MULTIFIDELITY STRATEGIES IN UQ: AN OVERVIEW ON SOME RECENT TRENDS IN SAMPLING BASED APPROACHES

SAND2020-9951C

Gianluca Geraci[1], Michael S. Eldred[1], Alex A. Gorodetsky[2]
and John D. Jakeman[1]

[1]Sandia National Laboratories, Albuquerque [2]Department of Aerospace Engineering, University of Michigan

Doctoral Course – Aeronautical and Space Engineering – Summer School 2020
University of Rome La Sapienza
September 18 2020

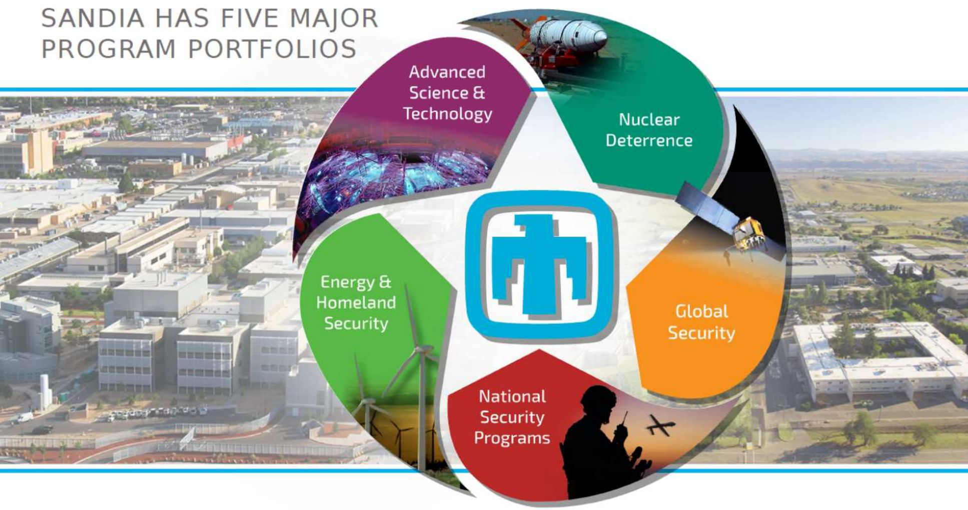
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Plan of the talk

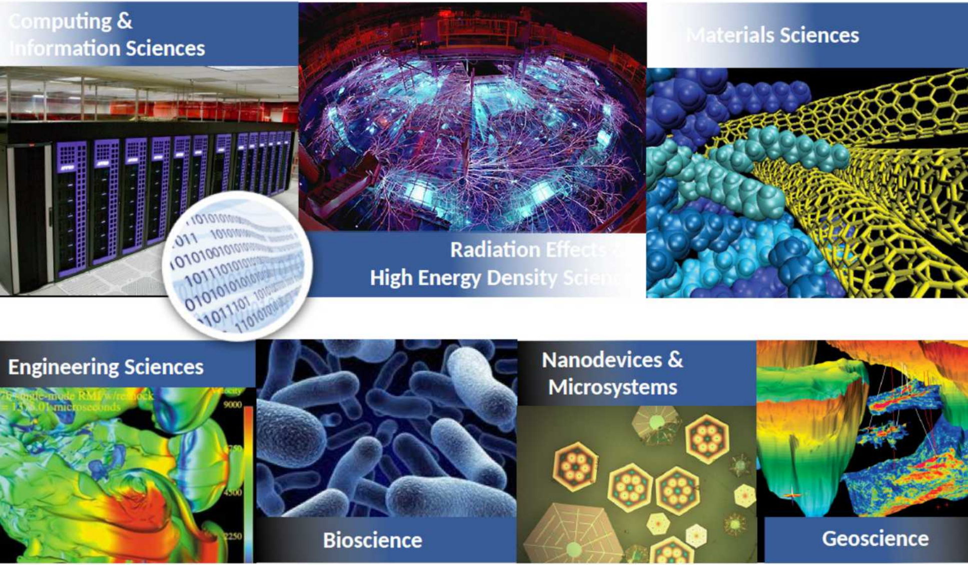
- UQ @ Sandia National Laboratories
- MF in UQ: motivation
- Topic I: Multifidelity sampling
- Topic II: Leveraging Active Directions for MF UQ
- Topic III: Multifidelity Bayesian Calibration
- Conclusions

UQ @ Sandia National Laboratories

Sandia National Laboratories
Main role and areas of interest



Sandia National Laboratories
Advanced Science & Technology



Sandia National Laboratories

Algorithms R&D: From Core Solvers to Modeling and Simulation Applications

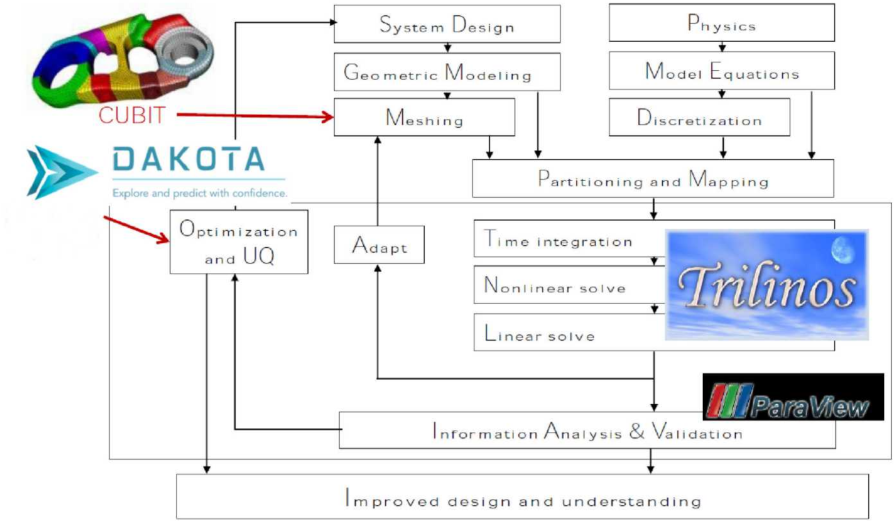


Figure: Courtesy of Brian Adams

Sandia National Laboratories
DAKOTA – Explore and Design with confidence

Algorithms for **design exploration** and **simulation credibility**

- Suite of iterative mathematical and statistical methods that interface to computational models
- Makes sophisticated parametric exploration of simulations practical for a computational design-analyze-test cycle

Features

- **Sensitivity:** Which are the crucial factors/parameters?
- **Uncertainty:** How safe, reliable, or robust is my system?
- **Optimization:** What is the best performing design or control?
- **Calibration/Parameter Estimation:** What models and parameters best match data?

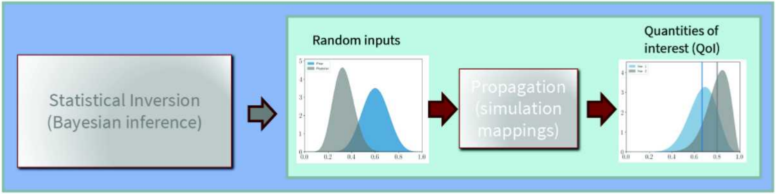
Credible Prediction

- **Verification:** Is the model implemented correctly, converging as expected?
- **Validation:** How does the model compare to experimental data, including uncertainties?

DAKOTA
Explore and predict with confidence.
https://dakota.sandia.gov/

Uncertainty Quantification

The complete workflow



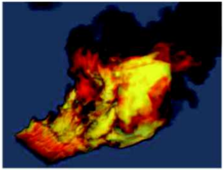
Notes:

- Prior **distributions based on a priori knowledge**
- From observational data (experiments, reference solutions, etc.) **we can infer posterior distributions via Bayes rule**
- Use of **data can reduce uncertainty** in parameter to QoI mapping (priors are constrained)
- Design using **prior uncertainties can be overly conservative**
- Reduced uncertainty of **data-informed UQ can produce designs with greater performance**

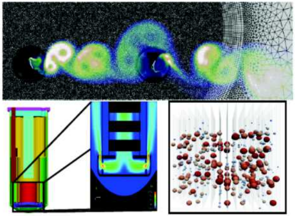
Why multifidelity in Uncertainty Quantification?

Uncertainty Quantification
DoE and DoD deployment activities

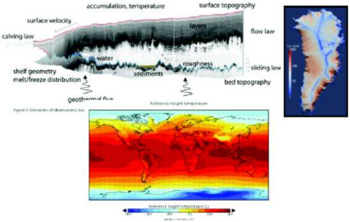
Stewardship (NNSA ASC)
Safety in abnormal environments



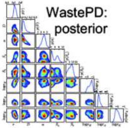
Energy (ASCR, EERE, NE)
Wind turbines, nuclear reactors

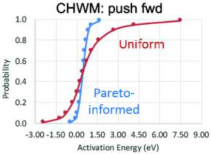


Climate (SciDAC, CSSEF, ACME)
Ice sheets, CISM, CESM, ISSM, CSDMS



Addtnl. Office of Science:
(SciDAC, EFRC)
Comp. Matls: waste forms / hazardous matls (WastePD, CHWM)
MHD: Tokamak disruption (TDS)





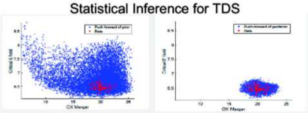


FIGURE: Courtesy of Mike Eldred

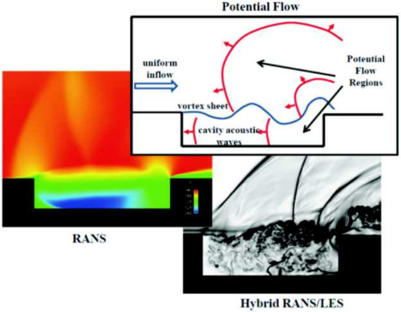
High-fidelity state-of-the-art modeling and simulations with HPC
- **Severe** simulations **budget constraints**
- **Significant dimensionality** driven by model complexity

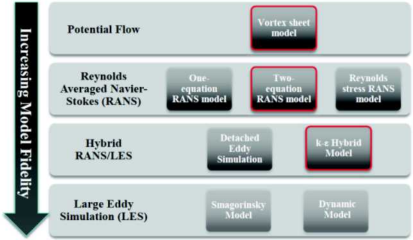
Uncertainty Quantification

Rich set of modeling choices – Discretization Vs Fidelity

Multi-fidelity: several accuracy levels available

- Physical models (Laminar/Turbulent, Reacting/non-reacting, viscous/inviscid...)
- Numerical methods (high/low order, Euler/RANS/LES, *etc*...)
- Numerical discretization (fine/coarse mesh...)
- Quality of statistics (long/short time history for turbulent flow...)





Relationships amongst models can be difficult to anticipate

- A simple **hierarchical sequence** can correspond to strict modeling choices (*e.g.* discretization levels)
- More often, for some QoI, we can have **peer models**

Topic I

MF Sampling-based approaches

Uncertainty Quantification
Forward Propagation – Why Sampling Methods?

UQ context at a glance:

- High-dimensionality, non-linearity and possibly non-smooth responses
- Rich physics and several discretization levels/models available

Natural candidate:

- **Sampling**-based (MC-like) approaches because they are **non-intrusive**, **robust** and **flexible**...
- **Drawback:** Slow convergence $\mathcal{O}(N^{-1/2}) \rightarrow$ many realizations to build reliable statistics

Goal of the talk: **Reducing the computational cost** of obtaining MC reliable statistics

Pivotal idea:

- Simplified (**low-fidelity**) models are **inaccurate** but **cheap**
  - **low-variance** estimates
- **High-fidelity** models are **costly**, but **accurate**
  - **low-bias** estimates

Monte Carlo

Monte Carlo

A brief of its history (1/2)

Halton (1970): *representing the solution of a problem as a parameter of a hypothetical population, and using a random sequence of numbers to construct a sample of the population, from which statistical estimates of the parameter can be obtained.*

- One of the first documented MC experiments is Buffon's needle experiment which Laplace (1812) suggested can be used to approximate π (Johansen and Evers, 2007)

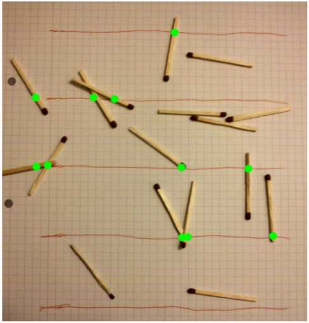


Figure: Buffon's needle experiment based on 17 throws. (Source: Wikipedia)

$$\pi \approx \frac{2Nl}{Pt},$$

where

- N: number of needles
- l: length of the needles
- P: number of needles crossing the lines
- t: distance between the lines

Monte Carlo

A brief of its history (2/2) – *Los Alamos Science No. 15, Special Issue 1987 – In honor of Stan Ulam*

Around 1940:

- ENIAC: first electronic computer at the University of Pennsylvania

[...] Stan's (Stanislaw Ulam) extensive mathematical background made him aware that statistical sampling techniques had fallen into desuetude because of the length and tediousness of the calculations. But with this miraculous development of the ENIAC, [...] it occurred to him that statistical techniques should be resuscitated, and he discussed this idea with von Neumann. Thus was triggered the spark that led to the Monte Carlo method.

- The name: *Ulam had a uncle who would borrow money from relatives because he "just had to go to Monte Carlo"*

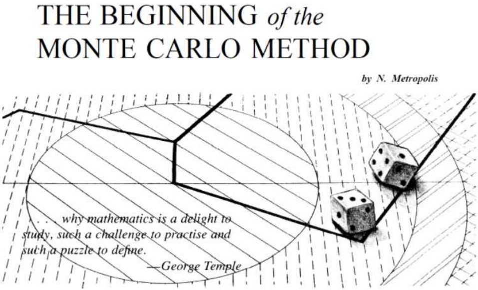


FIGURE: Metropolis' contribution to the Los Alamos Science Special Issue, 1987

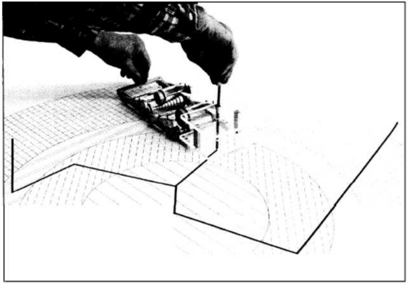


FIGURE: Analog device dubbed FERMIAC, Image from Los Alamos Science No. 15, 1987

Sampling Methods

How are sampling methods used within UQ?

- There are several applications for the MC method
- In Uncertainty Quantification (UQ) we are often concerned with the computation of a the expected value of a function (or higher moments)[1]

$$\mathbb{E}\left[f(\xi)\right] = \int_{\Xi} f(\xi) p(\xi) \mathrm{d}\xi$$

- Therefore one of the tasks to be performed in UQ is the quadrature in (very often) high-dimension ($\Xi \subset \mathbb{R}^d$)

The **Monte Carlo** method is based upon three main steps:

- **Pre-processing:** generation of random numbers
- **Evaluation step:** Computation of the Quantity of Interest from the computational code
- **Post-processing:** Estimator and confidence interval evaluation

[1] UQ is a much richer area than 'just' numerical quadrature, but nevertheless this is an important task

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Two main estimator's properties

- **Unbiased** (for each choice of N!): $\mathbb{E}\left[\hat{Q}_N^{MC}\right] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[Q^{(i)}\right] = \mathbb{E}\left[Q\right]$
- **Convergent (Strong law of large numbers)**: $\lim_{N \to \infty} \hat{Q}_N^{\mathrm{MC}} = \mathbb{E}\left[Q\right]$ a.s.

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Two main estimator's properties

- **Unbiased** (for each choice of N!): $\mathbb{E}\left[\hat{Q}_N^{MC}\right] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[Q^{(i)}\right] = \mathbb{E}\left[Q\right]$
- **Convergent (Strong law of large numbers)**: $\lim_{N \to \infty} \hat{Q}_N^{\mathrm{MC}} = \mathbb{E}\left[Q\right]$ a.s.

Main mathematical tool used for the analysis is the **Central Limit Theorem (CLT)**

- Let's define the error $e_N = \mathbb{E}\left[Q\right] - \hat{Q}_N^{MC}$
- Let's assume $\mathbb{V}ar\left[Q\right]$ is finite, then for $N \to \infty$

$$\frac{e_N}{\sqrt{\mathbb{V}ar\left[\hat{Q}_N^{\mathrm{MC}}\right]}} \sim \mathcal{N}(0, 1),$$

where

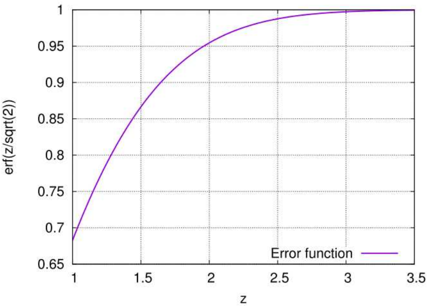
$$\mathbb{V}ar\left[\hat{Q}_N^{\mathrm{MC}}\right] = \frac{\mathbb{V}ar\left[Q\right]}{N}$$

Central Limit Theorem
Confidence interval

CLT is the fundamental result that enable us to obtain a confidence interval for MC

- $\mathrm{P}\left(N^{1/2}\dfrac{e_N}{\mathbb{V}ar^{1/2}(Q)} \leq z\right) = F_Z(z)$, for $Z \sim \mathcal{N}(0,1)$
- $F_Z(z) = \dfrac{1}{2}\left(1 + \mathrm{erf}\left(\dfrac{z}{\sqrt{2}}\right)\right)$
- We want to control the probability of $\left|N^{1/2}\dfrac{e_N}{\mathbb{V}ar^{1/2}(Q)}\right|$, therefore

$$\mathrm{P}\left(\left|N^{1/2}\frac{e_N}{\mathbb{V}ar^{1/2}(Q)}\right| \leq z\right) = 1 - 2F_Z(z) = \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)$$



z	$1-2F_Z(z)$
1	0.683
2	0.954
3	0.997

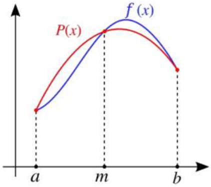
Monte Carlo
Target accuracy

We can **use the distribution of e_N to estimate the number of simulations** required.

- Let's assume we want an estimator accurate at the 99.7% with error $e_N = \varepsilon$
- We need to select $z = 3$ (from the previous table)
- $N = 9 \dfrac{\mathbb{V}ar\left[Q\right]}{\varepsilon^2}$

Few additional comments:

- The number of samples scales as ε^2, *i.e.* one order of increased accuracy is obtained with 100 times more samples
- Error is **not a function of the dimension** ($e_N \propto N^{-1/2}$)
- Error is **not a function of the regularity** of the quantity Q
- On the contrary the error for a composite (Cavalieri,Kepler-)Simpson's rule ($[0, 1]$) is bounded by



$$\frac{h^4}{180} \max_{x\in[0,1]} f^{(4)}(x), \quad \text{therefore} \quad e_N \propto N_{1D}^{-4} = N^{-4/d}$$

(MC integration is competitive for $d > 8$ w.r.t. Simpson's rule)

Figure: https://en.wikipedia.org/wiki/Simpson%27s_rule

Monte Carlo
Estimator variance: derivation

In summary we have seen so far:

- CLT provide a rigorous way to assess the accuracy of a MC simulation
- $e_N \sim \sqrt{\mathbb{V}ar\left[\hat{Q}_N^{\mathrm{MC}}\right]}\mathcal{N}(0,1)$
- $e_N \propto N^{-1/2}$ and (numerical cost) is $\mathcal{C}^{\mathrm{MC}} \propto N$, therefore $\mathcal{C}^{\mathrm{MC}} \propto e_N^{-2}$
- MC **convergence is independent from the dimensionality** of the problem (indeed more efficient w.r.t. other strategies as d increases)
- MC **does not require a certain degree of regularity** to maintain its properties

Monte Carlo
Estimator variance: derivation

In summary we have seen so far:

- CLT provide a rigorous way to assess the accuracy of a MC simulation
- $e_N \sim \sqrt{\mathbb{V}ar\left[\hat{Q}_N^{\mathrm{MC}}\right]}\mathcal{N}(0,1)$
- $e_N \propto N^{-1/2}$ and (numerical cost) is $\mathcal{C}^{\mathrm{MC}} \propto N$, therefore $\mathcal{C}^{\mathrm{MC}} \propto e_N^{-2}$
- MC **convergence is independent from the dimensionality** of the problem (indeed more efficient w.r.t. other strategies as d increases)
- MC **does not require a certain degree of regularity** to maintain its properties

Variance of a MC estimator is

$$\begin{aligned}
\mathbb{V}ar\left[\hat{Q}_N^{\mathrm{MC}}\right] &= \mathbb{V}ar\left[\frac{1}{N}\sum_{i=1}^{N} Q^{(i)}\right] \\
&= \frac{1}{N^2}\mathbb{V}ar\left[\sum_{i=1}^{N} Q^{(i)}\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{V}ar\left[Q\right] \\
&= \frac{1}{N}\mathbb{V}ar\left[Q\right]
\end{aligned}$$

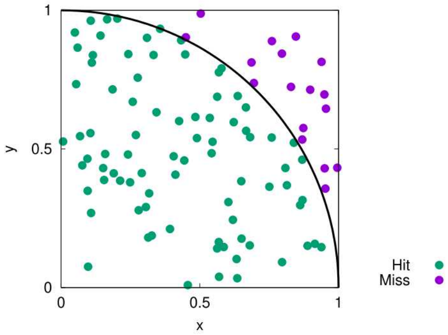
Monte Carlo

Estimator Variance: a simple demonstration

Let consider a random variable Q, we want to compute **its expected value** $\mathbb{E}[Q]$ (or some high-order moment):

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Let's use MC to compute the value $\pi \propto \frac{\#\mathrm{Hit}}{N}$



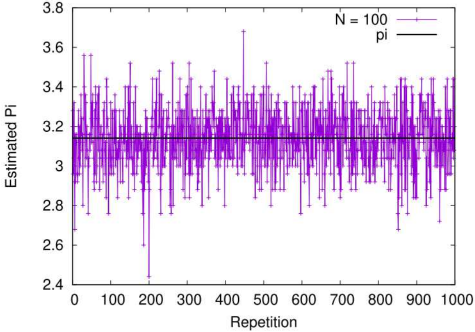
Monte Carlo

Estimator Variance: a simple demonstration

Let consider a random variable Q, we want to compute its **expected value** $\mathbb{E}\left[Q\right]$ (or some high-order moment):

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} Q^{(i)}$$

Let's use MC to compute the value $\pi \propto \frac{\#\mathrm{Hit}}{N}$



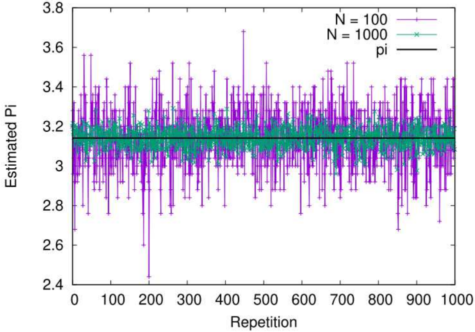
Monte Carlo

Estimator Variance: a simple demonstration

Let consider a random variable Q, we want to compute its **expected value** $\mathbb{E}[Q]$ (or some high-order moment):

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Let's use MC to compute the value $\pi \propto \dfrac{\#\mathrm{Hit}}{N}$



Variance reduction strategies

An (incomplete list)

Variance of the estimator:

$$\mathbb{V}ar\left[\hat{Q}_N^{MC}\right] = \frac{\mathbb{V}ar\left[Q\right]}{N}$$

What can we do to drive down the variance of the estimator?

