Scientific Computing
Department of Informatics
Technical University of Munich

SAND2021-2257C

Multifidelity Monte Carlo Estimators for Robust Formulations in Optimization under Uncertainty

Friedrich Menhorn[1], Gianluca Geraci[2], Daniel T. Seidl [2], Michael S. Eldred[2], Ryan N. King [3], Hans-Joachim Bungartz[1], Youssef M. Marzouk[4]

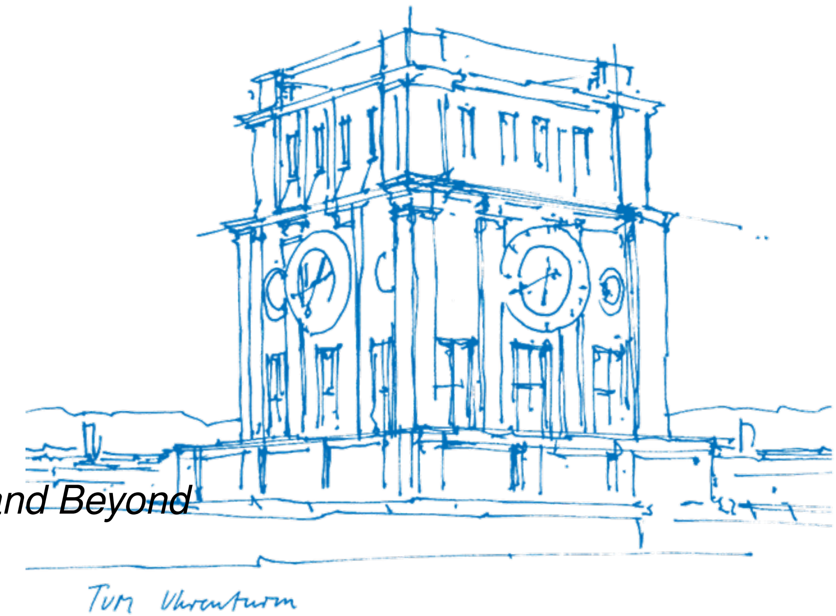
menhorn@in.tum.de

[1]Technical University of Munich, GER

[2]Sandia National Laboratories, USA

[3]National Renewable Laboratory, USA

[4]Massachusetts Institute of Technology, USA



SIAM CSE 2021

MS18

Models, Layers, Levels, Fidelities: Hierarchies in UQ and Beyond

Virtual, Mar 1st, 2021

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: Design optimization of a wind power plant

Provided by NREL

Vattenfall's Horns Rev wind farm off Denmark*

Setup:

- Wakes of turbine upstream interfere with turbines downstream
- Task: Steer turbines to maximize total power production P_{total}

Challenges:

- Complex, computationally expensive CFD black box code
- Uncertain conditions
- No gradients available

*Figure from https://www.rechargenews.com/wind/will-wind-wake-slow-industrys-ambitions-offshore-/2-1-699430

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Design optimization of a wind power plant

Provided by NREL

Vattenfall's Horns Rev wind farm off Denmark

Mean in OUU:

$$\max_{\gamma} \mathscr{R}_{\text{Mean}} := \max_{\gamma} \mathbb{E}[P_{\text{total}}(\gamma, \theta)]$$

Scalarization in OUU:

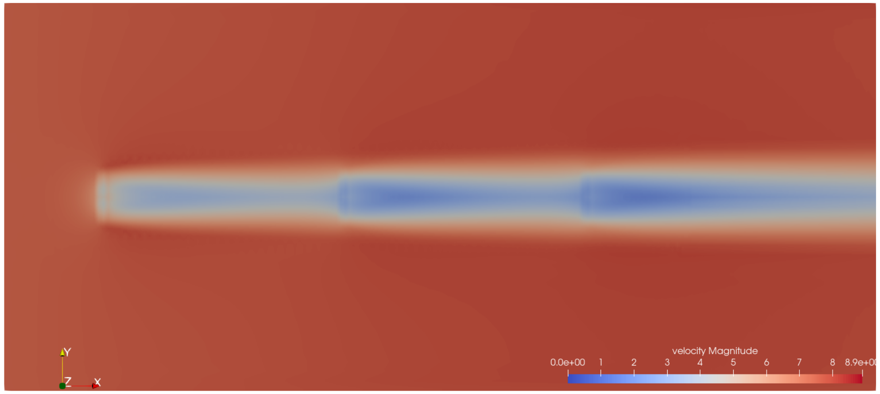
$$\max_{\gamma} \mathscr{R}_{\text{Pback}} := \max_{\gamma} (\mathbb{E}[P_{\text{total}}(\gamma, \theta)] - 3\sigma[P_{\text{total}}(\gamma, \theta)])$$

- γ: yaw angles of turbines
- θ: uncertain/stochastic inputs (wind speed, sensor errors, etc.)
- $\mathscr{R}$: measure of robustness/reliability

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Design optimization of a wind power plant

Provided by NREL



Mean in OUU:

$$\max_{\gamma} \mathbb{E}[P_{\text{total}}(\gamma, \theta)]$$
$$\approx \max_{\gamma} \frac{1}{N} \sum_{i=1}^{N} [P_{\text{total}}(\gamma, \theta_i)]$$

Scalarization in OUU:

$$\max_{\gamma} \mathbb{E}[P_{\text{total}}(\gamma, \theta)] - 3\sigma[P_{\text{total}}(\gamma, \theta)]$$
$$\approx \max_{\gamma} \frac{1}{N} \sum_{i=1}^{N} P_{\text{total}}(\gamma, \theta_i) - 3\sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (P_{\text{total}}(\gamma, \theta_i) - \hat{P})^2}$$

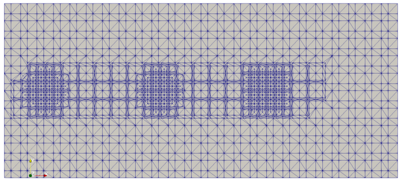
MC property: Error $\sim \mathcal{O}(\frac{1}{\sqrt{N}}) \Rightarrow N$ should be high but evaluations computationally expensive!

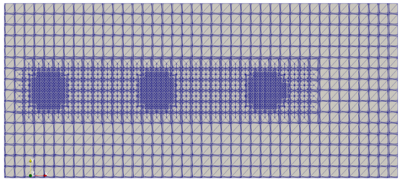
Motivation: Design optimization of a wind power plant

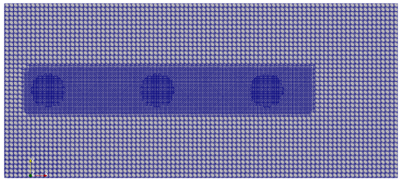
Provided by NREL

Hierarchies of levels and fidelities (future work) available:

- Discretization level:







- Model level:

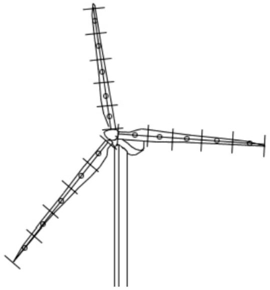


Image: Yu, Zheng, Ma, https://nwtc.nrel.gov/OpenFAST – Jason Jonkman, et al.

