MLMC for the computation of CVaR and its sensitivities in PDE-constrained risk-averse optimization

Fabio Nobile

CSQI - Institute of Mathematics, EPFL, Switzerland

Joint work with: Sundar Ganesh (EPFL/Deutch Bank),

Acknowlegments: Sebastian Krumscheid (KIT), Michele Pisaroni (Credit Suisse)

14 International Conference on Monte Carlo Methods and Applications
Paris, 26-30 June 2023

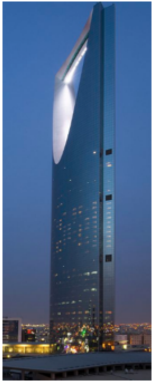
ExaQUte
Exascale Quantification of Uncertainties for Technology and Science Simulation

H2020 grant 800898: *EXAscale Quantification of Uncertainties for Technology and Science Simulation*

EPFL

Outline

Uncertainty quantification and robust design in civil engineering design



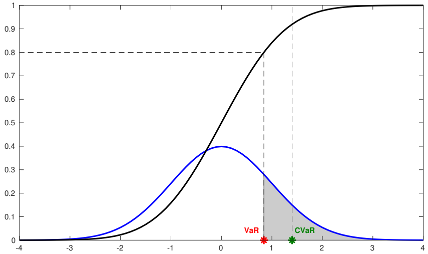
- Uncertainty Quantification of wind load on tall buildings
- Shape optimization under uncertainty

ExaQUte
Exascale Quantification of Uncertainties for Technology and Science Simulation

Outline

CVaR as risk measure

Given a random variable Q with cumulative distribution function F_Q



- τ-quantile (VaR):

$$q_\tau = \inf\{\theta : F_Q(\theta) \geq \tau\}$$

- τ-Conditional Value at Risk (CVaR)

$$\zeta_\tau = \frac{1}{1-\tau}\int_{q_\tau}^{\infty} x\, dF_Q(x) = \mathbb{E}[Q \mid Q \geq q_\tau]$$

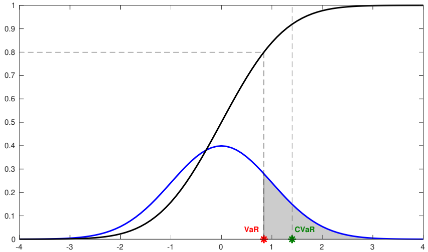
Define the function $\quad \Phi(\theta) = \mathbb{E}[\phi(\theta, Q)], \qquad \phi(\theta, Q) = \theta + \frac{1}{1-\tau}(Q-\theta)_+$

If Q has no atoms at q_τ then it holds ([Rockafellar-Uryasev 2000])

$$q_\tau = \underset{\theta\in\mathbb{R}}{\operatorname{argmin}}\,\Phi(\theta) \qquad \zeta_\tau = \min_{\theta\in\mathbb{R}} \Phi(\theta)$$

CVaR as risk measure

Given a random variable Q with cumulative distribution function F_Q



- τ-quantile (VaR):

$$q_\tau = \inf\{\theta : F_Q(\theta) \geq \tau\}$$

- τ-Conditional Value at Risk (CVaR)

$$\zeta_\tau = \frac{1}{1-\tau}\int_{q_\tau}^{\infty} x \, dF_Q(x) = \mathbb{E}[Q \mid Q \geq q_\tau]$$

Define the function $\quad \Phi(\theta) = \mathbb{E}[\phi(\theta, Q)], \qquad \phi(\theta, Q) = \theta + \frac{1}{1-\tau}(Q-\theta)_+$

If Q has no atoms at q_τ then it holds ([Rockafellar-Uryasev 2000])

$$q_\tau = \operatorname*{argmin}_{\theta\in\mathbb{R}} \Phi(\theta) \qquad \zeta_\tau = \min_{\theta\in\mathbb{R}} \Phi(\theta)$$

CVaR optimization – Problem formulation

$$
\begin{aligned}
(\Omega, \mathcal{F}, \mathbb{P}) &\rightarrow \text{Complete probability space} \\
z \in \mathbb{R}^d &\rightarrow \text{Design variables} \\
\omega \in \Omega &\rightarrow \text{Random elementary event} \\
u(z,\omega) &\rightarrow \text{Solution of underlying PDE} \\
Q(z,\omega) = \tilde{Q}(u(z,\omega)) &\rightarrow \text{Output functional to be optimized} \\
\Phi(\theta; z) = \theta + \frac{\mathbb{E}[(Q(z,\cdot)-\theta)_+]}{1-\tau} &\rightarrow \text{Parametric expectation. } \tau\text{-CVaR: } \zeta_\tau(z) = \min_\theta \Phi(\theta; z)
\end{aligned}
$$

Risk-averse optimization problem: find deterministic design z which minimizes τ-CVaR of $Q(z,\omega)$

$$
\begin{aligned}
\mathcal{J}^* &= \min_{z\in\mathbb{R}^d} \left(\min_{\theta\in\mathbb{R}} \Phi(\theta; z) \right) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2 \\
&= \min_{\substack{z\in\mathbb{R}^d \\ \theta\in\mathbb{R}}} \mathcal{J}(\theta, z), \qquad \mathcal{J}(\theta, z) := \Phi(\theta; z) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2
\end{aligned}
$$

Interested in gradient based algorithms. What can be said about $\mathcal{J}$?

CVaR optimization – Problem formulation

$$
\begin{aligned}
(\Omega, \mathcal{F}, \mathbb{P}) \quad &\rightarrow \quad \text{Complete probability space} \\
z \in \mathbb{R}^d \quad &\rightarrow \quad \text{Design variables} \\
\omega \in \Omega \quad &\rightarrow \quad \text{Random elementary event} \\
u(z, \omega) \quad &\rightarrow \quad \text{Solution of underlying PDE} \\
Q(z, \omega) = \tilde{Q}(u(z, \omega)) \quad &\rightarrow \quad \text{Output functional to be optimized} \\
\Phi(\theta; z) = \theta + \frac{\mathbb{E}[(Q(z, \cdot) - \theta)_+]}{1 - \tau} \quad &\rightarrow \quad \text{Parametric expectation. } \tau\text{-CVaR: } \zeta_\tau(z) = \min_\theta \Phi(\theta; z)
\end{aligned}
$$

Risk-averse optimization problem: find deterministic design z which minimizes τ-CVaR of $Q(z, \omega)$

$$
\begin{aligned}
\mathcal{J}^* &= \min_{z \in \mathbb{R}^d} \left(\min_{\theta \in \mathbb{R}} \Phi(\theta; z) \right) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2 \\
&= \min_{\substack{z \in \mathbb{R}^d \\ \theta \in \mathbb{R}}} \mathcal{J}(\theta, z), \qquad \mathcal{J}(\theta, z) := \Phi(\theta; z) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2
\end{aligned}
$$

Interested in gradient based algorithms. What can be said about $\mathcal{J}$?

CVaR optimization – Problem formulation

$$
\begin{aligned}
(\Omega, \mathcal{F}, \mathbb{P}) &\rightarrow \text{Complete probability space} \\
z \in \mathbb{R}^d &\rightarrow \text{Design variables} \\
\omega \in \Omega &\rightarrow \text{Random elementary event} \\
u(z,\omega) &\rightarrow \text{Solution of underlying PDE} \\
Q(z,\omega) = \tilde{Q}(u(z,\omega)) &\rightarrow \text{Output functional to be optimized} \\
\Phi(\theta; z) = \theta + \frac{\mathbb{E}[(Q(z,\cdot)-\theta)_+]}{1-\tau} &\rightarrow \text{Parametric expectation. } \tau\text{-CVaR: } \zeta_\tau(z) = \min_\theta \Phi(\theta; z)
\end{aligned}
$$

Risk-averse optimization problem: find deterministic design z which minimizes τ-CVaR of $Q(z,\omega)$

$$
\begin{aligned}
\mathcal{J}^* &= \min_{z\in\mathbb{R}^d} \left(\min_{\theta\in\mathbb{R}} \Phi(\theta; z) \right) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2 \\
&= \min_{\substack{z\in\mathbb{R}^d \\ \theta\in\mathbb{R}}} \mathcal{J}(\theta, z), \qquad \mathcal{J}(\theta, z) := \Phi(\theta; z) + \kappa \left\| z - z_{ref} \right\|_{l^2}^2
\end{aligned}
$$

Interested in gradient based algorithms. What can be said about $\mathcal{J}$?

