SAND2022-4364C

Scientific Computing
Department of Informatics
Technical University of Munich

Multilevel Monte Carlo estimators for derivative-free optimization under uncertainty

Friedrich Menhorn[1], Gianluca Geraci[2], Daniel T. Seidl [2], Youssef M. Marzouk[3], Michael S. Eldred[2], Hans-Joachim Bungartz[1]

menhorn@in.tum.de

[1]Technical University of Munich, GER

[2]Sandia National Laboratories, USA

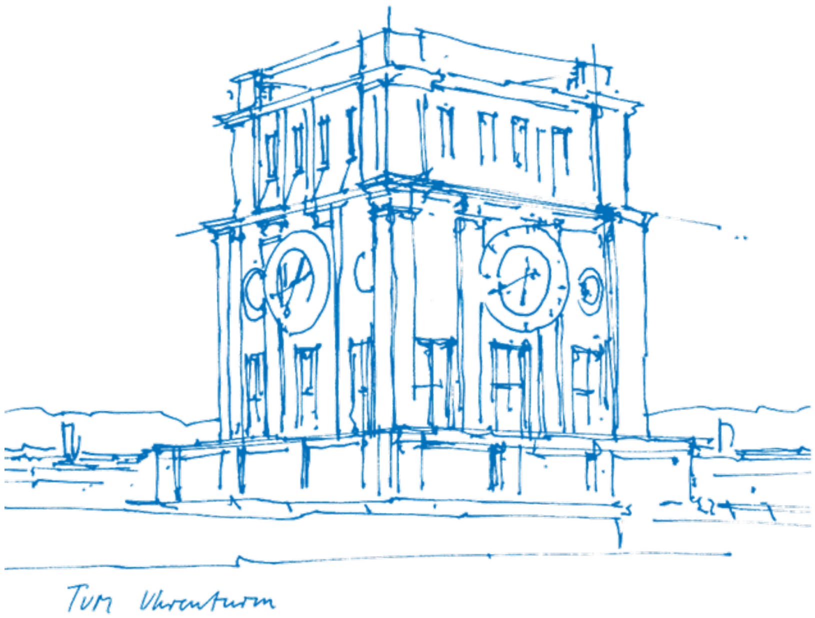
[3]Massachusetts Institute of Technology, USA

SIAM UQ 22, April 12-15, 2022

MS82

PDE-Constrained Optimization Under Uncertainty

April 13th, 2022



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^N Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^N (Q_i - \widehat{\mu_1})^2}}$$

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^N Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^N (Q_i - \widehat{\mu_1})^2}}$$

BUT: Monte Carlo is very expensive ($\mathscr{O}(\frac{1}{\sqrt{N}})$)

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^{N} Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha\ \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Q_i-\widehat{\mu_1})^2}}$$

BUT: Monte Carlo is very expensive ($\mathscr{O}(\frac{1}{\sqrt{N}})$)

$\Rightarrow$ Use MLMC estimators† if hierarchy available

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^{N} Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Q_i-\widehat{\mu_1})^2}}$$

BUT: Monte Carlo is very expensive ($\mathscr{O}(\frac{1}{\sqrt{N}})$)

$\Rightarrow$ Use MLMC estimators† if hierarchy available

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^{N} Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Q_i-\widehat{\mu_1})^2}}$$

BUT: Monte Carlo is very expensive ($\mathscr{O}(\frac{1}{\sqrt{N}})$)

$\Rightarrow$ Use MLMC estimators† if hierarchy available

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

BUT: Original MLMC estimator not sufficient for $\mathscr{R}_{\text{Scalar}}$ formulation

*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Motivation

Robustness in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} = \min_x \mathbb{E}[Q(x,\theta)]$$

Reliability in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} = \min_x \mathbb{E}[Q(x,\theta)] + \alpha\sigma[Q(x,\theta)]$$

Solve using derivative-free optimization method SNOWPAC*

$\Rightarrow$ SNOWPAC requires sampling estimator with samples $Q_i := Q(x,\theta_i)$:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \underbrace{\widehat{\mu_1}[Q(x,\theta)]}_{\frac{1}{N}\sum_{i=1}^N Q_i}$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1}[Q(x,\theta)] + \alpha \underbrace{\widehat{\sigma_{\text{biased}}}[Q(x,\theta)]}_{\sqrt{\frac{1}{N-1}\sum_{i=1}^N (Q_i - \widehat{\mu_1})^2}}$$

BUT: Monte Carlo is very expensive ($\mathscr{O}(\frac{1}{\sqrt{N}})$)

$\Rightarrow$ Use MLMC estimators† if hierarchy available

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

BUT: Original MLMC estimator not sufficient for $\mathscr{R}_{\text{Scalar}}$ formulation

$\Rightarrow$ Need MLMC estimators for higher order moments

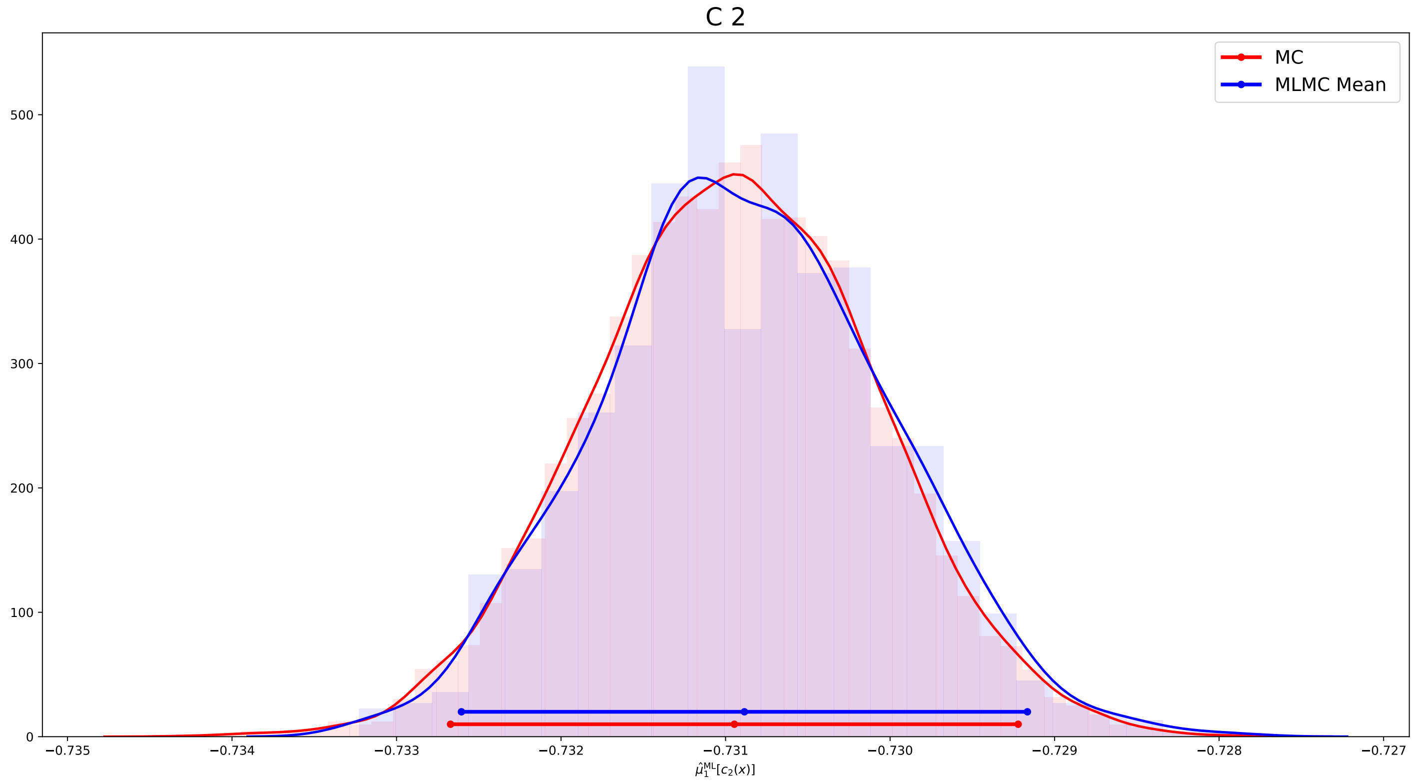
*Menhorn, F. et al. (2022). A trust-region method for derivative-free nonlinear constrained stochastic optimization. (arXiv)

†Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Example

MLMC mean estimator for $\mathscr{R}_{\text{Mean}} \approx \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$

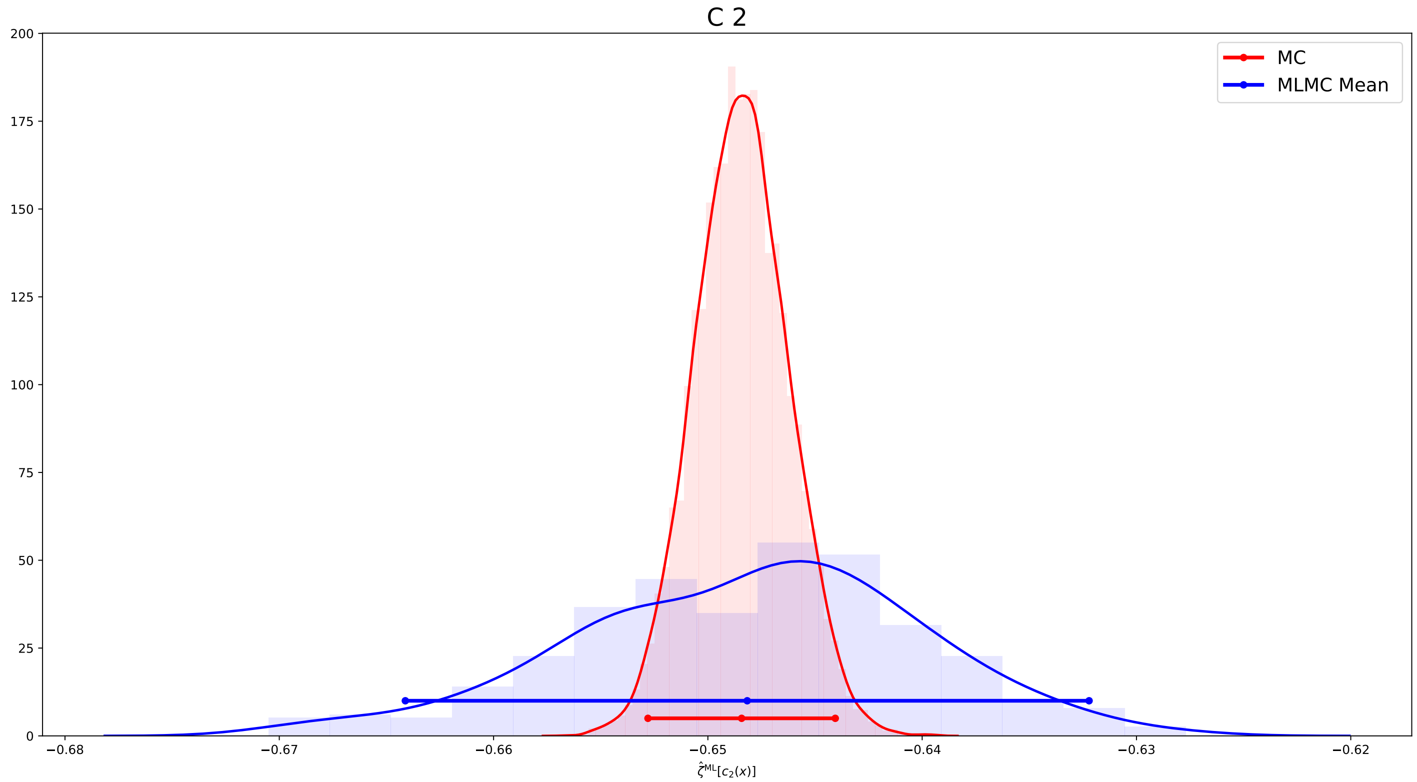


- Good match between standard MC estimator and MLMC.

