2) \\ Z_1(\omega_2^2) \\ Z_2(\omega_2^2) \\ Z_1(\omega_1^1) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_1 \\ \mu_2 \\ \mu_1 \end{bmatrix} + \begin{bmatrix} Z_1(\omega_1^2) - \mu_1 \\ Z_2(\omega_1^2) - \mu_2 \\ Z_1(\omega_2^2) - \mu_1 \\ Z_2(\omega_2^2) - \mu_2 \\ Z_1(\omega_1^1) - \mu_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \eta^2(\omega_1^2) \\ \eta^2(\omega_2^2) \\ \eta^1(\omega_1^1) \end{bmatrix}$$

Linear Model (1/3)

- Model groups $S^k \in 2^{\{1,\dots,L\}} \setminus \{\emptyset\}$, $k = 1, \dots, 2^L - 1 =: K$

Linear model within group

$$Z^k = R^k \mu + \eta^k, \qquad \text{(Sample linear regression problem)}$$

$$Z^k := (Z_\ell)_{\ell \in S^k}, \qquad \text{(Observations)}$$

$$\eta^k := (Z_\ell - \mu_\ell)_{\ell \in S^k}, \qquad \text{(Mean zero noise, Covariance } C_{S^k,S^k}\text{)}$$

$$R^k v := (v_\ell)_{\ell \in S^k}, \quad \text{for all } v \in \mathbb{R}^L \qquad \text{(Restriction operator, Design matrix)}$$

- Idea: Collect group linear models and samples in a (huge) block linear model

Block linear model

$m_k \in \mathbb{N}_0$ samples of models in $S^k \in 2^{\{1,\dots,L\}} \setminus \{\emptyset\}$ with independent events ω_i^k

$$Y = H\mu + \varepsilon, \qquad \text{(Linear model)}$$

$$Y := ((Z^k(\omega_i^k))_{i=1}^{m_k})_{k=1}^K, \qquad \text{(Observations)}$$

$$H := ((R^k)_{i=1}^{m_k})_{k=1}^K \qquad \text{(Design matrix)}$$

$$\varepsilon := ((\eta^k(\omega_i^k))_{i=1}^{m_k})_{k=1}^K \qquad \text{(Mean zero noise)}$$

Linear Model (2/3)

- $Y = H\mu + \varepsilon$
- $y := H^\top \widehat{C}^{-1} Y = \sum_{k=1}^{K} P^k (C^k)^{-1} \sum_{i=1}^{m_k} Z^k(\omega_i^k)$
- $\Psi := H^\top \widehat{C}^{-1} H = \sum_{k=1}^{K} m_k P^k (C^k)^{-1} R^k$
- **BLUE**
  $\widehat{\mu}_L^B := (H^\top \widehat{C}^{-1} H)^{-1} H^\top \widehat{C}^{-1} Y = e_L^T \Psi^{-1} y$

- $C^k := C_{S^k, S^k}$
- $P^k := (R^k)^T$

Theorem (Gauss-Markov-Aitken)

Let C be positive definite and assume each model Z_ℓ is evaluated at least once

$$\{\ell \in \{1, \ldots, L\} \,|\, \exists m_k > 0 \text{ with } \ell \in S^k\} = \{1, \ldots, L\}.$$

Then $\widehat{\mu}_L^B$ is well defined and is the (unique) Best Linear Unbiased Estimator **(BLUE)** with variance $\mathrm{Var}(\widehat{\mu}_L^B) = e_L^T \Psi^{-1} e_L$.