Problem formulation - Sensitivities

Proposition [Ganesh-N. 2022]

Under mild assumptions on $z \mapsto Q(z,\cdot)$ ($Q(z) \in L^p(\Omega)$ for some $p \geq 1$ and is differentiable in z; $Q(z)$ has a Lebesque density for all z) $\mathcal{J}$ is Fréchet differentiable, with partial derivatives given by:

$$\mathcal{J}_\theta(\theta, z) = 1 - \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta}\right]}{1-\tau},$$

$$\mathcal{J}_{z^k}(\theta, z) = \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta} Q_{z^k}(z,\cdot)\right]}{1-\tau} + 2\kappa(z^k - z^k_{ref}), \quad k = 1, \dots, d$$

Use MLMC to estimate the expectations directly? Several issues...

- Variance of $\mathbb{1}_{Q_\ell \geq \theta} - \mathbb{1}_{Q_{\ell-1}\geq\theta}$ decays at reduced rate compared to $Q_\ell - Q_{\ell-1}$
- Accurate estimation requires impractically large number of samples

Possible remedies: [Giles-Nagapetyan-Ritter 2015, 2017]: smoothing: $\mathrm{CDF}_\epsilon(\theta) = \mathbb{E}[\rho_\epsilon * \mathbb{1}_{\{Q\leq\theta\}}]$; [Bayer-BenHammouda-Tempone, 2020]: numerical smoothing by integrating out one variable; [HajiAli-Spence-Teckentrup, 2023] for single θ: adaptive computation of $\mathbb{1}_{\{Q_\ell\leq\theta\}} - \mathbb{1}_{\{Q_{\ell-1}\leq\theta\}}$; accuracy increased when close to threshold; [Elferson-Hellman-Malqvist, 2016] for a single θ: similar as above using almost sure error bounds.

Problem formulation - Sensitivities

Proposition [Ganesh-N. 2022]

Under mild assumptions on $z \mapsto Q(z,\cdot)$ ($Q(z) \in L^p(\Omega)$ for some $p \geq 1$ and is differentiable in z; $Q(z)$ has a Lebesque density for all z) $\mathcal{J}$ is Fréchet differentiable, with partial derivatives given by:

$$\mathcal{J}_\theta(\theta, z) = 1 - \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta}\right]}{1-\tau},$$

$$\mathcal{J}_{z^k}(\theta, z) = \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta} Q_{z^k}(z,\cdot)\right]}{1-\tau} + 2\kappa(z^k - z^k_{ref}), \quad k = 1, \dots, d$$

Use MLMC to estimate the expectations directly? Several issues...

- Variance of $\mathbb{1}_{Q_\ell \geq \theta} - \mathbb{1}_{Q_{\ell-1}\geq\theta}$ decays at reduced rate compared to $Q_\ell - Q_{\ell-1}$
- Accurate estimation requires impractically large number of samples

Possible remedies: [Giles-Nagapetyan-Ritter 2015, 2017]: smoothing: $\mathrm{CDF}_\epsilon(\theta) = \mathbb{E}[\rho_\epsilon * \mathbb{1}_{\{Q\leq\theta\}}]$; [Bayer-BenHammouda-Tempone, 2020]: numerical smoothing by integrating out one variable; [HajiAli-Spence-Teckentrup, 2023] for single θ: adaptive computation of $\mathbb{1}_{\{Q_\ell\leq\theta\}} - \mathbb{1}_{\{Q_{\ell-1}\leq\theta\}}$; accuracy increased when close to threshold; [Elferson-Hellman-Malqvist, 2016] for a single θ: similar as above using almost sure error bounds.

Problem formulation - Sensitivities

Proposition [Ganesh-N. 2022]

Under mild assumptions on $z \mapsto Q(z,\cdot)$ ($Q(z) \in L^p(\Omega)$ for some $p \geq 1$ and is differentiable in z; $Q(z)$ has a Lebesque density for all z) $\mathcal{J}$ is Fréchet differentiable, with partial derivatives given by:

$$\mathcal{J}_\theta(\theta, z) = 1 - \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta}\right]}{1-\tau},$$

$$\mathcal{J}_{z^k}(\theta, z) = \frac{\mathbb{E}\left[\mathbb{1}_{Q(z,\cdot)\geq\theta} Q_{z^k}(z,\cdot)\right]}{1-\tau} + 2\kappa(z^k - z^k_{ref}), \quad k = 1, \dots, d$$

Use MLMC to estimate the expectations directly? Several issues...

- Variance of $\mathbb{1}_{Q_\ell \geq \theta} - \mathbb{1}_{Q_{\ell-1}\geq\theta}$ decays at reduced rate compared to $Q_\ell - Q_{\ell-1}$
- Accurate estimation requires impractically large number of samples

Possible remedies: [Giles-Nagapetyan-Ritter 2015, 2017]: smoothing: $\mathrm{CDF}_\epsilon(\theta) = \mathbb{E}[\rho_\epsilon * \mathbb{1}_{\{Q\leq\theta\}}]$; [Bayer-BenHammouda-Tempone, 2020]: numerical smoothing by integrating out one variable; [HajiAli-Spence-Teckentrup, 2023] for single θ: adaptive computation of $\mathbb{1}_{\{Q_\ell\leq\theta\}} - \mathbb{1}_{\{Q_{\ell-1}\leq\theta\}}$; accuracy increased when close to threshold; [Elferson-Hellman-Malqvist, 2016] for a single θ: similar as above using almost sure error bounds.

Outline

Problem formulation - Sensitivities using parametric expectations

Idea: rewrite sensitivities as θ-derivatives of parametric expectations

$$\mathcal{J}_\theta(\theta, z) = \partial_\theta \Phi(\theta; z), \qquad \Phi(\theta; z) = \mathbb{E}\left[\phi(\theta, Q(z, \cdot))\right] \qquad \phi(\theta, Q) = \theta + \frac{(Q-\theta)_+}{1-\tau}$$

$$\mathcal{J}_{z^k}(\theta, z) = \partial_\theta \Psi_k(\theta; z) + 2\kappa(z^k - z^k_{ref}), \quad \Psi_k(\theta; z) = \mathbb{E}[\psi(\theta, Q, Q_{z^k})], \quad \psi(\theta, Q, Q_{z^k}) = -\frac{(Q-\theta)_+ Q_{z^k}}{1-\tau}$$

Proposed procedure: Let $f(\theta; z)$ be either $\Phi(\theta; z)$ or $\Psi_k(\theta; z)$.

- estimate $F(\theta_i; z)$ by MLMC on a grid of points $\vec{\theta} = \{\theta_1, ..., \theta_n\} \in \Theta \subset \mathbb{R}$
- interpolate the obtained values e.g. with cubic splines
- differentiate numerically to estimate $\mathcal{J}_\theta(\theta, z)$ and $\mathcal{J}_z(\theta, z)$

Advantages:

- MLMC applied on ϕ, ψ, whcih are Lipschitz continuous in Q
- Direct minimization in θ possible for given z

Problem formulation - Sensitivities using parametric expectations

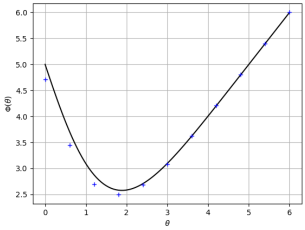
Idea: rewrite sensitivities as θ-derivatives of parametric expectations

$$\mathcal{J}_\theta(\theta, z) = \partial_\theta \Phi(\theta; z), \qquad \Phi(\theta; z) = \mathbb{E}\left[\phi(\theta, Q(z, \cdot))\right] \qquad \phi(\theta, Q) = \theta + \frac{(Q-\theta)_+}{1-\tau}$$

$$\mathcal{J}_{z^k}(\theta, z) = \partial_\theta \Psi_k(\theta; z) + 2\kappa(z^k - z^k_{ref}), \quad \Psi_k(\theta; z) = \mathbb{E}[\psi(\theta, Q, Q_{z^k})], \quad \psi(\theta, Q, Q_{z^k}) = -\frac{(Q-\theta)_+ Q_{z^k}}{1-\tau}$$

Proposed procedure: Let $f(\theta; z)$ be either $\Phi(\theta; z)$ or $\Psi_k(\theta; z)$.