#0 **Increase the number of samples** $\rightarrow$ this is going to cost us too much for HF applications

#1 **Replace the HF model with a computational cheapest one**, *e.g.* Reduced Order Models (ROMs)

#2 **Act on the sampling (Stratification, Important Sampling etc.)**

#2 Replace the original QoI with a lower variance alternative (with the same mean)

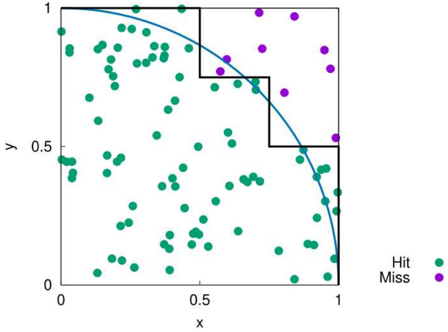
Sampling-based variance reduction techniques:

- **Importance sampling**
  - Very useful when the main contribution to $\mathbb{E}\left[Q\right]$ comes from rare events
- **Stratified sampling**
  - Very effective in 1D, not clear how to extend to multiple dimensions
- **Latin hypercube**
  - Effective if the function can be decomposed into a sum of 1D functions
- **(Randomized) quasi-MC**
  - Possibly provides better error than MC, but need to be randomized to get the confidence interval

Monte Carlo
Introducing the notion of fidelity: bias of the estimator

Numerical problems **cannot be resolved with infinite accuracy**: a discretization/numerical error is often introduced

$$\hat{Q}^{MC}_{M,N} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} Q_M^{(i)}$$

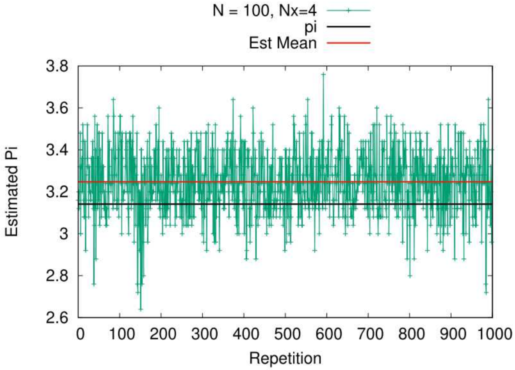


Monte Carlo

Introducing the notion of fidelity: bias of the estimator

Numerical problems **cannot be resolved with infinite accuracy**: a discretization/numerical error is often introduced

$$\hat{Q}^{MC}_{M,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} Q_M^{(i)}$$

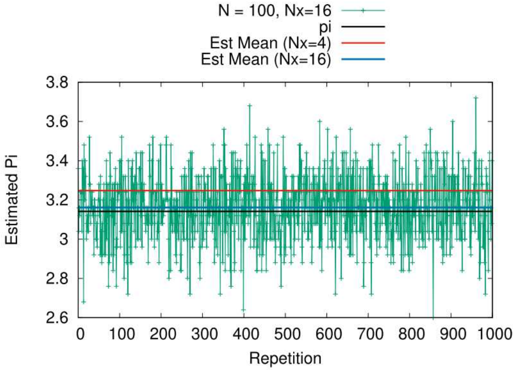


Monte Carlo

Introducing the notion of fidelity: bias of the estimator

Numerical problems **cannot be resolved with infinite accuracy**: a discretization/numerical error is often introduced

$$\hat{Q}^{MC}_{M,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} Q_M^{(i)}$$



Monte Carlo simulation
Introducing the spatial discretization

Problem statement: We are interested in the statistics of a functional (linear or non-linear) Q_M of the solution $\mathbf{u}_M$

$$Q_M = \mathcal{G}(\mathbf{u}_M) \quad \rightarrow \quad \mathbb{E}\left[Q_M\right]$$

- M is (related to) the number of **spatial** degrees of freedom
- $\mathbb{E}\left[Q_M\right] \xrightarrow{M \rightarrow \infty} \mathbb{E}\left[Q\right]$ for some RV $Q : \Omega \rightarrow \mathbb{R}$

$$\hat{Q}_{M,N}^{MC} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} Q_M^{(i)},$$

Looking at the **Mean Square Error (MSE)**:

$$\begin{aligned}
\mathbb{E}\left[(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q\right])^2\right] &= \mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q_M\right] + \mathbb{E}\left[Q_M\right] - \mathbb{E}\left[Q\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\mathbf{Q}}_{\mathbf{M,N}}^{\mathbf{MC}} - \mathbb{E}\left[\mathbf{Q}_{\mathbf{M}}\right]\right)^{\mathbf{2}}\right] + 2\mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q_M\right]\right)\left(\mathbb{E}\left[Q_M\right] - \mathbb{E}\left[Q\right]\right)\right] \\
&+ \mathbb{E}\left[\left(\mathbb{E}\left[\mathbf{Q}_{\mathbf{M}}\right] - \mathbb{E}\left[\mathbf{Q}\right]\right)^{\mathbf{2}}\right] \\
&= \mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}_{\mathbf{M,N}}^{\mathbf{MC}}\right] + \left(\mathbb{E}\left[\mathbf{Q}_{\mathbf{M}} - \mathbf{Q}\right]\right)^{\mathbf{2}}
\end{aligned}$$

Monte Carlo
Overall estimator error

Two sources of error in the **Mean Square Error**:

$$\mathbb{E}\left[\left(\hat{Q}^{MC}_{M,N} - \mathbb{E}\left[Q\right]\right)^2\right] = \mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}^{\mathbf{MC}}_{\mathbf{M,N}}\right] + \left(\mathbb{E}\left[\mathbf{Q_M} - \mathbf{Q}\right]\right)^{\mathbf{2}}$$

- **Sampling error:** replacing the expected value by a (finite) sample average, *i.e.*

$$\mathbb{V}ar\left[\hat{Q}^{MC}_{M,N}\right] = \frac{\mathbb{V}ar\left[Q\right]}{N}$$

  From the CLT, for $N \to \infty$

$$\left(\hat{Q}^{MC}_{M,N} - \mathbb{E}\left[Q\right]\right) \sim \sqrt{\frac{\mathbb{V}ar\left[Q\right]}{N}}\mathcal{N}(0,1)$$

- **Model fidelity (e.g. discretization):** finite accuracy

Monte Carlo
Overall estimator error

Two sources of error in the **Mean Square Error**:

$$\mathbb{E}\left[\left(\hat{Q}^{MC}_{M,N} - \mathbb{E}\left[Q\right]\right)^2\right] = \mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}^{\mathbf{MC}}_{\mathbf{M,N}}\right] + \left(\mathbb{E}\left[\mathbf{Q_M} - \mathbf{Q}\right]\right)^{\mathbf{2}}$$

- **Sampling error:** replacing the expected value by a (finite) sample average, *i.e.*

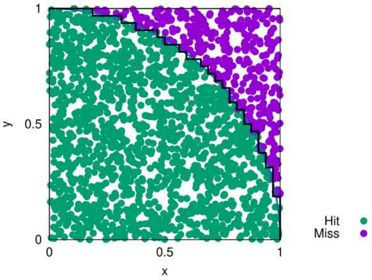
$$\mathbb{V}ar\left[\hat{Q}^{MC}_{M,N}\right] = \frac{\mathbb{V}ar\left[Q\right]}{N}$$

  From the CLT, for $N \to \infty$

$$\left(\hat{Q}^{MC}_{M,N} - \mathbb{E}\left[Q\right]\right) \sim \sqrt{\frac{\mathbb{V}ar\left[Q\right]}{N}}\mathcal{N}(0,1)$$

- **Model fidelity (e.g. discretization):** finite accuracy

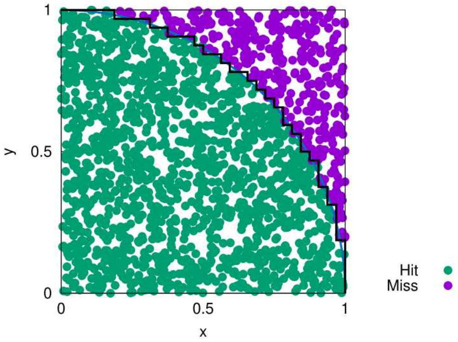
Accurate estimation $\Rightarrow$ **Large number** of samples evaluated for the high fidelity model



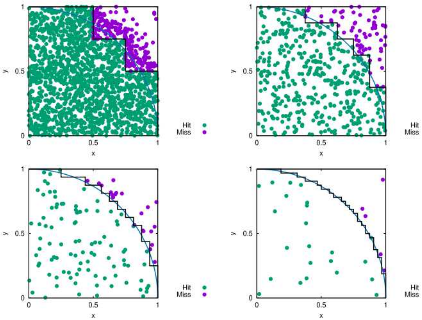
Accelerating Monte Carlo

Bringing multiple fidelity models into the picture

Single Fidelity



Multi Fidelity



Pivotal idea:

- **High-fidelity** models are **costly**, but **accurate**
  - **low-bias** estimates
- **Low-fidelity** models are **inaccurate** but **cheap-to-evaluate**
  - **low-variance** estimates

Multifidelity challenges:

- How do you arrange the information sources?
- How do you optimally allocate samples among models?

Why is this supposed to work?

A prototypical estimator: the difference estimator

Ingredients:

- High-fidelity: Q
- Low-fidelity: P

$$\mathbb{E}\left[Q\right] = \mathbb{E}\left[P + (Q - P)\right] = \mathbb{E}\left[P\right] + \mathbb{E}\left[Q - P\right] \approx \frac{1}{N_P}\sum_{i=1}^{N_P} P^{(i)} + \frac{1}{N_Q}\sum_{j=1}^{N_Q}\left(Q^{(j)} - P^{(j)}\right)$$

Properties of the difference estimator

- Unbiased
- Variance

$$\frac{\mathbb{V}ar\left[P\right]}{N_P} + \frac{\mathbb{V}ar\left[Q - P\right]}{N_Q} = \frac{\mathbb{V}ar\left[P\right]}{\mathbf{N_P}} + \frac{1}{N_Q}\left(\mathbb{V}\mathbf{ar}\left[\mathbf{Q}\right] + \mathbb{V}\mathbf{ar}\left[\mathbf{P}\right] - \mathbf{2}\mathrm{Cov}\left(\mathbf{Q}, \mathbf{P}\right)\right)$$

NOTE: The **negative** term can help you if the cost of computing P is low and if $\mathbb{V}ar\left[P\right]$ approaches $\mathbb{V}ar\left[Q\right]$

Control variate
Can we do slightly better?

A **Control Variate** MC estimator (function Q_1 with μ_1 **known**)

$$\hat{Q}_N^{CV} = \hat{Q} - \beta\left(\hat{Q}_1 - \mu_1\right), \quad \beta \in \mathbb{R}$$

NOTE-1: $\hat{Q}$ is the MC estimator of the HF and $\hat{Q}_1$ is the MC estimator of the LF
NOTE-2: $\hat{Q}$ and $\hat{Q}_1$ are obtained with the same samples

Properties:

- Unbiased, *i.e.* $\mathbb{E}\left[\hat{Q}_N^{CV}\right] = \mathbb{E}\left[\hat{Q}\right] = \mathbb{E}[Q]$ (for any β)
- $\underset{\beta}{\text{argmin}}\, \mathbb{V}ar\left[\hat{Q}_N^{CV}\right] \rightarrow \beta = -\rho \dfrac{\mathbb{V}ar^{1/2}(Q)}{\mathbb{V}ar^{1/2}(Q_1)}$
- Pearson's $\rho = \dfrac{\text{Cov}(Q, Q_1)}{\mathbb{V}ar^{1/2}(Q)\,\mathbb{V}ar^{1/2}(Q_1)}$ where $|\rho| < 1$

$$\boxed{\mathbb{V}ar\left[\hat{Q}_N^{CV}\right] = \mathbb{V}ar\left[\hat{Q}\right]\left(1 - \rho^2\right)}$$

Let's consider:

- $\mathbb{V}ar[Q_1] \approx \mathbb{V}ar[Q]$
- $\rho \approx 1$
- It follows that $\beta \approx -1$

NOTE: In reality β is estimated by a finite number of samples, therefore the variance is slightly higher and there is a small bias (that can be quantified)...

Multifidelity Monte Carlo

Multifidelity

Practical implications of unknown low-fidelity statistics

Let's modify the high-fidelity $\hat{Q}$, to decrease its variance

$$\hat{Q}_N^{CV} = \hat{Q} + \beta \left(\hat{Q}_1 - \hat{\mu}_1 \right).$$

Multifidelity

Practical implications of unknown low-fidelity statistics

Let's modify the high-fidelity $\hat{Q}$, to decrease its variance

$$\hat{Q}_N^{CV} = \hat{Q} + \beta \left(\hat{Q}_1 - \hat{\mu}_1 \right).$$

In practical situations

- the term $\hat{\mu}_1$ is unknown (low fidelity $\neq$ analytic function)
- we use an additional and independent set $\Delta^{\mathrm{LF}} = (\mathbf{r} - 1) N^{\mathrm{HF}}$

$$\hat{\mu}_1 \simeq \frac{1}{\mathbf{r} N^{\mathrm{HF}}} \sum_{i=1}^{\mathbf{r} N^{\mathrm{HF}}} Q_1^{(i)}.$$

Finally the variance is

$$\boxed{\mathbb{V}ar\left[\hat{Q}_N^{CV}\right] = \mathbb{V}ar\left[\hat{Q}\right] \left(1 - \frac{\mathbf{r} - \mathbf{1}}{\mathbf{r}} \rho_1^2 \right)}$$

[1] Pasupathy, R., Taaffe, M., Schmeiser, B. W. & Wang, W., Control-variate estimation using estimated control means. *IIE Transactions*, **44**(5), 381–385, 2012

[2] Ng, L.W.T. & Willcox, K. Multifidelity Approaches for Optimization Under Uncertainty. *Int. J. Numer. Meth. Engng* 100, no. 10, pp. 746772, 2014.

[3] Peherstorfer, B., Willcox, K. & Gunzburger, M., Optimal Model Management for Multifidelity Monte Carlo Estimation. *SIAM J. Sci. Comput.* 38(5), A3163A3194, 2016.

Multifidelity Estimator

How do we select the important parameters?

$$\mathbb{V}ar\left[\hat{Q}_N^{CV}\right] = \mathbb{V}ar\left[\hat{Q}_N\right]\left(1 - \frac{\mathbf{r}-\mathbf{1}}{\mathbf{r}}\rho_1^2\right)$$

Two questions:

1. How do I pick β?
2. How many samples do I need to evaluate for each model?

Q: If $\frac{\mathbf{r}-\mathbf{1}}{\mathbf{r}} \to 1$, why don't we use a very large r for the estimator? (Remember, $N^{\mathrm{LF}} = rN^{\mathrm{HF}}$)

Multifidelity Estimator

How do we select the important parameters?

$$\mathbb{V}ar\left[\hat{Q}_N^{CV}\right] = \mathbb{V}ar\left[\hat{Q}_N\right]\left(1 - \frac{\mathbf{r}-\mathbf{1}}{\mathbf{r}}\rho_1^2\right)$$

Two questions:

1 How do I pick β?

2 How many samples do I need to evaluate for each model?

Q: If $\frac{\mathbf{r}-\mathbf{1}}{\mathbf{r}} \to 1$, why don't we use a very large r for the estimator? (Remember, $N^{\mathrm{LF}} = rN^{\mathrm{HF}}$)

A: An **optimal solution for r** exists if we try to minimize the overall estimator cost for a certain target variance

Multifidelity Estimator

How do we select the important parameters?

$$\mathbb{V}ar\left[\hat{Q}_N^{CV}\right] = \mathbb{V}ar\left[\hat{Q}_N\right]\left(1 - \frac{\mathbf{r-1}}{\mathbf{r}}\rho_1^2\right)$$

Two questions:

1 How do I pick β?

2 How many samples do I need to evaluate for each model?

Q: If $\frac{\mathbf{r-1}}{\mathbf{r}} \to 1$, why don't we use a very large r for the estimator? (Remember, $N^{\mathrm{LF}} = rN^{\mathrm{HF}}$)

A: An **optimal solution for** r exists if we try to minimize the overall estimator cost for a certain target variance

Let's introduce the following notation

- Cost of one low-fidelity realization: $\mathcal{C}^{\mathrm{LF}}$
- Cost of one high-fidelity realization: $\mathcal{C}^{\mathrm{HF}}$
- Total cost: $\mathcal{C}^{tot}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + rN^{\mathrm{HF}}\mathcal{C}^{\mathrm{LF}}$

Remember...

$$\mathbb{E}\left[(\hat{Q}_M^{\mathrm{HF},CV} - \mathbb{E}[Q])^2\right] = \mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}_{\mathbf{N},\mathbf{M}}^{\mathbf{CV}}\right] + (\mathbb{E}[\mathbf{Q_M} - \mathbf{Q}])^{\mathbf{2}}$$

Additional considerations:

- Let's assume someone is giving us the weak error $\mathbb{E}[\mathbf{Q_M} - \mathbf{Q}]$ committed on the resolution level M
- Let's call $(\mathbb{E}[\mathbf{Q_M} - \mathbf{Q}])^{\mathbf{2}} = \varepsilon^2/2$ for simplicity

Multifidelity Estimator

Minimization of the computational cost (Problem definition)

We want to solve the following problem:

- Minimization of the **total computational cost**: $\mathcal{C}^{tot}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + rN^{\mathrm{HF}}\mathcal{C}^{\mathrm{LF}}$
- We want to reach a **target MSE** of ε^2, therefore $\mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}^{\mathbf{CV}}_{\mathbf{N},\mathbf{M}}\right] = \varepsilon^2/2$
- The cost ratio between the two models is: $w = \mathcal{C}^{\mathrm{HF}}/\mathcal{C}^{\mathrm{LF}}$

Multifidelity Estimator

Minimization of the computational cost (Problem definition)

We want to solve the following problem:

- Minimization of the **total computational cost**: $\mathcal{C}^{tot}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + rN^{\mathrm{HF}}\mathcal{C}^{\mathrm{LF}}$
- We want to reach a **target MSE** of ε^2, therefore $\mathbb{V}\mathbf{ar}\left[\hat{\mathbf{Q}}^{\mathbf{CV}}_{\mathbf{N,M}}\right] = \varepsilon^2/2$
- The cost ratio between the two models is: $w = \mathcal{C}^{\mathrm{HF}}/\mathcal{C}^{\mathrm{LF}}$

More formally, let's define our optimization problem (Lagrange constrain optimization)

$$\underset{N^{\mathrm{HF}}, r, \lambda}{\operatorname{argmin}} (\mathcal{L}) \quad \mathcal{L} = \mathcal{C}^{tot} - \lambda \left(\frac{1}{N^{\mathrm{HF}}} \mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right] \Lambda(r) - \frac{\varepsilon^2}{2} \right)$$

$$\begin{aligned}
\mathcal{C}^{tot}\left(N^{\mathrm{HF}}, r\right) &= N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + rN^{\mathrm{HF}}\mathcal{C}^{\mathrm{LF}} \\
&= N^{\mathrm{HF}}\left(\mathcal{C}^{\mathrm{HF}} + r\mathcal{C}^{\mathrm{LF}}\right) \\
&= N^{\mathrm{HF}}\mathcal{C}^{\mathrm{eq}}(r) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}}\Gamma(r) \\
\Lambda(r) &= 1 - \frac{r-1}{r}\rho_1^2.
\end{aligned}$$

Multifidelity

Minimization of the computational cost (Optimal solution)

The solution of the optimization problem is obtained as

$$r^{\star} = \sqrt{\frac{w\rho^2}{1-\rho^2}}$$

$$N^{\mathrm{HF},\star} = \frac{\mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right]}{\varepsilon^2/2}\Lambda(r^{\star}),$$

Multifidelity

Minimization of the computational cost (Optimal solution)

The solution of the optimization problem is obtained as

$$r^{\star} = \sqrt{\frac{w\rho^2}{1-\rho^2}}$$

$$N^{\mathrm{HF},\star} = \frac{\mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right]}{\varepsilon^2/2}\Lambda(r^{\star}),$$

How this compare to MC?