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Example

MLMC mean estimator for $\mathscr{R}_{\text{Scalar}} \approx \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$

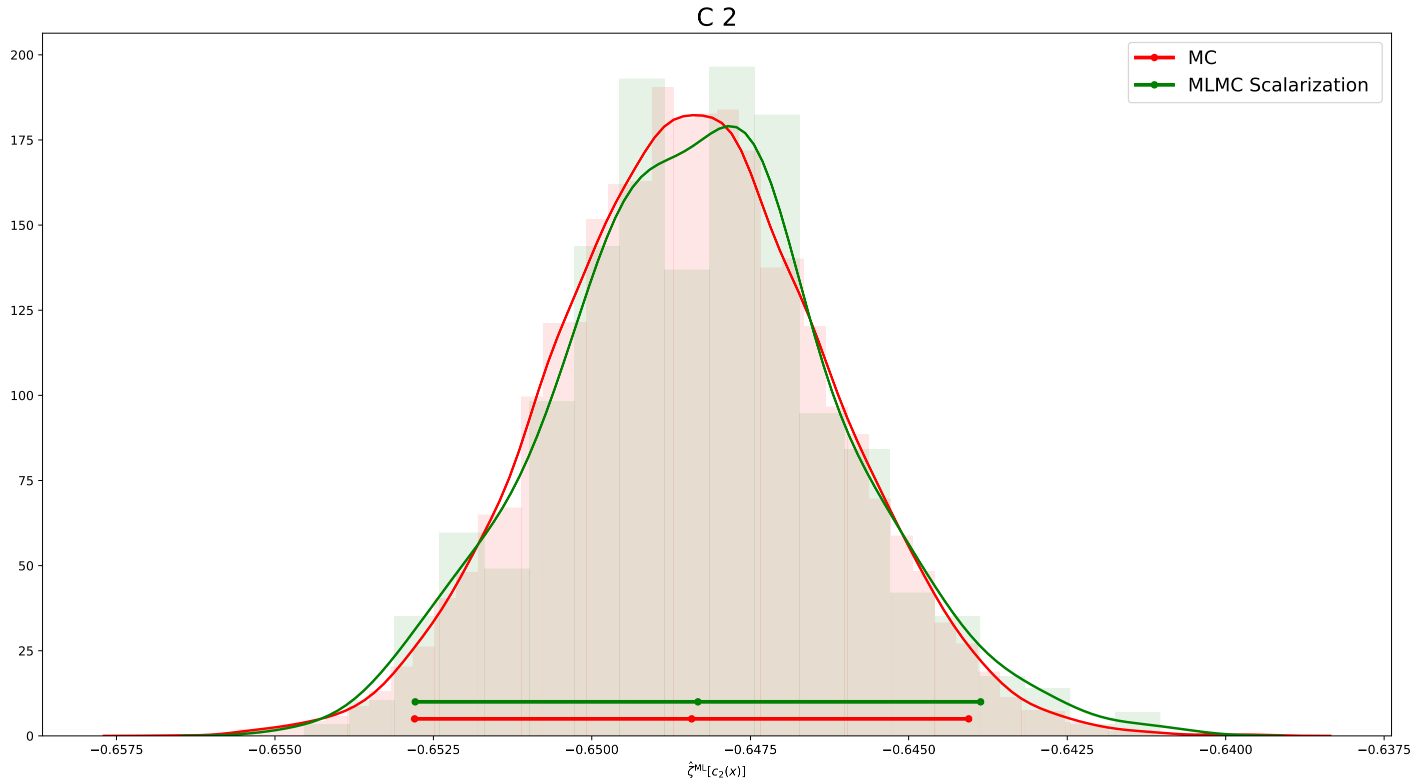


- MLMC estimator targeting the mean underestimates compared to MC.

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Goal

MLMC scalarization estimator for $\mathscr{R}_{\text{Scalar}} \approx \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$



- New MLMC estimator targeting the scalarization matches MC.

Up next:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

Up next:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

Up next:

Up next:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \boxed{\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]}$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

Up next:

1. Recap (from literature): $\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$

Up next:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \boxed{\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]}$$

Up next:

1. Recap (from literature): $\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$
2. (Our contribution) ML for Standard deviation: $\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$

Up next:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \boxed{\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]}$$

Up next:

1. Recap (from literature): $\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$
2. (Our contribution) ML for Standard deviation: $\widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$
3. (Our contribution) ML for Scalarization: $\widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$

Multilevel Monte Carlo (MLMC) Estimator

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Mean

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

‡Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Mean

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

- Estimator[‡]:

$$\mathbb{E}[Q_L] = \mu_1^{\text{ML}}[Q_L] \approx \widehat{\mu_1^{\text{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_1}[Q^{(\ell)} - Q^{(\ell-1)}]}_{\text{telescopic sum}} = \sum_{\ell=0}^{L} \underbrace{\frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - Q_{i,\ell}^{(\ell-1)})}_{\text{estimator expansion}}, \quad Q_{i,0}^{(-1)} := 0$$

[‡]Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

MLMC estimator: Mean

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

- Estimator[‡]:

$$\mathbb{E}[Q_L] = \mu_1^{\text{ML}}[Q_L] \approx \widehat{\mu_1^{\text{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_1}[Q^{(\ell)} - Q^{(\ell-1)}]}_{\text{telescopic sum}} = \underbrace{\sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - Q_{i,\ell}^{(\ell-1)})}_{\text{estimator expansion}}, \quad Q_{i,0}^{(-1)} := 0$$

- Sample allocation:

$$\min_{N_\ell^{\mathbb{E}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{E}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \sum_{\ell=0}^{L} \mathbb{V}[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Q^{(\ell)} - Q^{(\ell-1)}]}{N_\ell}$$

[‡]Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

MLMC estimator: Mean

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

- Estimator[‡]:

$$\mathbb{E}[Q_L] = \mu_1^{\text{ML}}[Q_L] \approx \widehat{\mu_1^{\text{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_1}[Q^{(\ell)} - Q^{(\ell-1)}]}_{\text{telescopic sum}} = \underbrace{\sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - Q_{i,\ell}^{(\ell-1)}),}_{\text{estimator expansion}} \quad Q_{i,0}^{(-1)} := 0$$

- Sample allocation:

$$\min_{N_\ell^{\mathbb{E}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{E}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \sum_{\ell=0}^{L} \mathbb{V}[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Q^{(\ell)} - Q^{(\ell-1)}]}{N_\ell}$$

- Solution:

$$N_\ell^{\mathbb{E}} = \left\lceil \lambda \sqrt{\frac{\mathbb{V}[Q_\ell - Q_{\ell-1}]}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathbb{V}[Q_\ell - Q_{\ell-1}] C_\ell}$$

[‡]Giles, M.B., "Multilevel Monte Carlo methods," Acta Numerica, Vol.24, 2015, p.259–328

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Standard deviation in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

$$\sigma[Q_L] = \sqrt{\mathbb{V}[Q_L]} \approx \sqrt{\widehat{\mu_2^{\mathrm{ML}}}} := \widehat{\sigma^{\mathrm{ML}}_{\mathrm{biased}}}$$

where

$$\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\mathrm{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}]}_{\text{telescopic sum}}$$

$$= \sum_{\ell=0}^{L} \underbrace{\frac{1}{N_\ell - 1}\left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2\right)}_{\text{estimator expansion}}$$

$$= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

$$\sigma[Q_L] = \sqrt{\mathbb{V}[Q_L]} \approx \sqrt{\widehat{\mu_2^{\text{ML}}}} := \widehat{\sigma^{\text{ML}}_{\text{biased}}}$$

where

$$\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\text{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}]}_{\text{telescopic sum}}$$

$$= \sum_{\ell=0}^{L} \underbrace{\frac{1}{N_\ell - 1}\left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2\right)}_{\text{estimator expansion}}$$

$$= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0$$

- Sample allocation:

$$\min_{N_\ell^\sigma} \sum_{\ell=0}^{L} C_\ell N_\ell^\sigma,$$

$$\text{s.t. } \mathbb{V}[\widehat{\sigma^{\text{ML}}_{\text{biased}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\sigma^{\text{ML}}_{\text{biased}}}] \approx \frac{1}{4}\frac{\mathbb{V}[\widehat{\mu_2^{\text{ML}}}]}{\widehat{\mu_2^{\text{ML}}}} \text{ (Delta Method)}$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

$$\sigma[Q_L] = \sqrt{\mathbb{V}[Q_L]} \approx \sqrt{\widehat{\mu_2^{\mathrm{ML}}}} := \widehat{\sigma^{\mathrm{ML}}_{\mathrm{biased}}}$$

where

$$\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\mathrm{ML}}}[Q_L] = \sum_{\ell=0}^{L} \underbrace{\widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}]}_{\text{telescopic sum}}$$

$$= \sum_{\ell=0}^{L} \underbrace{\frac{1}{N_\ell - 1}\left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2\right)}_{\text{estimator expansion}}$$

$$= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0$$

- Sample allocation:

$$\min_{N_\ell^\sigma} \sum_{\ell=0}^{L} C_\ell N_\ell^\sigma,$$

$$\text{s.t. } \mathbb{V}[\widehat{\sigma^{\mathrm{ML}}_{\mathrm{biased}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\sigma^{\mathrm{ML}}_{\mathrm{biased}}}] \approx \frac{1}{4}\frac{\mathbb{V}[\widehat{\mu_2^{\mathrm{ML}}}]}{\widehat{\mu_2^{\mathrm{ML}}}} \text{ (Delta Method)}$$

- How to solve this?

MLMC sample allocation for higher order terms

1. Analytic approximation [§]
   - **Obs**: Decrease of variance of MLMC estimator $\mathcal{O}(\frac{1}{N})$ true also for higher order moments
   - **Idea**: Use similar analytic solution approach as for MLMC Mean:

$$\min_{N_\ell^{\mathbb{E}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{E}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Q^{(\ell)} - Q^{(\ell-1)}]}{N_\ell}$$

$$N_\ell^{\mathbb{E}} = \left\lceil \lambda \sqrt{\frac{\mathbb{V}[Q_\ell - Q_{\ell-1}]}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathbb{V}[Q_\ell - Q_{\ell-1}] C_\ell}$$

[§]Pisaroni, M., Krumscheid, S., and Nobile, F., "MATHICSE Technical Report : Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part I: Central moment estimation," 2017, p. 29.