- estimate $F(\theta_i; z)$ by MLMC on a grid of points $\vec{\theta} \equiv \{\theta_1, \ldots, \theta_n\} \in \Theta \subset \mathbb{R}$
- interpolate the obtained values e.g. with cubic splines
- differentiate numerically to estimate $\mathcal{J}_\theta(\theta, z)$ and $\mathcal{J}_z(\theta, z)$



Advantages:

- MLMC applied on ϕ, ψ, whcih are Lipschitz continuous in Q
- Direct minimization in θ possible for given z

Problem formulation - Sensitivities using parametric expectations

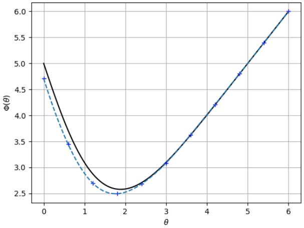
Idea: rewrite sensitivities as θ-derivatives of parametric expectations

$$\mathcal{J}_\theta(\theta,z) = \partial_\theta \Phi(\theta;z), \qquad \Phi(\theta;z) = \mathbb{E}\left[\phi(\theta, Q(z,\cdot))\right] \qquad \phi(\theta,Q) = \theta + \frac{(Q-\theta)_+}{1-\tau}$$

$$\mathcal{J}_{z^k}(\theta,z) = \partial_\theta \Psi_k(\theta;z) + 2\kappa(z^k - z^k_{ref}), \quad \Psi_k(\theta;z) = \mathbb{E}[\psi(\theta,Q,Q_{z^k})], \quad \psi(\theta,Q,Q_{z^k}) = -\frac{(Q-\theta)_+ Q_{z^k}}{1-\tau}$$

Proposed procedure: Let $f(\theta;z)$ be either $\Phi(\theta;z)$ or $\Psi_k(\theta;z)$.

- estimate $F(\theta_i;z)$ by MLMC on a grid of points $\vec{\theta} \equiv \{\theta_1, \dots, \theta_n\} \in \Theta \subset \mathbb{R}$
- interpolate the obtained values e.g. with cubic splines
- differentiate numerically to estimate $\mathcal{J}_\theta(\theta,z)$ and $\mathcal{J}_z(\theta,z)$



Advantages:

- MLMC applied on ϕ, ψ, whcih are Lipschitz continuous in Q
- Direct minimization in θ possible for given z

Problem formulation - Sensitivities using parametric expectations

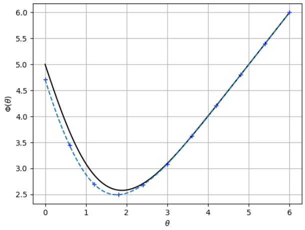
Idea: rewrite sensitivities as θ-derivatives of parametric expectations

$$\mathcal{J}_\theta(\theta,z) = \partial_\theta \Phi(\theta;z), \qquad \Phi(\theta;z) = \mathbb{E}\left[\phi(\theta, Q(z,\cdot))\right] \qquad \phi(\theta,Q) = \theta + \frac{(Q-\theta)_+}{1-\tau}$$

$$\mathcal{J}_{z^k}(\theta,z) = \partial_\theta \Psi_k(\theta;z) + 2\kappa(z^k - z^k_{ref}), \quad \Psi_k(\theta;z) = \mathbb{E}[\psi(\theta,Q,Q_{z^k})], \quad \psi(\theta,Q,Q_{z^k}) = -\frac{(Q-\theta)_+ Q_{z^k}}{1-\tau}$$

Proposed procedure: Let $f(\theta;z)$ be either $\Phi(\theta;z)$ or $\Psi_k(\theta;z)$.

- estimate $F(\theta_i;z)$ by MLMC on a grid of points $\vec{\theta} \equiv \{\theta_1,\dots,\theta_n\} \in \Theta \subset \mathbb{R}$
- interpolate the obtained values e.g. with cubic splines
- differentiate numerically to estimate $\mathcal{J}_\theta(\theta,z)$ and $\mathcal{J}_z(\theta,z)$



Advantages:

- MLMC applied on ϕ,ψ, whcih are Lipschitz continuous in Q
- Direct minimization in θ possible for given z

MLMC for parametric expectations

- Let $Q_0, Q_1, \ldots, Q_L$ be a sequence of approximations of Q with increasing accuracy and cost.
- Let $f(\theta; Q_\ell(z))$ be either $\phi(\theta, Q_\ell(z))$ or $\psi(\theta, Q_\ell(z), \partial_{z^k} Q_\ell(z))$
- Pointwise MLMC estimator (same samples for every θ_i)

$$\hat{F}_L^{MLMC}(\theta_i; z) = \frac{1}{N_0}\sum_{k=1}^{N_0} f(\theta_i, Q_0^{(k,0)}(z)) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\left[f(\theta_i, Q_\ell^{(k,\ell)}(z)) - f(\theta_i, Q_{\ell-1}^{(k,\ell)}(z))\right]$$

with $(Q_\ell(z)^{(k,\ell)}, Q_{\ell-1}^{(k,\ell)}(z)) \overset{iid}{\sim} (Q_\ell(z), Q_{\ell-1}(z))$.

- MLMC estimator of parametric extectation: spline interpolant

$$\hat{F}_L = \mathcal{S}_n\left(\hat{F}_L^{MLMC}\right)$$

- Postprocess: compute $\hat{F}_L' = \dfrac{d\hat{F}_L}{d\theta} = \dfrac{d}{d\theta}\mathcal{S}_n\left(F_L^{MLMC}\right)$, $\quad \min_{\theta\in\Theta}\hat{F}_L$, $\quad$ etc.

MLMC for parametric expectations

- Let $Q_0, Q_1, \dots, Q_L$ be a sequence of approximations of Q with increasing accuracy and cost.
- Let $f(\theta; Q_\ell(z))$ be either $\phi(\theta, Q_\ell(z))$ or $\psi(\theta, Q_\ell(z), \partial_{z^k} Q_\ell(z))$
- Pointwise MLMC estimator (same samples for every θ_i)

$$\hat{F}_L^{MLMC}(\theta_i; z) = \frac{1}{N_0}\sum_{k=1}^{N_0} f(\theta_i, Q_0^{(k,0)}(z)) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\left[f(\theta_i, Q_\ell^{(k,\ell)}(z)) - f(\theta_i, Q_{\ell-1}^{(k,\ell)}(z))\right]$$

with $(Q_\ell(z)^{(k,\ell)}, Q_{\ell-1}^{(k,\ell)}(z)) \overset{iid}{\sim} (Q_\ell(z), Q_{\ell-1}(z))$.

- MLMC estimator of parametric extectation: spline interpolant

$$\hat{F}_L = \mathcal{S}_n\left(\hat{F}_L^{MLMC}\right)$$

- Postprocess: compute $\hat{F}_L' = \frac{d\hat{F}_L}{d\theta} = \frac{d}{d\theta}\mathcal{S}_n\left(F_L^{MLMC}\right)$, $\quad \min_{\theta\in\Theta}\hat{F}_L$, $\quad$ etc.

MLMC for parametric expectations

- Let $Q_0, Q_1, \ldots, Q_L$ be a sequence of approximations of Q with increasing accuracy and cost.
- Let $f(\theta; Q_\ell(z))$ be either $\phi(\theta, Q_\ell(z))$ or $\psi(\theta, Q_\ell(z), \partial_{z^k} Q_\ell(z))$
- Pointwise MLMC estimator (same samples for every θ_i)

$$\hat{F}_L^{MLMC}(\theta_i; z) = \frac{1}{N_0}\sum_{k=1}^{N_0} f(\theta_i, Q_0^{(k,0)}(z)) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\left[f(\theta_i, Q_\ell^{(k,\ell)}(z)) - f(\theta_i, Q_{\ell-1}^{(k,\ell)}(z))\right]$$

with $(Q_\ell(z)^{(k,\ell)}, Q_{\ell-1}^{(k,\ell)}(z)) \overset{iid}{\sim} (Q_\ell(z), Q_{\ell-1}(z))$.

- MLMC estimator of parametric extectation: spline interpolant

$$\hat{F}_L = \mathcal{S}_n\left(\hat{F}_L^{MLMC}\right)$$

- Postprocess: compute $\hat{F}_L' = \frac{d\hat{F}_L}{d\theta} = \frac{d}{d\theta}\mathcal{S}_n\left(F_L^{MLMC}\right)$, $\quad \min_{\theta\in\Theta}\hat{F}_L$, $\quad$ etc.