Courtesy of Barone and Maniaci, SNL

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Design optimization of a wind power plant

Provided by NREL

Leverage hierarchy using multilevel Monte Carlo estimators:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] \approx \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Push back:

$$\min_x \mathscr{R}_{\text{Pback}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] - 3\sigma[P_{\text{total}}(\gamma,\theta)] \approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)] - 3\widehat{\sigma^{\text{ML}}_{\text{biased}}}[P_{\text{total}}(\gamma,\theta)]$$

Motivation: Design optimization of a wind power plant

Provided by NREL

Leverage hierarchy using multilevel Monte Carlo estimators:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] \approx \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Push back:

$$\min_x \mathscr{R}_{\text{Pback}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] - 3\sigma[P_{\text{total}}(\gamma,\theta)] \approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)] - 3\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Up next:

Motivation: Design optimization of a wind power plant

Provided by NREL

Leverage hierarchy using multilevel Monte Carlo estimators:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma, \theta)] \approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma, \theta)]$$

Push back:

$$\min_x \mathscr{R}_{\text{Pback}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma, \theta)] - 3\sigma[P_{\text{total}}(\gamma, \theta)] \approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma, \theta)] - 3\widehat{\sigma^{\text{ML}}_{\text{biased}}}[P_{\text{total}}(\gamma, \theta)]$$

Up next:

1. Recap: $\widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma, \theta)]$

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Design optimization of a wind power plant

Provided by NREL

Leverage hierarchy using multilevel Monte Carlo estimators:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)]$$
$$\approx \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Push back:

$$\min_x \mathscr{R}_{\text{Pback}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] - 3\sigma[P_{\text{total}}(\gamma,\theta)]$$
$$\approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)] - 3\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Up next:

1. Recap: $\widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$
2. (Our contribution) ML for Standard deviation: $\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$

Scientific Computing
Department of Informatics
Technical University of Munich

Motivation: Design optimization of a wind power plant

Provided by NREL

Leverage hierarchy using multilevel Monte Carlo estimators:

Mean in OUU:

$$\min_x \mathscr{R}_{\text{Mean}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] \approx \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Push back:

$$\min_x \mathscr{R}_{\text{Pback}} := \max_\gamma \mathbb{E}[P_{\text{total}}(\gamma,\theta)] - 3\sigma[P_{\text{total}}(\gamma,\theta)] \approx \max_\gamma \widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)] - 3\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$$

Up next:

1. Recap: $\widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$
2. (Our contribution) ML for Standard deviation: $\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$
3. (Our contribution) ML for Scalarization: $\widehat{\mu_1^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)] - 3\widehat{\sigma_{\text{biased}}^{\text{ML}}}[P_{\text{total}}(\gamma,\theta)]$

Multilevel Monte Carlo (MLMC) Estimator

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

MLMC estimator: Mean

Mean in OUU:

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

- Estimator[†]:

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

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

MLCC estimator: Mean

Mean in OUU:

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

- Estimator[†]:

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

- 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[†]:

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

- 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

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:

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

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

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

where

$$\begin{aligned}
\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\text{ML}}}[Q_L] &= \sum_{\ell=0}^{L} \widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}] \\
&= \sum_{\ell=0}^{L} \frac{1}{N_\ell - 1} \left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2 \right) \\
&= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0
\end{aligned}$$

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

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

where

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

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

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

- Sample allocation:

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

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

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimator: Standard deviation

- Estimator:

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

where

$$\begin{aligned}\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\text{ML}}}[Q_L] &= \sum_{\ell=0}^{L} \widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}] \\ &= \sum_{\ell=0}^{L} \frac{1}{N_\ell - 1} \left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2 \right) \\ &= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0\end{aligned}$$

- Sample allocation:

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

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

- Solution: Numerical Optimization

MLMC estimator: Standard deviation

- Estimator:

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

where

$$\begin{aligned}\mathbb{V}[Q_L] \approx \widehat{\mu_2^{\mathrm{ML}}}[Q_L] &= \sum_{\ell=0}^{L} \widehat{\mu_2}[Q^{(\ell)}] - \widehat{\mu_2}[Q^{(\ell-1)}] \\ &= \sum_{\ell=0}^{L} \frac{1}{N_\ell - 1}\left(\sum_{i=1}^{N_\ell} (Q_i^{(\ell)} - \widehat{\mu_1^{(\ell)}})^2 - (Q_{i,\ell}^{(\ell-1)} - \widehat{\mu_{1,\ell}^{(\ell-1)}})^2 \right) \\ &= \sum_{\ell=0}^{L} (\widehat{\mu_2^{(\ell)}} - \widehat{\mu_{2,\ell}^{(\ell-1)}}), \quad Q_{i,0}^{(-1)} := 0\end{aligned}$$

- Sample allocation:

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

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

- Solution: Numerical Optimization

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:

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

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:

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

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

- Solution: Numerical Optimization

MLMC estimators: 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:

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

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

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

From work for higher order moments:

- 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.
- (Bierig, C., Chernov, A., “Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method,” Stochastics and Partial Differential Equations Analysis and Computations, Vol. 4, No. 1, 2016, pp. 3-40.)

Scientific Computing
Department of Informatics
Technical University of Munich

MLMC estimators: 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: $\mathcal{V}_\ell = \mathbb{V}[\widehat{\mu_1^{(\ell)}}] N_\ell^{\mathbb{E}} \Leftrightarrow \mathbb{V}[\widehat{\mu_1^{(\ell)}}] = \frac{\mathcal{V}_\ell}{N_\ell}$

From work for higher order moments:

- 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.
- (Bierig, C., Chernov, A., "Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method," Stochastics and Partial Differential Equations Analysis and Computations, Vol. 4, No. 1, 2016, pp. 3-40.)

MLMC estimators: 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: $\mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_x^{(\ell)}}] N_\ell^{\mathbb{X}} \Leftrightarrow \mathbb{V}[\widehat{\mu_x^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell}$

From work for higher order moments:

- 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.
- (Bierig, C., Chernov, A., “Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method,” Stochastics and Partial Differential Equations Analysis and Computations, Vol. 4, No. 1, 2016, pp. 3-40.)

MLMC estimators: Analytic approximation

- **Obs**: Decrease of variance of MLMC estimator $\mathscr{O}(\frac{1}{N})$ 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: $\mathscr{V}_\ell = \mathbb{V}[\widehat{\mu_x^{(\ell)}}] N_\ell^{\mathbb{X}} \Leftrightarrow \mathbb{V}[\widehat{\mu_x^{(\ell)}}] = \frac{\mathscr{V}_\ell}{N_\ell}$
- Disregard higher order terms in $\mathscr{V}_\ell \Rightarrow$ Analytic approximation

From work for higher order moments:

- 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.
- (Bierig, C., Chernov, A., "Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method," Stochastics and Partial Differential Equations Analysis and Computations, Vol. 4, No. 1, 2016, pp. 3-40.)