- Total cost of MC: $\mathcal{C}_{tot}^{\mathrm{MC}} = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} = \frac{\mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right]}{\varepsilon^2/2}\mathcal{C}^{\mathrm{HF}}$
- Total cost MF: $\mathcal{C}^{tot} = N^{\mathrm{HF},\star}\mathcal{C}^{\mathrm{eq}}(r^{\star}) = \frac{\mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right]}{\varepsilon^2/2}\mathcal{C}^{\mathrm{HF}}\Theta(w,\rho^2)$, where

$$\Theta(w,\rho^2) \stackrel{\mathrm{def}}{=} \Lambda(r^{\star})\Gamma(r^{\star})$$

measures the efficiency of the method (w.r.t. MC, *i.e.* we want $\Theta(w,\rho^2) < 1$)

Multifidelity

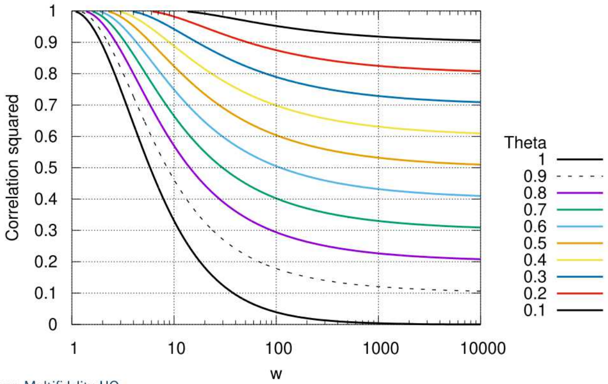
Minimization of the computational cost (Optimal solution)

The solution of the optimization problem is obtained as ($w = \mathcal{C}_{\mathrm{HF}}/\mathcal{C}_{\mathrm{LF}}$)

$$r^\star = \sqrt{\frac{w\rho^2}{1-\rho^2}}$$

$$N^{\mathrm{HF},\star} = \frac{\mathbb{V}ar\left[Q_M^{\mathrm{HF}}\right]}{\varepsilon^2}\left(1 - \frac{r^\star - 1}{r^\star}\rho^2\right)$$

$$\mathcal{C}_{tot} = N^{\mathrm{HF},\star}\mathcal{C}_{\mathrm{HF}}\left(1+\frac{r^\star}{w}\right) = N_{MC}\mathcal{C}_{\mathrm{HF}}\left(\mathbf{1}+\frac{\mathbf{r}^\star}{\mathbf{w}}\right)\left(\mathbf{1}-\frac{\mathbf{r}^\star-\mathbf{1}}{\mathbf{r}^\star}\rho^{\mathbf{2}}\right) = N_{MC}\mathcal{C}_{\mathrm{HF}}\Theta(w,\rho^2)$$



Multilevel Monte Carlo

Geometrical MLMC

Accelerating the Monte Carlo method with multilevel strategies

Multilevel MC: Sampling from **several** approximations Q_M of Q (Multigrid...)

Ingredients:

- $\{M_\ell : \ell = 0, \dots, L\}$ with $M_0 < M_1 < \cdots < M_L \stackrel{\text{def}}{=} M$
- Estimation of $\mathbb{E}[Q_M]$ by means of **correction** w.r.t. the next lower level

$$Y_\ell \stackrel{\text{def}}{=} \begin{cases} Q_{M_\ell} - Q_{M_{\ell-1}} & \ell > 0 \\ Q_0 & \ell = 0 \end{cases} \xrightarrow{\textit{linearity}} \mathbb{E}[Q_M] = \mathbb{E}\left[Q_{M_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[Q_{M_\ell} - Q_{M_{\ell-1}}\right] = \sum_{\ell=0}^{L} \mathbb{E}[\mathbf{Y}_\ell]$$

- Multilevel Monte Carlo estimator

$$\hat{Q}_M^{\text{ML}} \stackrel{\text{def}}{=} \sum_{\ell=0}^{L} \hat{\mathbf{Y}}_{\ell,\mathbf{N}_\ell}^{\text{MC}} = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\mathbf{Q}_{\mathbf{M}_\ell}^{(\mathbf{i})} - \mathbf{Q}_{\mathbf{M}_{\ell-1}}^{(\mathbf{i})} \right)$$

- The Mean Square Error is

$$\mathbb{E}\left[(\hat{Q}_M^{ML} - \mathbb{E}[Q])^2\right] = \sum_{\ell=0}^{\mathbf{L}} \mathbf{N}_\ell^{-1} \mathbb{V}\mathbf{ar}[\mathbf{Y}_\ell] + (\mathbb{E}[\mathbf{Q_M} - \mathbf{Q}])^2$$

Note If $Q_M \to Q$ (in a mean square sense), then $\mathbb{V}\mathbf{ar}[\mathbf{Y}_\ell] \xrightarrow{\ell \to \infty} 0$

Geometrical MLMC

Designing a MLMC simulation: cost estimation

Let us consider the **numerical cost** of the estimator

$$\mathcal{C}(\hat{Q}_M^{ML}) = \sum_{\ell=0}^{L} N_\ell \mathcal{C}_\ell$$

Determining the **ideal number of samples** per level (i.e. minimum cost at fixed variance)

$$\left.\begin{aligned} \mathcal{C}(\hat{Q}_M^{ML}) &= \sum_{\ell=0}^{L} N_\ell \mathcal{C}_\ell \\ \sum_{\ell=0}^{L} N_\ell^{-1} \mathbb{V}ar\left[Y_\ell\right] &= \varepsilon^2/2 \end{aligned}\right\} \xrightarrow{\textit{Lagrange multiplier}} \boxed{N_\ell = \frac{2}{\varepsilon^2}\left[\sum_{k=0}^{L} \left(\mathbb{V}ar\left[Y_k\right]\mathcal{C}_k\right)^{1/2}\right]\sqrt{\frac{\mathbb{V}ar\left[Y_\ell\right]}{\mathcal{C}_\ell}}}$$

$$\boxed{\mathbb{V}ar\left[\hat{Q}_M^{ML}\right] = \sum_{\boldsymbol{\ell=0}}^{\mathbf{L}} N_\ell^{-1} \mathbb{V}ar\left(Y_\ell\right).}$$

- **MLMC** can be reinterpreted as a particular instance of **recursive control variate** (more on this later)
- MLMC has been originally introduced for problems for which it is possible to control the highest resolution (full MSE control)
- No need to estimate coefficients, but optimal for very controlled scenarios (*i.e.* discretization level)

[1] Giles, M.B., Multilevel Monte Carlo path simulation. *Oper. Res.* **56**, 607-617, 2008.

[2] Haji-Ali, A., Nobile, F., Tempone, R. Multi Index Monte Carlo: When Sparsity Meets Sampling, *Numerische Mathematik*, Vol. 132, 767–806, 2016.

Multilevel-Multifidelity Monte Carlo[2]

[2] In Collaboration with Prof. Gianluca Iaccarino (Stanford)

Multilevel-Multifidelity approach

Combination of discretization and model form

- OUTER SHELL – **Multi-level**: no need to estimate coefficient (mesh based, high correlation)

$$\mathbb{E}\left[Q_M^{\mathrm{HF}}\right] = \sum_{l=0}^{L_{\mathrm{HF}}} \mathbb{E}\left[Y_\ell^{\mathrm{HF}}\right] = \sum_{l=0}^{L_{\mathrm{HF}}} \hat{Y}_\ell^{\mathrm{HF}}$$

- INNER BLOCK – **Multi-fidelity** (*i.e.* control variate on each level)

$$Y_\ell^{\mathrm{HF},\star} = \hat{Y}_\ell^{\mathrm{HF}} + \alpha_\ell \left(\hat{\mathbf{Y}}_\ell^{\mathrm{LF}} - \mathbb{E}\left[\mathbf{Y}_\ell^{\mathrm{LF}}\right]\right)$$

Final properties of the estimator

$$\hat{Q}_M^{MLMF} = \sum_{l=0}^{L_{\mathrm{HF}}} \left[\hat{Y}_\ell^{\mathrm{HF}} + \alpha_\ell \left(\hat{\mathbf{Y}}_\ell^{\mathrm{LF}} - \mathbb{E}\left[\mathbf{Y}_\ell^{\mathrm{LF}}\right]\right)\right]$$

and

$$\mathbb{V}ar\left[\hat{Q}_M^{MLMF}\right] = \sum_{l=0}^{L_{\mathrm{HF}}} \left(\frac{1}{N_\ell^{\mathrm{HF}}} \mathbb{V}ar\left[Y_\ell^{\mathrm{HF}}\right] \left(1 - \frac{\mathbf{r}_\ell - \mathbf{1}}{\mathbf{r}_\ell} \rho_\ell^{\mathbf{2}}\right)\right)$$

Multilevel-Multifidelity

Optimal allocation across discretization and model forms

- Target accuracy for the estimator: ε^2
- Cost per level is now $\mathcal{C}_\ell^{\mathrm{eq}} = \mathcal{C}_\ell^{\mathrm{HF}} + \mathcal{C}_\ell^{\mathrm{LF}} r_\ell$
- the (constrained) optimization problem is

$$\underset{N_\ell^{\mathrm{HF}}, r_\ell, \lambda}{\operatorname{argmin}} \; (\mathcal{L}), \quad \text{where} \quad \mathcal{L} = \sum_{\ell=0}^{L_{\mathrm{HF}}} N_\ell^{\mathrm{HF}} \mathcal{C}_\ell^{\mathrm{eq}} + \lambda \left(\sum_{\ell=0}^{L_{\mathrm{HF}}} \frac{1}{N_\ell^{\mathrm{HF}}} \mathbb{V}ar \left[Y_\ell^{\mathrm{HF}} \right] \Lambda_\ell(r_\ell) - \varepsilon^2/2 \right)$$

- $\Lambda_\ell(r_\ell) = 1 - \rho_\ell^2 \dfrac{r_\ell - 1}{r_\ell}$

After the first iteration the algorithm can adjust the number of samples on the HF or LF side depending on the correlation properties discovered on flight

After the minimization ($N_\ell^{\mathrm{LF}} = N_\ell^{\mathrm{HF}} + \Delta_\ell^{\mathrm{LF}} = N_\ell^{\mathrm{HF}} r_\ell$)

$$\begin{cases} r_\ell^\star = \sqrt{\dfrac{\rho_\ell^2}{1 - \rho_\ell^2} w_\ell}, \quad \text{where} \quad w_\ell = \mathcal{C}_\ell^{\mathrm{HF}} / \mathcal{C}_\ell^{\mathrm{LF}} \\ N_\ell^{\mathrm{HF},\star} = \dfrac{2}{\varepsilon^2} \left[\displaystyle\sum_{k=0}^{L_{\mathrm{HF}}} \left(\dfrac{\mathbb{V}ar\left[Y_\ell^{\mathrm{HF}}\right] \mathcal{C}_\ell^{\mathrm{HF}}}{\mathbf{1} - \boldsymbol{\rho}_\ell^{\mathbf{2}}} \right)^{1/2} \Lambda_\ell \right] \sqrt{\left(\mathbf{1} - \boldsymbol{\rho}_\ell^{\mathbf{2}}\right) \dfrac{\mathbb{V}ar\left[Y_\ell^{\mathrm{HF}}\right]}{\mathcal{C}_\ell^{\mathrm{HF}}}} \end{cases}$$

Enhancing the CV effect
Maximizing the correlation for a fixed LF model (1/2)

Possible cures for **low-correlation** (of the discrepancy terms):

- Iteration with the application team to identify the lack of convergence
  - **LF model improvement**
- Algorithmic-contained correlation improvement
  - **Reformulation of the LF discrepancy** to gain optimality

$$\mathring{Y}_\ell^{\mathrm{LF}} = \gamma_\ell Q_\ell^{\mathrm{LF}} - Q_{\ell-1}^{\mathrm{LF}},$$

where γ_ℓ is chosen in order to maximize the correlation between Y_ℓ^{HF} and $\mathring{Y}_\ell^{\mathrm{LF}}$

Following the same MLMF approach

$$\mathbb{V}ar\left[\hat{Q}_M^{MLMF}\right] = \sum_{l=0}^{L_{\mathrm{HF}}} \left(\frac{1}{N_\ell^{\mathrm{HF}}} \mathbb{V}ar\left[Y_\ell^{\mathrm{HF}}\right] \left(1 - \frac{r_\ell - 1}{r_\ell} \rho_\ell^2 \frac{\theta_\ell^{\mathbf{2}}}{\tau_\ell}\right)\right), \quad \text{where}$$

$$\theta_\ell = \frac{\mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, \mathring{Y}_\ell^{\mathrm{LF}}\right)}{\mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, Y_\ell^{\mathrm{LF}}\right)} \qquad \tau_\ell = \frac{\mathbb{V}ar\left(\mathring{Y}_\ell^{\mathrm{LF}}\right)}{\mathbb{V}ar\left(Y_\ell^{\mathrm{LF}}\right)}$$

Enhancing the CV

Maximizing the correlation for a fixed LF model (2/2)

The optimal LF model coefficient γ_ℓ **can be computed analytically**:

$$\gamma_\ell^\star = \frac{\mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, Q_{\ell-1}^{\mathrm{LF}}\right) \mathrm{Cov}\left(Q_\ell^{\mathrm{LF}}, Q_{\ell-1}^{\mathrm{LF}}\right) - \mathbb{V}ar\left(Q_{\ell-1}^{\mathrm{LF}}\right) \mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, Q_\ell^{\mathrm{LF}}\right)}{\mathbb{V}ar\left(Q_\ell^{\mathrm{LF}}\right) \mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, Q_{\ell-1}^{\mathrm{LF}}\right) - \mathrm{Cov}\left(Y_\ell^{\mathrm{HF}}, Q_\ell^{\mathrm{LF}}\right) \mathrm{Cov}\left(Q_\ell^{\mathrm{LF}}, Q_{\ell-1}^{\mathrm{LF}}\right)}.$$

The resulting optimal allocation of samples across levels and model forms is given by

$$r_\ell^\star = \sqrt{\frac{\rho_l^2 \dfrac{\theta_\ell^2}{\tau_\ell}}{1 - \rho_\ell^2 \dfrac{\theta_\ell^2}{\tau_\ell}} w_\ell}, \quad \text{where} \quad w_\ell = \mathcal{C}_\ell^{\mathrm{HF}} / \mathcal{C}_\ell^{\mathrm{LF}}$$

$$\Lambda_\ell = 1 - \rho_\ell^2 \frac{\theta_\ell^2}{\tau_\ell} \frac{r_\ell^\star - 1}{r_\ell^\star}$$

$$N_\ell^{\mathrm{HF},\star} = \frac{2}{\varepsilon^2} \left[\sum_{k=0}^{L_{\mathrm{HF}}} \left(\frac{\mathbb{V}ar\left(Y_k^{\mathrm{HF}}\right) \mathcal{C}_k^{\mathrm{HF}}}{1 - \rho_\ell^2 \dfrac{\theta_\ell^2}{\tau_\ell}} \right)^{1/2} \Lambda_k(r_k^\star) \right] \sqrt{\left(1 - \rho_\ell^2 \frac{\theta_\ell^2}{\tau_\ell}\right) \frac{\mathbb{V}ar\left(Y_\ell^{\mathrm{HF}}\right)}{\mathcal{C}_\ell^{\mathrm{HF}}}}$$

[1] G. Geraci, M.S. Eldred & G. Iaccarino, A multifidelity control variate approach for the multilevel Monte Carlo technique. *Center for Turbulence Research, Annual Research Briefs 2015*, pp. 169–181.

[2] G. Geraci, M.S. Eldred & G. Iaccarino, A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications *19th AIAA Non-Deterministic Approaches Conference, AIAA SciTech Forum, (AIAA 2017-1951)*

Practical implementation
Budget-constrained optimization

1 (Coupled) Pilot runs for LF and HF

$$\begin{cases} r_\ell^\star = \sqrt{\frac{\rho_\ell^2}{1-\rho_\ell^2} w_\ell}, \quad \text{where} \quad w_\ell = \mathcal{C}_\ell^{\text{HF}} / \mathcal{C}_\ell^{\text{LF}} \\ N_\ell^{\text{HF},\star} = \frac{2}{\varepsilon^2} \left[\sum_{k=0}^{L_{\text{HF}}} \left(\frac{\mathbb{V}ar\left[Y_\ell^{\text{HF}}\right] \mathcal{C}_\ell^{\text{HF}}}{1-\rho_\ell^2} \Lambda_\ell \right)^{1/2} \right] \sqrt{\left(1-\rho_\ell^2\right) \frac{\mathbb{V}ar\left[Y_\ell^{\text{HF}}\right]}{\mathcal{C}_\ell^{\text{HF}}}} \end{cases}$$

2 Optimal ratio sequence (ε independent!)

$$\frac{N_\ell^{\text{HF},\star}}{N_{\ell-1}^{\text{HF},\star}} = \sqrt{\frac{(1-\rho_\ell^2)\mathbb{V}ar\left[Y_\ell^{\text{HF}}\right]}{(1-\rho_{\ell-1}^2)\mathbb{V}ar\left[Y_{\ell-1}^{\text{HF}}\right]} \frac{\mathcal{C}_{\ell-1}^{\text{HF}}}{\mathcal{C}_\ell^{\text{HF}}}}$$

$$\tau^\star = \left(\tau_1^\star = \frac{N_1^{\text{HF},\star}}{N_0^{\text{HF},\star}}, \tau_2^\star = \frac{N_2^{\text{HF},\star}}{N_1^{\text{HF},\star}}, \ldots \tau_{L-1}^\star = \frac{N_{L-1}^{\text{HF},\star}}{N_{L-2}^{\text{HF},\star}}, \tau_L^\star = \frac{N_L^{\text{HF},\star}}{N_{L-1}^{\text{HF},\star}} \right)$$

3 Given the target number $N_{\text{target}}^{\text{HF}}$ of HF runs at finer resolution L

$$\hat{N}_\ell^{\text{HF},\star} = N_{\text{target}}^{\text{HF}} \Big/ \left(\prod_{q=0}^{L-\ell-1} \tau_{L-q}^\star \right)$$

4 Optimal low fidelity simulations $N_\ell^{\text{LF}} = r_\ell^\star \hat{N}_\ell^{\text{HF},\star}$

Test problems

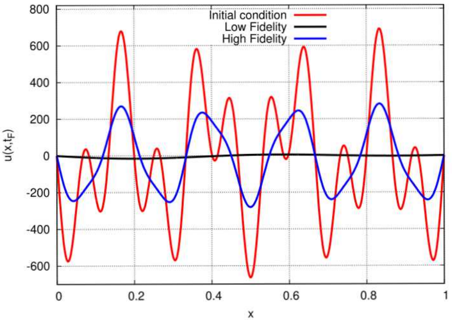
Heat equation – Parabolic 1D

Heat equation

Verification test case (we know the exact solution)

Heat-equation in presence of uncertain thermal diffusivity and initial condition:

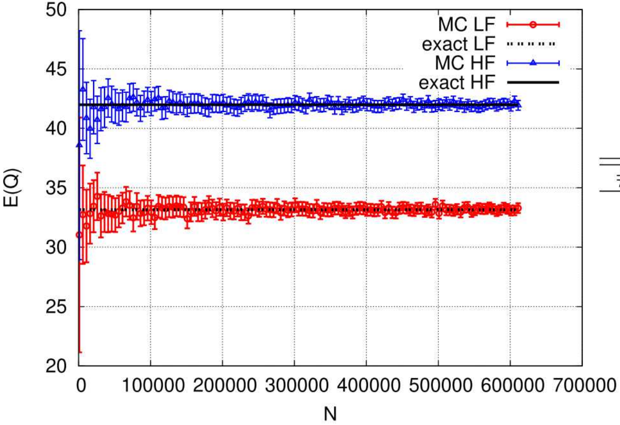
$$\begin{cases} \dfrac{\partial u(x,\boldsymbol{\xi},t)}{\partial t} - \alpha(\boldsymbol{\xi}) \dfrac{\partial^2 u(x,\boldsymbol{\xi},t)}{\partial x^2} = 0, \quad \alpha > 0,\ x \in [0,L] = \Omega \subset \mathbb{R} \\ u(x,\boldsymbol{\xi},0) = u_0(x,\boldsymbol{\xi}), \quad t \in [0,t_F] \quad \text{and} \quad \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^d \\ u(x,\boldsymbol{\xi},t)|_{\partial\Omega} = 0 \\ u_0(x,\boldsymbol{\xi}) = \mathcal{G}(\boldsymbol{\xi})\mathcal{F}_1(\mathbf{x}) + \mathcal{I}(\boldsymbol{\xi})\mathcal{F}_2(\mathbf{x}) \end{cases}$$



- **Low-fidelity:** $\bar{n}_{\text{low}} = \{1,2,3\} \to \mathbb{E}[Q_{\text{low}}] = 33.15$
- **High-fidelity:** $\bar{n}_{\text{high}} = \bar{n}_{\text{low}} \cup \{9,21\} \to \mathbb{E}[Q_{\text{high}}] = 41.98$
- Discrepancy $\mathbb{E}[Q_{\text{high}}] - \mathbb{E}[Q_{\text{low}}] = 8.83$ (21%)

Numerical results

Designing a challenging test case – MC on $N_x = 1000$

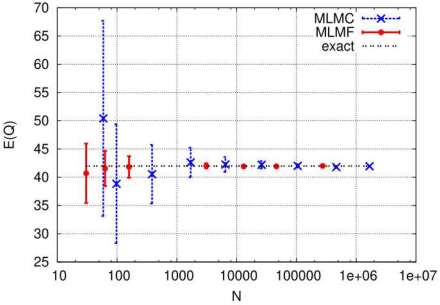


	LF	HF	
# modes	3	21	
	N_x		$\mathbf{w}_\ell$
$\ell = 0$	5	30	42
$\ell = 1$	15	60	28
$\ell = 2$	30	100	23
$\ell = 3$	60	200	23

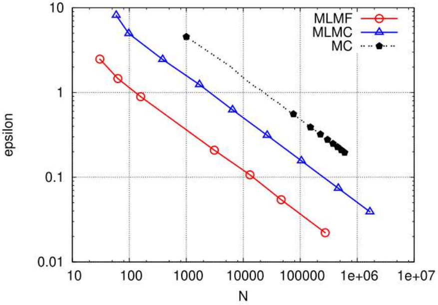
The LF cannot increase the overall accuracy because it is heavily biased...