MLMC for parametric expectations

- Let $Q_0, Q_1, \ldots, Q_L$ be a sequence of approximations of Q with increasing accuracy and cost.
- Let $f(\theta; Q_\ell(z))$ be either $\phi(\theta, Q_\ell(z))$ or $\psi(\theta, Q_\ell(z), \partial_{z^k} Q_\ell(z))$
- Pointwise MLMC estimator (same samples for every θ_i)

$$\hat{F}_L^{MLMC}(\theta_i; z) = \frac{1}{N_0}\sum_{k=1}^{N_0} f(\theta_i, Q_0^{(k,0)}(z)) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\left[f(\theta_i, Q_\ell^{(k,\ell)}(z)) - f(\theta_i, Q_{\ell-1}^{(k,\ell)}(z))\right]$$

with $(Q_\ell(z)^{(k,\ell)}, Q_{\ell-1}^{(k,\ell)}(z)) \overset{iid}{\sim} (Q_\ell(z), Q_{\ell-1}(z))$.

- MLMC estimator of parametric extectation: spline interpolant

$$\hat{F}_L = \mathcal{S}_n\left(\hat{F}_L^{MLMC}\right)$$

- Postprocess: compute $\hat{F}_L' = \frac{d\hat{F}_L}{d\theta} = \frac{d}{d\theta}\mathcal{S}_n\left(F_L^{MLMC}\right)$, $\quad \min_{\theta\in\Theta} \hat{F}_L$, $\quad$ etc.

MLMC for parametric expectations

- Let $Q_0, Q_1, \ldots, Q_L$ be a sequence of approximations of Q with increasing accuracy and cost.
- Let $f(\theta; Q_\ell(z))$ be either $\phi(\theta, Q_\ell(z))$ or $\psi(\theta, Q_\ell(z), \partial_{z^k} Q_\ell(z))$
- Pointwise MLMC estimator (same samples for every θ_i)

$$\hat{F}_L^{MLMC}(\theta_i; z) = \frac{1}{N_0}\sum_{k=1}^{N_0} f(\theta_i, Q_0^{(k,0)}(z)) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\left[f(\theta_i, Q_\ell^{(k,\ell)}(z)) - f(\theta_i, Q_{\ell-1}^{(k,\ell)}(z))\right]$$

with $(Q_\ell(z)^{(k,\ell)}, Q_{\ell-1}^{(k,\ell)}(z)) \overset{iid}{\sim} (Q_\ell(z), Q_{\ell-1}(z))$.

- MLMC estimator of parametric extectation: spline interpolant

$$\hat{F}_L = \mathcal{S}_n\left(\hat{F}_L^{MLMC}\right)$$

- Postprocess: compute $\hat{F}_L' = \dfrac{d\hat{F}_L}{d\theta} = \dfrac{d}{d\theta}\mathcal{S}_n\left(F_L^{MLMC}\right)$, $\quad \min_{\theta\in\Theta}\hat{F}_L$, $\quad$ etc.

A posteriori error estimation

The MSE of $\hat{F}_L^{(m)}$, $m = 0, 1$ naturally splits as:

$$\mathrm{MSE}\left(\hat{F}_L^{(m)}\right) := \mathbb{E}\left[\left\|\hat{F}_L^{(m)} - F^{(m)}\right\|^2_{\mathrm{L}^\infty(\Theta)}\right] \leq \text{spline interp. error} \qquad (e_i^{(m)})^2 = C_{ie}\| F^{(m)} - \mathcal{S}_n^{(m)}(F)\|^2_{L^\infty}$$

$$+ \text{discr. error} \qquad (e_b^{(m)})^2 = C_{be}\|\mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}]\|^2_{L^\infty}$$

$$+ \text{Statistical error} \qquad (e_s^{(m)})^2 = C_{se}\mathbb{E}[\|\hat{F}^{(m)} - \mathbb{E}[\hat{F}_L^{(m)}]\|^2_{L^\infty}]$$

- How to estimate the three error contributions?
- In particular, how to split the statistical error $e_s^{(m)}$ over the levels for optimal MLMC tuning?

EPFL MATHICSE

A posteriori error estimation

The MSE of $\hat{F}_L^{(m)}$, $m = 0, 1$ naturally splits as:

$$\mathrm{MSE}\left(\hat{F}_L^{(m)}\right) := \mathbb{E}\left[\left\|\hat{F}_L^{(m)} - F^{(m)}\right\|^2_{L^\infty(\Theta)}\right] \leq \text{spline interp. error} \quad (e_i^{(m)})^2 = C_{ie}\|F^{(m)} - \mathcal{S}_n^{(m)}(F)\|^2_{L^\infty}$$

$$+ \text{discr. error} \quad (e_b^{(m)})^2 = C_{be}\|\mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}]\|^2_{L^\infty}$$

$$+ \text{Statistical error} \quad (e_s^{(m)})^2 = C_{se}\mathbb{E}[\|\hat{F}^{(m)} - \mathbb{E}[\hat{F}_L^{(m)}]\|^2_{L^\infty}]$$

- How to estimate the three error contributions?
- In particular, how to split the statistical error $e_s^{(m)}$ over the levels for optimal MLMC tuning?

Error estimator - Bias

Bias error $e_b^{(m)} \lesssim \left\| \mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}] \right\|_{\mathrm{L}^\infty} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f - f_L\right]\right) \right\|_{\mathrm{L}^\infty}$, with $f_L := f(\cdot, Q_L)$

In practice we estimate $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{\mathrm{L}^\infty}$

- Naive estimator: replace expectation with sample average $\rightsquigarrow$ performs poorly for $m \geq 1$: do not get the right decay
- Inverse inequality + sample average

$$\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{\mathrm{L}^\infty} \approx C(n,m) \max_j |\hat{\mathbb{E}}[f_L(\theta_j) - f_{L-1}(\theta_j)]|$$

$\rightsquigarrow$ gives right decay but constants are too large

- KDE smoothing: compute $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}^{kde}[f_L - f_{L-1}]\right) \right\|_{\mathrm{L}^\infty}$ where we smooth the empirical measure by kerned density estimation.

Empirical PDF

KDE-smoothened PDF

Error estimator - Bias

Bias error $e_b^{(m)} \lesssim \left\| \mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}] \right\|_{L^\infty} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f - f_L]) \right\|_{L^\infty}$, with $f_L := f(\cdot, Q_L)$

In practice we estimate $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f_L - f_{L-1}]) \right\|_{L^\infty}$

- Naive estimator: replace expectation with sample average $\rightsquigarrow$ performs poorly for $m \geq 1$: do not get the right decay
- Inverse inequality + sample average

$$\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f_L - f_{L-1}]) \right\|_{L^\infty} \approx C(n,m) \max_j |\hat{\mathbb{E}}[f_L(\theta_j) - f_{L-1}(\theta_j)]|$$

$\rightsquigarrow$ gives right decay but constants are too large

- KDE smoothing: compute $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}^{kde}[f_L - f_{L-1}]) \right\|_{L^\infty}$ where we smooth the empirical measure by kerned density estimation.

Empirical PDF

KDE-smoothened PDF

Error estimator - Bias

Bias error $e_b^{(m)} \lesssim \left\| \mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}] \right\|_{L^\infty} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f - f_L\right]\right) \right\|_{L^\infty}$, with $f_L := f(\cdot, Q_L)$

In practice we estimate $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{L^\infty}$

- Naive estimator: replace expectation with sample average $\rightsquigarrow$ performs poorly for $m \geq 1$: do not get the right decay
- Inverse inequality + sample average

$$\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{L^\infty} \approx C(n, m) \max_j \left| \hat{\mathbb{E}}[f_L(\theta_j) - f_{L-1}(\theta_j)] \right|$$

$\rightsquigarrow$ gives right decay but constants are too large

- KDE smoothing: compute $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}^{kde}[f_L - f_{L-1}]\right) \right\|_{L^\infty}$ where we smooth the empirical measure by kerned density estimation.

Empirical PDF

KDE-smoothened PDF

Error estimator - Bias

Bias error $e_b^{(m)} \lesssim \left\| \mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}] \right\|_{L^\infty} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f - f_L]) \right\|_{L^\infty}$, with $f_L := f(\cdot, Q_L)$

In practice we estimate $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f_L - f_{L-1}]) \right\|_{L^\infty}$

- Naive estimator: replace expectation with sample average $\rightsquigarrow$ performs poorly for $m \geq 1$: do not get the right decay
- Inverse inequality + sample average

$$\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}[f_L - f_{L-1}]) \right\|_{L^\infty} \approx C(n, m) \max_j |\hat{\mathbb{E}}[f_L(\theta_j) - f_{L-1}(\theta_j)]|$$

$\rightsquigarrow$ gives right decay but constants are too large

- KDE smoothing: compute $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}(\mathbb{E}^{kde}[f_L - f_{L-1}]) \right\|_{L^\infty}$ where we smooth the empirical measure by kerned density estimation.