MLMC sample allocation for higher order terms

1. Analytic approximation [§]
   - **Obs**: Decrease of variance of MLMC estimator $\mathcal{O}(\frac{1}{N})$ true also for higher order moments
   - **Idea**: Use similar analytic solution approach as for MLMC Mean:

$$\min_{N_\ell^{\mathbb{E}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{E}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] = \sum_{\ell=0}^{L} \frac{\mathbb{V}[Q^{(\ell)} - Q^{(\ell-1)}]}{N_\ell}$$

$$N_\ell^{\mathbb{E}} = \left\lceil \lambda \sqrt{\frac{\mathbb{V}[Q_\ell - Q_{\ell-1}]}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathbb{V}[Q_\ell - Q_{\ell-1}] C_\ell}$$

   - Introduce helper variance: $\mathbb{V}[\widehat{\mu_1^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell} \Leftrightarrow \mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_1^{(\ell)}}] N_\ell^{\mathbb{E}}$

[§] Pisaroni, M., Krumscheid, S., and Nobile, F., "MATHICSE Technical Report : Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part I: Central moment estimation," 2017, p. 29.

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC sample allocation for higher order terms

1. Analytic approximation [§]
   - **Obs**: Decrease of variance of MLMC estimator $\mathcal{O}(\frac{1}{N})$ true also for higher order moments
   - **Idea**: Use similar analytic solution approach as for MLMC Mean:

$$\min_{N_\ell^{\mathbb{X}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{X}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_x^{\mathrm{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_x^{\mathrm{ML}}}] = \sum_{\ell=0}^{L} \frac{\mathscr{V}_\ell}{N_\ell}$$

$$N_\ell^{\mathbb{X}} = \left\lceil \lambda \sqrt{\frac{\mathscr{V}_\ell}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathscr{V}_\ell C_\ell}$$

   - Introduce helper variance: $\mathbb{V}[\widehat{\mu_x^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell} \Leftrightarrow \mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_x^{(\ell)}}] N_\ell^{\mathbb{X}}$

[§] Pisaroni, M., Krumscheid, S., and Nobile, F., “MATHICSE Technical Report : Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part I: Central moment estimation,” 2017, p. 29.

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC sample allocation for higher order terms

1. Analytic approximation [§]
   - **Obs**: Decrease of variance of MLMC estimator $\mathcal{O}(\frac{1}{N})$ true also for higher order moments
   - **Idea**: Use similar analytic solution approach as for MLMC Mean:

$$\min_{N_\ell^{\mathbb{X}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{X}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_x^{\mathrm{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_x^{\mathrm{ML}}}] = \sum_{\ell=0}^{L} \frac{\mathscr{V}_\ell}{N_\ell}$$

$$N_\ell^{\mathbb{X}} = \left\lceil \lambda \sqrt{\frac{\mathscr{V}_\ell}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathscr{V}_\ell C_\ell}$$

   - Introduce helper variance: $\mathbb{V}[\widehat{\mu_x^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell} \Leftrightarrow \mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_x^{(\ell)}}] N_\ell^{\mathbb{X}}$
   - Disregard higher order terms in $\mathscr{V}_\ell \Rightarrow$ Analytic approximation

[§] Pisaroni, M., Krumscheid, S., and Nobile, F., "MATHICSE Technical Report : Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part I: Central moment estimation," 2017, p. 29.

MLMC sample allocation for higher order terms

1. Analytic approximation [§]
   - **Obs**: Decrease of variance of MLMC estimator $\mathcal{O}(\frac{1}{N})$ true also for higher order moments
   - **Idea**: Use similar analytic solution approach as for MLMC Mean:

$$\min_{N_\ell^{\mathbb{X}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{X}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\mu_X^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\mu_X^{\text{ML}}}] = \sum_{\ell=0}^{L} \frac{\mathscr{V}_\ell}{N_\ell}$$

$$N_\ell^{\mathbb{X}} = \left\lceil \lambda \sqrt{\frac{\mathscr{V}_\ell}{C_\ell}} \right\rceil, \text{ where } \lambda = \varepsilon^{-2} \sum_{\ell=0}^{L} \sqrt{\mathscr{V}_\ell C_\ell}$$

   - Introduce helper variance: $\mathbb{V}[\widehat{\mu_X^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell} \Leftrightarrow \mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_X^{(\ell)}}] N_\ell^{\mathbb{X}}$
   - Disregard higher order terms in $\mathscr{V}_\ell \Rightarrow$ Analytic approximation
2. Numerical optimization
   - Solve $\min_{N_\ell^{\mathbb{X}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{X}}$, s.t. $\mathbb{V}[\widehat{\mu_X^{\text{ML}}}] = \varepsilon^2$ numerically

[§] Pisaroni, M., Krumscheid, S., and Nobile, F., "MATHICSE Technical Report : Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part I: Central moment estimation," 2017, p. 29.

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)]$$

Scalarization in OUU:

$$\min_x \mathscr{R}_{\text{Scalar}} \approx \min_x \widehat{\mu_1^{\text{ML}}}[Q(x,\theta)] + \alpha \widehat{\sigma_{\text{biased}}^{\text{ML}}}[Q(x,\theta)]$$

MLMC estimator: Scalarization

- Estimator:

$$\mathbb{S}[Q_L] := \mathbb{E}[Q_L] + \alpha\sigma[Q_L] \approx \widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}} := \widehat{\zeta}^{\text{ML}}$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization

- Estimator:

$$\mathbb{S}[Q_L] := \mathbb{E}[Q_L] + \alpha\sigma[Q_L] \approx \widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}} := \widehat{\zeta^{\text{ML}}}$$

- Sample allocation:

$$\min_{N_\ell^{\mathbb{S}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{S}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] \approx \mathbb{V}[\widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$

MLMC estimator: Scalarization

- Estimator:

$$\mathbb{S}[Q_L] := \mathbb{E}[Q_L] + \alpha\sigma[Q_L] \approx \widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}} := \widehat{\zeta^{\text{ML}}}$$

- Sample allocation:

$$\min_{N_\ell^{\mathbb{S}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{S}},$$

$$\text{s.t. } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] \approx \mathbb{V}[\widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$

- Variance of scalarization:

$$\begin{aligned}\mathbb{V}[\widehat{\zeta^{\text{ML}}}] &= \mathbb{V}[\widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}] \\ &= \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2\mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]\end{aligned}$$

- Solution: Numerical Optimization or analytic approximation

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization

- Estimator:

$$\mathbb{S}[Q_L] := \mathbb{E}[Q_L] + \alpha\sigma[Q_L] \approx \widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}} := \widehat{\zeta^{\text{ML}}}$$

- Sample allocation:

$$\min_{N_\ell^{\mathbb{S}}} \sum_{\ell=0}^{L} C_\ell N_\ell^{\mathbb{S}},$$
$$\text{s.t. } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \varepsilon^2, \text{ where } \mathbb{V}[\widehat{\zeta^{\text{ML}}}] \approx \mathbb{V}[\widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$

- Variance of scalarization:

$$\begin{aligned}\mathbb{V}[\widehat{\zeta^{\text{ML}}}] &= \mathbb{V}[\widehat{\mu_1^{\text{ML}}} + \alpha\widehat{\sigma_{\text{biased}}^{\text{ML}}}] \\ &= \underbrace{\mathbb{V}[\widehat{\mu_1^{\text{ML}}}]}_{\text{known}} + \alpha^2\underbrace{\mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}]}_{\text{known}} + 2\alpha\underbrace{\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]}_{\text{up next}}\end{aligned}$$

- Solution: Numerical Optimization or analytic approximation

MLMC estimator: Scalarization Covariance

$$\mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$

$$= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]$$

Three (independent) solution strategies:

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization Covariance

$$\mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$
$$= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]$$

Three (independent) solution strategies:

1. Pearson correlation $\rho[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$:

$$\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}] \leq \sqrt{\mathbb{V}[\widehat{\mu_1^{\text{ML}}}] \cdot \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}]}, \text{ since} \qquad -1 \leq \frac{\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]}{\sqrt{\mathbb{V}[\widehat{\mu_1^{\text{ML}}}]\mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}]}} \leq 1$$

MLMC estimator: Scalarization Covariance

$$\mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$
$$= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]$$

Three (independent) solution strategies:

1. Pearson correlation $\rho[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$:

$$\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}] \leq \sqrt{\mathbb{V}[\widehat{\mu_1^{\text{ML}}}] \cdot \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}]}, \text{ since} \qquad -1 \leq \frac{\mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]}{\sqrt{\mathbb{V}[\widehat{\mu_1^{\text{ML}}}]\mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}]}} \leq 1$$

2. Bootstrapping:

$$\mathbb{C}\text{ov}[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}}] \approx \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{\mu_{1,b}^{(\ell)}} - \overline{\widehat{\mu_{1,b}^{(\ell)}}})(\widehat{\sigma_{\text{biased},b}^{(\ell)}} - \overline{\widehat{\sigma_{\text{biased},b}^{(\ell)}}})$$

$$\text{where e.g.} \widehat{\mu_{1,b}^{(\ell)}} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} Q_i^* \text{ and } \overline{\widehat{\mu_{1,b}^{(\ell)}}} = \frac{1}{B} \sum_{b=1}^{B} \widehat{\mu_{1,b}^{(\ell)}} \text{ and } Q_i^* \text{ iid from } \{Q_i\}_{i=1}^{N_\ell}$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization Covariance

$$\mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$
$$= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]$$

Three (independent) solution strategies:

1. Pearson correlation: (easy to compute, **inequality**)
2. Bootstrapping (more expensive, **approximation**)

MLMC estimator: Scalarization Covariance

$$
\begin{aligned}
\mathbb{V}[\widehat{\zeta^{\text{ML}}}] &= \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}] \\
&= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]
\end{aligned}
$$

Three (independent) solution strategies:

1. Pearson correlation: (easy to compute, **inequality**)
2. Bootstrapping (more expensive, **approximation**)
3. Correlation Lift:

$$
\text{Assume: } \rho[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(k)}}] \approx \rho[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(k)}}], (\ell - k) \in \{-1, 0, 1\}
$$

$$
\text{Then: } \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(k)}}\right] \approx \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(k)}}\right] \sqrt{\frac{\mathbb{V}[\widehat{\mu_2^{(k)}}]}{\mathbb{V}[\widehat{\sigma_{\text{biased}}^{(k)}}]}},
$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Scalarization Covariance

$$\mathbb{V}[\widehat{\zeta^{\text{ML}}}] = \mathbb{V}[\widehat{\mu_1^{\text{ML}}}] + \alpha^2 \mathbb{V}[\widehat{\sigma_{\text{biased}}^{\text{ML}}}] + 2\alpha \mathbb{C}\text{ov}[\widehat{\mu_1^{\text{ML}}}, \widehat{\sigma_{\text{biased}}^{\text{ML}}}]$$
$$= \sum_{\ell=0}^{L} \mathbb{V}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}\right] + \alpha^2 \mathbb{V}\left[\widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right] + 2\alpha \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}} - \widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\sigma_{\text{biased}}^{(\ell)}} - \widehat{\sigma_{\text{biased},\ell}^{(\ell-1)}}\right]$$