Numerical results

Multi-level Multi-fidelity (comparison with MLMC and MC)



Expected Value



Accuracy ε

Non-linear elastic waves propagation – Hyperbolic CLAWs 1D

Elastic waves propagation in a composite material
28 uncertain variables

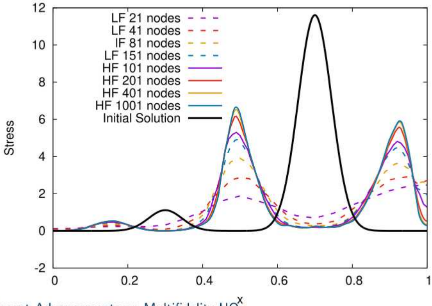
- Rod constituted by **50 layers**, two alternated materials (A and B) with constitutive laws

$$\begin{cases} \sigma_A = K_1^A \epsilon + K_2^A \epsilon^2, & K_1^A = 1 \quad \text{and} \quad K_2^A = \xi_j \quad \xi_j \sim \mathcal{U}(0.01, 0.02) \\ \sigma_B = K_1^B \epsilon + K_2^B \epsilon^2, & K_1^A = 1.5 \quad \text{and} \quad K_2^A = 0.8 \end{cases}$$

- Uncertain **initial static** $(u(x, t = 0) = 0)$ **pre-loading** state:

$$\sigma(x) = \begin{cases} \xi_3 \exp\left(-\frac{(x-0.35)(x-0.25)}{2 \times 0.002}\right) & \text{if} \quad 0 < x < 1/2 \quad \xi_3 \sim \mathcal{U}(0.5, 2) \\ \xi_2 \exp\left(-\frac{(x-0.65)(x-0.75)}{2 \times 0.002}\right) & \text{if} \quad 1/2 < x < 1 \quad \xi_2 \sim \mathcal{U}(0.5, 6.5) \end{cases}$$

- Spatially varying **uncertain density**: $\rho(x) = \xi_1 + 0.5 \sin(2\pi x), \xi_1 \sim \mathcal{U}(1.5, 2)$
- **Clamped rod** as B.C.

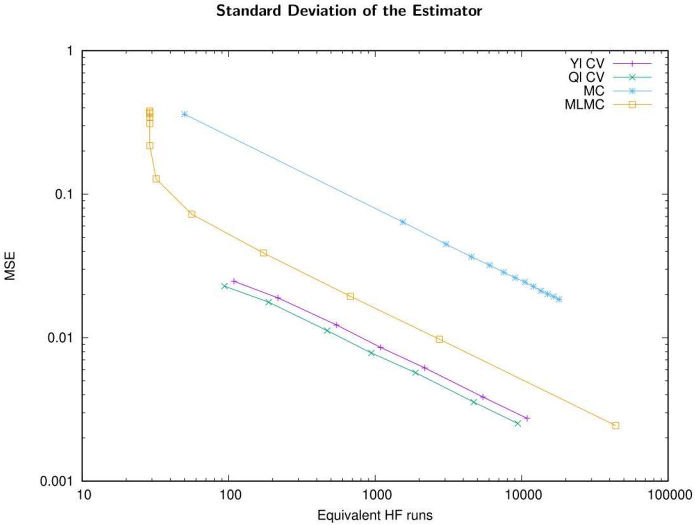


	N_x	N_t	Δ_t
Low-fidelity (GODUNOV)	21	50	3.6×10^{-3}
	41	100	1.8×10^{-3}
	81	150	1.2×10^{-3}
	151	288	6.25×10^{-4}
High-fidelity (MUSCL–van Leer)	101	200	9×10^{-4}
	201	400	4.5×10^{-4}
	401	900	2×10^{-4}
	1001	2000	9×10^{-5}

TABLE: Low- and high- fidelity simulations

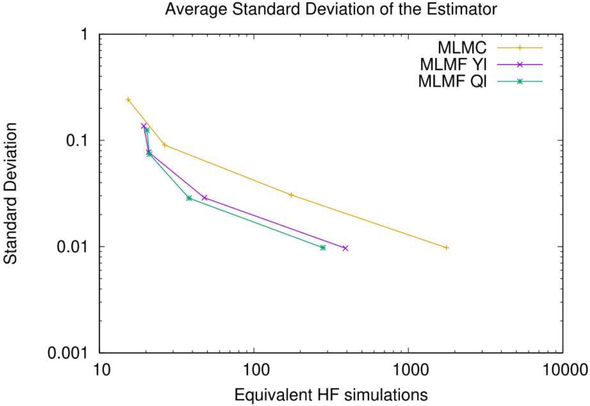
Elastic waves propagation in a composite material

28 uncertain variables



Elastic waves propagation in a composite material

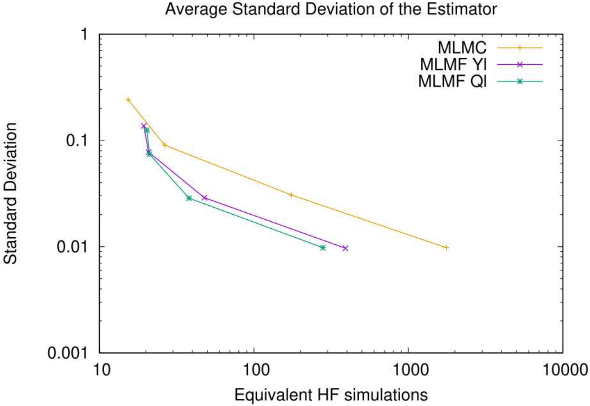
28 uncertain variables – Average of 50 realizations



Level	MLMC	MLMF-YI				MLMF-QI			
	N_ℓ	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2
0	80029	5960	243178	40	0.97	4682	192090	40	0.97
1	6282	2434	12487	4	0.49	1049	13781	12	0.83
2	1271	262	3877	14	0.82	151	3657	23	0.92
3	212	47	966	19	0.84	34	754	21	0.86

Elastic waves propagation in a composite material

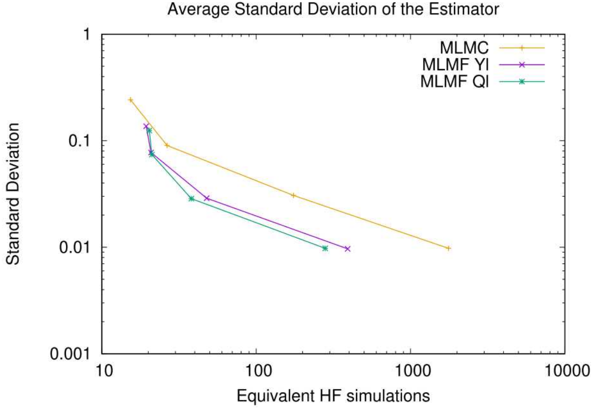
28 uncertain variables – Average of 50 realizations



Level	MLMC	MLMF-YI				MLMF-QI			
	N_ℓ	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2
0	80029	5960	243178	40	0.97	4682	192090	40	0.97
1	6282	2434	12487	4	0.49	1049	13781	12	0.83
2	1271	262	3877	14	0.82	151	3657	23	0.92
3	212	47	966	19	0.84	34	754	21	0.86

Elastic waves propagation in a composite material

28 uncertain variables – Average of 50 realizations



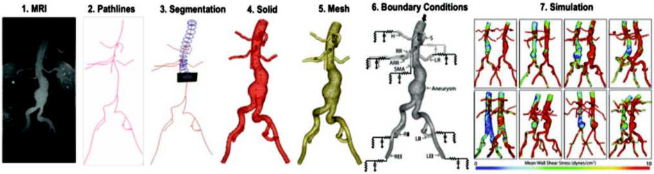
Level	MLMC	MLMF-YI				MLMF-QI			
	N_ℓ	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2	N_ℓ^{HF}	N_ℓ^{LF}	r_ℓ	ρ_ℓ^2
0	80029	5960	243178	40	0.97	4682	192090	40	0.97
1	6282	2434	12487	4	0.49	1049	13781	12	0.83
2	1271	262	3877	14	0.82	151	3657	23	0.92
3	212	47	966	19	0.84	34	754	21	0.86

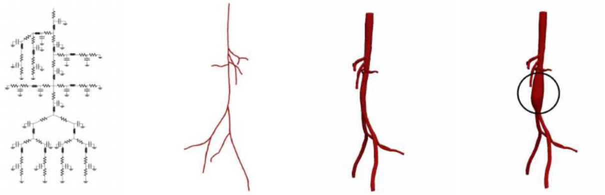
Representative Applications

Cardiovascular flow – Flow/Structure interaction

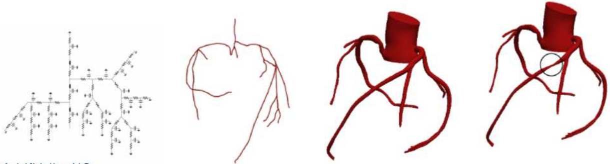
Cardiovascular Flow

In collaboration with Fleeter and Prof. Mardsen (Stanford) and Prof. Schiavazzi (Notre Dame)





(a) (b) (c) (d)



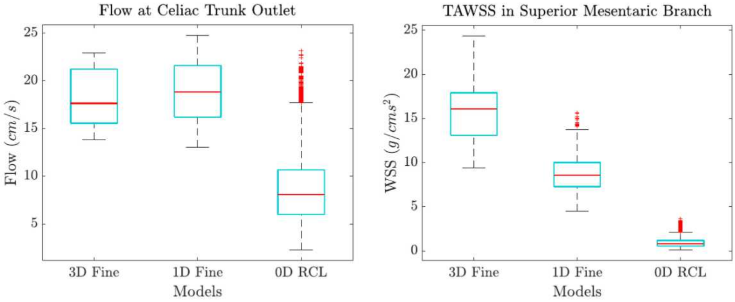
Cardiovascular Flow
Computational setting and UQ setup

	Aorto-Femoral Ranges		Coronary Ranges	
Uncertain Parameter	Min	Max	Min	Max
BC: Total R	1.0079×10^3	1.8718×10^3	1.0500×10^3	1.9500×10^3
BC: Total C	7.0000×10^{-4}	1.3000×10^{-3}	7.0000×10^{-4}	1.3000×10^{-3}
BC: Ratio of R_p/R_{total}	3.9200×10^{-2}	7.2800×10^{-2}	6.3000×10^{-2}	1.1700×10^{-1}
BC: Ratio of R_p/R_{total} (renal arteries)	1.9600×10^{-1}	3.6400×10^{-1}	—	—
Young's Modulus	4.9700×10^5	9.2300×10^5	4.9700×10^5	9.2300×10^5
Young's Modulus (coronary arteries)	—	—	8.0500×10^5	1.4950×10^6
Inlet waveform total flow	5.8333×10^1	1.0833×10^2	6.3490×10^1	1.1791×10^2
Blood Density	7.4200×10^{-1}	1.3780	7.4200×10^{-1}	1.3780
Blood Viscosity	2.8000×10^{-2}	5.2000×10^{-2}	2.8000×10^{-2}	5.2000×10^{-2}

	Aorto-Femoral Healthy		Aorto-Femoral Diseased		Coronary Healthy		Coronary Diseased	
Fidelity & Level	Cost	Effective Cost	Cost	Effective Cost	Cost	Effective Cost	Cost	Effective Cost
3D Fine Mesh	870.80 h	1	667.23 h	1	2 164.61 h	1	1 198.48 h	1
3D Medium Mesh	228.44 h	2.62×10^{-1}	157.05 h	2.35×10^{-1}	497.23 h	2.30×10^{-1}	286.88 h	2.39×10^{-1}
3D Coarse Mesh	98.02 h	1.13×10^{-1}	56.21 h	8.42×10^{-2}	78.65 h	3.63×10^{-2}	120.63 h	1.01×10^{-1}
1D Fine Mesh	11.60 m	2.22×10^{-4}	11.87 m	2.96×10^{-4}	4.33 m	3.34×10^{-5}	4.78 m	6.65×10^{-5}
1D Medium Mesh	2.95 m	5.65×10^{-5}	2.62 m	6.54×10^{-5}	1.90 m	1.46×10^{-5}	2.00 m	2.78×10^{-5}
1D Coarse Mesh	1.90 m	3.64×10^{-5}	1.52 m	3.79×10^{-5}	1.08 m	8.34×10^{-6}	1.13 m	1.58×10^{-5}
0D Full Model	0.49 m	3.64×10^{-6}	0.50 m	1.25×10^{-5}	0.17 m	7.66×10^{-5}	0.16 m	1.36×10^{-4}
0D Simple Model	0.03 m	6.60×10^{-7}	0.03 m	7.60×10^{-7}	0.03 m	2.51×10^{-7}	0.03 m	4.72×10^{-7}

Cardiovascular Flow

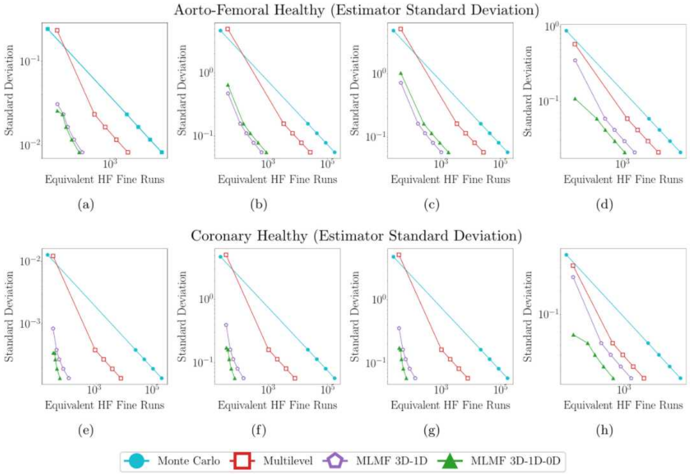
UQ Results – for more see Fleeter, Geraci *et al.*, CMAME, Volume 365, 15 June 2020, 113030



(a) (b)

Cardiovascular Flow

UQ Results – for more see Fleeter, Geraci *et al.*, CMAME, Volume 365, 15 June 2020, 113030



Legend:

- (a)-(e) Outlet flow
- (b)-(f) Outlet pressure
- (c)-(g) Time-averaged pressure
- (d)-(h) TAWSS

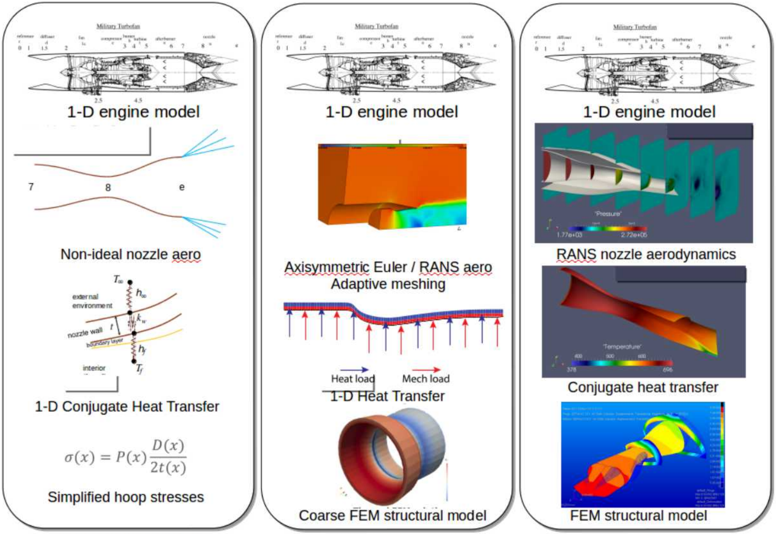
Nozzle design – Aero-Thermo-Structural interaction

Aero-Thermo-Structural analysis
Problem description

(a) X47B UCAS (b) Nozzle close-up

Figure: Northrop Grumman X-47B UCAS and close up of its nozzle (Source: `http://www.northropgrumman.com/MediaResources/Pages/MediaGallery.aspx?ProductId=UC-10028`)

Aero-Thermo-Structural analysis
Computational setting



Low-fidelity model | Medium-fidelity model | High-fidelity model

Aero-Thermo-Structural analysis
15 Uncertain Parameters

Parameter	Range
Inlet stagnation temperature [K]	897.75-992.25
Atmospheric Temperature [K]	248.9-275.1
Inlet stagnation pressure [Pa]	216,000-264,000
Atmospheric Pressure [Pa]	57,000-63,000
Thermal conductivity [W/m K]	8.064-9.856
Elastic modulus [Pa]	7.38e10-9.02e10
Thermal expansion coefficient [1/K]	1.8e-6-2.2e-6
lower Bspline 1 [-]	0.005-0.03
lower Bspline 2 [-]	0.005-0.03
lower Bspline 3 [-]	0.005-0.03
lower Bspline 4 [-]	0.005-0.03
upper Bspline 1 [-]	0.005 -0.03
upper Bspline 2 [-]	0.005-0.03
upper Bspline 3 [-]	0.005-0.03
upper Bspline 4 [-]	0.005-0.03

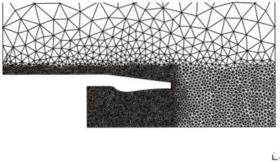
Table: Uncertain parameters for the nozzle problem.

- **HF**
  Flow: Euler
  Thermal/Stress: FEM
- **LF**
  Flow: 1D non-ideal nozzle
  Thermal/Stress: Thermal resistances and hoop model
- **LF (updated)**
  Flow: 1D non-ideal nozzle
  Thermal/Stress: FEM

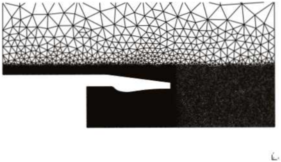

Control variate only at coarsest level!

Aero-Thermo-Structural analysis

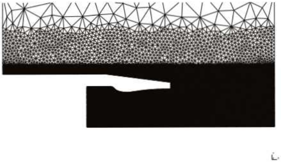
Mesh discretization hierarchy



(a) Coarse



(b) Medium



(c) Fine

FIGURE: Close up of the meshes.

	Triangles
Coarse	6,119
Medium	29,025
Fine	142,124

TABLE: Number of triangles.

	LF	HF
Coarse	0.016	0.053
Medium	N/A	0.253
Fine	N/A	1.0

TABLE: Computational cost.