Empirical PDF

KDE-smoothened PDF

Error estimator - Bias

Bias error $e_b^{(m)} \lesssim \left\| \mathcal{S}_n^{(m)}(F) - \mathbb{E}[\hat{F}_L^{(m)}] \right\|_{L^\infty} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f - f_L\right]\right) \right\|_{L^\infty}$, with $f_L := f(\cdot, Q_L)$

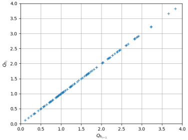
In practice we estimate $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{L^\infty}$

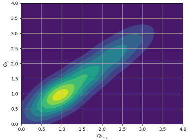
- Naive estimator: replace expectation with sample average $\rightsquigarrow$ performs poorly for $m \geq 1$: do not get the right decay
- Inverse inequality + sample average

$$\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}\left[f_L - f_{L-1}\right]\right) \right\|_{L^\infty} \approx C(n, m) \max_j \left| \hat{\mathbb{E}}[f_L(\theta_j) - f_{L-1}(\theta_j)] \right|$$

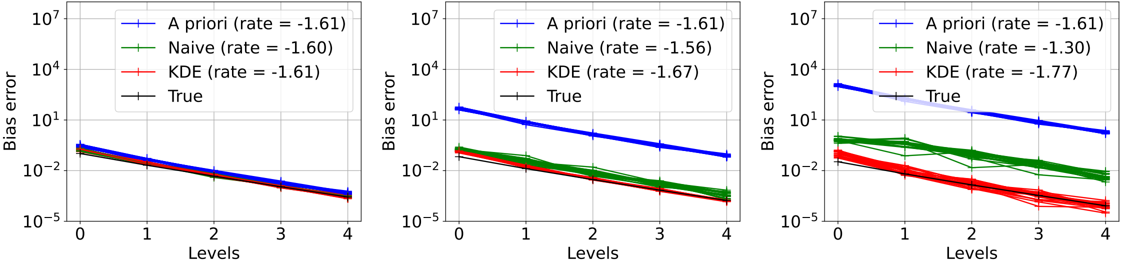
$\rightsquigarrow$ gives right decay but constants are too large

- KDE smoothing: compute $\hat{e}_b^{(m)} = \left\| \mathcal{S}_n^{(m)}\left(\mathbb{E}^{kde}[f_L - f_{L-1}]\right) \right\|_{L^\infty}$ where we smooth the empirical measure by kerned density estimation.





Error estimators - Bias estimator performance



$m = 0$ $\quad$ $m = 1$ $\quad$ $m = 2$

Comparison of bias error estimators with 100 samples per level.

Error estimator - Statistical error

Statistical error $(e_s^{(m)})^2 \lesssim \mathbb{E}[\|\hat{F}_L^{(m)} - \mathbb{E}[\hat{F}_L^{(m)}]\|_{L^\infty}^2] = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|_{L^\infty}^2]$

- Inverse inequality and ℓ^∞ bounds [Krumscheid-N., 2018]

$$(e_s^{(m)})^2 = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|_{L^\infty}^2] \leq C(n,m)\mathbb{E}[\max_j |\hat{F}_L^{MLMC}(\theta_j) - \mathbb{E}[\hat{F}_L^{MLMC}(\theta_j)]|^2]$$

$$\leq C(n,m)\log(n)\sum_{\ell=0}^{L}\frac{1}{N_\ell}\mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$$

$V_\ell = \mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$ can be estimated easily by sample averages and decay at same rate as $\mathbb{E}[(Q_\ell - Q_{\ell-1})^2]$. However, the constants are too large !

Error estimator - Statistical error

Statistical error $(e_s^{(m)})^2 \lesssim \mathbb{E}[\|\hat{F}_L^{(m)} - \mathbb{E}[\hat{F}_L^{(m)}]\|^2_{L^\infty}] = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|^2_{L^\infty}]$

- Inverse inequality and ℓ^∞ bounds [Krumscheid-N., 2018]

$$(e_s^{(m)})^2 = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|^2_{L^\infty}] \leq C(n,m)\mathbb{E}[\max_j |\hat{F}_L^{MLMC}(\theta_j) - \mathbb{E}[\hat{F}_L^{MLMC}(\theta_j)]|^2]$$

$$\leq C(n,m)\log(n)\sum_{\ell=0}^{L}\frac{1}{N_\ell}\mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$$

$V_\ell = \mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$ can be estimated easily by sample averages and decay at same rate as $\mathbb{E}[(Q_\ell - Q_{\ell-1})^2]$. However, the constants are too large !

Error estimator - Statistical error

Statistical error $(e_s^{(m)})^2 \lesssim \mathbb{E}[\|\hat{F}_L^{(m)} - \mathbb{E}[\hat{F}_L^{(m)}]\|_{L^\infty}^2] = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|_{L^\infty}^2]$

- Inverse inequality and ℓ^∞ bounds [Krumscheid-N., 2018]

$$(e_s^{(m)})^2 = \mathbb{E}[\|\mathcal{S}_n^{(m)}\left(\hat{F}_L^{MLMC} - \mathbb{E}[\hat{F}_L^{MLMC}]\right)\|_{L^\infty}^2] \leq C(n,m)\mathbb{E}[\max_j |\hat{F}_L^{MLMC}(\theta_j) - \mathbb{E}[\hat{F}_L^{MLMC}(\theta_j)]|^2]$$
$$\leq C(n,m)\log(n)\sum_{\ell=0}^{L}\frac{1}{N_\ell}\mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$$

$V_\ell = \mathbb{E}[\max_j |f_\ell(\theta_j) - f_{\ell-1}(\theta_j)|^2]$ can be estimated easily by sample averages and decay at same rate as $\mathbb{E}[(Q_\ell - Q_{\ell-1})^2]$. However, the constants are too large !

Error estimator - Statistical error

- Alternative: estimate $(e_s^{(m)})^2$ by bootstrap (resample each level of the MLMC estimator).

$$(e_s^{(m)})^2 \approx (\hat{e}_{bs}^{(m)})^2 := \frac{1}{N_{bs}} \sum_{p=1}^{N_{bs}} \|\mathcal{S}_n^{(m)} \left(\hat{F}_L^{MLMC,p} - \overline{\hat{F}_L^{MLMC}} \right) \|_{L^\infty}^2 .$$

But how to localize over levels? **Idea**: still use V_ℓ as local indicators, rescaled so that the total error matches the bootstrap estimator:

$$(\hat{e}_s^{(m)})^2 := \sum_{\ell=0}^{L} \frac{\tilde{V}_\ell}{N_\ell}, \qquad \tilde{V}_\ell = \frac{(\hat{e}_{bs}^{(m)})^2}{\sum_{k=0}^{L} V_k / N_k} V_\ell$$

$m = 0$ $\quad m = 1$ $\quad m = 2$

Comparison of statistical error estimators for $N_l = N_0 \times 2^{-l}$, $L = 5$.

Error estimator - Statistical error

- Alternative: estimate $(e_s^{(m)})^2$ by bootstrap (resample each level of the MLMC estimator).

$$(e_s^{(m)})^2 \approx (\hat{e}_{bs}^{(m)})^2 := \frac{1}{N_{bs}} \sum_{p=1}^{N_{bs}} \|\mathcal{S}_n^{(m)} \left(\hat{F}_L^{MLMC,p} - \overline{\hat{F}_L^{MLMC}} \right) \|_{\mathrm{L}^\infty}^2 .$$

But how to localize over levels? Idea: still use V_ℓ as local indicators, rescaled so that the total error matches the bootstrap estimator:

$$(\hat{e}_s^{(m)})^2 := \sum_{\ell=0}^{L} \frac{\tilde{V}_\ell}{N_\ell}, \qquad \tilde{V}_\ell = \frac{(\hat{e}_{bs}^{(m)})^2}{\sum_{k=0}^{L} V_k/N_k} V_\ell$$

$m = 0$ $m = 1$ $m = 2$

Comparison of statistical error estimators for $N_l = N_0 \times 2^{-l}$, $L = 5$.

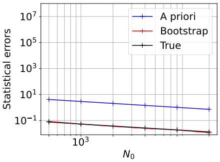
Error estimator - Statistical error

- Alternative: estimate $(e_s^{(m)})^2$ by bootstrap (resample each level of the MLMC estimator).