Three (independent) solution strategies:

1. Pearson correlation: (easy to compute, **inequality**)
2. Bootstrapping (more expensive, **approximation**)
3. Correlation Lift:

$$\text{Assume: } \rho[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(k)}}] \approx \rho[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(k)}}], (\ell - k) \in \{-1, 0, 1\}$$

$$\text{Then: } \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(k)}}\right] \approx \mathbb{C}\text{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(k)}}\right] \sqrt{\frac{\mathbb{V}[\widehat{\mu_2^{(k)}}]}{\mathbb{V}[\widehat{\sigma_{\text{biased}}^{(k)}}]}},$$

MLMC estimator: Scalarization Covariance

- Same level[¶]:

$$\mathbb{C}\mathrm{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(\ell)}}\right] = \frac{\mu_3^{(\ell)}}{N_\ell}$$

- Lower level variance:

$$\mathbb{C}\mathrm{ov}\left[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_{2,\ell}^{(\ell-1)}}\right] =$$

$$\frac{1}{N_\ell}\left[\mathbb{E}[Q^{(\ell)}(Q^{(\ell-1)})^2] - \mathbb{E}[Q^{(\ell)}]\mathbb{E}[(Q^{(\ell-1)})^2] - 2\mathbb{E}[Q^{(\ell-1)}]\mathbb{E}[Q^{(\ell)}Q^{(\ell-1)}] + 2\mathbb{E}[Q^{(\ell)}]\mathbb{E}[Q^{(\ell-1)}]^2\right]$$

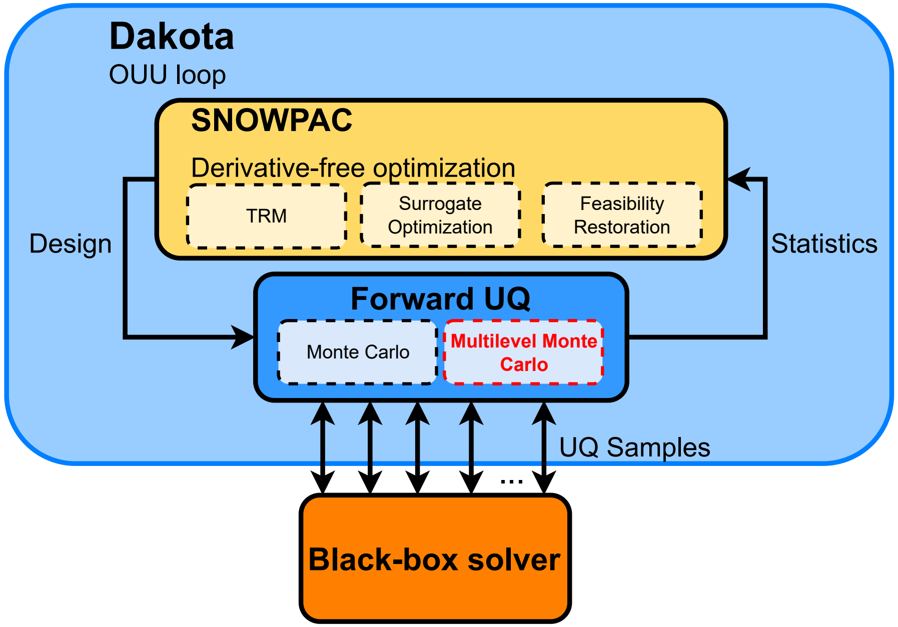
- Lower level mean:

$$\mathbb{C}\mathrm{ov}\left[\widehat{\mu_{1,\ell}^{(\ell-1)}}, \widehat{\mu_2^{(\ell)}}\right] =$$

$$\frac{1}{N_\ell}\left[\mathbb{E}[Q^{(\ell-1)}(Q^{(\ell)})^2] - \mathbb{E}[Q^{(\ell-1)}]\mathbb{E}[(Q^{(\ell)})^2] - 2\mathbb{E}[Q^{(\ell)}]\mathbb{E}[Q^{(\ell-1)}Q^{(\ell)}] + 2\mathbb{E}[Q^{(\ell-1)}]\mathbb{E}[Q^{(\ell)}]^2\right]$$

[¶]Dodge Y., Rousson V., The Complications of the Fourth Central Moment, The American Statistician, 1999

SNOWPAC in Dakota using MLMC



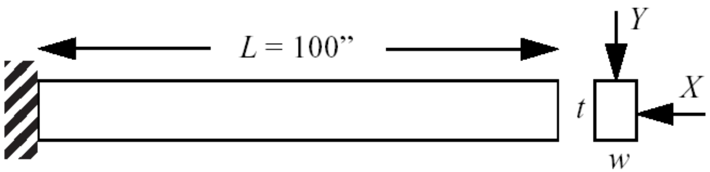
Algorithmic details:

- Iterative sample allocation
- Underrelaxation
- Choice of sample allocation strategy: Analytic approximation vs. numerical optimization
- Choice of covariance approximation: Pearson vs. Bootstrap vs. Correlation Lift

Results

Cantilever beam problem:

2 **design** parameters: w, t



Adams, B.M., et. al., Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User's Manual

Objective: $\min_{w,t} wt$

4 **uncertain** parameters:

- Yield stress: $R \sim \mathcal{N}(40000, 200)$
- Young's modulus: $E \sim \mathcal{N}(2.9e7, 1.45e6)$
- Horizontal load: $X \sim \mathcal{N}(500, 100)$
- Vertical load: $Y \sim \mathcal{N}(1000, 100)$

- Level 3: Rectangle (Cost = 1)

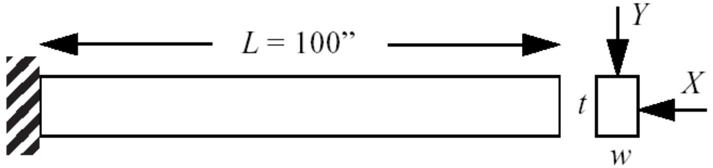
$$\text{stress} : (\frac{600}{wt^2}Y + \frac{600}{w^2 t}X)/R - 1 \leq 0$$

$$\text{displacement} : (\frac{4L^3}{Ewt}\sqrt{(\frac{Y}{t^2})^2 + (\frac{X}{w^2})^2})/D_0 - 1 \leq 0$$

- Level 2: Ellipse instead of rectangle (Cost = 1e-2)
- Level 1: Circle inscribed in rectangle instead of rectangle (Cost = 1e-4)
- Level 0: Circle with same area instead of rectangle (Cost = 1e-6)

Cantilever beam problem:

2 **design** parameters: w, t



Adams, B.M., et. al., Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User's Manual

Objective: ~~$\min_{w,t}$~~ w t (Sampling only)

4 **uncertain** parameters:

- Yield stress: $R \sim \mathcal{N}(40000, 200)$
- Young's modulus: $E \sim \mathcal{N}(2.9e7, 1.45e6)$
- Horizontal load: $X \sim \mathcal{N}(500, 100)$
- Vertical load: $Y \sim \mathcal{N}(1000, 100)$

- Level 3: Rectangle (Cost = 1)

$$\text{stress}: \left(\frac{600}{wt^2}Y + \frac{600}{w^2 t}X\right)/R - 1 \leq 0$$

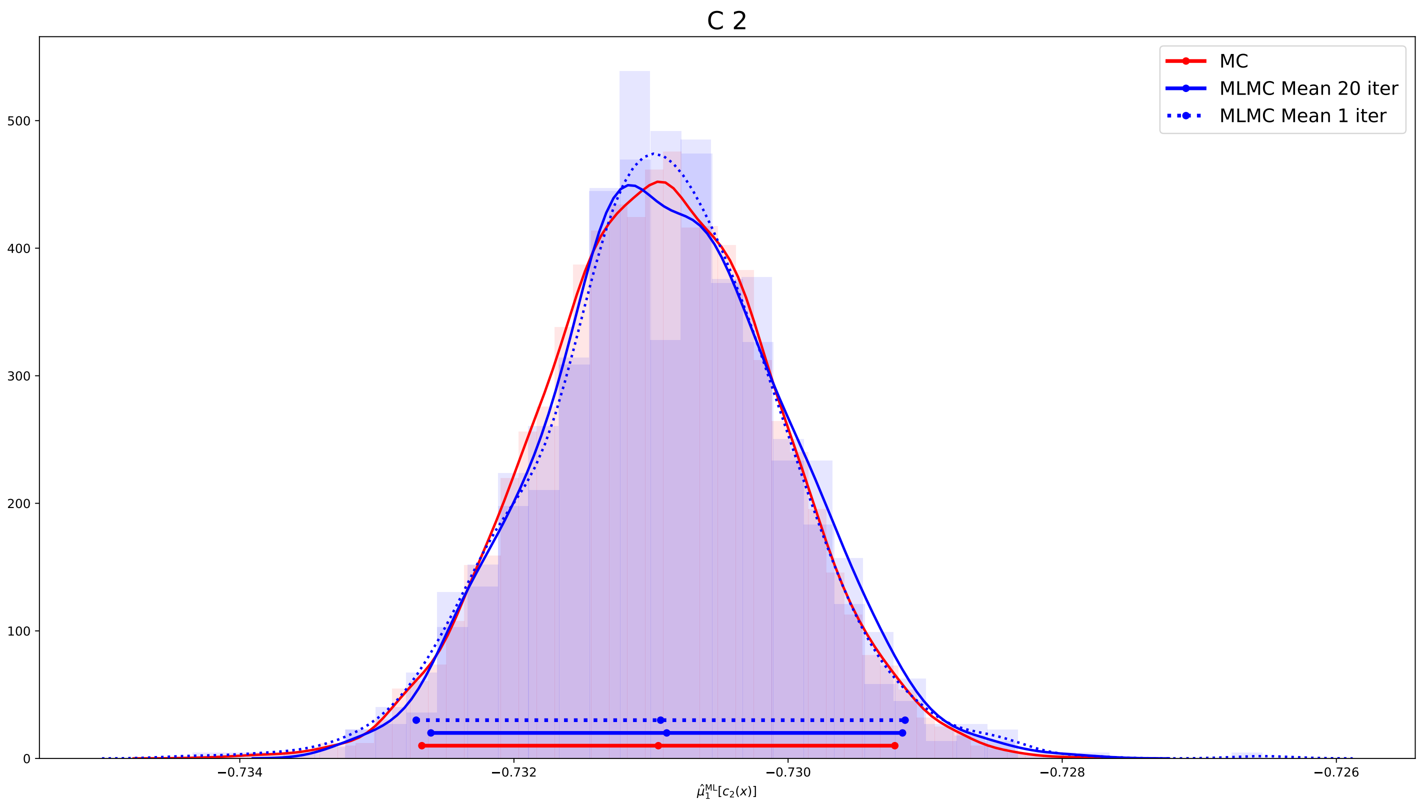
$$\text{displacement}: \left(\frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2}\right)/D_0 - 1 \leq 0$$