Aero-Thermo-Structural analysis
Correlation and variance reduction

	LF		LF (updated)	
	correlation	Variance reduction [%]	correlation	Variance reduction [%]
Thrust	0.997	91.42	0.996	94.2
Mechanical Stress	**2.31e-5**	**2.12e-3**	**0.944**	**89.2**
Thermal Stress	**0.391**	**12.81**	**0.987**	**93.4**

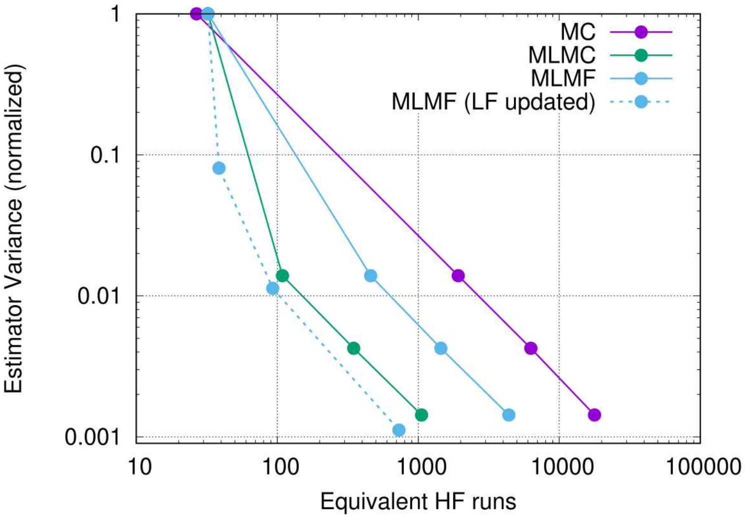
Table: Correlations and variance reduction for $\varepsilon^2/\varepsilon_0^2 = 0.001$.

Accuracy ($\varepsilon^2/\varepsilon_0^2$)	LF	HF			LF (updated)	MF		
	Coarse	Coarse	Medium	Fine	Coarse	Coarse	Medium	Fine
0.1	N/A	N/A	N/A	N/A	404	20	20	20
0.01	21,143	1,757	20	20	3,091	177	31	20
0.003	69,580	5,775	36	20	N/A	N/A	N/A	N/A
0.001	212,828	17,715	109	34	32,433	1,773	314	20

Table: Sample profiles for the LF and HF model as function of the normalized accuracy $\varepsilon^2/\varepsilon_0^2$.

Aero-Thermo-Structural analysis

Multilevel/Multifidelity efficiency



Scramjet – 2D/3D LES (Combustion)

Scramjet engines

A little bit of context: Opportunities and Challenges

Supersonic combustion ramjet (Scramjet) engines

- are propulsion systems for **hypersonic flight**
- aim at directly utilize atmospheric air for **stable combustion while maintaining supersonic airflow**
- obviates the need to carry **on-board oxidizer**
- overcome the losses from **slowing flows** to subsonic speeds (no rotating element)

Several challenges

- characterizing and predicting **combustion properties** for multiscale and multiphysical turbulent flows (under extreme environments)
- **low throughput time** vs need for mixture and self-ignition
- **stable combustion** for constant thrust

Designing an optimal engine requires

- Maximization of the combustion efficiency
- Minimization of the pressure losses, thermal loading
- Reducing the risk of unstart and flame blow-out
- Accomplishing these tasks under uncertain operational conditions (robustness and reliability)

From Jurzay (2018): *The challenge of enterprising supersonic combustion in scramjet is [...] as difficult as lighting a match in a hurricane.*

[1] Urzay, J., Supersonic Combustion in Air-Breathing Propulsion Systems for Hypersonic Flight, Annual Review of Fluid Mechanics, Vol. 50, No. 1, 2018, pp. 593627. doi:10.1146/annurev-fluid-122316-045217.

[2] Leyva, I., The relentless pursuit of hypersonic flight, Physics Today, Vol. 70, No. 11, 2017, pp. 3036. doi:10.1063/PT.3.3762.

Hypersonic International Flight Research and Experimentation (HIFiRE)

Problem description and computational setup

- The HIFiRE project studied a cavity-based hydrocarbon-fueled dual-mode scramjet configuration
- Ground test rig, HIFiRE Direct Connect Rig (HDCR), built to replicated the isolator/combustion section

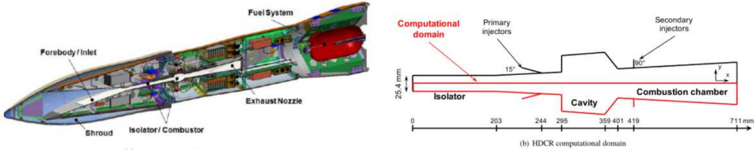


Figure: Left: HIFiRE Flight 2 payload [1]. Right: HDCR schematic.

Computational setup

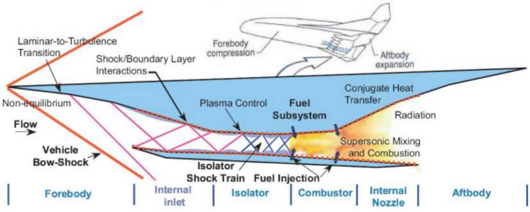
- A reduced three-step mechanism to characterize the combustion process
- Arrhenius formulations of the kinetic reaction rates (parameters are fixed at values that retain robust and stable combustion)
- Large Eddy Simulations carried out by using RAPTOR code (Prof. Joe Oefelein)

SNL LES code RAPTOR

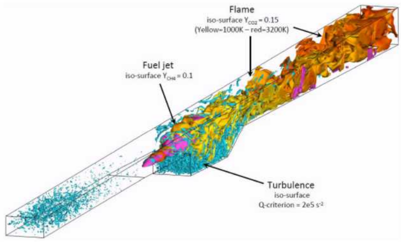
- Fully coupled conservation equations of mass, momentum, total-energy, and species for a chemically reacting flow
- can handles high Reynolds numbers
- real gas effects
- robust over wide range of Mach numbers
- non-dissipative, discretely conservative, staggered finite-volume schemes

[1] Jackson, K. R., Gruber, M. R., and Buccellato, S., HIFiRE Flight 2 Overview and Status Uptate 2011, 17th AIAA International Space Planes and Hypersonic Systems and Technologies Conference, AIAA Paper 2011–2202, San Francisco, CA, 2011. doi:10.2514/6.2011-2202.

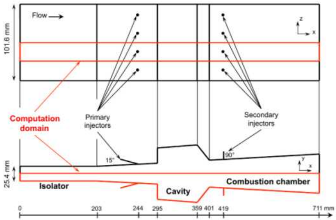
Supersonic Combusting Ramjet
Problem description

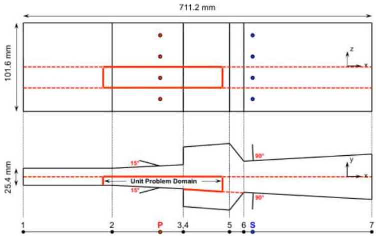


In flight



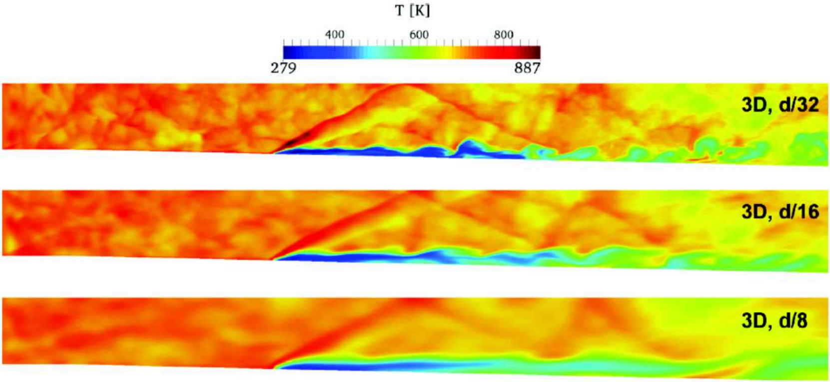
Numerical model





Scramjet

Instantaneous temperature field over different mesh resolutions



Scramjet
24 Uncertain parameters

Parameter	Symbol	Range
Inflow boundary conditions		
Inlet		
Stagnation pressure	$p_{0,i}$	1.48 MPa $\pm$ 5%
Stagnation temperature	$T_{0,i}$	1550 K $\pm$ 5%
Mach number	M_i	$2.51 \pm 10\%$
Turbulence intensity	$I_i = u_i^{'}/U_i$	$[0.0 \ - \ 0.05]$
Turbulence intensity ratio	$I_r = v_i^{'}/u_i^{'}$	1.0
Turbulence length scale	L_i	$[0.0 \ - \ 8.0]$mm
Boundary layer thickness	δ_i	$[2.0 \ - \ 6.0]$mm
Fuel injection $(36\%CH_4, 64\%C_2H_4)$		
Mass flux	$\dot{m}_f$	7.37×10^{-3} kg/s $\pm$ 10%
Static Temperature	T_f	300.0 K $\pm$ 5%
Mach Number	M_f	$1.0 \pm 5\%$
Turbulence intensity	$I_f = u_f^{'}/U_f$	$[0.025 \ - \ 0.075]$
Turbulence length scale	L_f	$[0.02 \ - \ 1.0]$ mm
Wall boundary conditions		
Wall Temperature	T_w	Profile from KLE Expansion (10 params)
Turbulence model parameters		
Static Smagorinsky		
Modified Smagorinsky constant	C_R	$[0.01 - 0.016]$
Turbulent Prandtl number	Pr_t	$[0.5 - 1.7]$
Turbulent Schmidt number	Sc_t	$[0.5 - 1.7]$

TABLE: Summary of the uncertain parameters for the SCRAMJET problem.

Scramjet

UQ results

	correlation		Variance reduction [%]	
	Coarse	Fine	Coarse	Fine
$P_{0,mean}$	0.997	0.761	93	50
$P_{0,rms,mean}$	0.875	0.593	72	30
M_{mean}	0.975	0.649	89	36
TKE_{mean}	0.824	0.454	64	17
χ_{mean}	0.450	0.714	19	44

Table: Correlations and variance reduction.

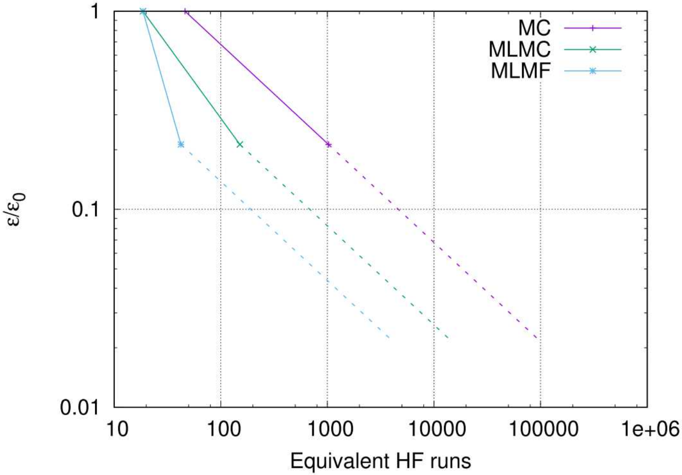
	2D	3D
$d/8$	5E-4	0.11
$d/16$	0.014	1

Table: Computational cost.

	2D	3D
$d/8$	4,191	263
$d/16$	68	**9**

Table: LES simulations (**target of 9 runs** at 3D $d/16$ and $\varepsilon^2/\varepsilon_0^2 = 0.045$).

Scramjet
UQ setting



Multilevel, Multifidelity and MLMF

Relatively large experience with realistic problems

Success stories

- PSAAP II – particle laden turbulence flow in radiative environment (collaborators: Gianluca Iaccarino, Alireza Doostan, Lluis Jofre, Hillary Fairbanks)
- Cardiovascular flows – fluid-structure (collaborators: Casey Fleeter, Daniele Schiavazzi, Alison Marsden)
- Aero-thermo-structural analysis for nozzle devices (collaborators: Juan Alonso, Gianluca Iaccarino, Paul Constantine)
- SCRAMJET engine (collaborators: Habib Najm, Cosmin Safta, Xun Huan)
- Large Eddy Simulations for wind plants (collaborators: David Maniaci, Ryan King)
- Computer networks (collaborators: Laura Swiler, Jonathan Crussell, Bert Debusschere)

Does MLMF always work better than MLMC?

- It cannot be worse than MLMC (except for the cost of the pilot samples), but not always better than MC if MLMC is outside the 'design conditions' (more on this later on)
- An example: Wind turbine analysis with LES where 3-level MLMC performed worse than a 2-level MLMC

Q: How do we ensure that our sequence of models is 'optimal'?
A: Very often you can only control the way in which you fuse information...

Approximate Control Variate

Optimal Control Variate

M low-fidelity models with known expected value

Let's consider M **low-fidelity models with known mean**. The Optimal Control Variate (OCV) is generated by adding M unbiased terms to the MC estimator

$$\hat{Q}^{\mathrm{CV}} = \hat{Q} + \sum_{i=1}^{M} \alpha_i \left(\hat{Q}_i - \mu_i \right)$$

- $\hat{Q}_i$ MC estimator for the i**th low-fidelity model**
- μ_i **known expected value** for the ith low-fidelity model
- $\underline{\alpha} = [\alpha_i, \ldots, \alpha_M]^{\mathrm{T}}$ set of **weights** (to be determined)

Let's define

- The **covariance matrix** among all the low-fidelity models: $\mathbf{C} \in \mathbb{R}^{M \times M}$
- The **vector of covariances** between the high-fidelity Q and each low-fidelity Q_i: $\mathbf{c} \in \mathbb{R}^{M}$
- $\bar{\mathbf{c}} = \mathbf{c} / \mathbb{V}ar\left[Q\right] = \left[\rho_1 \mathbb{V}ar\left[Q_1\right], \ldots, \rho_M \mathbb{V}ar\left[Q_M\right]\right]^{\mathrm{T}}$, where ρ_i is the correlation coefficient (Q, Q_i)

The optimal weights are obtained as $\underline{\alpha}^{\star} = -\mathbf{C}^{-1}\mathbf{c}$ and the variance of the OCV estimator

$$\begin{aligned} \mathbb{V}ar\left[\hat{Q}^{\mathrm{CV}}\right] &= \mathbb{V}ar\left[\hat{Q}\right] \left(1 - \bar{\mathbf{c}}^{\mathrm{T}} \mathbf{C}^{-1} \bar{\mathbf{c}}\right) \\ &= \mathbb{V}ar\left[\hat{Q}\right] (1 - R^2_{OCV}), \quad 0 \leq R^2_{OCV} \leq 1. \end{aligned}$$

For a single low-fidelity model: $R^2_{OCV-1} = \rho_1^2$

Approximate Control Variate
M low-fidelity models with unknown expected value

For complex engineering models the **expected values of the M low-fidelity models are unknown** *a priori*

- Let's define the **set of sample** used for the **high-fidelity** model: $\mathbf{z}$
- Let's consider N_i **ordered evaluations** for Q_i: $\mathbf{z_i}$ (we assume $N_i = \lceil r_i N \rceil$)
- Let's partition $\mathbf{z_i}$ in two ordered subsets $\mathbf{z_i^1} \bigcup \mathbf{z_i^2} = \mathbf{z_i}$ (note that in general $\mathbf{z}_i^1 \bigcap \mathbf{z}_i^2 \neq \emptyset$)

The **generic Approximate Control Variate** is defined as

$$\tilde{Q}\left(\underline{\alpha}, \underline{\mathbf{z}}\right) = \hat{Q}(\mathbf{z}) + \sum_{i=1}^{M} \alpha_i \left(\hat{Q}_i(\mathbf{z_i^1}) - \hat{\mu}_i(\mathbf{z_i^2}) \right) = \hat{Q}(\mathbf{z}) + \sum_{i=1}^{M} \alpha_i \Delta_i(\mathbf{z_i}) = \hat{Q} + \underline{\alpha}^{\mathrm{T}} \underline{\Delta},$$

The **optimal weights** and **variance** can be obtained as

$$\underline{\alpha}^{\mathrm{ACV}} = -\mathbb{C}ov\left[\underline{\Delta}, \underline{\Delta}\right]^{-1} \mathbb{C}ov\left[\underline{\Delta}, \hat{Q}\right]$$

$$\mathbb{V}ar\left[\tilde{Q}\left(\underline{\alpha}^{ACV}\right)\right] = \mathbb{V}ar\left[\hat{Q}\right] \left(1 - \mathbb{C}ov\left[\underline{\Delta}, \hat{Q}\right]^{\mathrm{T}} \frac{\mathbb{C}ov\left[\underline{\Delta}, \underline{\Delta}\right]^{-1}}{\mathbb{V}ar\left[\hat{Q}\right]} \mathbb{C}ov\left[\underline{\Delta}, \hat{Q}\right]\right)$$

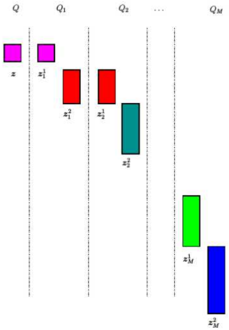
$$= \mathbb{V}ar\left[\hat{Q}\right] \left(1 - R^2_{\mathrm{ACV}}\right).$$

For a single low-fidelity model: $R^2_{ACV-1} = \frac{r_1 - 1}{r_1} \rho_1^2$ (this result does not depend on the partitioning of $\mathbf{z_1}$)

NOTES: we are going from $\mathbb{C}ov\left[Q_i, Q_j\right]$ to $\mathbb{C}ov\left[\Delta_i, \Delta_j\right]$

Recursive Difference Estimator

A recursive partitioning with independent estimators (Equivalent to MLMC for fixed bias)



MLMC can be obtained from ACV with

- $\mathbf{z_i^1} = \mathbf{z}$
- $\mathbf{z_i^2} = \mathbf{z_{i+1}^1}$ for $i = 1, \dots, M-1$
- $\alpha_\mathbf{i} = -\mathbf{1}$ for all i

$$\hat{Q}^{\text{MLMC}}(\mathbf{z}) = \hat{Q} + \sum_{i=1}^{M} (-\mathbf{1}) \left(\hat{Q}_i(\mathbf{z_i^1}) - \hat{\mu}_i(\mathbf{z_i^2}) \right)$$

$$\mathbb{V}ar\left[\hat{Q}^{\text{MLMC}}\right] = \mathbb{V}ar\left[\hat{Q}\right]\left(1 - R^2_{\text{RDiff}}\right)$$

$$R^2_{RDiff} = -\alpha_1^2\tau_1^2 - 2\alpha_1\rho_1\tau_1 - \alpha_M^2\frac{\tau_M}{\eta_M} - \sum_{1=2}^{M}\frac{1}{\eta_{i-1}}\left(\alpha_i^2\tau_i^2 + \tau_{i-1}^2\tau_{i-1}^2 - 2\alpha_i\alpha_{i-1}\rho_{i,i-1}\tau_i\tau_{i-1}\right),$$

where

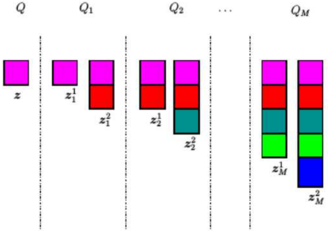
- $\tau_i = \frac{\text{Var}^{1/2}(Q_i)}{\text{Var}^{1/2}(Q)}$ and $\eta_i = \left|z_i^2\right|/N$ is the ratio between the cardinality of the sets z_i^2 and z.

NOTES

- Given the recursive nature of RDiff, we can show that $R^2_{\text{RDiff}} < \rho_1^2$ (as $r_i \to \infty$ and N is fixed)
- It is actually possible to compute an optimal set of weights instead of using $\alpha_\mathbf{i} = -\mathbf{1}$ (w-RDiff)

Multifidelity Monte Carlo (Peherstorfer, Willcox and Gunzburger, 2016)

An approximated control variate with a recursive partitioning



MFMC can be obtained from ACV with

- $\mathbf{z_i^1} = \mathbf{z_{i-1}}$ and $\mathbf{z_i^2} = \mathbf{z_i}$ for $i = 2, \ldots, M$
- $\mathbf{z_1^1} = \mathbf{z}$ and $\mathbf{z_1^2} = \mathbf{z_1}$

$$\alpha_i^{\mathrm{MFMC}} = -\frac{\mathbb{C}ov\left[Q, Q_i\right]}{\mathbb{V}ar\left[Q_i\right]}, \quad \text{for} \quad i = 1, \ldots, M,$$

and the variance of the estimator is

$$\mathbb{V}ar\left[\underline{\alpha}^{\mathrm{MFMC}}\right] = \mathbb{V}ar\left[\hat{Q}\right]\left(1 - R^2_{\mathrm{MFMC}}\right)$$

$$R^2_{\mathrm{MFMC}} = \sum_{i=1}^{M} \frac{r_i - r_{i-1}}{r_i r_{i-1}} \rho_i^2 = \rho_1^2 \left(\frac{r_1 - 1}{r_1} + \sum_{i=2}^{M} \frac{r_i - r_{i-1}}{r_i r_{i-1}} \frac{\rho_i^2}{\rho_1^2}\right).$$

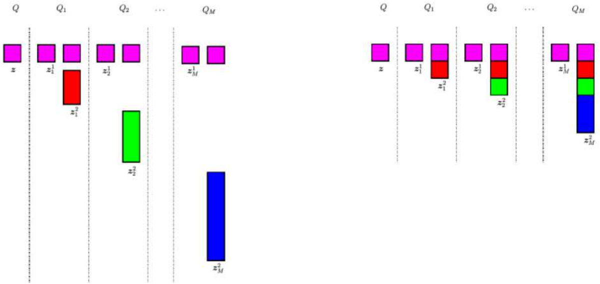
NOTES

- Given the recursive nature of MFMC, we can show that $R^2_{\mathrm{MFMC}} < \rho_1^2$ (as $r_i \to \infty$ and N is fixed)
- Surprisingly, the covariance matrix $\mathbb{C}ov\left[\underline{\Delta}, \underline{\Delta}\right]$ **is diagonal** $\to$ you can compute in close form the optimal weights, but the ability to leverage correlations among all the models is lost

Examples of Convergent Estimators

Is it possible to overcome the limitation of the recursive sampling schemes?