Linear Model (3/3)

Properties of the BLUE: $\widehat{\mu}_L^B = e_L^T \Psi^{-1} y$

- Normal equations (small!): $\Psi \widehat{\mu}^B = y, \quad \Psi \in \mathbb{R}^{L \times L}, \; y \in \mathbb{R}^L$
- $\widehat{\mu}_L^B$ is the last component of the solution vector of the normal equations
- $\mathrm{Var}(\widehat{\mu}_L^B) \leq \mathrm{Var}(\widehat{\mu}_L)$ for any linear unbiased estimator $\widehat{\mu}_L$ with the same sample sequence $(m_1, \ldots, m_K)$
- **Linear combination** of Monte Carlo estimators:

$$\widehat{\mu}_L^B = \sum_{k=1}^{K} \sum_{\ell \in S^k} \beta_\ell^k \frac{1}{m_k} \sum_{i=1}^{m_k} Z_\ell(\omega_i^k)$$

Outline

- Linear models, best linear unbiased estimators (BLUEs)
- **Variance bound, monomial example**
- Sample allocation, PDE example
- Asymptotic analysis of multilevel BLUEs

Lower bound for estimator variance

- Follow setup in [Gorodetsky et al., 2018]
- Split indices: $Q, Q_\infty \subseteq \{1, \dots, L\}$, $Q \cup Q_\infty = \{1, \dots, L\}$, $Q \subsetneq Q_\infty$
- Model groups in Q are evaluated M-times
- Model groups in Q_∞ are evaluated N-times, consider limit $N \to +\infty$
- Lower bound for variance of BLUE for $Q = \{1, \dots, L\}$ and $Q_\infty = \{1, \dots, L-1\}$:

$$\mathrm{Var}(\widehat{\mu}_L^B) \geq \left(C_{L,L} - C_{L,1:L-1}\, C_{1:L-1,1:L-1}^{-1}\, C_{1:L-1,L}\right)/M$$

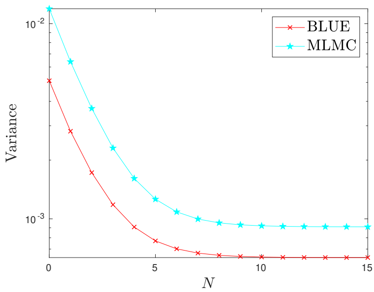
(same as for Approximate Control Variates in [Gorodetsky et al., 2018])

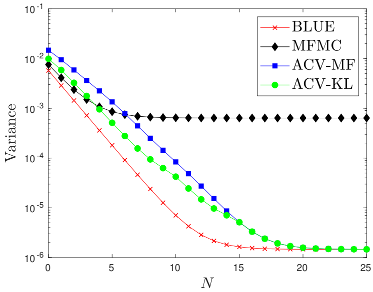
Monomial Example (1/3)

[Gorodetsky et al., 2018]

- $Z_\ell(\omega) = \omega^\ell$ for $\ell = 1, \ldots, 5$ with $\omega \sim U(0,1)$.
- Fix the total number of model evaluations

$$n_\ell = 2^N 2^{L-\ell}, \quad \text{for } Z_1, \ldots, Z_4,$$
$$n_L = 1, \quad \text{for } Z_5.$$



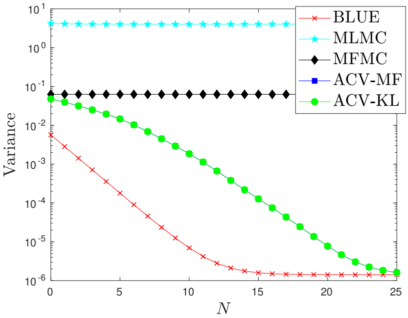


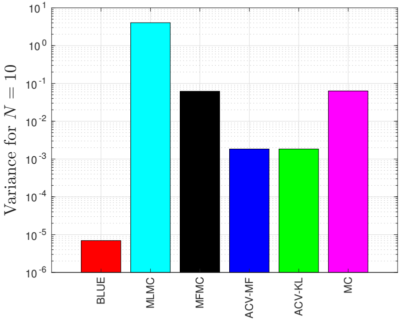
Monomial Example (2/3)