- Level 2: Ellipse instead of rectangle (Cost = 1e-2)
- Level 1: Circle inscribed in rectangle instead of rectangle (Cost = 1e-4)
- Level 0: Circle with same area instead of rectangle (Cost = 1e-6)

Test results

1. Present sampling results by resampling estimators and plot histogram of **estimators**
2. Compare three strategies for estimating $\mathbb{C}$ov-term:
   - Pearson
   - Bootstrap
   - using $\rho[\widehat{\mu_1^{(\ell)}}, \widehat{\sigma_{\text{biased}}^{(k)}}] = \rho[\widehat{\mu_1^{(\ell)}}, \widehat{\mu_2^{(k)}}]$ (CorrLift)
3. Compare algorithmic strategies:
   - Iterations vs. no iterations
   - Analytics approximation (only) vs. numerical optimization

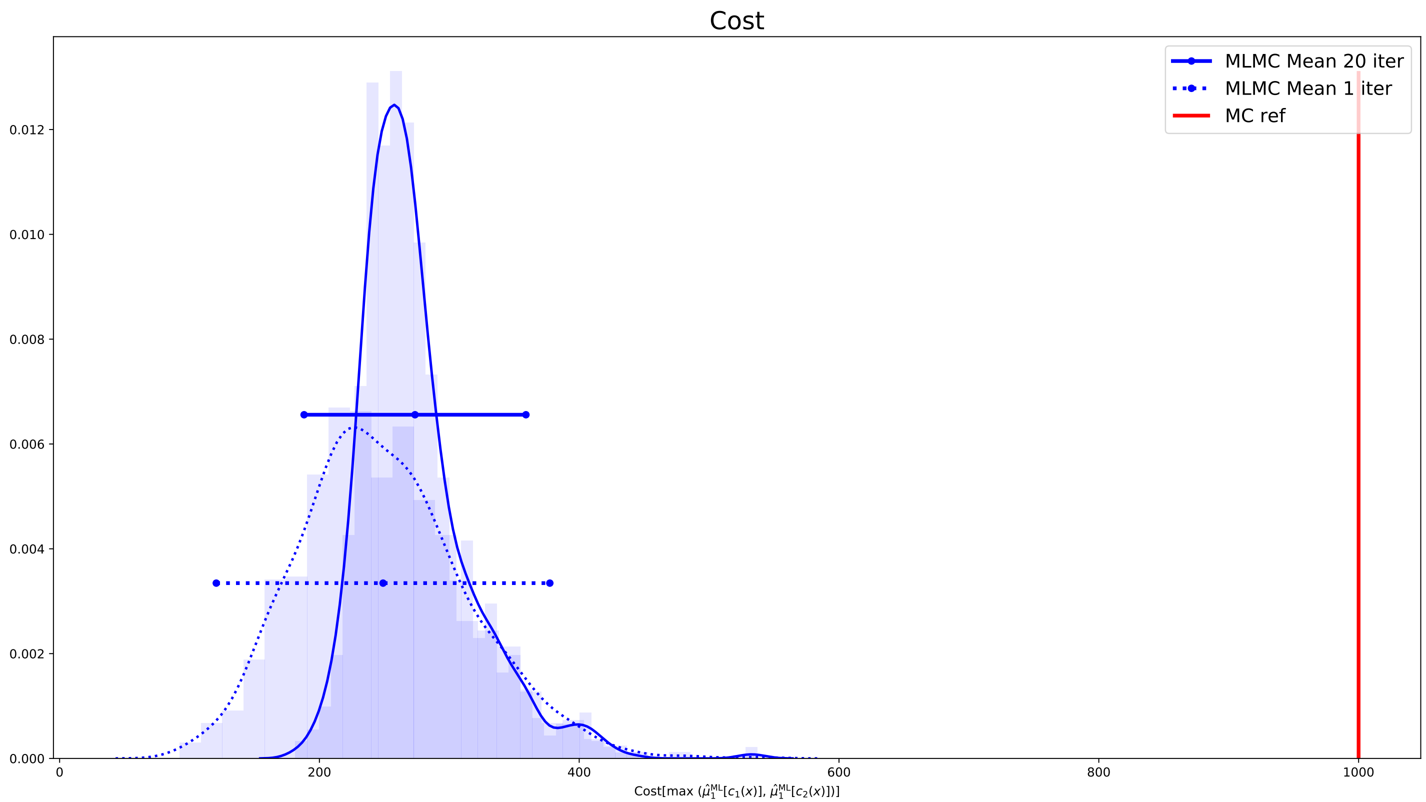
Sampling Results: Mean



- MLMC Mean distribution (blue) consistent with MC reference (red) for lower cost.

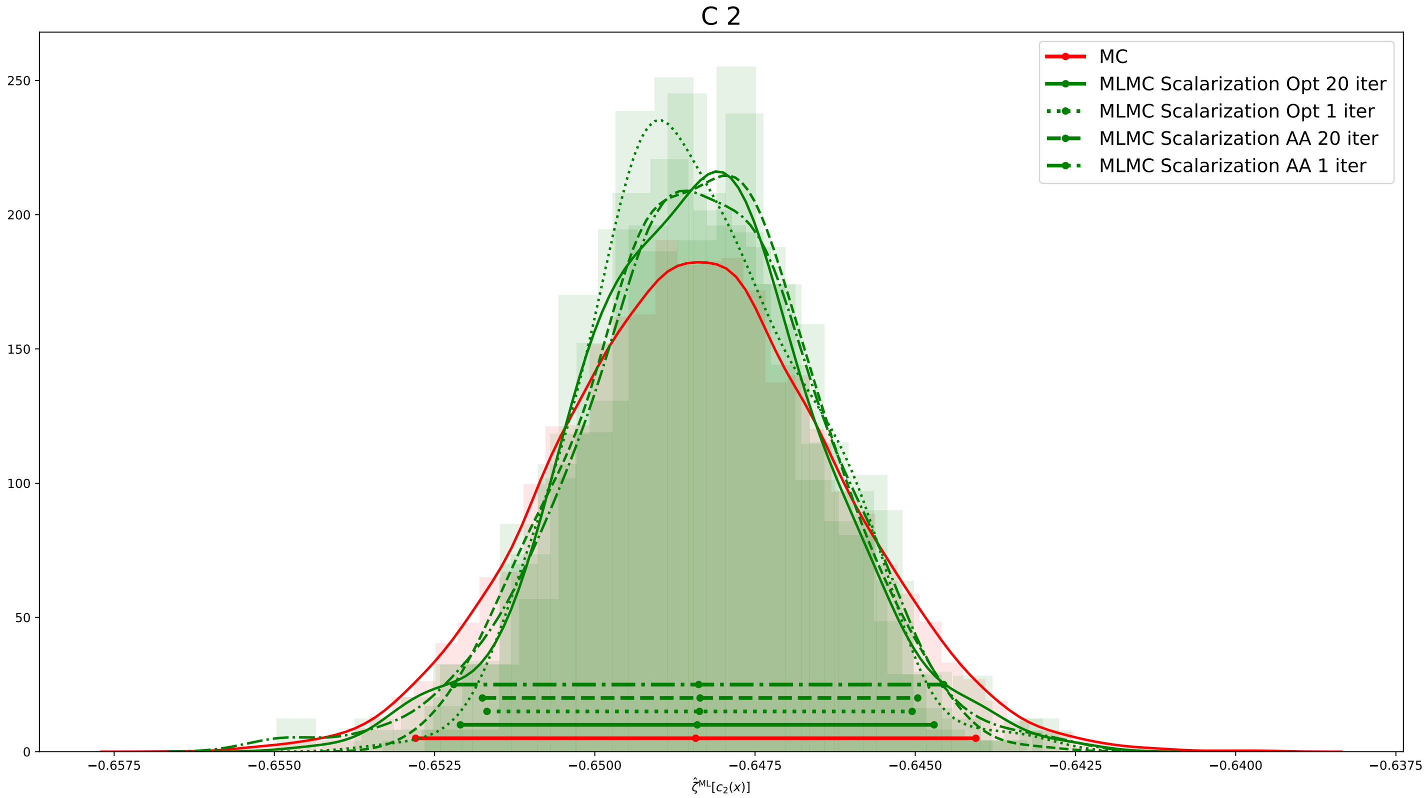
Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results: Mean



- MLMC Mean distribution (blue) consistent with MC reference (red) for lower cost.

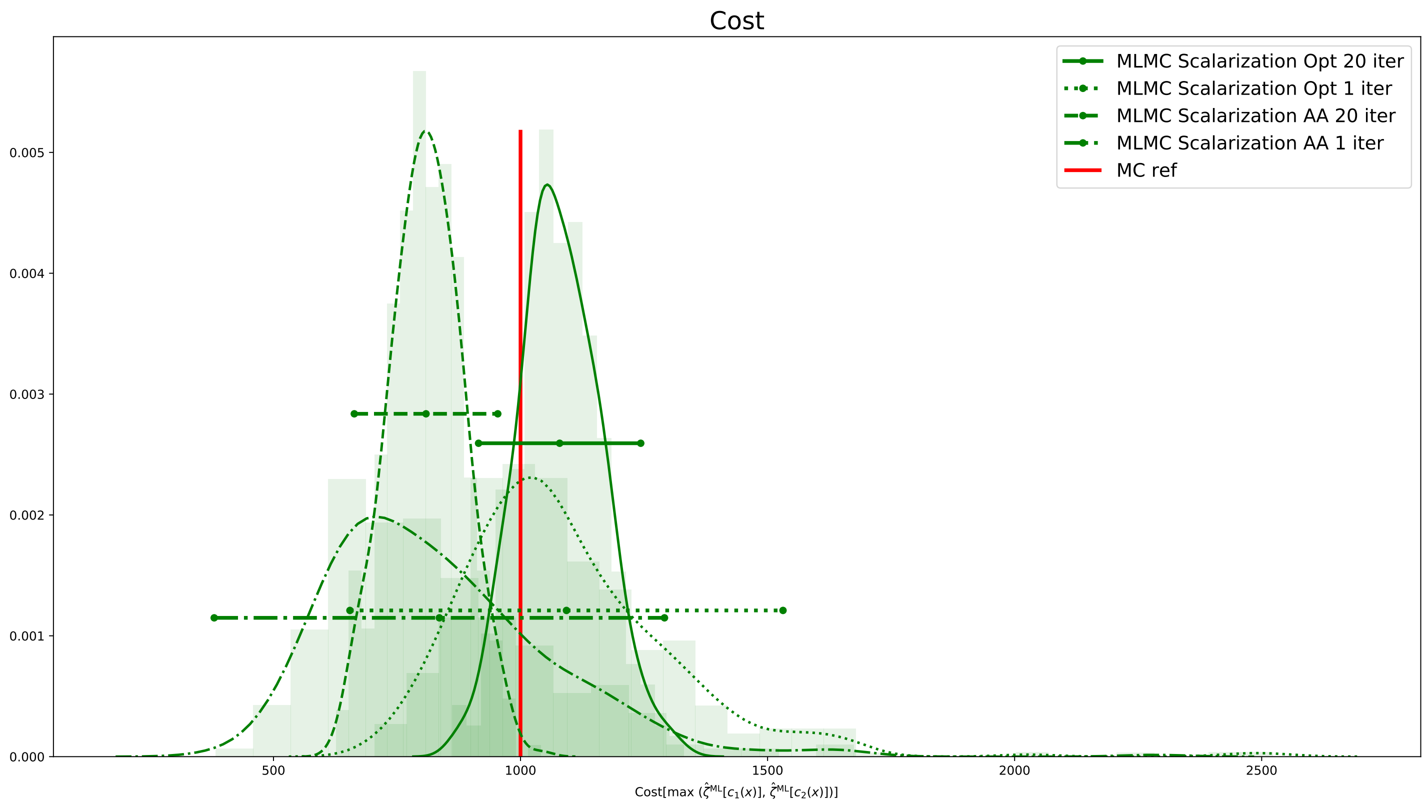
Sampling Results: Scalarization (Pearson)



- MLMC Scalarization (green) close but overresolves to MC reference case (red).