We proposed two sampling strategies that overcome the limitation of the recursive schemes



(a) ACV-IS sampling strategy. (b) ACV-MF sampling strategy.

As an example, let's consider the **ACV-MF estimator**

$$R^2_{\mathrm{ACV-MF}} = \left[\mathrm{diag}\left(\mathbf{F}^{(\mathrm{MF})}\right) \circ \bar{\mathbf{c}}\right]^{\mathrm{T}} \left[\mathbf{C} \circ \mathrm{diag}\left(\mathbf{F}^{(\mathrm{MF})}\right)\right]^{-1} \left[\mathrm{diag}\left(\mathbf{F}^{(\mathrm{MF})}\right) \circ \bar{\mathbf{c}}\right].$$

The matrix $\mathbf{F}^{(\mathrm{MF})} \in \mathbb{R}^{M \times M}$ **encodes the particular sampling strategy** and is defined as

$$\mathbf{F}^{(\mathrm{MF})}_{ij} = \begin{cases} \frac{\min(r_i, r_j) - 1}{\min(r_i, r_j)} & \text{if } i \neq j \\ \frac{r_i - 1}{r_i} & \text{otherwise} \end{cases}, \quad \text{for} \quad \mathbf{r_i} \to \infty, \quad \mathbf{F}^{(\mathrm{MF})} \to \mathbf{1_M} \quad \text{and} \quad \mathbf{R^2_{\mathrm{ACV-MF}}} \to \mathbf{R^2_{\mathrm{OCV}}}$$

NOTE

- **No closed form** for the optimal weights and the samples allocation per model

A parametric model problem

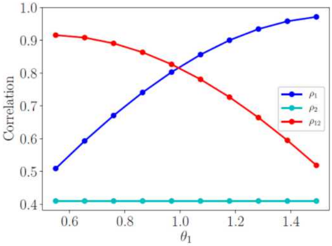
What happens for a limited number of low-fidelity simulations?

We designed a parametric test problem to explore different cost and correlation scenarios $(x, y \sim \mathcal{U}(-1, 1))$

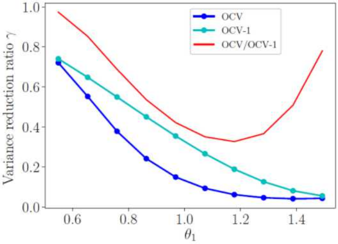
$$Q = A\left(\cos\theta\, x^5 + \sin\theta\, y^5\right)$$
$$Q_1 = A_1\left(\cos\theta_1\, x^3 + \sin\theta_1\, y^3\right)$$
$$Q_2 = A_2\left(\cos\theta_2\, x + \sin\theta_2\, y\right)$$

We use the following definitions

- $A = \sqrt{11}$, $A_1 = \sqrt{7}$, and $A_2 = \sqrt{3}$ (give unitary variance for each model)
- $\theta = \pi/2$ and $\theta_2 = \pi/6$ and θ_1 varies uniformly in the bounds $\theta_2 < \theta_1 < \theta$
- We consider a fixed cost ratio between models, *i.e.* a relative cost of 1 for Q, $1/w$ for Q_1 and $1/w^2$ for Q_2



(a) Correlations



(b) Var. reduction ratios

A parametric model problem

Comparison of different estimators (Eq. Cost 100 HF)

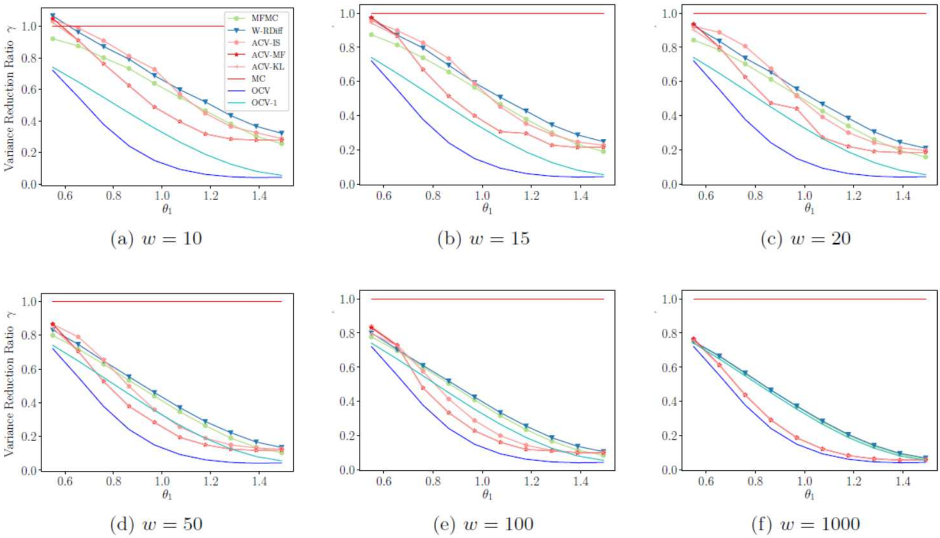


FIGURE: Variance reduction for cost ratios of $[1, 1/w, 1/w^2]$ for Q, Q_1, and Q_2

Non-linear elasticity in heterogeneous media – Hyperbolic 2D CLAWs

Non-linear elasticity in heterogeneous media
Problem setup

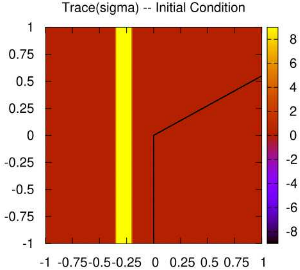
Hyperbolic system of equations describing the elastic wave propagation (normal and shear components) in two spatial dimensions for a domain with two materials

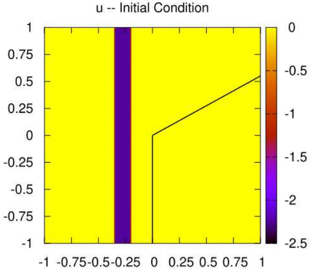
$$q_t + Aq_x + Bq_y = 0, \quad \text{where}$$

$$A = -\begin{bmatrix} 0 & 0 & 0 & (\lambda+2\mu) & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & mu \\ \frac{1}{\rho} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\rho} & 0 & 0 \end{bmatrix}, \quad B = -\begin{bmatrix} 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & (\lambda+2\mu) \\ 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & \frac{1}{\rho} & 0 & 0 \\ 0 & \frac{1}{\rho} & 0 & 0 & 0 \end{bmatrix}$$

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)} \quad \text{and} \quad \mu = \frac{E}{2(1+\nu)},$$

Parameters	ρ_l	λ_l	μ_l	ρ_r	λ_r	μ_r
Distribution	$\mathcal{U}(0.5, 1.5)$	$\mathcal{U}(3.0, 5.0)$	$\mathcal{U}(0.25, 0.75)$	$\mathcal{U}(0.5, 1.5)$	$\mathcal{U}(1.0, 3.0)$	$\mathcal{U}(0.5, 1.5)$





Non-linear elasticity in heterogeneous media

Deterministic results – CLAWPACK http://www.clawpack.org (Ver. 5.x)

	I order					II order				
Resolution	200	100	50	25	10	200	100	50	25	10
Norm. Cost	1.000	0.147	0.026	0.009	0.002	0.498	0.080	0.013	0.004	0.002

TABLE: Normalized cost with respect to the cost of the second order 200×200 resolution.

HF: top row – **LF:** bottom row

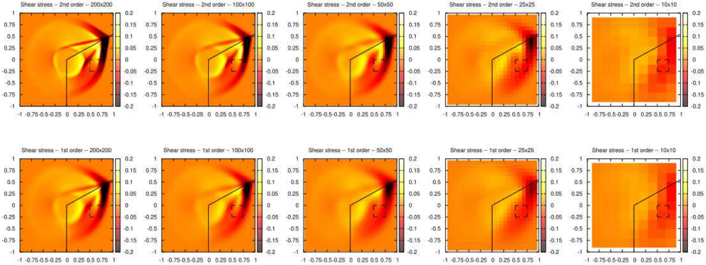


FIGURE: Shear stress at final time 0.5 for the two model fidelities (top and bottom rows) and the five discretization levels (200 x 200, 100 x 100, 50 x 50, 25 x 25, 10 x 10 from left to right) corresponding to the mean values of the random parameters. The QoI is the average value of the shear in the dashed region within the right material.

Non-linear elasticity in heterogeneous media

Correlation matrix

200 (II)	100 (II)	50 (II)	25 (II)	10 (II)	200 (I)	100 (I)	50 (I)	25 (I)	10 (I)
1.00000	0.99838	0.99245	0.96560	0.70267	0.99312	0.98333	0.93857	0.85400	0.56719
0.99838	1.00000	0.99092	0.96461	0.69060	0.99160	0.98380	0.93360	0.84743	0.55127
0.99245	0.99092	1.00000	0.98759	0.76255	0.99866	0.99484	0.96738	0.89785	0.63184
0.96560	0.96461	0.98759	1.00000	0.83904	0.98697	0.99400	0.99102	0.94874	0.71607
0.70267	0.69060	0.76255	0.83904	1.00000	0.76356	0.79165	0.89148	0.96032	0.96725
0.99312	0.99160	0.99866	0.98697	0.76356	1.00000	0.99700	0.96965	0.90058	0.63184
0.98333	0.98380	0.99484	0.99400	0.79165	0.99700	1.00000	0.98022	0.92207	0.66156
0.93857	0.93360	0.96738	0.99102	0.89148	0.96965	0.98022	1.00000	0.97785	0.78607
0.85400	0.84743	0.89785	0.94874	0.96032	0.90058	0.92207	0.97785	1.00000	0.89023
0.56719	0.55127	0.63184	0.71607	0.96725	0.63184	0.66156	0.78607	0.89023	1.00000

Table 6: Correlation matrix for the ten models used in the elastic equation problem Equation (45). The second-order (II) and the first-order (I) schemes both employ five different resolution levels.

Non-linear elasticity in heterogeneous media
Algorithms performance under three realistic scenarios

- **Single fidelity (coarsening only): HF:** 200 (II), **LF:** 100 (II), 50 (II), 25 (II), 10 (II)
- **MultiFidelity + Coarsening: HF:** 200 (II), **LF:** 100 (I), 50 (I), 25 (I), 10 (I)
- **MultiFidelity + Aggressive Coarsening: HF:** 200 (II), **LF:** 50 (I), 25 (I), 10 (I)

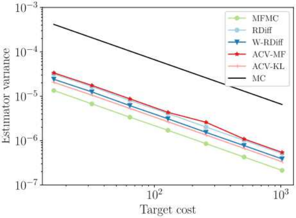


FIGURE: **Coarsening only**

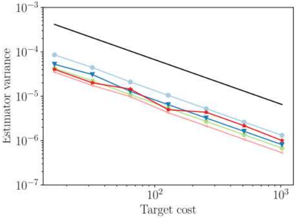


FIGURE: **MF + Coarsening**

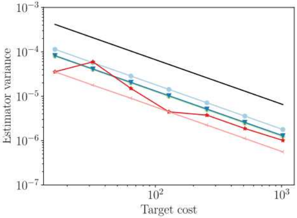


FIGURE: **MF + Aggr. Coarsening**

Nozzle design – A more realistic engineering example

Aero-thermo-structural analysis of a jet engine nozzle

Computational setup (Data courtesy of Jeff Hokanson and Paul Constantine, CU Boulder)

Operative conditions

- Reconnaissance mission for an high-subsonic aircraft
- Most critical condition is the top-of-climb (Required thrust is 21 500 N) @ 40 000 ft and Mach 0.51

Nozzle structure Two layers separated by an air gap

- Inner thermal layer: ceramic matrix composite
- Outer load layer: composite sandwich material (titanium honeycomb between two layers of graphite-bismaleimide Gr/BMI)

Uncertain parameters 40 uncertain parameters – mix of uniform and log-normal variables

- 35 material properties variables
- 2 atmospheric conditions
- 2 inlet conditions
- 1 heat transfer coefficient

Quantities of Interest (QoIs)

- **Mass** as a surrogate for the cost of the device
- **Thrust** for the aerodynamics performance
- A temperature failure criterion in the inner load layer (**Thermal stresses**)
- A strain failure criterion in the thermal layer (**Mechanical stresses**)

Numerical exploration of the OCV/ACV performance

Computational setup (Data courtesy of Jeff Hokanson and Paul Constantine, CU Boulder)

- Exploration of the theoretical performance for ACV, *i.e.* $R^2_{OCV} > R^2_{OCV-1}$

CFD	FEM (Thermal/Structural)	Cost
1D	COARSE	2.63e-04
Euler 2D COARSE	COARSE (axisymmetric)	9.69e-04
Euler 2D MEDIUM	MEDIUM (axisymmetric)	3.18e-03
Euler 2D FINE	FINE (axisymmetric)	9.05e-03
Euler 3D COARSE	COARSE	1.16e-02
Euler 3D MEDIUM	MEDIUM	3.58e-02
RANS 3D COARSE	COARSE	1.00

Table: Relative computational cost for several model fidelities for the nozzle problem. All the cost are normalized with respect to the 3D RANS solver.

	Variance reduction		
QoI	**OCV**	**OCV-1**	**Ratio OCV/OCV-1**
Thrust	0.020595	0.050432	0.41
Thermal stresses	0.0043612	0.0075662	0.58
Mechanical stresses	6.2981e-04	0.011720	0.05

Table: Performance of OCV and OCV-1 for the nozzle problem and three different QoIs.

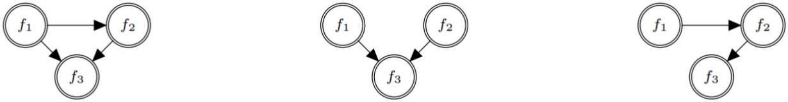
- A separation between OCV and OCV-1 exists for all QoIs
- OCV-1 attains more than one order of magnitude reduction over MC
- For Thrust and Thermal stresses an additional 60% and 40% reduction can be gained with OCV
- For the Mechanical stresses the additional benefit is larger than 90%

How do we cover even more arbitrary models relationships?

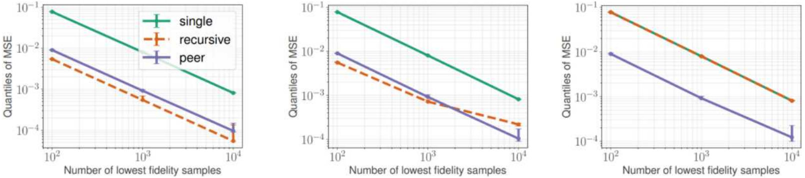
MFNets: Multifidelity networks

A framework to encode arbitrary relationships between information sources

A simple three model case



(a) Full (b) Peer (c) Hierarchical



(a) No corrupted data (b) 1% corruption (c) 10% corruption

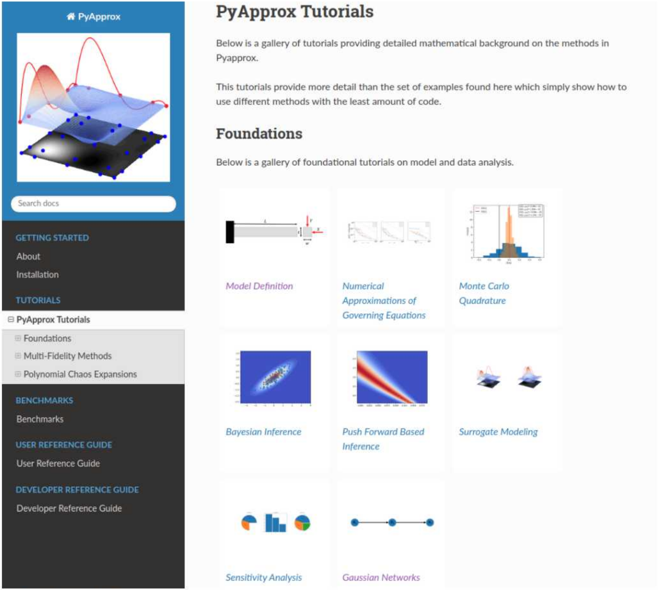
[1] Gorodetsky, Jakeman, Geraci, Eldred, MFNets: Multi-fidelity data-driven networks for Bayesian learning and prediction. *International Journal for Uncertainty Quantification*, In press, 2020.

[2] Gorodetsky, Jakeman, Geraci, MFNets: Learning network representations for multifidelity surrogate modeling *Journal of Computational Physics*, Under reviw, 2020.

A note on software

PyApprox: a research-oriented software for UQ

Software and tutorials on Multifidelity UQ – https://sandialabs.github.io/pyapprox/index.html



PyApprox

Search docs

GETTING STARTED

About

Installation

TUTORIALS

⊟ PyApprox Tutorials

⊞ Foundations

⊞ Multi-Fidelity Methods

⊞ Polynomial Chaos Expansions

BENCHMARKS

Benchmarks

USER REFERENCE GUIDE

User Reference Guide

DEVELOPER REFERENCE GUIDE

Developer Reference Guide

PyApprox Tutorials

Below is a gallery of tutorials providing detailed mathematical background on the methods in Pyapprox.

This tutorials provide more detail than the set of examples found here which simply show how to use different methods with the least amount of code.

Foundations

Below is a gallery of foundational tutorials on model and data analysis.

Model Definition

Numerical Approximations of Governing Equations

Monte Carlo Quadrature

Bayesian Inference

Push Forward Based Inference

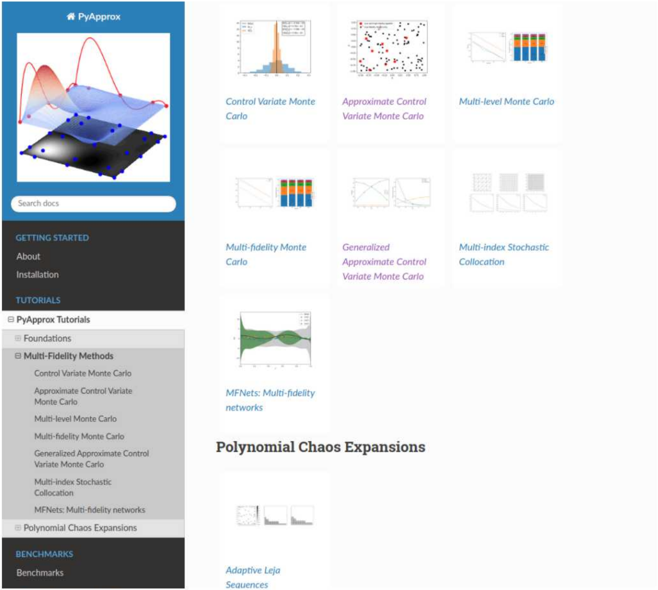
Surrogate Modeling

Sensitivity Analysis

Gaussian Networks

PyApprox: a research-oriented software for UQ

Software and tutorials on Multifidelity UQ – https://sandialabs.github.io/pyapprox/index.html



Topic II

Leveraging Active Directions for Multifidelity UQ

Can we enhance correlation between models?
Multifidelity UQ on the reduced (shared) space

Core Question
Q: Can we identify a shared space between models (possibly with independent/non-shared parameterization) where the correlation is higher?
A: Active Subspace method seems well suited for this (but this idea is not limited to it)

Pivotal idea and its main features

- For each model one can search for **Active Directions** independently
- If the **input variables** of a models are **standard Gaussian variables** then the Active Variables are also standard Gaussian variables
- Therefore, for each model the QoI can be represented on a (possibly reduced) space characterized by a join standard Gaussian distribution
- We can sample along these shared Active Directions and **'map back'** to the original coordinates of **each model separately**

Some Questions:

- How do we treat the inactive variables?
- What if the model input are not Gaussian variables?
- What does it happen if the Active Directions are different between models? We expect this to happen often in practice
- Why is this even supposed to work from a physical standpoint?