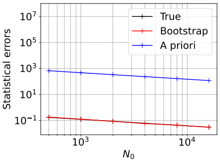
$$(e_s^{(m)})^2 \approx (\hat{e}_{bs}^{(m)})^2 := \frac{1}{N_{bs}} \sum_{p=1}^{N_{bs}} \| \mathcal{S}_n^{(m)} \left(\hat{F}_L^{MLMC,p} - \overline{\hat{F}_L^{MLMC}} \right) \|_{L^\infty}^2 .$$

But how to localize over levels? **Idea**: still use V_ℓ as local indicators, rescaled so that the total error matches the bootstrap estimator:

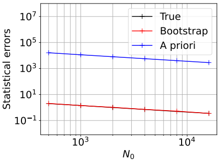
$$(\hat{e}_s^{(m)})^2 := \sum_{\ell=0}^{L} \frac{\tilde{V}_\ell}{N_\ell}, \qquad \tilde{V}_\ell = \frac{(\hat{e}_{bs}^{(m)})^2}{\sum_{k=0}^{L} V_k / N_k} V_\ell$$



$m = 0$



$m = 1$



$m = 2$

Comparison of statistical error estimators for $N_l = N_0 \times 2^{-l}$, $L = 5$.

Outline

Alternating Minimization Gradient Descent (AMGD) algorithm

$$
\begin{aligned}
j &\rightarrow \text{Optimization iteration} \\
\hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}(\theta, z) \text{ for fized } z \\
\hat{\mathcal{J}}_\theta^{(j)} = \partial_\theta \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_\theta(\theta, z) \text{ for fixed } z \\
\tilde{\mathcal{J}}_z^{(j)} \neq \partial_z \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_z(\theta, z) \text{ for fized } z \\
\Theta \subset \mathbb{R} &\rightarrow \text{Subset on which } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)} \text{ are constructed}
\end{aligned}
$$

Alternating Minimization Gradient Descent (AMGD) algorithm:

$$\theta_j \in \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{\mathcal{J}}^{(j)}(\theta, z_j) \qquad \text{(need MLMC estimator accurate for } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)})$$

$$z_{j+1} = z_j - \alpha \tilde{\mathcal{J}}_z^{(j)}(\theta_j, z_j) \qquad \text{(no further MLMC estimation)}$$

Do the iterates converge to (θ^*, z^*)? How fast in j?

Alternating Minimization Gradient Descent (AMGD) algorithm

$$
\begin{aligned}
j &\rightarrow \text{Optimization iteration} \\
\hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}(\theta, z) \text{ for fized } z \\
\hat{\mathcal{J}}_\theta^{(j)} = \partial_\theta \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_\theta(\theta, z) \text{ for fixed } z \\
\tilde{\mathcal{J}}_z^{(j)} \neq \partial_z \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_z(\theta, z) \text{ for fized } z \\
\Theta \subset \mathbb{R} &\rightarrow \text{Subset on which } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)} \text{ are constructed}
\end{aligned}
$$

Alternating Minimization Gradient Descent (AMGD) algorithm:

$$\theta_j \in \operatorname*{argmin}_{\theta \in \Theta} \hat{\mathcal{J}}^{(j)}(\theta, z_j) \qquad \text{(need MLMC estimator accurate for } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)})$$

$$z_{j+1} = z_j - \alpha \tilde{\mathcal{J}}_z^{(j)}(\theta_j, z_j) \qquad \text{(no further MLMC estimation)}$$

Do the iterates converge to (θ^*, z^*)? How fast in j?

Alternating Minimization Gradient Descent (AMGD) algorithm

$$\begin{aligned}
j &\rightarrow \text{Optimization iteration} \\
\hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}(\theta, z) \text{ for fized } z \\
\hat{\mathcal{J}}_\theta^{(j)} = \partial_\theta \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_\theta(\theta, z) \text{ for fixed } z \\
\tilde{\mathcal{J}}_z^{(j)} \neq \partial_z \hat{\mathcal{J}}^{(j)} &\rightarrow \text{Approximation to } \theta \mapsto \mathcal{J}_z(\theta, z) \text{ for fized } z \\
\Theta \subset \mathbb{R} &\rightarrow \text{Subset on which } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)} \text{ are constructed}
\end{aligned}$$

Alternating Minimization Gradient Descent (AMGD) algorithm:

$$\theta_j \in \operatorname*{argmin}_{\theta \in \Theta} \hat{\mathcal{J}}^{(j)}(\theta, z_j) \qquad \text{(need MLMC estimator accurate for } \hat{\mathcal{J}}^{(j)}, \hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)})$$

$$z_{j+1} = z_j - \alpha \tilde{\mathcal{J}}_z^{(j)}(\theta_j, z_j) \qquad \text{(no further MLMC estimation)}$$

Do the iterates converge to (θ^*, z^*)? How fast in j?

Convergence of AMGD [Ganesh-N. 2023]

- Notation: $w = (\theta, z) \in \mathbb{R} \times \mathbb{R}^d$ and $\hat{\mathcal{J}}_w^{(j)} = (\hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)})$.
- Assume $\mathcal{J}$ strongly convex and with Lipschitz continuous gradients.
- Let $\hat{\mathcal{J}}^{(j)} \in \Theta \times \mathbb{R}^d$ satisfy the following accuracy condition for some $\eta > 0$:

$$\mathrm{MSE}\left(\hat{\mathcal{J}}_w^{(j)}(\cdot, z_j)\right) := \mathbb{E}_j\left[\left\|\hat{\mathcal{J}}_\theta^{(j)}(\cdot, z_j) - \mathcal{J}_\theta(\cdot, z_j)\right\|^2_{L^\infty(\Theta)}\right] + \sum_{k=1}^{d} \mathbb{E}_j\left[\left\|\tilde{\mathcal{J}}_{z^k}^{(j)}(\cdot, z_j) - \mathcal{J}_{z^k}(\cdot, z_j)\right\|^2_{L^\infty(\Theta)}\right] \leq \eta^2 \left\|\mathcal{J}_w(\theta_{j-1}, z_j)\right\|^2_{l^2},$$

where $\mathbb{E}_j$ is expectation conditional on $\{z_0, \theta_0, z_1, \theta_1, \dots, \theta_{j-1}, z_j\}$

Then:

$$\mathbb{E}\left[\|z_{j+1} - z^*\|^2_{l^2} + C_1(\theta_j - \theta^*)^2\right] \leq \xi \mathbb{E}\left[\|z_j - z^*\|^2_{l^2} + C_1(\theta_{j-1} - \theta^*)^2\right]$$

for some constants $C_1 > 0$ and $0 < \xi < 1$.

Based on analysis from [Beiser-Keith-Urbainczyk-Wohlmuth, 2020]

Convergence of AMGD [Ganesh-N. 2023]

- Notation: $w = (\theta, z) \in \mathbb{R} \times \mathbb{R}^d$ and $\hat{\mathcal{J}}_w^{(j)} = (\hat{\mathcal{J}}_\theta^{(j)}, \tilde{\mathcal{J}}_z^{(j)})$.
- Assume $\mathcal{J}$ strongly convex and with Lipschitz continuous gradients.
- Let $\hat{\mathcal{J}}^{(j)} \in \Theta \times \mathbb{R}^d$ satisfy the following accuracy condition for some $\eta > 0$:

$$\mathrm{MSE}\left(\hat{\mathcal{J}}_w^{(j)}(\cdot, z_j)\right) := \mathbb{E}_j\left[\left\|\hat{\mathcal{J}}_\theta^{(j)}(\cdot, z_j) - \mathcal{J}_\theta(\cdot, z_j)\right\|^2_{L^\infty(\Theta)}\right] + \sum_{k=1}^{d} \mathbb{E}_j\left[\left\|\tilde{\mathcal{J}}_{z^k}^{(j)}(\cdot, z_j) - \mathcal{J}_{z^k}(\cdot, z_j)\right\|^2_{L^\infty(\Theta)}\right] \leq \eta^2 \left\|\mathcal{J}_w(\theta_{j-1}, z_j)\right\|^2_{l^2},$$

where $\mathbb{E}_j$ is expectation conditional on $\{z_0, \theta_0, z_1, \theta_1, \ldots, \theta_{j-1}, z_j\}$

Then:

$$\mathbb{E}\left[\|z_{j+1} - z^*\|^2_{l^2} + C_1(\theta_j - \theta^*)^2\right] \leq \xi \mathbb{E}\left[\|z_j - z^*\|^2_{l^2} + C_1(\theta_{j-1} - \theta^*)^2\right]$$

for some constants $C_1 > 0$ and $0 < \xi < 1$.