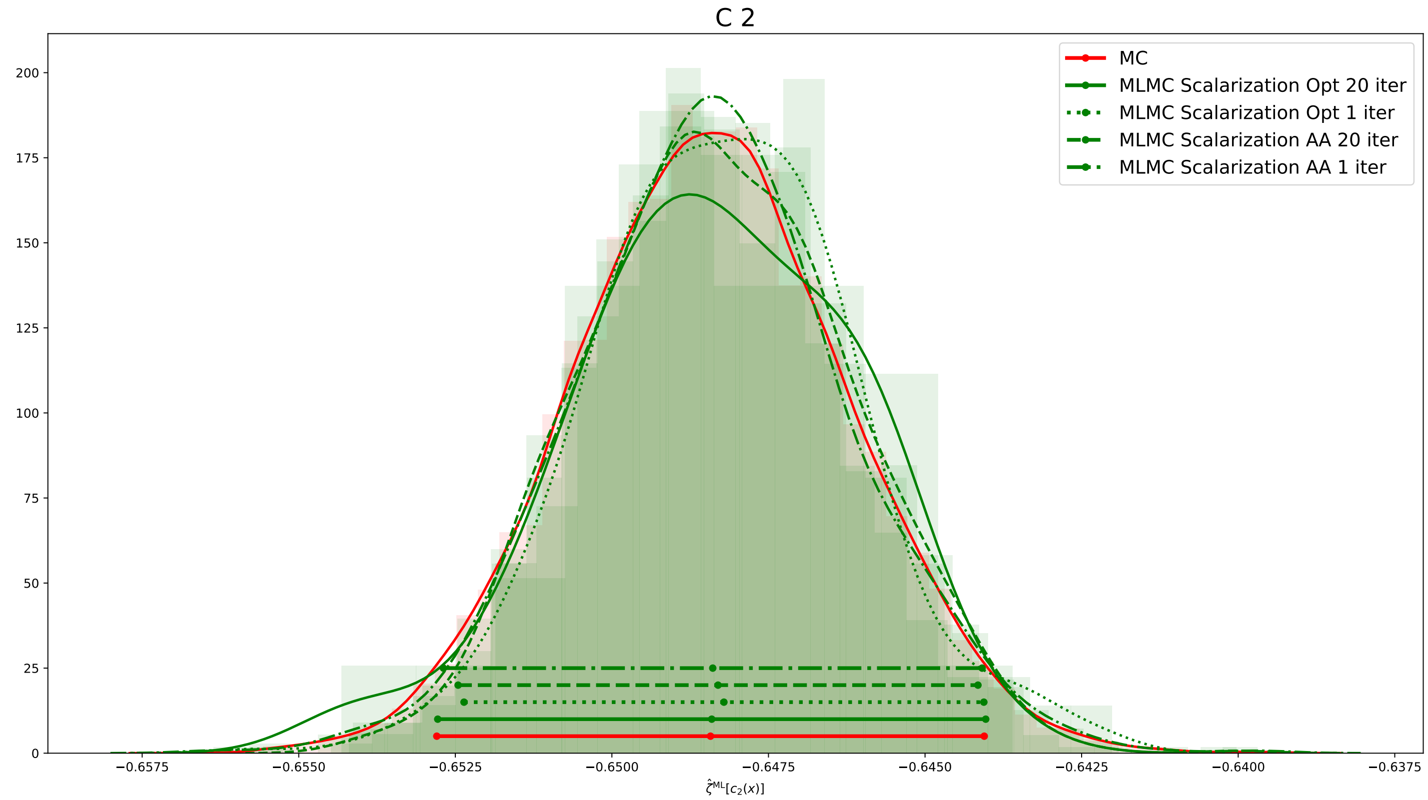
Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results: Scalarization (Pearson)



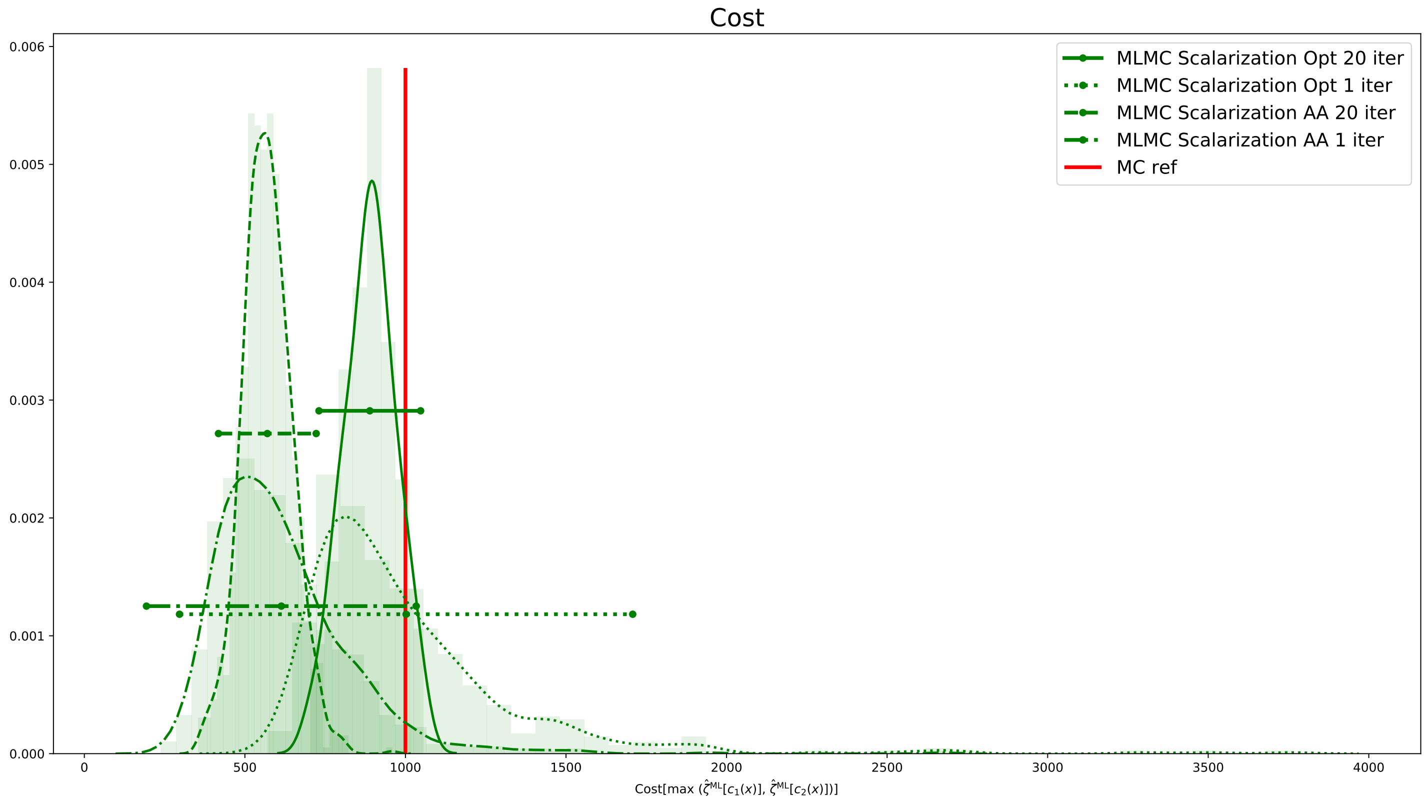
- MLMC Scalarization (green) close but overresolves to MC reference case (red).
- Results in higher computational cost than MC reference.

Sampling Results: Scalarization (Bootstrap)



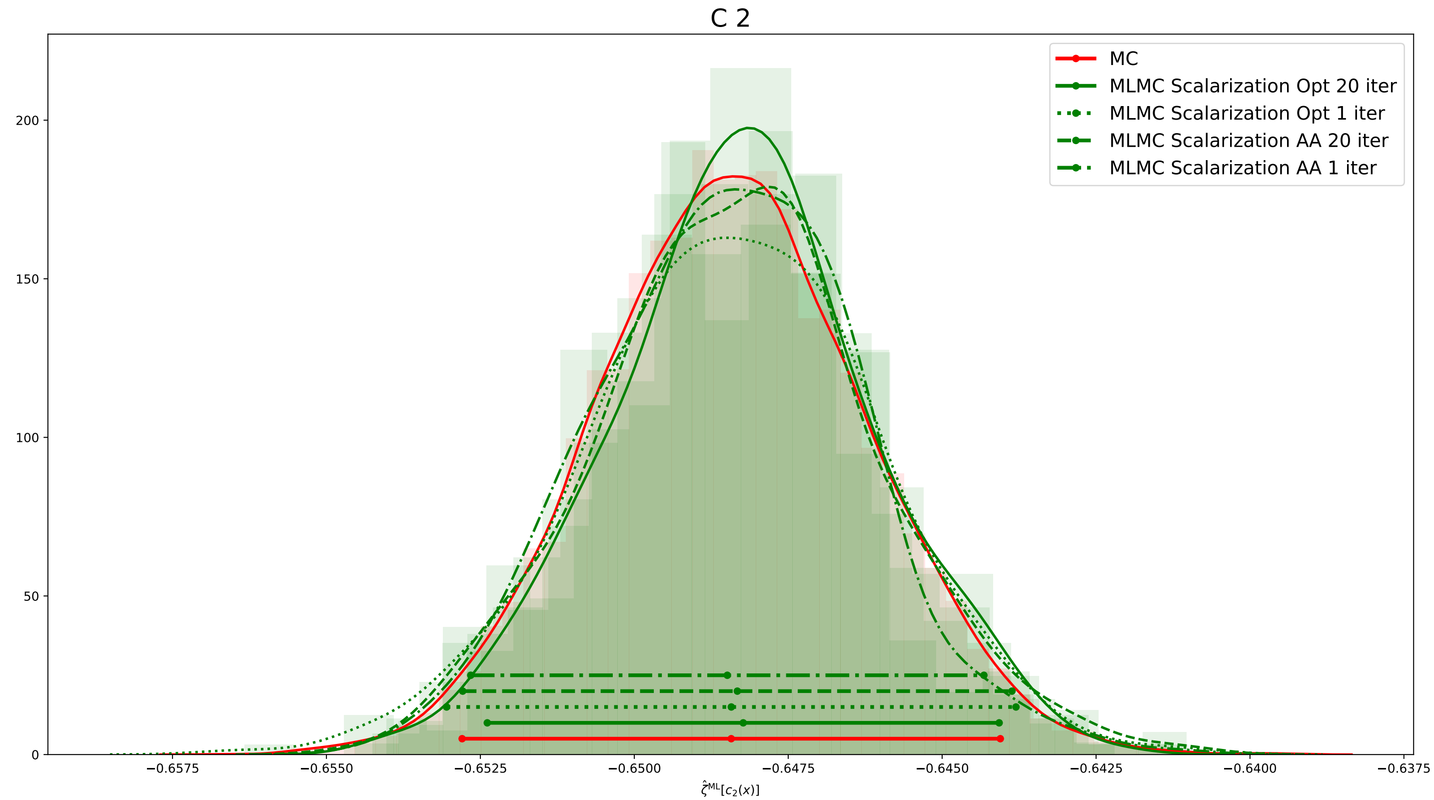
- MLMC Scalarization (green) consistent with MC reference (red) using Bootstrap.
- Cost reduced compared to Pearson
- BUT bootstrapping itself becomes computational bottleneck