Active Subspaces in a Nutshell

(Almost) everything you need to know to use it with Multifidelity – see Constantine (2015) for more

We consider a black-box approach, *i.e.* the QoI Q is obtained through a computational model f given a vector of input parameters $\mathbf{x}$

$$\mathbf{x} \rightarrow \boxed{f(\mathbf{x})} \rightarrow Q$$

- Vector of Input parameters: $\mathbf{x} \in \mathbb{R}^m$ with joint distribution $\rho(\mathbf{x})$
- Let's introduce the $m \times m$ matrix $\mathbf{C}$

$$\mathbf{C} = \int \left(\vec{\nabla} f\right) \left(\vec{\nabla} f\right)^{\mathrm{T}} \rho(\mathbf{x}) \mathrm{d}\,\mathbf{x}$$

- Since $\mathbf{C}$ is I) Positive semidefinite and II) Symmetric, it exists a real eigenvalue decomposition

$$\mathbf{C} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{\mathrm{T}}, \text{ where}$$

- $\mathbf{W}$ is the $m \times m$ orthogonal matrix whose columns are the normalized eigenvectors
- $\mathbf{\Lambda} = \mathrm{diag}\left\{\lambda_1, \ldots, \lambda_m\right\}$ and $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$

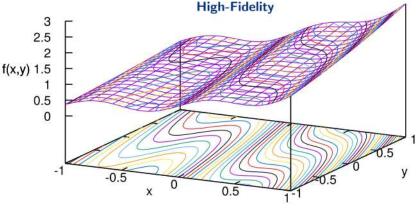
Let's define two sets of variables

$$\begin{cases} \mathbf{y} = \mathbf{W}_A^{\mathrm{T}}\mathbf{x} \in \mathbb{R}^n & \text{(Active)} \\ \mathbf{z} = \mathbf{W}_I^{\mathrm{T}}\mathbf{x} \in \mathbb{R}^{(m-n)} & \text{(Inactive)} \end{cases} \implies \mathbf{x} = \mathbf{W}_A\mathbf{y} + \mathbf{W}_I\mathbf{z} \approx \mathbf{W}_A\mathbf{y}$$

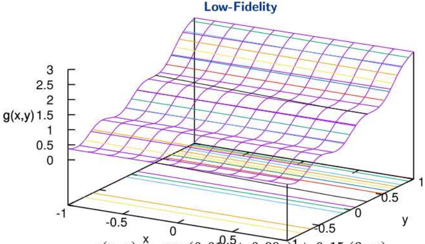
Linearity: $\boxed{\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})}$ $\left(\mathcal{X} = \mathbb{R}^m\right)$ then $\mathcal{Y} = \left\{\mathbf{y} \in \mathbb{R}^n, \mathbf{y} = \mathbf{W}_A^{\mathrm{T}}\mathbf{x}, \mathbf{x} \in \mathbb{R}^m\right\}$ and $\boxed{\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbb{I})}$

This is true for each model, i.e. there will always be a shared space between different models (even if they have a different parameterization)

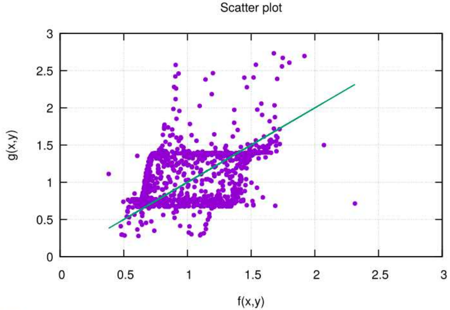
A Quick Demonstration – Gaussian Input
Low-correlated models (Correlation squared 0.05)



$$f(x, y) = \exp(0.7x + 0.3y) + 0.15\,(2\pi x)$$



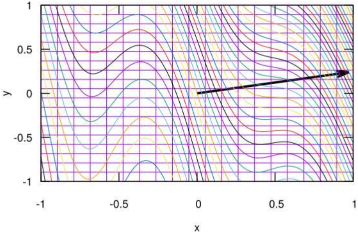
$$g(x, y) = \exp(0.01x + 0.99y) + 0.15\,(3\pi y)$$

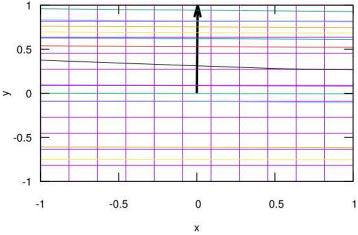


A Quick Demonstration – Gaussian Input

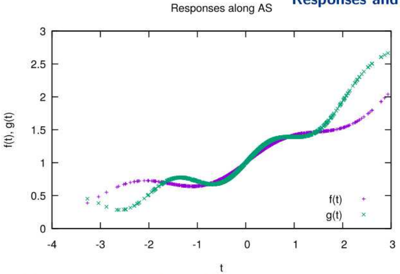
Important Directions in Actions (Correlation Squared from 0.05 to 0.9)

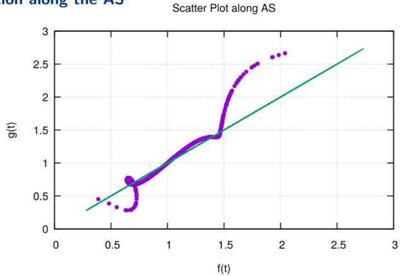
Independent Important Directions





Responses and Correlation along the AS





A Quick Demonstration – Gaussian Input

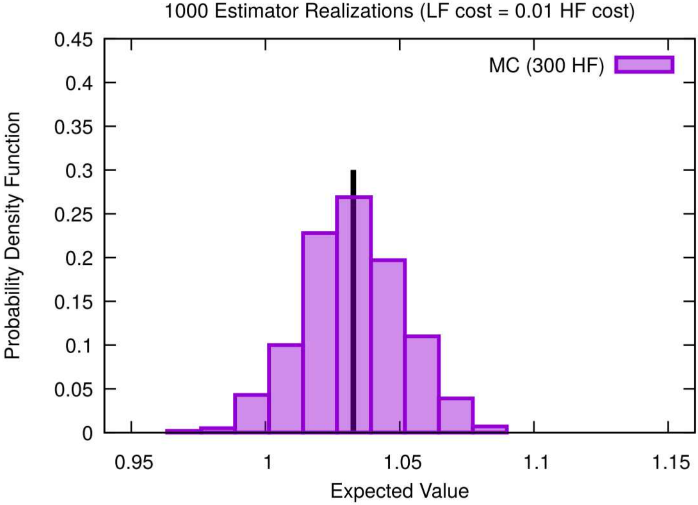
Numerical Experiment Setup

We performed the following numerical experiment:

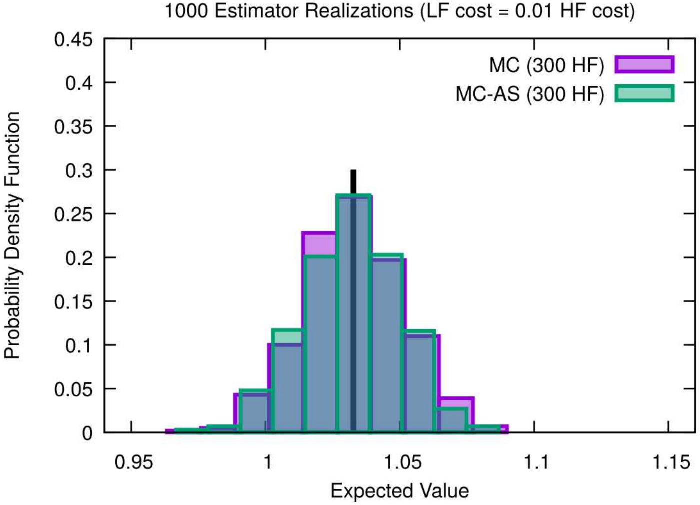
- We **fix a computational budget** (300 HF runs)
- We compute **1000 realizations for each estimator**
- For MF estimator the cost of the total set of HF+LF runs is considered
- We report the pdf of the estimated Expected Value

NOTE 1: For this problem the expected value is known
NOTE 2: In this example the AS are searched for each estimator realization during the pilot sample phase (this cost is not included, but they can be reused if needed...)

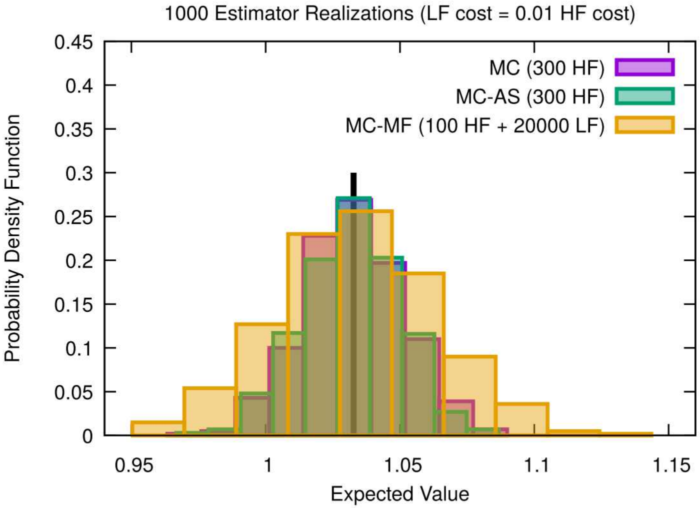
A Quick Demonstration
Monte Carlo versus Control Variate



A Quick Demonstration
Monte Carlo versus Control Variate

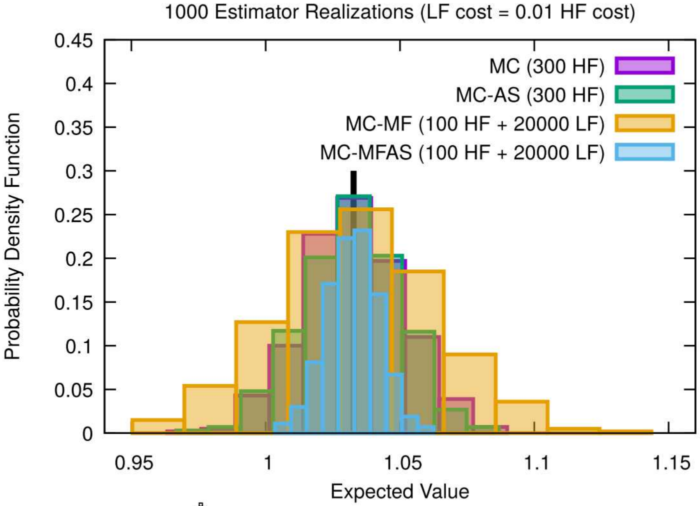


A Quick Demonstration
Monte Carlo versus Control Variate



A Quick Demonstration

Monte Carlo versus Control Variate



Same computational cost for all the estimators!

Why is this supposed to work from a physical point-of-view?

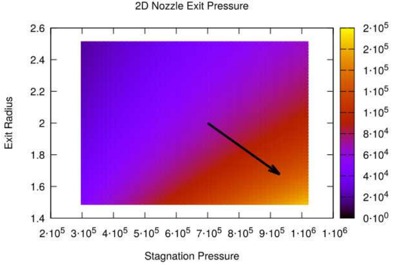
Active Directions let emerge the underlying physics

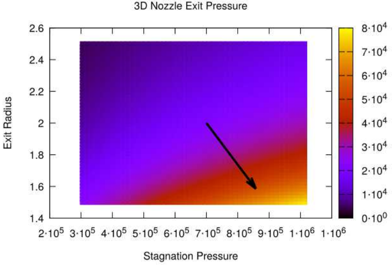
As an example, let consider the **supersonic isoentropic flow** in a diverging nozzle (sonic throat)

$$P_e = P_0 \left(1 + \frac{\gamma - 1}{2} M_e^2\right)^{-\frac{\gamma}{\gamma - 1}}, \quad \text{where}$$

$$\underset{M_e}{\text{argmin}}\, \mathcal{L} = f(M_e) - \frac{A_e}{A^\star} \quad \text{with} \quad f(M_e) = \frac{1}{M_e}\left[\frac{2}{\gamma+1}\left(1 + \frac{\gamma-1}{2} M_e^2\right)\right]^{\frac{\gamma+1}{2(\gamma-1)}}$$

- Given the shape of the nozzle (and its exit radius h_e), we can imagine 2 possible choices: 3D axisymmetric and 2D planar
- The area ratio $(A_e/A^\star)$ is linear in the 2D case (h_e/h_t) and quadratic in the 3D case (h_e^2/h_t^2)
- Given the same longitudinal shape, the 3D nozzle lets the fluid expands more than the 2D nozzle





Why is this supposed to work from a physical point-of-view?

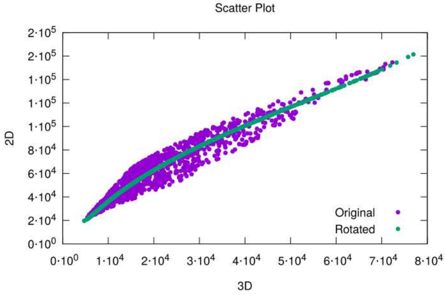
Active Directions let emerge the underlying physics ($\rho^2 = 0.9 \rightarrow 0.99$)

As an example, let consider the **supersonic isoentropic flow** in a diverging nozzle (sonic throat)

$$P_e = P_0 \left(1 + \frac{\gamma - 1}{2} M_e^2\right)^{-\dfrac{\gamma}{\gamma - 1}}, \quad \text{where}$$

$$\underset{M_e}{\text{argmin}}\, \mathcal{L} = f(M_e) - \frac{A_e}{A^\star} \quad \text{with} \quad f(M_e) = \frac{1}{M_e}\left[\frac{2}{\gamma + 1}\left(1 + \frac{\gamma - 1}{2} M_e^2\right)\right]^{\dfrac{\gamma + 1}{2(\gamma - 1)}}$$

- Given the shape of the nozzle (and its exit radius h_e), we can imagine 2 possible choices: 3D axisymmetric and 2D planar
- The area ratio $(A_e/A^\star)$ is linear in the 2D case (h_e/h_t) and quadratic in the 3D case (h_e^2/h_t^2)
- Given the same longitudinal shape, the 3D nozzle lets the fluid expands more than the 2D nozzle



Lid- and Buoyancy-driven cavity flow – A CFD example

Lid- and Buoyancy-driven Cavity Flow
Test case generalities

Physical test case

- **Combination** of the Lid- and Buoyancy-driven test cases
- **Navier-Stokes** equations for a fluid with density ρ and kinematic viscosity ν enclosed in a square cavity of size L
- **Top wall sliding** with velocity U_L
- Top and bottom walls held at **different temperature** $\rightarrow$ **net body force** (buoyancy term via Boussinesq approx.)
- Adiabatic side walls
- Cavity immersed in a gravity field with components g_h and g_v
- Nominal conditions: $Re = 1000$ and $Ra = 100000$ for air $Pr = 0.71$ (constant)

Non-dimensional parameters

$$Re = \frac{U_L L}{\nu}$$

$$Gr = |g| \frac{\beta(T_h - T_c)L^3}{\nu^2}$$

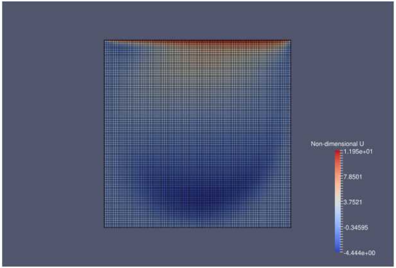
$$Pr = \frac{\nu}{\alpha}$$

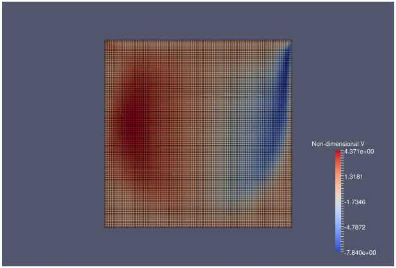
$$Ra = Pr\, Gr$$

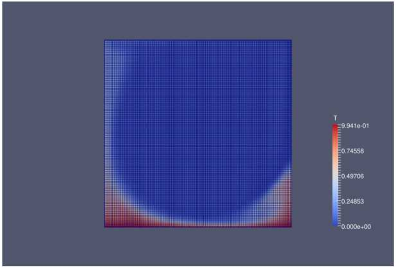
Numerical approach

- Implicit FV code on structured mesh with pressure-based SIMPLE discretization and dual-time stepping
- BC imposed via ghost cells

Lid- and Buoyancy-driven Cavity Flow
Flow field for the nominal conditions







Lid- and Buoyancy-driven Cavity Flow

Multifidelity UQ case

- **HF:** 101×101 spatial cells, $T = 80$ and $Dt = 0.25 \rightarrow \mathcal{C}^{\mathrm{HF}} = 1$
- **LF:** 21×21 spatial cells, $T = 15$ and $Dt = 0.5 \rightarrow \mathcal{C}^{\mathrm{LF}} = 0.00107$

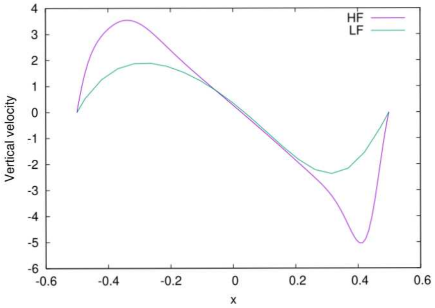


FIGURE: Vertical velocity profile at the horizontal mid-plane of the cavity for the reference condition for both HF and LF models.

Lid- and Buoyancy-driven Cavity Flow
Multifidelity parametrization

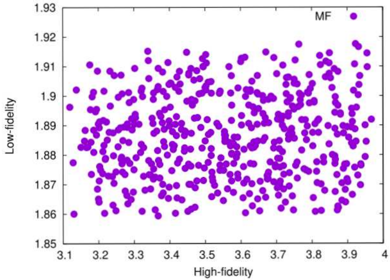
Parameter	Min	Max	Mean
ν	**0.009**	**0.011**	**0.01**
ΔT	**9**	**11**	**10**
g_v	8.1	9.9	9
g_h	3.6	4.4	4
U_L	9	11	10

TABLE: Ranges for the uniform variables of the cavity problem.

Let's have a look at the non-dimensional numbers (Pr is constant and $Gr = Gr(Ra, Re)$ for this case)

$$Re = Re(\nu, U_L)$$

$$Ra = Ra(g_v, g_h, \Delta T, \nu)$$



Lid- and Buoyancy-driven Cavity Flow
Multifidelity parametrization

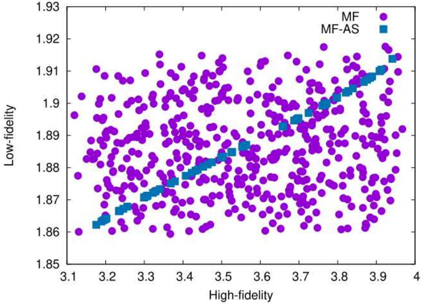


Figure: Scatter plot corresponding to 500 realizations of the HF and LF model with samples drawn in the physical space and 60 samples drawn along the common active direction.

Variable	Model	
	HF	LF
ν	-0.0860585	-0.31282
ΔT	-0.0036777	0.94981
g_v	-0.0057946	-
g_h	-0.0144436	-
U_l	0.9961617	-

Table: Dominant eigenvectors for the cavity problem.

Lid- and Buoyancy-driven Cavity Flow
Numerical test for Multifidelity

1 Fixed number of pilot samples equal to 30 samples (in the **physical space**)

2 AS evaluated (first order regression, no derivatives) from the pilot samples and **this sample set is discarded**

3 Initialization of the MF algorithm with 30 samples in the Active variables to estimate the correlation

4 Optimal oversampling ratio for the LF and perform the mean estimation

- Items (1-4) are **repeated 300 times** and the estimated mean are reported
- In mean we used an equivalent cost of **34 HF samples per estimator realization** (this number is used **for MC**, 300 repetitions)
- Variance of the mean estimator reduced by one order of magnitude

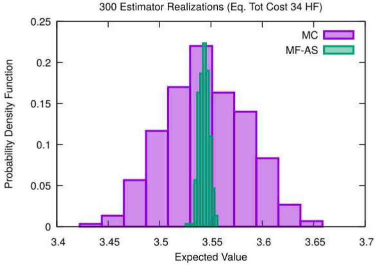


FIGURE: Probability density function for the estimators computed with 300 independent realizations.

Nozzle design – Aero-thermo-structural analysis

Preliminary results for the Sequoia problem
Problem setup

- We only consider the ACV-1 estimator here, but the extension to ACV is straightforward
- The high-fidelity model is 3D Euler with a COARSE mesh
- The low-fidelity model is 2D Euler with either a consistent or inconsistent parametrization, *i.e.* the area of the duct is forced to correspond to the one of 3D geometry

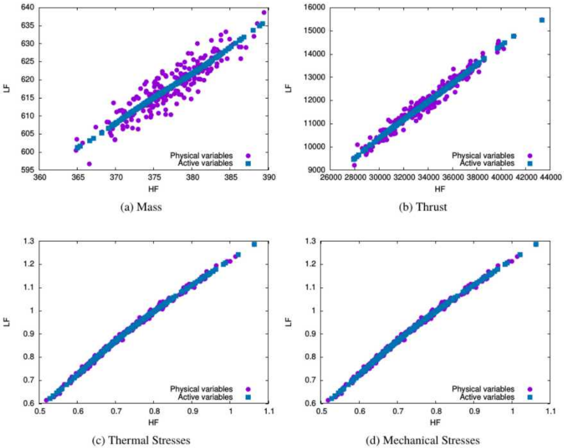
CFD	FEM (Thermal/Structural)	Parameterization	Cost
3D Euler COARSE	COARSE		1.00
2D Euler COARSE	COARSE (axisymmetric)	Consistent	0.201
2D Euler COARSE	COARSE (axisymmetric)	Inconstistent	0.135

Table: Relative computational cost for the models used for the Active Subspace tests for the nozzle problem. All the costs are normalized with respect to the 3D Euler COARSE solver.

We considered three scenarios

1. High- and low-fidelity model with **inconsistent parametrization** evaluated for the **same set of samples** (40 UQ parameters);
2. High- and low-fidelity model with **consistent parametrization** evaluated at an **independent set of samples** (40 UQ parameters);
3. High- and low-fidelity model with **inconsistent parametrization** evaluated for the same set of nominal samples **(96 + 40 UQ parameters)**.