Coupling ML with OUU

Scientific Computing
Department of Informatics
Technical University of Munich

SNOWPAC*

OUU problem statement

$$\mathcal{R}_\theta^* = \mathcal{R}^f(\mathbf{x}^*, \theta) = \min_{\mathcal{R}^c(\mathbf{x},\theta)\leq 0} \mathcal{R}^f(\mathbf{x}, \theta)$$

Features of SNOWPAC:

0. **Extension of NOWPAC**: Derivative-free nonlinear constraint optimization method using trust-regions (deterministic)
1. **Estimate robustness measures**: Use sampling, e.g.

$$\mathcal{R}_\theta^f = \mathbb{E}[f_\theta(\mathbf{x})] \approx R^f = \frac{1}{N}\sum_{i=1}^N f(\mathbf{x}, \theta_i) + \varepsilon_N$$

2. **Implement new trust region management**: Account for noise ε_N in objective/constraint evaluations
3. **Introduce Gaussian process surrogates**: Mitigate effect of noise ε_N
4. **Feasibility restoration mode**

*F. Augustin, Y. Marzouk, A trust-region method for derivative-free nonlinear constrained stochastic optimization. 2017

Scientific Computing
Department of Informatics
Technical University of Munich

SNOWPAC*

OUU problem statement

$$\mathscr{R}^*_\theta = \mathscr{R}^f(\mathbf{x}^*, \theta) = \min_{\mathscr{R}^c(\mathbf{x},\theta)\leq 0} \mathscr{R}^f(\mathbf{x}, \theta)$$

Features of SNOWPAC:

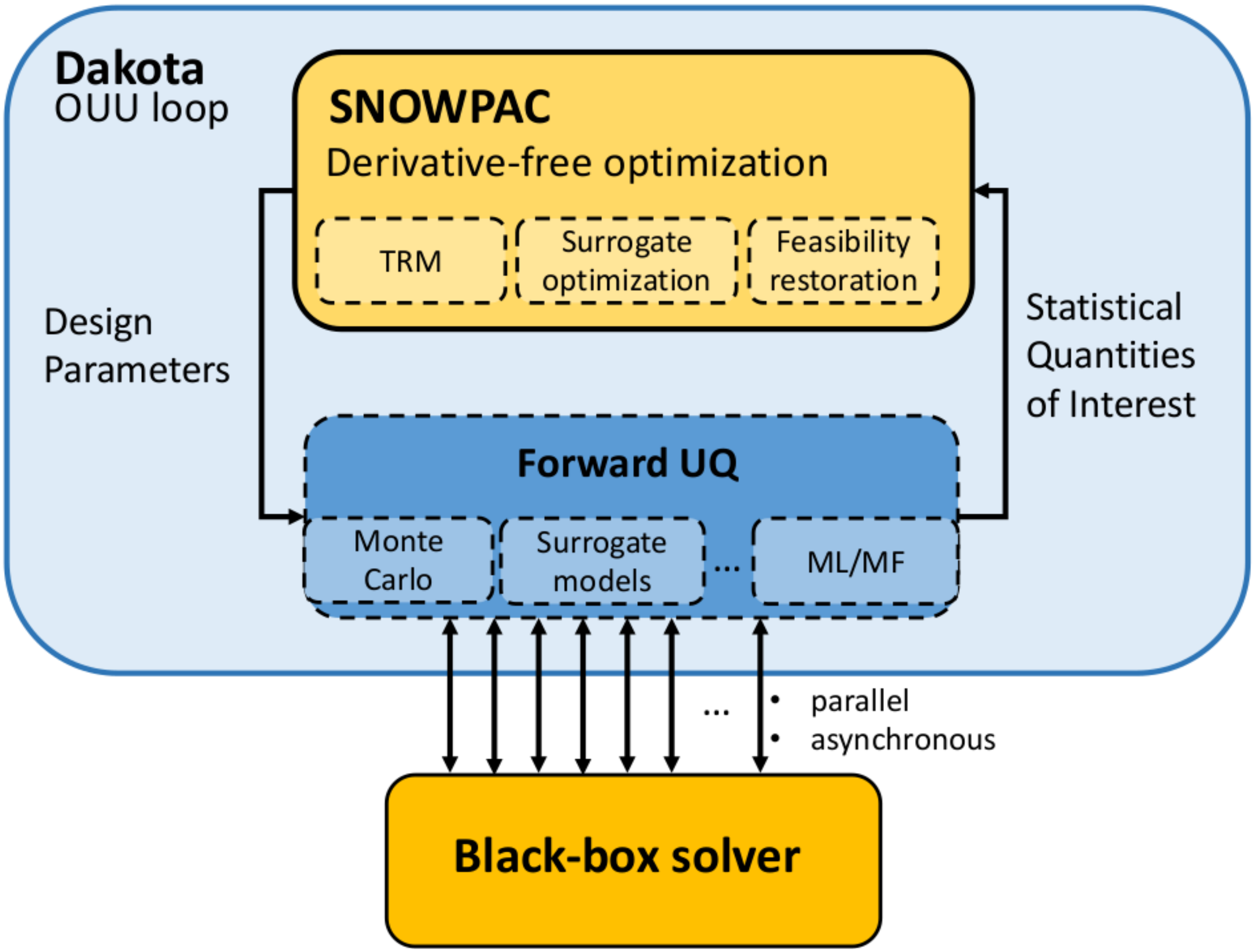
0. **Extension of NOWPAC**: Derivative-free nonlinear constraint optimization method using trust-regions (deterministic)
1. **Estimate robustness measures**: Use sampling, e.g.

$$\mathscr{R}^f_\theta = \mathbb{E}[f_\theta(\mathbf{x})] \approx R^f = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (f^\ell(\mathbf{x}, \theta_i) - f^{\ell-1}(\mathbf{x}, \theta_i)) + \varepsilon_N$$

2. **Implement new trust region management**: Account for noise ε_N in objective/constraint evaluations
3. **Introduce Gaussian process surrogates**: Mitigate effect of noise ε_N
4. **Feasibility restoration mode**

*F. Augustin, Y. Marzouk, A trust-region method for derivative-free nonlinear constrained stochastic optimization. 2017

SNOWPAC in Dakota using MLMC



Benchmark Results

Problem statement

Objective:

$$f(x) = \begin{cases} (x-2)^2 & \text{if } x \leq 3 \\ 2\log(x-2)+1 & \text{if } x > 3 \\ x \in [0,6] \end{cases}$$