Based on analysis from [Beiser-Keith-Urbainczyk-Wohlmuth, 2020]

AMGD - control on MSE

$$\left\|\hat{\mathcal{J}}_\theta^{(j)}(\cdot, z_j) - \mathcal{J}_\theta(\cdot, z_j)\right\|^2_{L^\infty(\Theta)} = \left\|\partial_\theta \hat{\Phi}_L(\cdot; z_j) - \partial_\theta \Phi(\cdot; z_j)\right\|^2_{L^\infty(\Theta)}$$
$$\left\|\tilde{\mathcal{J}}_{z^k}^{(j)}(\cdot, z_j) - \mathcal{J}_{z^k}(\cdot, z_j)\right\|^2_{L^\infty(\Theta)} = \left\|\partial_\theta \hat{\Psi}_{k,L}(\cdot; z_j) - \partial_\theta \Psi_k(\cdot; z_j)\right\|^2_{L^\infty(\Theta)}$$

Add together, take expectation $\mathbb{E}_j$, then we have:

$$\mathrm{MSE}\left(\hat{\mathcal{J}}_w^{(j)}(\cdot, z_j)\right) = \mathrm{MSE}\left(\partial_\theta \hat{\Phi}_L(\cdot, z_j)\right) + \sum_{k=1}^{d} \mathrm{MSE}\left(\partial_\theta \hat{\Psi}_{k,L}(\cdot, z_j)\right)$$
$$\leq \underbrace{\left((e_i^\Phi)^2 + \sum_{k=1}^{d}(e_i^{\Psi_k})^2\right)}_{\text{Squared interpolation error}} + \underbrace{\left((e_b^\Phi)^2 + \sum_{k=1}^{d}(e_b^{\Psi_k})^2\right)}_{\text{Squared bias error}} + \underbrace{\left((e_s^\Phi)^2 + \sum_{k=1}^{d}(e_s^{\Psi_k})^2\right)}_{\text{Squared statistical error}}.$$

We can use previous error estimators to optimally tune MLMC using a continuation algorithm [Collier-HajiAli-N.-vonSchwerin-Tempone 2015].

AMGD - control on MSE

$$\left\| \hat{\mathcal{J}}_\theta^{(j)}(\cdot, z_j) - \mathcal{J}_\theta(\cdot, z_j) \right\|^2_{L^\infty(\Theta)} = \left\| \partial_\theta \hat{\Phi}_L(\cdot; z_j) - \partial_\theta \Phi(\cdot; z_j) \right\|^2_{L^\infty(\Theta)}$$

$$\left\| \tilde{\mathcal{J}}_{z^k}^{(j)}(\cdot, z_j) - \mathcal{J}_{z^k}(\cdot, z_j) \right\|^2_{L^\infty(\Theta)} = \left\| \partial_\theta \hat{\Psi}_{k,L}(\cdot; z_j) - \partial_\theta \Psi_k(\cdot; z_j) \right\|^2_{L^\infty(\Theta)}$$

Add together, take expectation $\mathbb{E}_j$, then we have:

$$\mathrm{MSE}\left(\hat{\mathcal{J}}_w^{(j)}(\cdot, z_j)\right) = \mathrm{MSE}\left(\partial_\theta \hat{\Phi}_L(\cdot, z_j)\right) + \sum_{k=1}^{d} \mathrm{MSE}\left(\partial_\theta \hat{\Psi}_{k,L}(\cdot, z_j)\right)$$

$$\leq \underbrace{\left((e_i^{\Phi})^2 + \sum_{k=1}^{d} (e_i^{\Psi_k})^2\right)}_{\text{Squared interpolation error}} + \underbrace{\left((e_b^{\Phi})^2 + \sum_{k=1}^{d} (e_b^{\Psi_k})^2\right)}_{\text{Squared bias error}} + \underbrace{\left((e_s^{\Phi})^2 + \sum_{k=1}^{d} (e_s^{\Psi_k})^2\right)}_{\text{Squared statistical error}}.$$

We can use previous error estimators to optimally tune MLMC using a continuation algorithm [Collier-HajiAli-N.-vonSchwerin-Tempone 2015].

CMLMC-AMGD Algorithm

Algorithm 1 CMLMC-AMGD algorithm

Input: Initial design z_0, iterate $j = 0$, tolerance $0 < \epsilon < 1$, step size $\alpha > 0$ and $\eta > 0$.
Set residual $r = \epsilon + 1$
while $r > \epsilon$ **do**
    if $j = 0$ { Simulate screening hierarchy }
    else $\left\{ \text{Start CMLMC from the optimal hierarchy for } z_{j-1}\text{; Simulate CMLMC adapting hierarchy such that } \mathrm{MSE}\left(\hat{\mathcal{J}}_w^{(j)}(\cdot, z_j)\right) \leq \eta^2 \left\|\hat{\mathcal{J}}_w^{(j-1)}(w_{j-1})\right\|_{l^2}^2 \right\}$
    Compute minimiser $\theta_j \in \mathrm{argmin}_{\theta \in \Theta}\, \hat{\mathcal{J}}^{(j)}(\theta, z_j) = \hat{\Phi}_L(\theta, z_j)$
    Compute gradient $\tilde{\mathcal{J}}_z^{(j)}(\theta_j, z_j) = \partial_\theta \hat{\Psi}_L(\theta_j; z_j) + 2\kappa(z_j - z_{ref})$
    Compute gradient step $z_{j+1} = z_j - \alpha \tilde{\mathcal{J}}_z^{(j)}(\theta_j, z_j)$
    Set residual $r = \left\|\hat{\mathcal{J}}_w^{(j)}(w_j)\right\|_{l^2}^2 / \left\|\hat{\mathcal{J}}_w^{(0)}(w_0)\right\|_{l^2}^2$
    Update $j \leftarrow j + 1$
end while

Results - Pollutant transport

$$
\begin{aligned}
-\nabla \cdot (\epsilon \nabla u) + \mathbb{V} \cdot \nabla u &= f - B, \\
\mathbb{V} &:= \begin{bmatrix} b(\omega) - a(\omega) x_2 \\ a(\omega) x_1 \end{bmatrix}, \\
B &= \sum_{k=1}^{9} z^k \mathcal{N}(p_k, \sigma^2)
\end{aligned}
$$

Quantity of Interest -

$$
Q(z, \omega) := \frac{\kappa_s}{2} \int_{[0,1]^2} u^2 \, dx.
$$

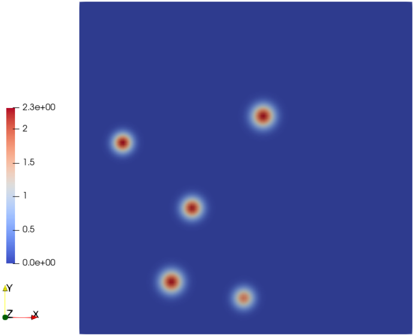


Figure: Source term f

Results - Pollutant transport

$$-\nabla \cdot (\epsilon \nabla u) + \mathbb{V} \cdot \nabla u = f - B,$$

$$\mathbb{V} := \begin{bmatrix} b(\omega) - a(\omega)x_2 \\ a(\omega)x_1 \end{bmatrix},$$

$$B = \sum_{k=1}^{9} z^k \mathcal{N}(p_k, \sigma^2)$$

Quantity of Interest -

$$Q(z, \omega) := \frac{\kappa_s}{2} \int_{[0,1]^2} u^2 dx.$$

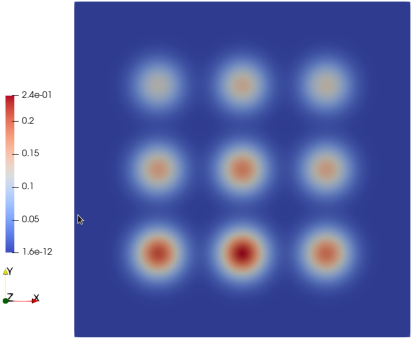


Figure: Control term B

Results - Pollutant transport

$$-\nabla \cdot (\epsilon \nabla u) + \mathbb{V} \cdot \nabla u = f - B,$$

$$\mathbb{V} := \begin{bmatrix} b(\omega) - a(\omega)x_2 \\ a(\omega)x_1 \end{bmatrix},$$