- Same example but with noise
- $Z_\ell(\omega, \xi) = \omega^{\ell-1} + \xi$ for $\ell = 1, \ldots, 5$ with $\omega \sim U(0,1)$, $\xi \sim N(0,2)$.
- $Z_6(\omega, \xi) = \omega^5$
- Small (empirical) correlation of $Z_1, \ldots, Z_5$ with Z_6 due to noise

Model	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
Z_1	1.0000	0.9898	0.9891	0.9902	0.9913	0.0012
Z_2	sym	1.0000	0.9993	0.9983	0.9974	0.1182
Z_3	sym	sym	1.0000	0.9997	0.9991	0.1374
Z_4	sym	sym	sym	1.0000	0.9998	0.1374
Z_5	sym	sym	sym	sym	1.0000	0.1319
Z_6	sym	sym	sym	sym	sym	1.0000

Monomial Example (3/3)





Outline

- Linear models, best linear unbiased estimators (BLUEs)
- Variance bound, monomial example
- **Sample allocation, PDE example**
- Asymptotic analysis of multilevel BLUEs

Sample Allocation

- Motivation: For fixed m_k the estimator $\widehat{\mu}_L^B$ is variance minimal. $m_k =?$
- Define cost $w_\ell := \text{Cost}(Z_\ell)$, $W^k := \text{Cost}(Z^k) = \sum_{\ell \in S^k} w_\ell$
- Fixed budget $p > 0$, exclude cost of estimating C
- Compute a minimizer m^* of the **Sample Allocation Problem**

$$\min_{m \in \mathbb{N}_0^K} \text{Var}(\widehat{\mu}_L^B) = e_L^T \Psi(m)^{-1} e_L$$

$$\text{s. t.} \sum_{k=1}^{K} m_k W^k \leq p.$$

- Define **S**ample **A**llocation **O**ptimal **B**LUE

$$\widehat{\mu}_L^{SAOB} := e_L^T \Psi(m^*)^{-1} y(m^*)$$

Properties of SAOB

- The SAOB has the **smallest variance** among all linear unbiased estimators that only use samples of $Z_1, \dots, Z_L$ and have costs bounded by p.
- The SAOB estimator has the **smallest cost** among all linear (unbiased) estimators that only use samples of $Z_1, \dots, Z_L$.
- There exists a solution m^* of the sample allocation problem using **at most L model groups**.

SAOB (1/2)

Theorem

Let C be positive definite, let $\widehat{\mu}_L$ be a linear unbiased estimator using only samples of $Z_1, \ldots, Z_L$ with cost bounded by $\mathrm{Cost}(\widehat{\mu}_L) \leq p$. Then

$$\mathrm{Var}(\widehat{\mu}_L) \geq \mathrm{Var}(\widehat{\mu}_L^{SAOB}).$$

Proof.

- Use Gauss-Markov-Aitken to show $\mathrm{Var}(\widehat{\mu}_L) \geq \mathrm{Var}(\widehat{\mu}_L^B(m))$ with the sample allocation of $m = m(\widehat{\mu}_L)$
- Since $\mathrm{Cost}(\widehat{\mu}_L) = \sum_{k=1}^K m_k W^k \leq p$, $m(\widehat{\mu}_L)$ is a feasible sample allocation. m^* is the minimizer, thus $\mathrm{Var}(\widehat{\mu}_L^B(m)) \geq \mathrm{Var}(\widehat{\mu}_L^B(m^*)) = \mathrm{Var}(\widehat{\mu}_L^{SAOB})$.

□

SAOB (2/2)

Corollary

Let C be positive definite, let $\widehat{\mu}_L$ be a linear unbiased estimator using only samples of $Z_1, \ldots, Z_L$ such that

$$\mathbb{E}[|\widehat{\mu}_L - \mathbb{E}[Z]|^2] \leq \varepsilon^2 \quad \text{with } \operatorname{Cost}(\widehat{\mu}_L) \leq \phi(\varepsilon^2).$$

Then the SAOB estimator satisifes

$$\mathbb{E}[|\widehat{\mu}_L^{SAOB} - \mathbb{E}[Z]|^2] \leq \varepsilon^2 \quad \text{with } \operatorname{Cost}(\widehat{\mu}_L^{SAOB}) \leq \phi(\varepsilon^2).$$

Proof.