OUU:

$$\min_x f(x)$$

Mean:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$

Push back:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$

Constraint:

$$g_{det}(x) = \frac{2.log(1.5)}{2.5}x - \frac{2.log(1.5)}{2.5}$$

$$g_H(x,\xi) = g_{det}(x) + \xi^3$$

$$g_L(x,\xi) = g_{det}(x) + A(x)\xi^3, \xi \sim \mathscr{U}(-0.5, 0.5)$$

$$A(x) = \frac{1}{12}x + 0.4$$

Algorithmic comparison:

- AA: Analytic Approximation
- Opt: Numerical Optimization

Scientific Computing
Department of Informatics
Technical University of Munich

Problem statement

Objective:

$$f(x) = \begin{cases} (x-2)^2 & \text{if } x \leq 3 \\ 2\log(x-2)+1 & \text{if } x > 3 \\ x \in [0,6] \end{cases}$$

OUU:

$$\min_x f(x)$$

Mean:

$$\text{s.t. } f(x) \geq \mathbb{E}[g_H(x,\xi)]$$

Push back:

$$\text{s.t. } f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$$

Algorithmic comparison:

- AA: Analytic Approximation
- Opt: Numerical Optimization

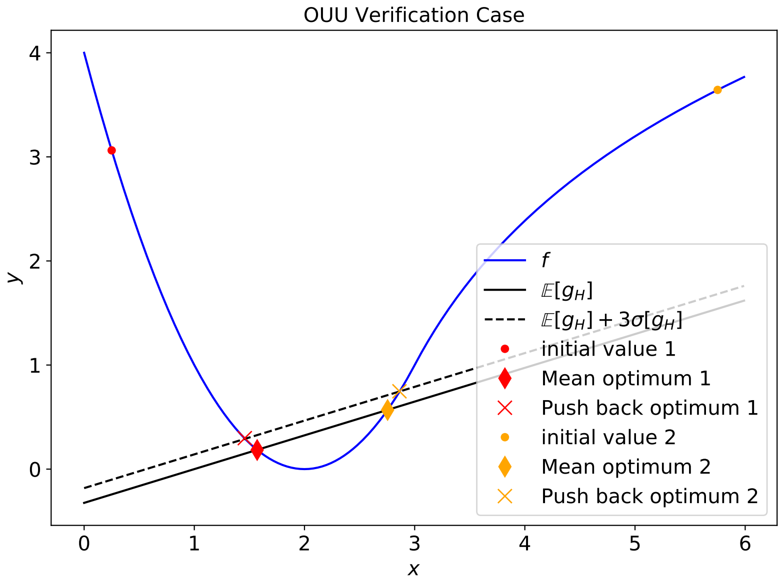
Constraint:

$$g_{det}(x) = \frac{2.log(1.5)}{2.5}x - \frac{2.log(1.5)}{2.5}$$

$$g_H(x,\xi) = g_{det}(x) + \xi^3$$

$$g_L(x,\xi) = g_{det}(x) + A(x)\xi^3, \xi \sim \mathscr{U}(-0.5, 0.5)$$

$$A(x) = \frac{1}{12}x + 0.4$$



Sampling Results: Sample allocation

- Mean

$$\mathbb{E}[g_H(x,\xi)] \approx \widehat{\mu_1^{\mathrm{ML}}}[g_{H/L}(x,\xi)]$$

N_{MC}		$x=1$	3	5
200	N_L	213	236	233
	N_H	69	39	16
300	N_L	309	354	350
	N_H	103	58	24

- Scalarization

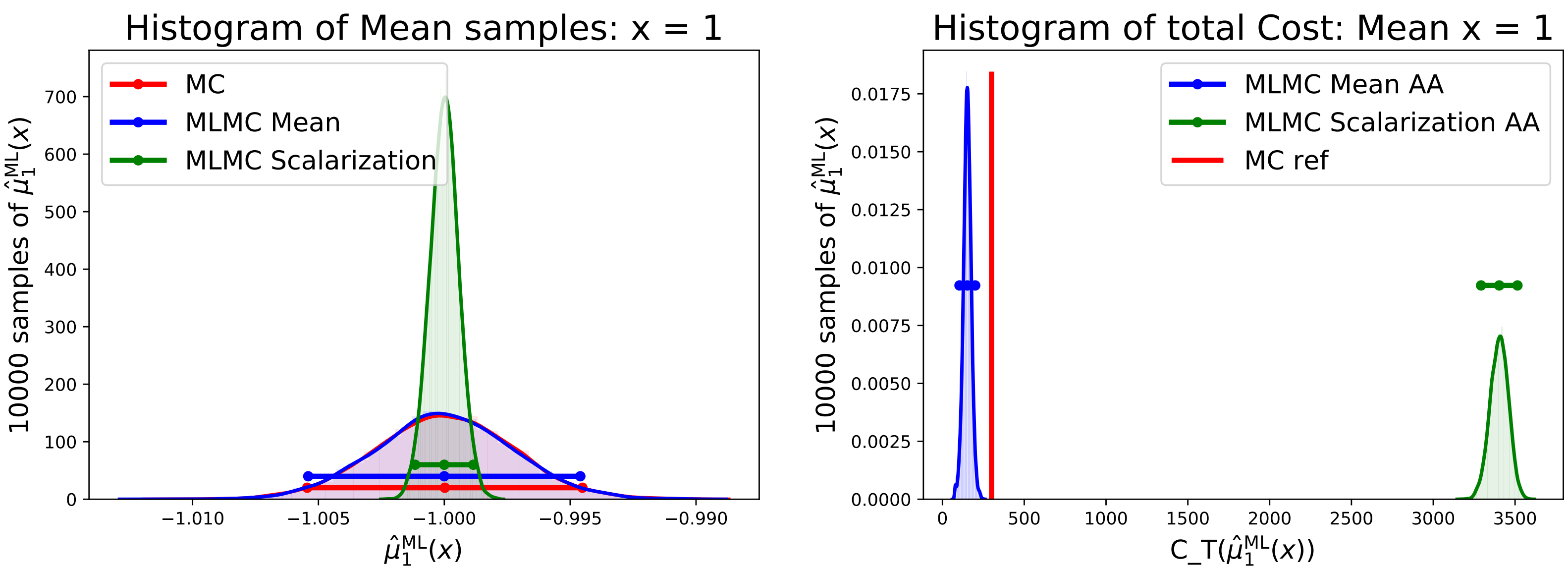
$$\mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)] \approx \widehat{\zeta^{\mathrm{ML}}}[g_{H/L}(x,\xi)]$$

N_{MC}		$x=1$	3	5
200	N_L	303	384	419
	N_H	190	121	63
300	N_L	456	579	626
	N_H	285	180	90

- Total computational cost: $C_T = 1.0 N_L + 1.1 N_H$
- Computational cost decrease for increasing x $\Leftrightarrow$ Higher correlation due to $A(x)$
- Overall higher computational cost for scalarization $\Leftrightarrow$ Need for higher resolution
- Cost for scalarization for $x = 1$ even higher than for standard MC $\Leftrightarrow$ Upper bound on $\mathbb{C}$ov

Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results

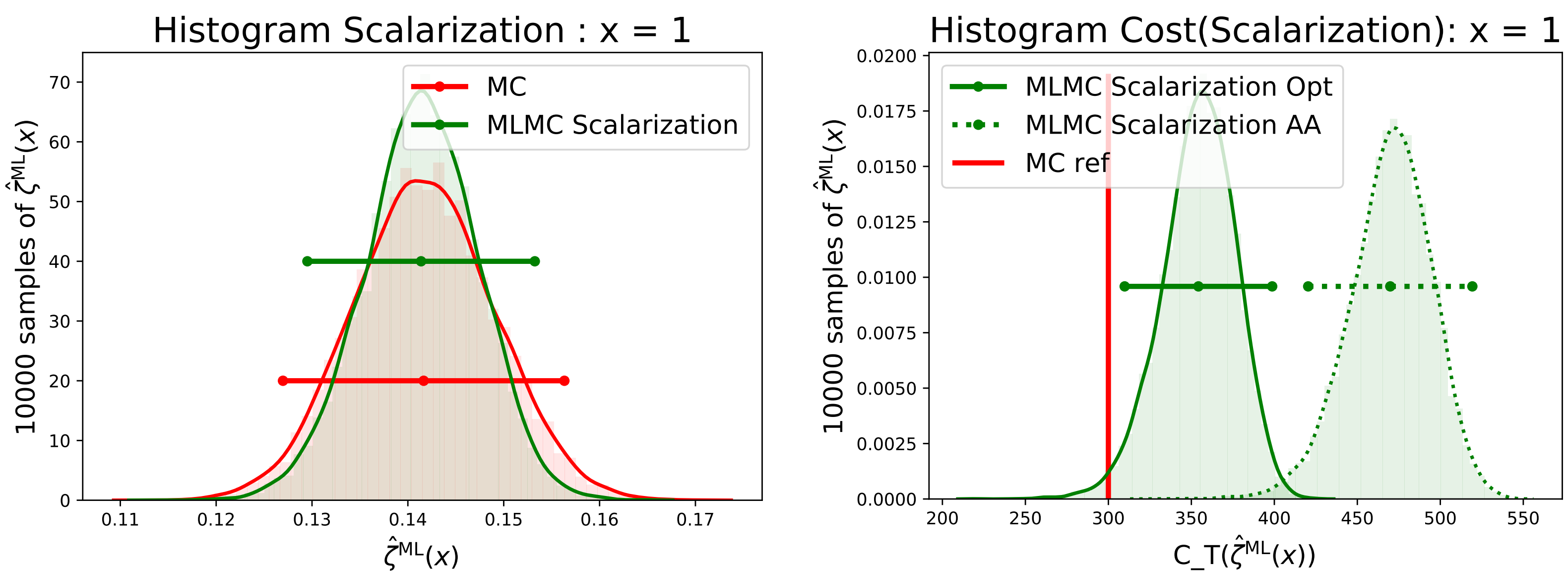


Mean: $\widehat{\mu_1^{ML}}[g_{H/L}(x,\xi)]$

- MLMC Mean distribution consistent with MC reference
- MLMC Scalarization overresolves, results in higher computational cost

Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results

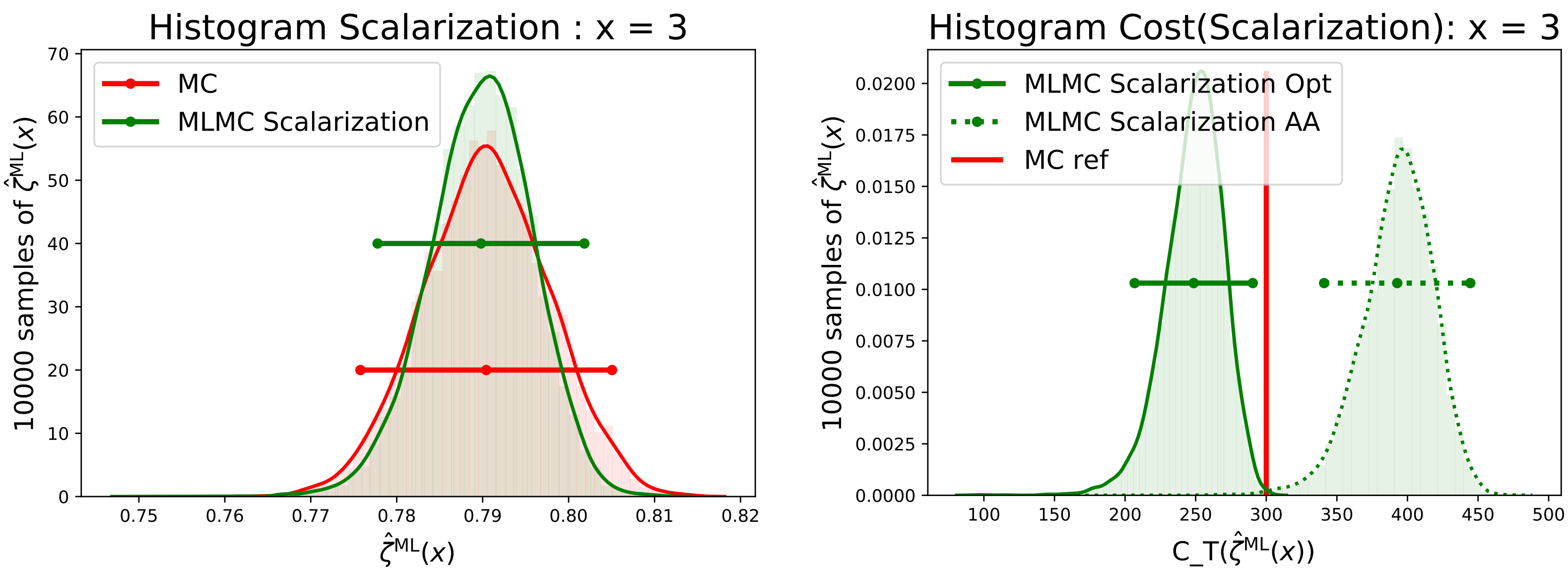


Scalarization: $\widehat{\zeta^{\mathrm{ML}}}[g_{H/L}(x,\xi)]$

- MLMC Scalarization close to MC reference case
- MLMC Scalarization overresolves due to upper bound in $\mathbb{C}$ov
- Numerical optimization reduce cost compared to analytical approximation
- Results in higher computational cost than MC reference for $x = 1$

Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results

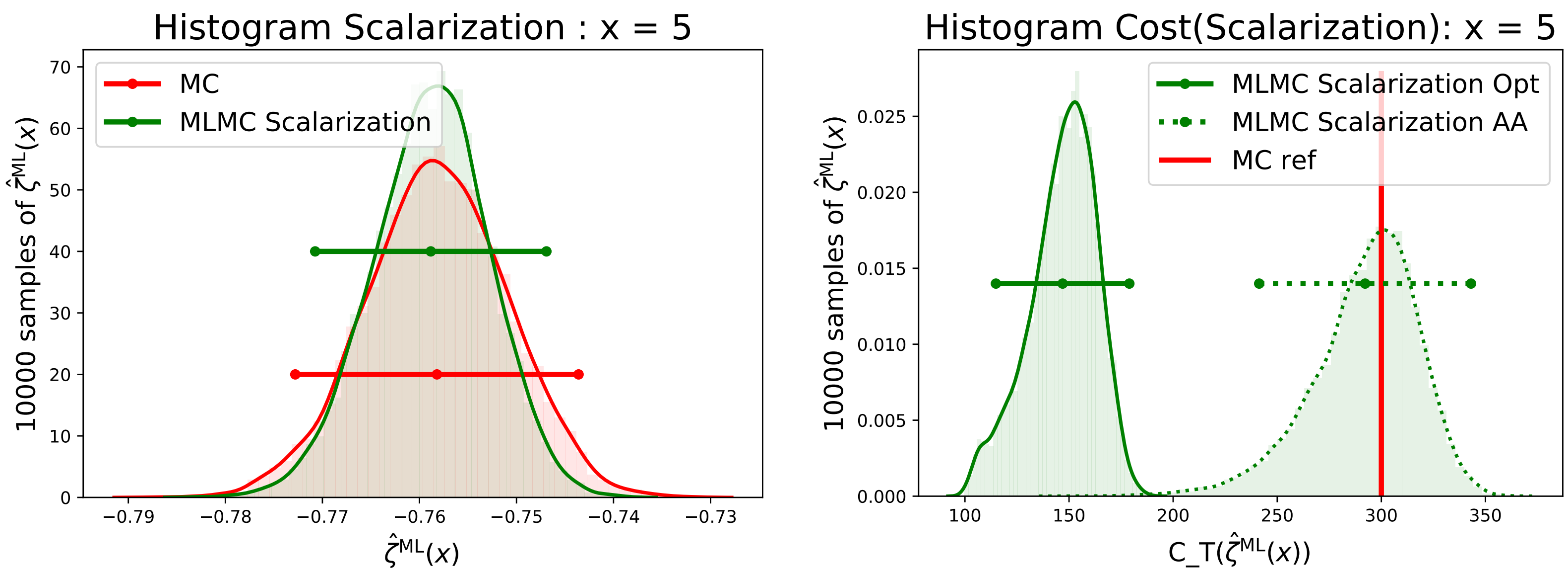


Scalarization: $\widehat{\zeta^{\mathrm{ML}}}[g_{H/L}(x,\xi)]$

- MLMC Scalarization close to MC reference case
- MLMC Scalarization overresolves due to upper bound in $\mathbb{C}$ov
- Numerical optimization reduce cost compared to analytical approximation
- Decrease in cost and lower than MC reference for increase in x

Scientific Computing
Department of Informatics
Technical University of Munich

Sampling Results



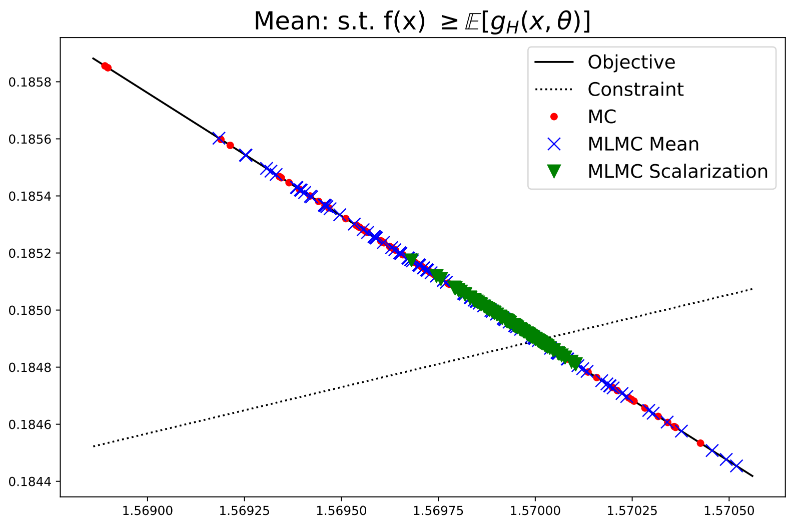
Scalarization: $\widehat{\zeta^{\text{ML}}}[g_{H/L}(x,\xi)]$

- MLMC Scalarization close to MC reference case
- MLMC Scalarization overresolves due to upper bound in $\mathbb{C}$ov
- Numerical optimization reduce cost compared to analytical approximation
- Decrease in cost and lower than MC reference for increase in x

OUU Results

Mean:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$



- Consistent match between MLMC Mean and MC
- Scalarization overresolves

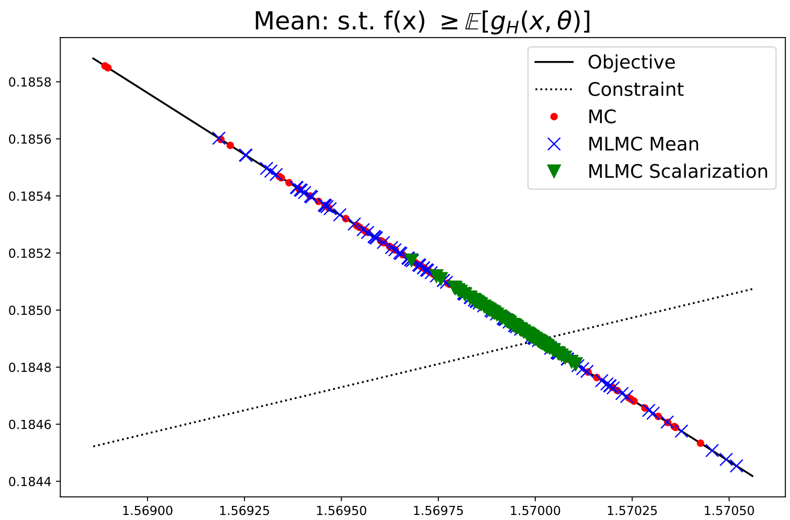
OUU Results

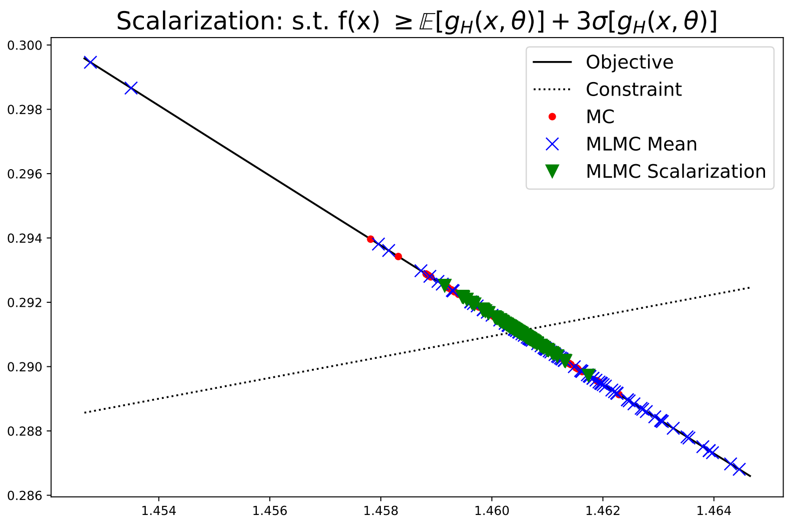
Mean:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$