$$B = \sum_{k=1}^{9} z^k \mathcal{N}(p_k, \sigma^2)$$

Quantity of Interest -

$$Q(z,\omega) := \frac{\kappa_s}{2} \int_{[0,1]^2} u^2 dx.$$

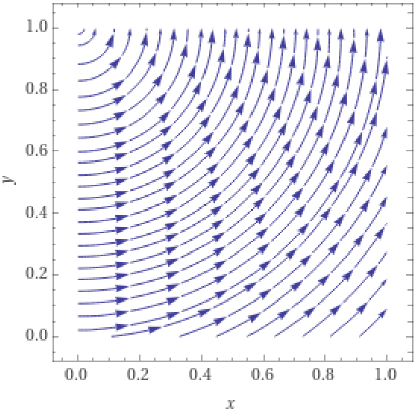


Figure: Velocity field $\mathbb{V}$

Results - Pollutant transport

$$-\nabla \cdot (\epsilon \nabla u) + \mathbb{V} \cdot \nabla u = f - B,$$

$$\mathbb{V} := \begin{bmatrix} b(\omega) - a(\omega)x_2 \\ a(\omega)x_1 \end{bmatrix},$$

$$B = \sum_{k=1}^{9} z^k \mathcal{N}(p_k, \sigma^2)$$

Quantity of Interest -

$$Q(z, \omega) := \frac{\kappa_s}{2} \int_{[0,1]^2} u^2 dx.$$

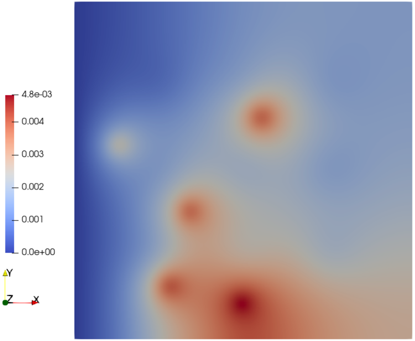
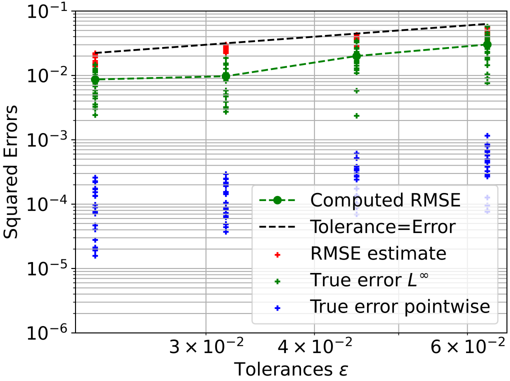
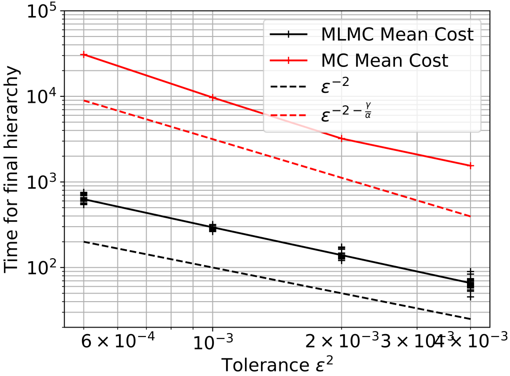


Figure: Pollutant concentration field u

Results - Pollutant transport sensitivity estimation



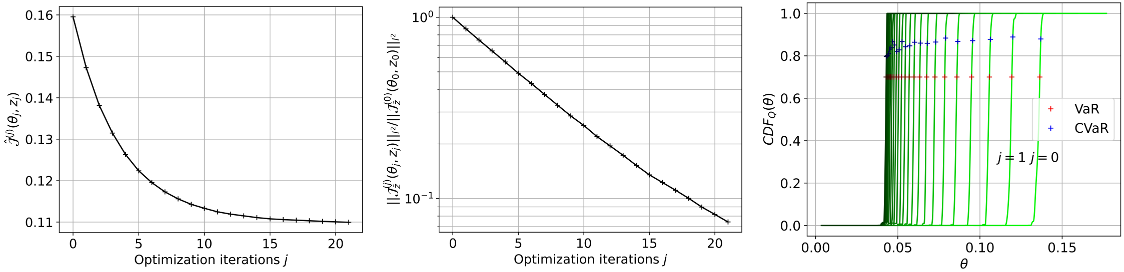
CVaR sensitivity reliability



CVaR sensitivity complexity

Figure: CMLMC performance for a given design z

Results - Pollutant transport optimization



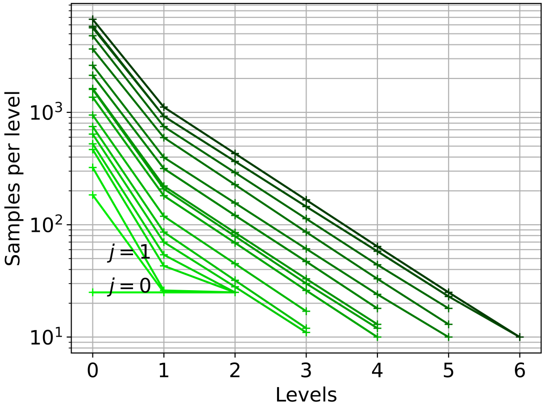
Objective function decay

Gradient ratio decay

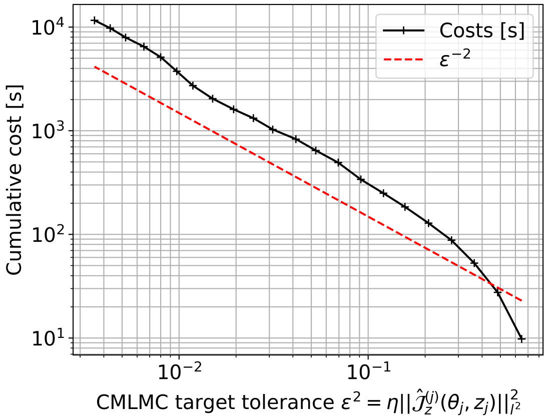
Change in CDF

Figure: Optimization performance over different iterations

Results - Pollutant transport



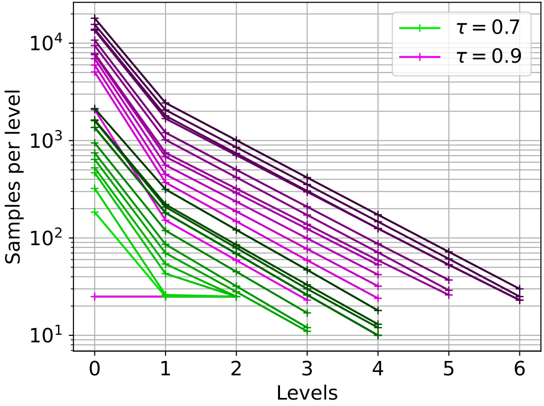
Level-wise sample sizes



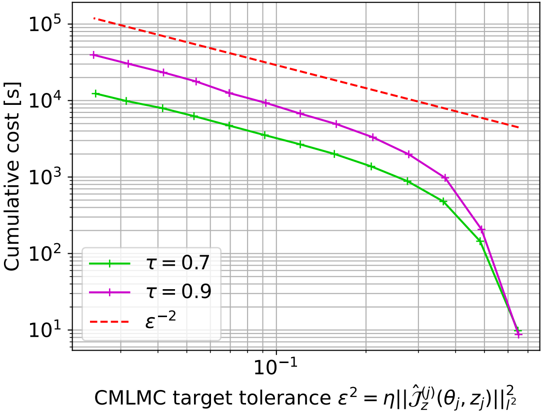
Cumulative cost

Figure: Hierarchy and complexity behaviour for different iterations

Results - Pollutant transport



Level-wise sample sizes



Cumulative cost

Figure: Performance comparison

Thank you for your attention!

References

S.S. Ganesh; F. Nobile
Gradient-based optimisation of the conditional-value-at-risk using the multi-level Monte Carlo method.
arXiv:2210.03485, 2022. https://doi.org/10.48550/arXiv.2210.03485

Q. Ayoul-Guilmard; S.S. Ganesh; S. Krumscheid; F. Nobile
Quantifying uncertain system outputs via the multi-level Monte Carlo method — distribution and robustness measures,
IJUQ, Vol. 13(5), pp. 61-98. 2023. httpr://doi.org/10.1615/Int.J.UncertaintyQuantification.2023045259

S. Krumscheid; F. Nobile
Multilevel Monte Carlo Approximation of Functions.
SIAM/ASA JUQ, Vol. 6(3), p. 1256-1293, 2018. https://doi.org/10.1137/17M1135566

S.S. Ganesh
Multi-Level Monte Carlo Methods for Uncertainty Quantification and Risk-Averse Optimisation.
PhD Thesis 10031, EPFL, 2022. https://doi.org/10.5075/epfl-thesis-10031