- Bias-variance decomposition $\mathbb{E}[|\widehat{\mu}_L - \mathbb{E}[Z]|^2] = |\mathbb{E}[\mu_L] - \mathbb{E}[Z]|^2 + \operatorname{Var}(\widehat{\mu}_L)$
- SAOB has same bias $\mathbb{E}[\widehat{\mu}_L^{SAOB}] = \mathbb{E}[\widehat{\mu}_L] = \mu_L$ but smaller or equal variance since budget $p = \operatorname{Cost}(\widehat{\mu}_L)$.

PDE Example (1/3)

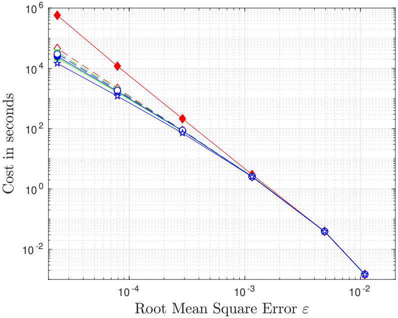
- $Z(\omega) = \frac{1}{|D_{obs}|} \int_{D_{obs}} y(x,\omega)dx$
- $D_{obs} := \left(\frac{3}{4}, \frac{7}{8}\right) \times \left(\frac{7}{8}, 1\right) \subseteq D := (0,1)^2$
- y solves elliptic PDE

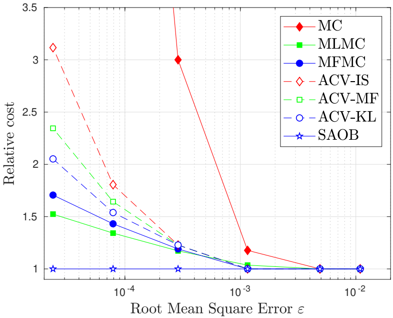
$$\begin{aligned} -\operatorname{div}(a(x,\omega)\nabla y(x,\omega)) &= 1, \quad x \in D, \\ y(x,\omega) &= 0, \quad x \in \partial D. \end{aligned}$$

- Lognormal diffusion coefficient $a(x,\omega) = \exp(\kappa(x,\omega))$
- κ mean zero Gaussian random field with Whittle–Matérn covariance function, smoothness $\nu = 3/2$, variance $\sigma^2 = 1$, correlation length $\rho = 0.1$.

PDE Example (2/3)

- Discretisations $Z_1, \ldots, Z_6$ with linear FEs, uniform mesh refinement
- Computed sample covariance $\approx C$ using 10^5 pilot samples
- Goal: Minimize root mean square error $\varepsilon = (\mathbb{E}[(\widehat{\mu}_\ell - \mathbb{E}[Z])^2])^{1/2}$





PDE Example (3/3)

- $\widehat{\mu}_L^{SAOB} = \sum_{k=1}^{K} \sum_{\ell \in S^k} \beta_\ell^k \frac{1}{m_k} \sum_{i=1}^{m_k} Z_\ell(\omega_i^k)$
- Coefficients for $\varepsilon = 2.5 \times 10^{-5}$

MC

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5						
Z_4						
Z_3						
Z_2						
Z_1						

MLMC

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5					1.00	-1.00
Z_4				1.00	-1.00	
Z_3			1.00	-1.00		
Z_2		1.00	-1.00			
Z_1	1.00	-1.00				

SAOB

	S^1	S^2	S^3	S^4	S^5	S^6
Z_6						1.00
Z_5					1.78	-1.78
Z_4				1.81	-2.79	0.98
Z_3			1.64	-2.66	1.24	-0.22
Z_2	1.18	0.03	-1.93	0.95	-0.26	0.02
Z_1		-0.05			0.04	0.00