Push back:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$





- Consistent match between MLMC Mean and MC
- Scalarization overresolves

- More accurate result for MLMC Scalarization cmp to MC
- MLMC Mean underresolves

Scientific Computing
Department of Informatics
Technical University of Munich

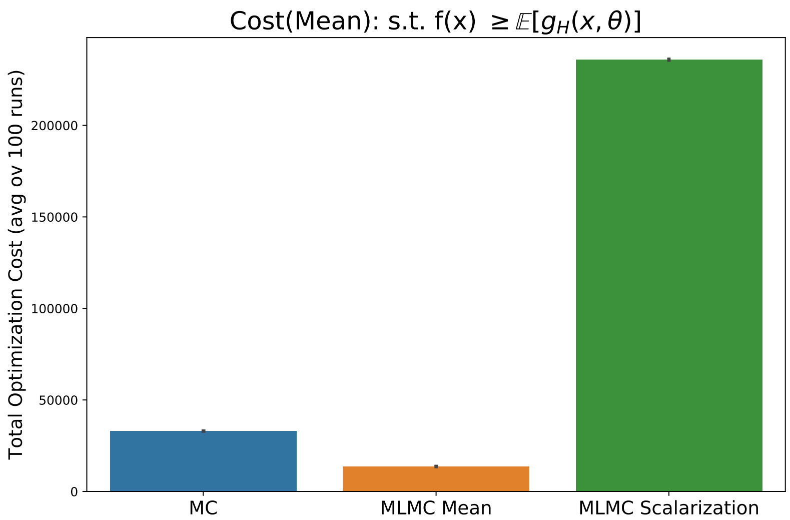
OUU Results

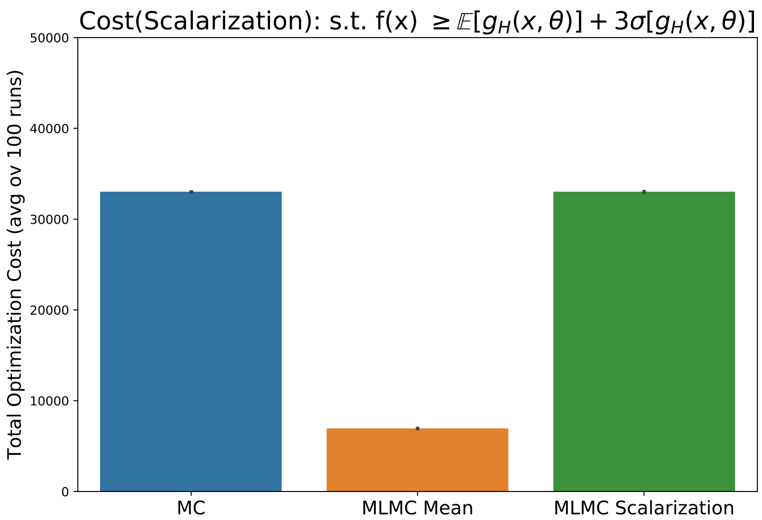
Mean:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$

Push back:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$





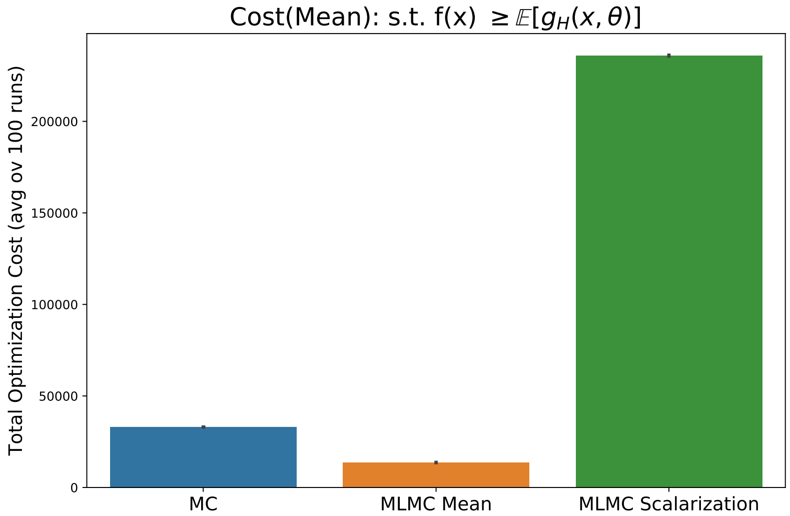
- Lower cost for MLMC Mean
- High cost for MLMC Scalarization

- Similar cost between Scalarization and MC (due to numerical opt)
- Too low cost for MLMC Mean

OUU Results

Mean:

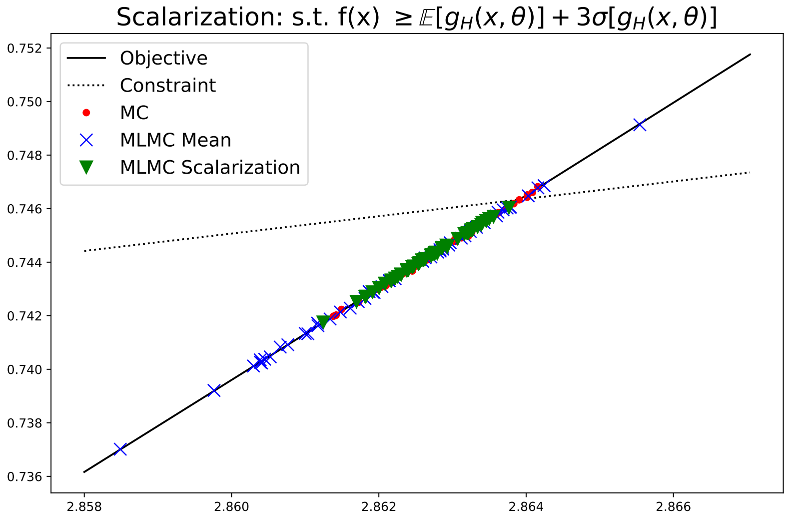
s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$