Sampling Results: Scalarization (Bootstrap)



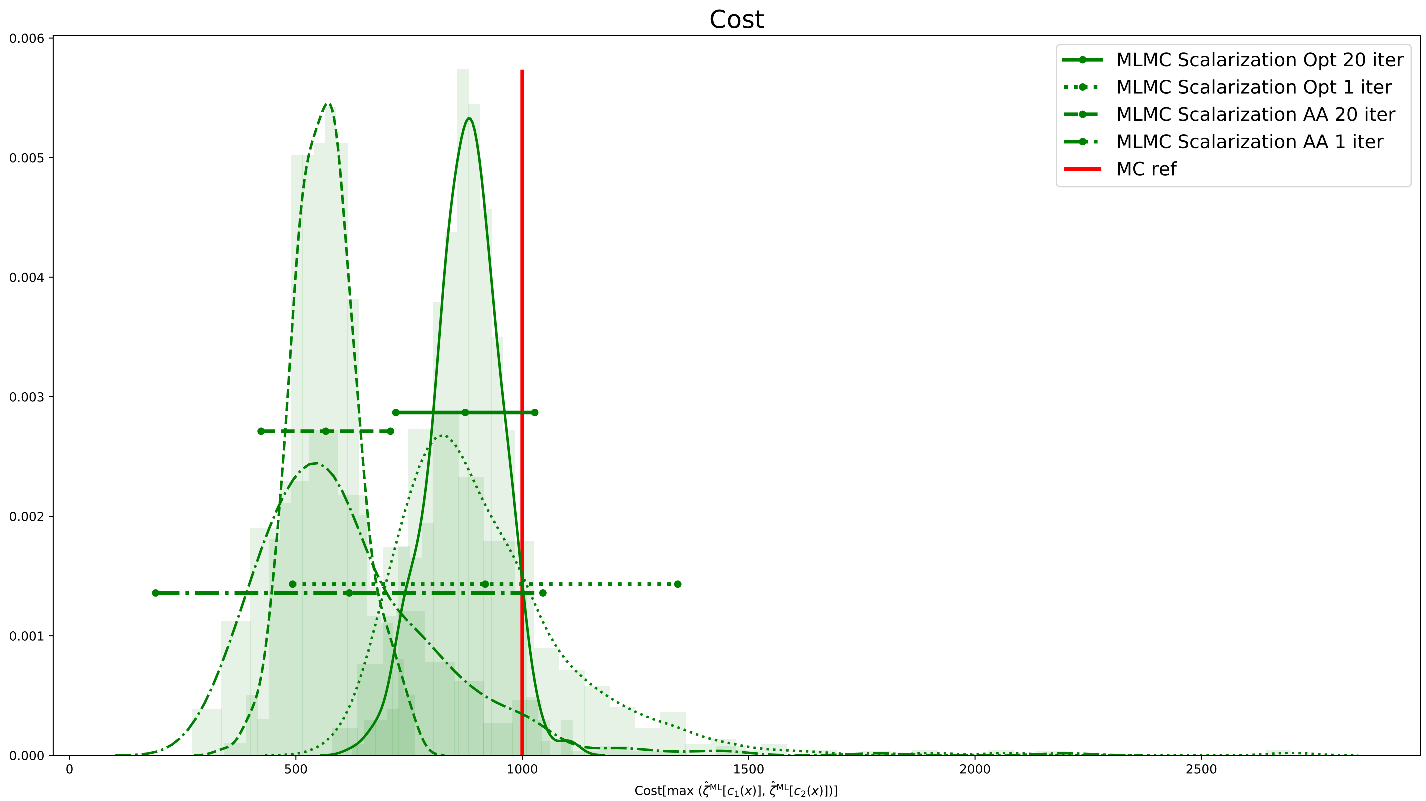
- MLMC Scalarization (green) consistent with MC reference (red) using Bootstrap.
- Cost reduced compared to Pearson
- BUT bootstrapping itself becomes computational bottleneck

Sampling Results: Scalarization (CorrLift)



- MLMC Scalarization (green) consistent with MC reference (red) (left) for lower cost (right) using CorrLift.
- Computational effort similar to Pearson and $<<$ Bootstrap

Sampling Results: Scalarization (CorrLift)



- MLMC Scalarization (green) consistent with MC reference (red) (left) for lower cost (right) using CorrLift.
- Computational effort similar to Pearson and $<<$ Bootstrap

Sampling Summary

- MLMC targeting the respective formulation necessary for optimal result
- Covariance term approximation crucial in scalarization
- Bootstrap leads to consistently better results but non-negligible computational cost
- ”Best of both worlds” achieved using CorrLift approach

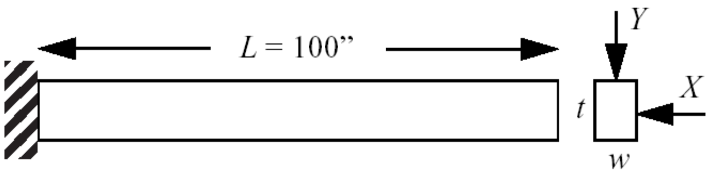
Sampling Summary

- MLMC targeting the respective formulation necessary for optimal result
- Covariance term approximation crucial in scalarization
- Bootstrap leads to consistently better results but non-negligible computational cost
- ”Best of both worlds” achieved using CorrLift approach

How does this translate to optimization?

Cantilever beam problem:

2 **design** parameters: w, t



Adams, B.M., et. al., Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User's Manual

Objective: $\min_{\mathbf{w},\mathbf{t}} wt$

4 **uncertain** parameters:

- Yield stress: $R \sim \mathcal{N}(40000, 200)$
- Young's modulus: $E \sim \mathcal{N}(2.9e7, 1.45e6)$
- Horizontal load: $X \sim \mathcal{N}(500, 100)$
- Vertical load: $Y \sim \mathcal{N}(1000, 100)$

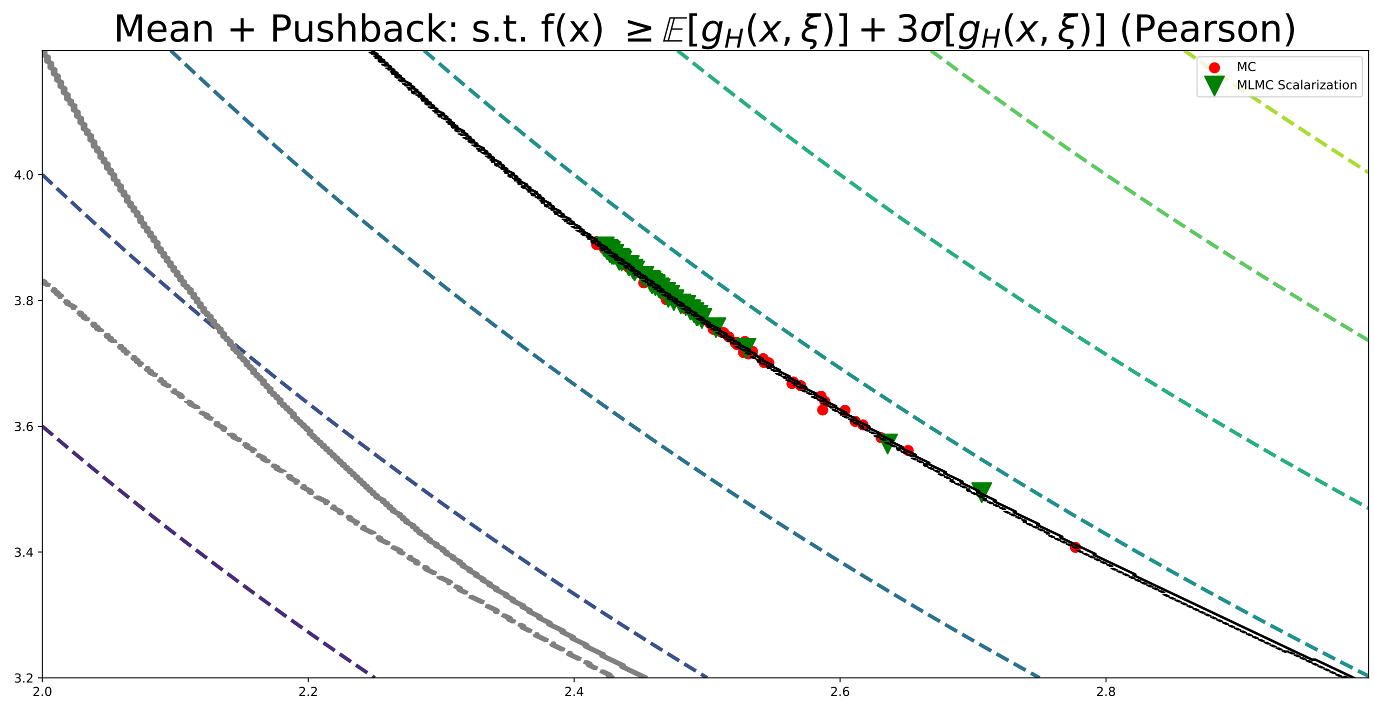
- Level 3: Rectangle (Cost = 1)

$$\text{stress}: \left(\frac{600}{wt^2}Y + \frac{600}{w^2t}X\right)/R - 1 \leq 0$$

$$\text{displacement}: \left(\frac{4L^3}{Ewt}\sqrt{\left(\frac{Y}{t^2}\right)^2 + \left(\frac{X}{w^2}\right)^2}\right)/D_0 - 1 \leq 0$$