Summary (1/2)

- Reformulated the estimation of $\mathbb{E}[Z]$ as linear regression problem
- Constructed novel multilevel BLUE $\longrightarrow$ minimal variance independent of number of samples
- Noisy monomial example shows high correlation with high fidelity model is not necessary for variance reduction
- Additional budget constraint leads to SAOB
- -50% comput. cost compared to MLMC for a scalar valued Quantity of Interest derived from an elliptic PDE with random diffusion coefficient

Outline

- Linear models, best linear unbiased estimators (BLUEs)
- Variance bound, monomial example
- Sample allocation, PDE example
- **Asymptotic analysis of multilevel BLUEs** $\longrightarrow$ **New preprint!**

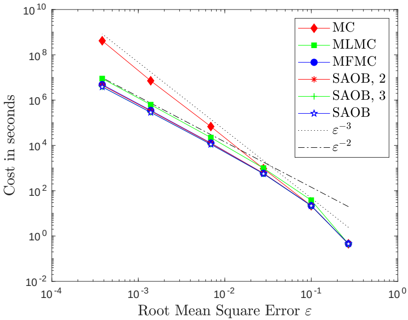
Parametric model family

- Discretization parameter $h_\ell = s^{-\ell+1}h$, $h > 0$, $s > 1$ fixed, $\ell = 1, \dots, L$
- $Z_\ell(\omega) \longleftrightarrow Z_\omega(h_\ell)$
- **Assumption:** Asymptotic expansion

$$Z_\omega(h) = Z_\omega(0) + \begin{cases} c_2(\omega) r^1(h), & q = 1, \\ \sum_{k=2}^{q} c_k(\omega) h^{\beta_k} + c_{q+1} r^q(h), & q \geq 2. \end{cases}$$

- Since the BLUE is a linear combination of model evaluations $Z_\omega(h_\ell)$, this assumption allows us to study the variance (and complexity) of the BLUE!

PDE Example – revisited



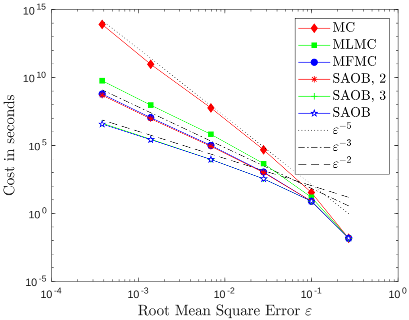


Figure: PDE example ($\rho = 0.5$, $\sigma^2 = 3$) with true cost rate $\beta^{\text{Cost}} = 2$ (left image) and manufactured cost rate $\beta^{\text{Cost}} = 6$ (right image). We can prove the complexity bound for the SAOB for the true cost rate $\beta^{\text{Cost}} = 2$, but not yet the complexity of the SAOB for the higher (manufactured) cost rate.

Summary (2/2)

- Complexity analysis of SAOB for PDE-based models with the help of an asymptotic expansion of the model family together with Richardson extrapolation
- Complexity of SAOB **provably** not worse than complexity of MLMC
- Manufactured example with optimal (and better than MLMC) complexity $O(\varepsilon^{-2})$ $\rightarrow$ Proof in progress

References

- A.A. Gorodetsky, G. Geraci, M. Eldred, J.D. Jakeman: *A Generalized Framework for Approximate Control Variates*, arXiv:1811.04988.
- D. Schaden, E. Ullmann: *On multilevel best linear unbiased estimators*, under review at JUQ, Preprint No. IGDK-2019-11, `https://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/1905_SchadenUllmann.pdf`
- D. Schaden, E. Ullmann: *Asymptotic analysis of multilevel best linear unbiased estimators*, Preprint No. IGDK-2019-22, `https://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/1911_saob_asymptotic_siam.pdf`, submitted.