- Lower cost for MLMC Mean
- High cost for MLMC Scalarization

Push back:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$



- Start at $x = 5.75$
- Consistent match between MLMC Scalarization and MC

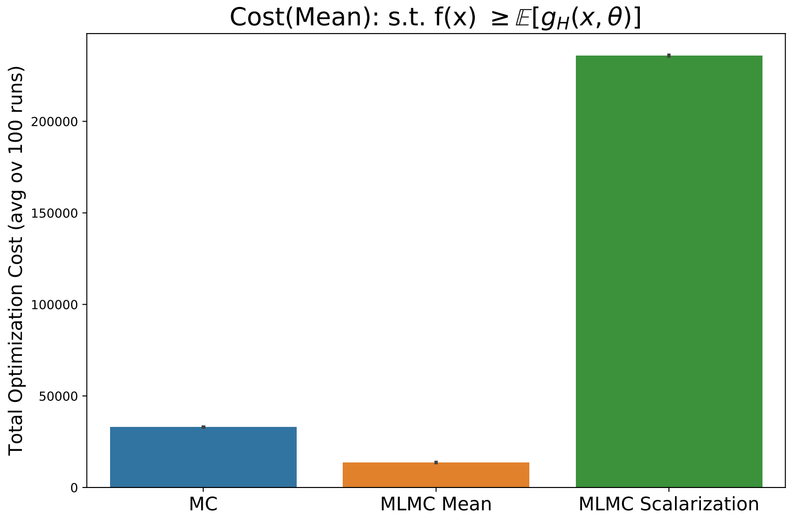
OUU Results

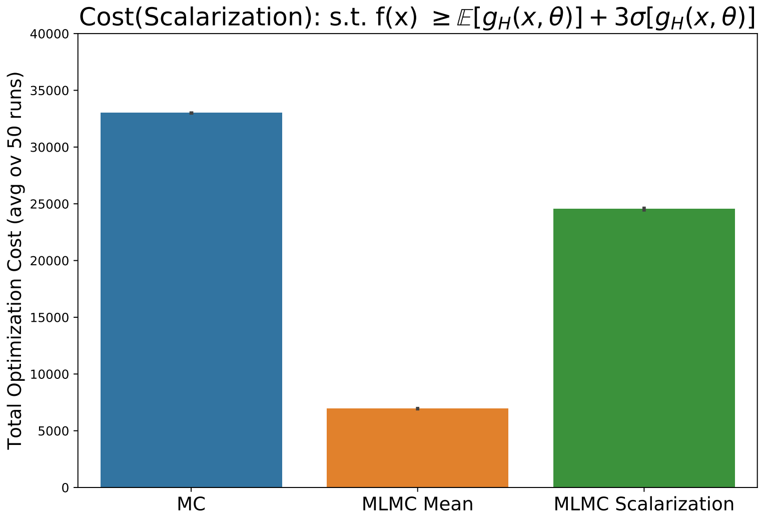
Mean:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)]$

Push back:

s.t. $f(x) \geq \mathbb{E}[g_H(x,\xi)] + 3\sigma[g_H(x,\xi)]$





- Lower cost for MLMC Mean
- High cost for MLMC Scalarization

- Start at $x = 5.75$
- Computational cost now lower than MC reference

Conclusion

- **NOWPAC** – Derivative-free trust region methods for constrained nonlinear optimization
- **SNOWPAC** – Stochastic derivative-free optimization using Gaussian process surrogates
- **DAKOTA** – Design Analysis Kit for Optimization and Terascale Applications

$\Rightarrow$ New MLMC estimators for Standard Deviation and Scalarization coupled with **SNOWPAC**.

Future work and open questions:

- Investigate bound for scalarization
- From MLMC to ACV

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

Multilevel Monte Carlo estimator: Standard deviation

Variance of variance:

$$\mathbb{V}[\widehat{\mu_2^{\text{ML}}}] = \sum_{\ell=0}^{L} \mathbb{V}[\widehat{\mu_2^{(\ell)}}] + \mathbb{V}[\widehat{\mu_{2,\ell}^{(\ell-1)}}] - 2\mathbb{C}\text{ov}[\widehat{\mu_2^{(\ell)}}, \widehat{\mu_{2,\ell}^{(\ell-1)}}]$$

Multilevel Monte Carlo estimator: Standard deviation

Variance of variance:

$$\mathbb{V}[\widehat{\mu_2^{\text{ML}}}] = \sum_{\ell=0}^{L} \mathbb{V}[\widehat{\mu_2^{(\ell)}}] + \mathbb{V}[\widehat{\mu_{2,\ell}^{(\ell-1)}}] - 2\mathbb{C}\text{ov}[\widehat{\mu_2^{(\ell)}}, \widehat{\mu_{2,\ell}^{(\ell-1)}}]$$

- $\mathbb{V}[\widehat{\mu_2}] = \frac{(N-1)}{N^2-2N+3}\left(\widehat{\mu_4} - \frac{N-3}{N-1}\widehat{\mu_2}^2\right)$

$$\text{where } \widehat{\mu_4} = \frac{1}{(N^2-3N+3) - \frac{(6N-9)(N^2-N)}{N(N^2-2N+3)}}\left(\frac{N^3}{N-1}\widehat{\mu_{4,\text{biased}}} - \frac{(6N-9)(N^2-N)}{N^2-2N+3}\widehat{\mu_2}^2\right)$$

- $\mathbb{C}\text{ov}[\widehat{\mu_2^{(\ell)}}, \widehat{\mu_{2,\ell}^{(\ell-1)}}] = \frac{1}{N_\ell}\mathbb{E}[\widehat{\mu_2^{(\ell)}}\widehat{\mu_{2,\ell}^{(\ell-1)}}] + \frac{1}{N_\ell(N_\ell-1)}(\widehat{\mu_1^{(\ell)}}[Q^{(\ell)}Q^{(\ell-1)}] - \widehat{\mu_1^{(\ell)}}\widehat{\mu_{1,\ell}^{(\ell-1)}})^2$

$$\begin{aligned}\text{where } \mathbb{E}[\widehat{\mu_2^{(\ell)}}\widehat{\mu_{2,\ell}^{(\ell-1)}}] = {} & \widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)2}Q^{(\ell-1)2}\right] - 2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)2}Q^{(\ell-1)}\right]\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)}\right] \\ & + 2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)}\right]^2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)2}\right] - 2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}\right]\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}Q^{(\ell-1)2}\right] \\ & + 4\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)}\right]\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}\right]\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}Q^{(\ell-1)}\right] + 2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}\right]^2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)2}\right] \\ & - 4\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)}\right]^2\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)}\right]^2 - \widehat{\mu_1^{(\ell)}}\left[Q^{(\ell)2}\right]\widehat{\mu_1^{(\ell)}}\left[Q^{(\ell-1)2}\right].\end{aligned}$$