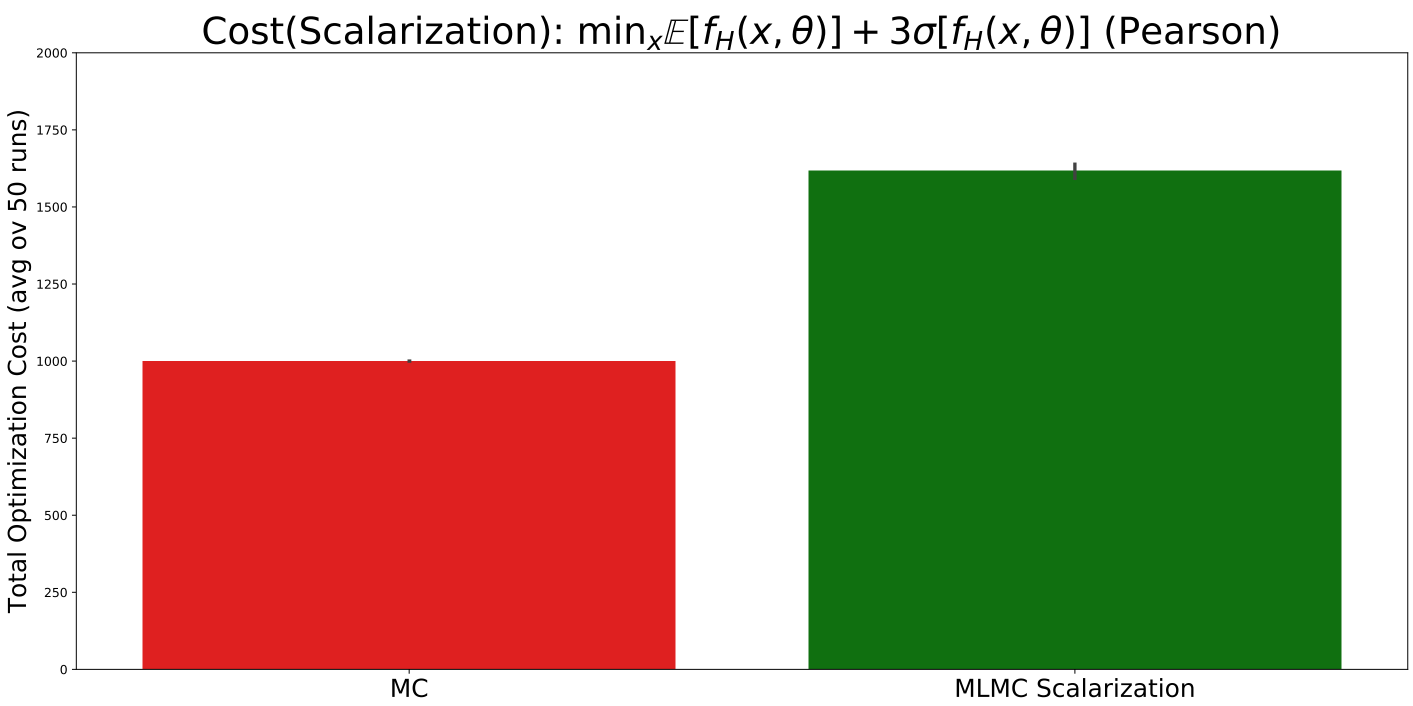
- Level 2: Ellipse instead of rectangle (Cost = 1e-2)
- ~~Level 1: Circle inscribed in rectangle instead of rectangle (Cost = 1e-4)~~
- ~~Level 0: Circle with same area instead of rectangle (Cost = 1e-6)~~

OUU Results (Pearson)



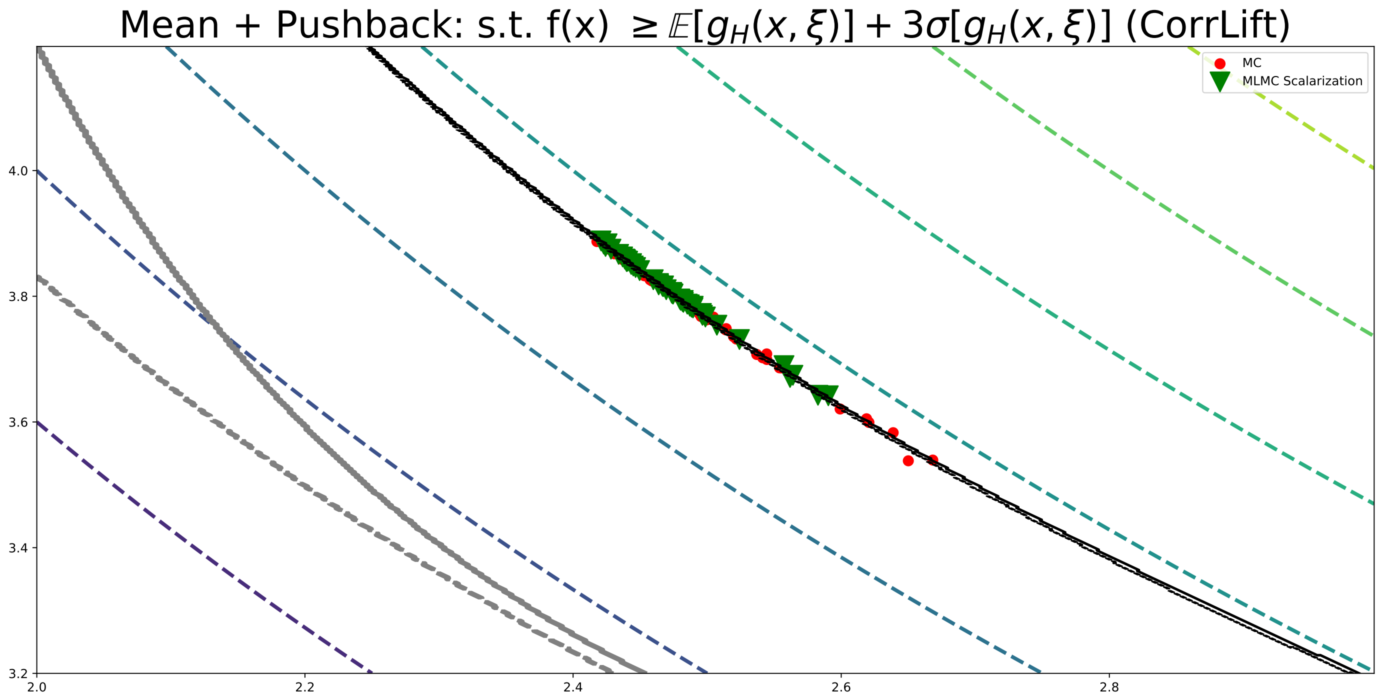
- Pearson approximation results in overestimation
- At the price of higher computational cost

OUU Results (Pearson)



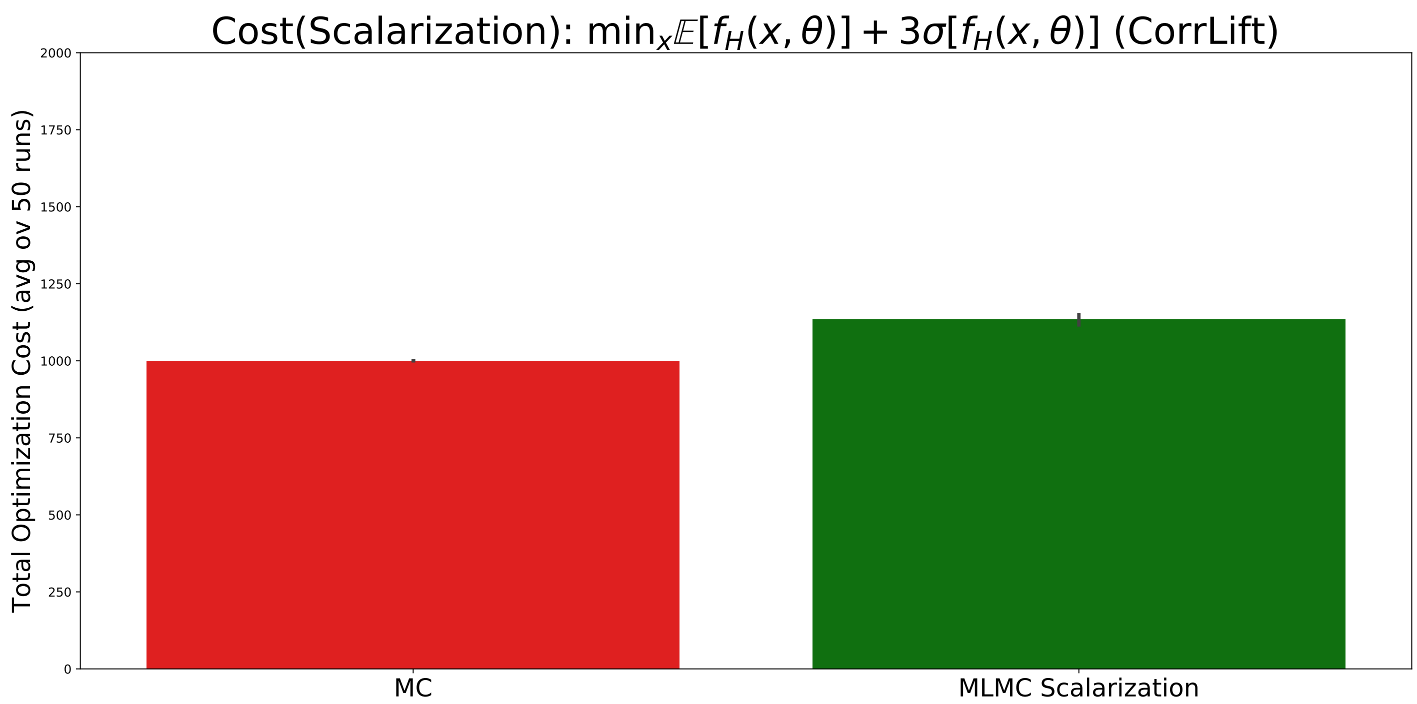
- Pearson approximation results in overestimation
- At the price of higher computational cost

OUU Results (CorrLift)



- CorrLift can reduce costs and improve the results
- Still higher computational cost compared to MC

OUU Results (CorrLift)



- CorrLift can reduce costs and improve the results
- Still higher computational cost compared to MC

Conclusion

⇒ New MLMC estimators for Standard Deviation and Scalarization coupled with **SNOWPAC**.

⇒ New approximations for $\mathbb{C}$ov-Term in Scalarization.

Future work and open questions:

- Investigate cost increase in Cantilever case
- Near future: Wind application

Links:

- SNOWPAC: github.com/snowpac/snowpac
- Dakota: dakota.sandia.gov

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References:

- FM, GG, DTS, MSE, RNK, HJB, YMM, Higher moment multilevel estimators for optimization under uncertainty applied to wind plant design. AIAA 2020
- F. Menhorn, F. Augustin, HJB, YMM, A trust-region method for derivative-free nonlinear constrained stochastic optimization. to be submitted