Preliminary results for the Sequoia problem
Scenario 3 – inconsistent parametrization and dimensionality 136 vs 40



(a) Mass (b) Thrust
(c) Thermal Stresses (d) Mechanical Stresses

Figure: QoIs w.r.t. the active variable for the nozzle problem in the case of inconsistent parameterization for both the original data and the PCE regression with respect to the active variable (Scenario 3).

Preliminary results for the Sequoia problem

Scenario 3 – inconsistent dimensionality 136 vs 40

QoIs	ρ^2	$\rho^2_{\mathbf{AS}}$	Estimator St.Dev		
			MC	OCV-1	OCV-1 (AS)
Mass	0.822	0.999	1	0.178	0.001
Thrust	0.956	0.998	1	0.044	0.002
Thermal Stress	0.982	0.998	1	0.018	0.002
Mechanical Stress	0.985	0.986	1	0.015	0.014

TABLE: (Estimated) Standard Deviation for OCV-1 and OCV-1 (AS) (normalized w.r.t. MC) for the Sequoia application problem in the case of inconsistent parameterization and uncertain design input in HF (Scenario 3).

These results are estimated through the PCE along the active directions. We need to confirm the results by running the model

Supersonic Combustion – A challenging multiphysics problem

RAPTOR code
Computational features

RAPTOR

- Fully coupled conservation equations of mass, momentum, total-energy, and species for a chemically reacting flow
- can handles high Reynolds numbers
- real gas effects
- robust over wide range of Mach numbers
- non-dissipative, discretely conservative, staggered finite-volume schemes

Numerical settings

- 2D simulations
- 3 grid resolutions where cell sizes are 1/8, 1/16, and 1/32 of the injector diameter $d = 3.175$ mm (denoted as $d/8$, $d/16$, and $d/32$)
- $63K$, $250K$ and $1M$ grid points, respectively
- adaptive time steps with approximately equal simulation physical time
- warm start from a quasi-steady state nominal condition run
- 1.7×10^3, 1.1×10^4, and 7.3×10^4 CPU hours per run, respectively
- Roughly a cost factor equal to 8 between resolution levels

RAPTOR CODE

EXAMPLE OF FLOW FIELDS

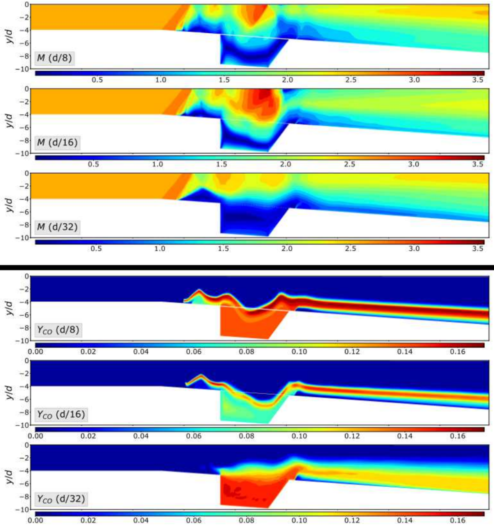


FIGURE: Solution fields of Mach number M (top three) and carbon monoxide mass fraction Y_{CO} (bottom three) simulated at a randomly sampled input settings using the three different grids.

SCRAMJET
Quantities of Interest (5)

- **Combustion efficiency (η_{comb})**, defined based on static enthalpy quantities

$$\eta_{\text{comb}} = \frac{H(T_{\text{ref}}, Y_{\text{e}}) - H(T_{\text{ref}}, Y_{\text{ref}})}{H(T_{\text{ref}}, Y_{\text{e,ideal}}) - H(T_{\text{ref}}, Y_{\text{ref}})}.$$

- **Burned equivalence ratio (ϕ_{burn})** is defined to be equal to $\phi_{\text{burn}} \equiv \phi_G \eta_{\text{comb}}$.
- **Stagnation pressure loss ratio (P_{stagloss})** is defined as

$$P_{\text{stagloss}} = 1 - \frac{P_{s,e}}{P_{s,i}}.$$

- **Maximum and average root-mean-square (RMS) pressures (max P_{rms} and ave P_{rms})** are, respectively, the maximum RMS pressure across the entire spatial domain, and the RMS pressure averaged across the spatial domain between two injectors:

$$\max P_{\text{rms}} = \max_{x,y} \sqrt{\overline{P(x,y)^2} - \left[\overline{P(x,y)}\right]^2},$$

$$\text{ave}\, P_{\text{rms}} = \frac{1}{\mathcal{V}} \int_{x,y} \sqrt{\overline{P(x,y)^2} - \left[\overline{P(x,y)}\right]^2}\, dx\, dy.$$

- **Initial shock location (x_{shock})** is the most upstream shock location.

SCRAMJET

Uncertain parameters (11)

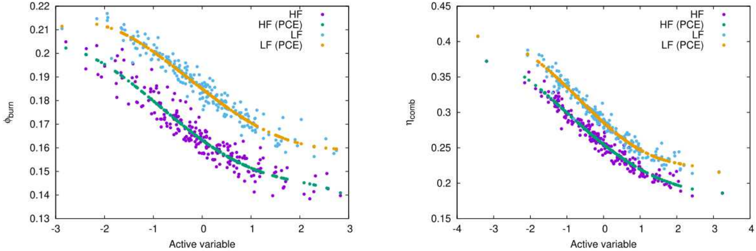
Parameter	Range	Description
Inlet boundary conditions:		
p_0	$[1.406, 1.554] \times 10^6$ Pa	Stagnation pressure
T_0	$[1472.5, 1627.5]$ K	Stagnation temperature
M_0	$[2.259, 2.761]$	Mach number
I_i	$[0, 0.05]$	Turbulence intensity horizontal component
R_i	$[0.8, 1.2]$	Ratio of turbulence intensity vertical to horizontal components
L_i	$[0, 8] \times 10^{-3}$ m	Turbulence length scale
Fuel inflow boundary conditions:		
I_f	$[0, 0.05]$	Turbulence intensity magnitude
L_f	$[0, 1] \times 10^{-3}$ m	Turbulence length scale
Turbulence model parameters:		
C_R	$[0.01, 0.06]$	Modified Smagorinsky constant
Pr_t	$[0.5, 1.7]$	Turbulent Prandtl number
Sc_t	$[0.5, 1.7]$	Turbulent Schmidt number

TABLE: Uncertain model input parameters. The uncertain distributions are assumed uniform across the ranges shown.

Scramjet dataset

Multifidelity approach from dataset

- 2 spatial resolutions
- 16 random variables (11 uncertainties + 5 design parameters)
- Dataset with 200 realizations (consistent parameterization)

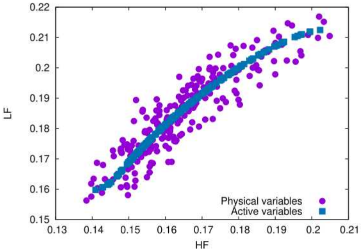


(a) ϕ_{burn} (b) η_{comb}

FIGURE: QoIs w.r.t. the active variables for the scramjet application problem.

Scramjet dataset
Multifidelity approach from dataset

- 2 spatial resolutions
- 16 random variables
- Dataset with 200 realizations (consistent parameterization)



(a) ϕ_{burn} (b) η_{comb}

FIGURE: Scatter plot for the active variables for the scramjet application problem.

Scramjet dataset
Multifidelity approach from dataset

- 2 spatial resolutions
- 16 random variables
- Dataset with 200 realizations (consistent parameterization)

QoIs			Estimator St.Dev		
	ρ^2	ρ^2_{AS}	MC	OCV-1	OCV-1 (AS)
ϕ_{burn}	0.802	0.967	1	0.198	0.033
η_{comb}	0.933	0.986	1	0.067	0.014

Table: (Estimated) Standard Deviation for MF and MF-AS (normalized w.r.t. MC) for the scramjet application problem.

TOPIC III

Surrogate-based UQ in action: Multifidelity Bayesian Calibration[3]

[3] In collaboration with: Tom Seidl (SNL), Friedrich Menhorn (TUM) and Ryan King (NREL)

Characterization and design of wind plan systems
Sandia National Laboratories Scaled Wind Farm Technology (SWiFT)

FIGURE: From
https://energy.sandia.gov/programs/renewable-energy/wind-power/wind_plant_opt/

Sandia National Laboratories Scaled Wind Farm Technology (SWiFT)

- Located at Texas Tech Universitys National Wind Institute Research Center in Lubbock, Texas
- Principal facility for investigating wind turbine wakes as part of the U.S. Department of Energy Atmosphere to Electrons research initiative (DOE-A2e)

Site features

- **Research-grade turbines**: three variable-speed variable pitch modified Vestas V27 wind turbines with full power conversion and extensive sensor suites
- **Highly characterized site:** more than two years of historical data

Computational tools

Wide range of model fidelities from Engineering Models to LES

Several computational models can be used for wind energy applications:

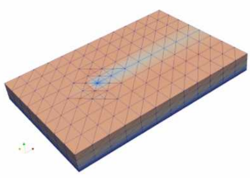
- **Nalu**: a generalized unstructured massively parallel low Mach flow code built on the Sierra Toolkit and Trilinos solver Tpetra solver stack
- **WindSE:** a python package that uses a FEniCS backend to perform wind farm simulations and optimization

WindSE

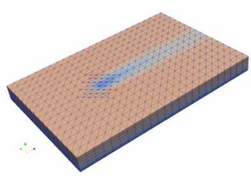
- Medium fidelity tool for 3D Reynolds-averaged Navier Stokes (RANS) simulations
- Turbines are represented by means of **non-rotating** actuator disks
- Turbulence closure via mixing length
- Based on FEniCS which enables easy user customization of finite elements, mesh discretizations, turbulence models, and turbine representation

Bayesian inversion

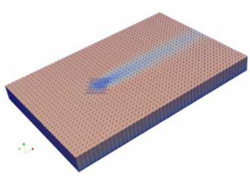
Wake characterization for a V27 rotor – Problem setup



(a) Coarse: 5252 DoFs



(b) Medium: 33428 DoFs



(c) Fine: 228064 DoFs

Model Resolution	N_x	$N_y = N_z$	Cost (s)
Coarse	12	8	8.51
Medium	24	16	60.4
Fine	48	32	1270

Table: Multilevel model hierarchy unrefined grid discretization and simulation cost.

Bayesian inversion

Wake characterization for a V27 rotor – Problem setup

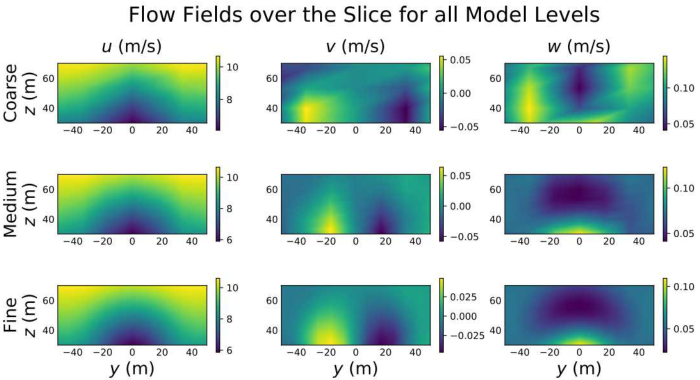


FIGURE: Nominal output for three velocity components u, v and w over all models.

Bayesian inversion

Wake characterization for a V27 rotor – Problem definition

Param	u_H ($\frac{m}{s}$)	α	θ_{wind} ($^\circ$)	Effective Thickness (m)	Axial Induction Factor	ℓ_{max} (m)
LB	8.25	0.02	-15	2.4	0.15	3.5
UB	8.75	0.5	15	15	0.9	15

Table: Uniform parameter bounds for the forward and inverse UQ studies.

"Experimental data" from Nalu

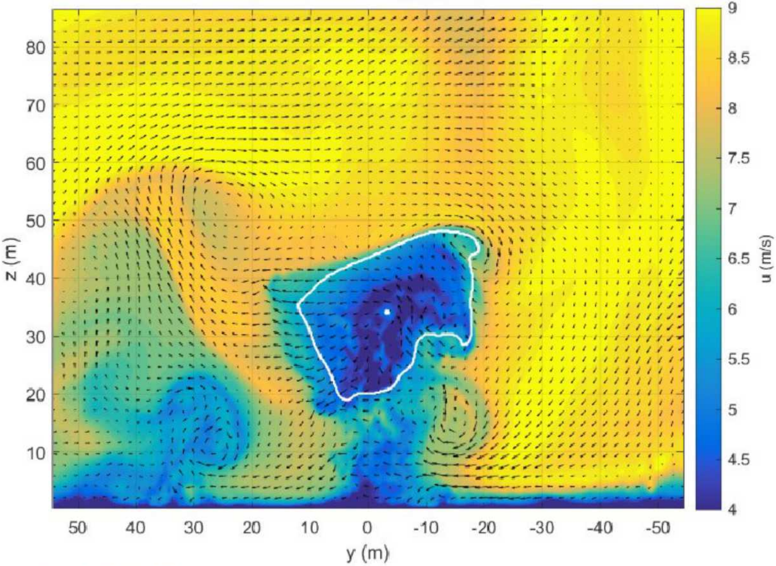
- 100 m $\times$110 m slice 5D downstream (135m)
- Data acquired each second for 10 minutes
- Reference data are averaged

RANS data

- First tests demonstrated that the misfit between the data was dominated by boundary layer data
- We truncated the spatial region of interest to $30 < z < 70$ (total of 131×161 points)
- The total number of QoIs to be considered is $31\,395$

Bayesian inversion

Wake characterization for a V27 rotor – HF (Nalu-Wind) snapshot



Bayesian inversion

ML PCE construction and performance

Tolerance	5e-4		5e-5		5e-6	
PCE Type	SF	ML	SF	ML	SF	ML
Coarse Evaluations	N/A	129	N/A	409	N/A	1201
Medium Evaluations	N/A	53	N/A	137	N/A	601
Fine Evaluations	81	13	209	17	433	61
Equivalent Fine Evaluations	17		27		99	
ML Speedup	4.9		8.0		4.4	

Table: Number of model evaluations for SF (single high-fidelity) and ML (multilevel) PCEs for three tolerances. The construction of each ML PCE requires less than a quarter of the cost of the corresponding SF model.

Bayesian inversion

ML PCE statistics

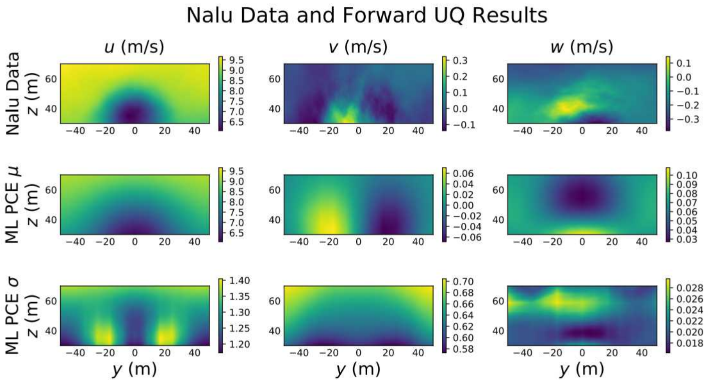
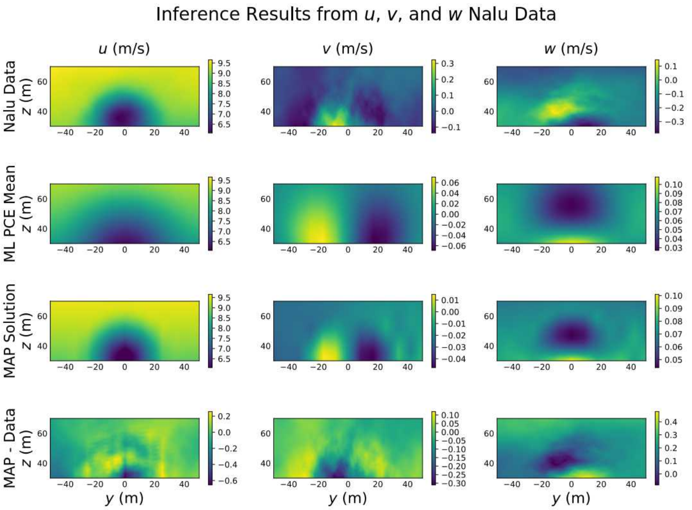


FIGURE: ML PCE built for all velocity components compared with the time-averaged Nalu slice data. The mean u component resembles the Nalu data but the other components do not due to the model error between WindSE and Nalu.

Bayesian inversion

Wake characterization for a V27 rotor – MAP solution



Bayesian inversion
Posterior distribution

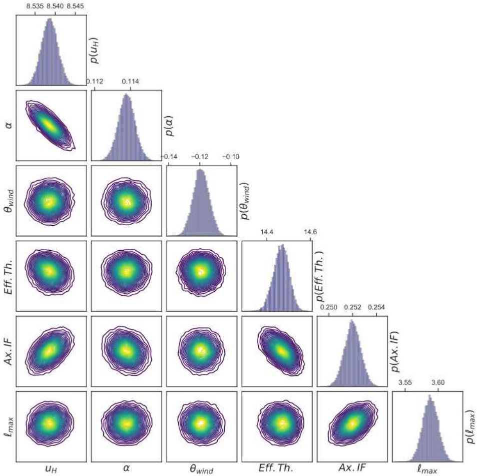


Figure: Visualization of the six-dimensional posterior distribution obtained through emulator-based inference from *all* velocity components. Marginal distributions are shown as histograms and pairwise joint distributions are displayed as contour plots.

Conclusions

Concluding Remarks
Open Challenges

We have both advanced the **state-of-the-art in multilevel/multifidelity UQ** and developed an experience in **deploying these techniques** to several application areas (aerospace, biomedical, energy, cybersecurity, *etc.*)

Can we still improve our frameworks/understanding?
A number of **outstanding challenges** still remain, a non exhaustive list:

1. How do we **exploit very large model ensemble** by efficiently discovering the relationships among models?
2. Can we take advantage of a **multi-physics** context?
3. **Optimization Under Uncertainty and Reliability/Safety** analysis require the **estimation of higher-order moments**, **rare events**, *etc.* (coll. with Prof. Marzouk, MIT and Friedrich Menhorn, TUM/MIT, Prof. Daniel Tartakovsky, Stanford)
4. **Global Sensitivity Analysis** (coll. with Prof. Gremaud and Michael Merritt, NCSU)
5. Can we integrate **online deterministic error** estimators in our multilevel/multifidelity workflow? (coll. with Prof. Guglielmo Scovazzi, Duke)
6. Can we extend our AS approach to other **dimension reduction strategies**? (coll. with Xiaoshu Zeng and Prof. Roger Ghanem, USC)
7. Can we **build low-fidelity models on-line** with a data-driven approach (*e.g.* ROM and Machine Learning)? (coll. with Dr. Patrick Blonigan, Francesco Rizzi, SNL and Ahmad Rushdi, SNL)

...

Concluding Remarks

Still an active research area

Summary:

- Multifidelity strategies are appealing techniques for UQ
- Hiererchical/Recursive estimators are limited in their performance
- ACV is a new framework to overcome this issue
- MFNets generalize this concept and enable to encode more flexible and arbitrary relationships
- Enhancing the correlation seems also possible by resorting to Active Directions/Latent Variables
- Sampling and surrogates are complementary tools, *e.g.* (MF) surrogates are very helpful for inference

(Incomplete) list of references:

- N. Metropolis, *The beginning of the Monte Carlo Method*, Los Alamos Science, No. 15, Special Issue 1987.
- Mike Giles' website: `https://people.maths.ox.ac.uk/gilesm/` (I've borrowed some material from his lectures)
- *Monte Carlo Methods* by Johansen and Evers, Lecture note. University of Bristol
- Pasupathy et al, *Control-variate estimation using estimated control means,* IIE Transactions **44**(5), 381–385, 2014.
- Halton, J. H., *A retrospective and prospective survey of the Monte Carlo method.* SIAM Review, 12, 163, 1970.
- G. Geraci, M.S. Eldred & G. Iaccarino, A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications *19th AIAA Non-Deterministic Approaches Conference, AIAA SciTech Forum, (AIAA 2017-1951)*
- A.A. Gorodetsky, G. Geraci, M.S. Eldred & J.D. Jakeman, A Generalized Framework for Approximate Control Variates. *Journal of Computational Physics*, 2020.
- G. Geraci, M.S. Eldred, Leveraging Intrinsic Principal Directions for Multifidelity Uncertainty Quantification. *Sandia Report SAND2018-10817*, 2018.
- G. Geraci, M.S. Eldred, A.A. Gorodetsky & J.D. Jakeman, Recent advancements in Multilevel-Multifidelity techniques for forward UQ in the DARPA Sequoia project. *AIAA Scitech 2019 Forum*
- G Geraci, F Menhorn, X Huan, C Safta, Y Marzouk, HN Najm, MS Eldred, Progress in Scramjet Design Optimization Under Uncertainty Using Simulations of the HIFiRE Direct Connect Rig. *AIAA Scitech 2019 Forum*
- A. Gorodetsky, J. Jakeman, G. Geraci, M. Eldred, MFNets: Multi-fidelity data-driven networks for Bayesian learning and prediction. *International Journal for Uncertainty Quantification*, In press, 2020.
- A. Gorodetsky, J. Jakeman, G. Geraci, MFNets: Learning network representations for multifidelity surrogate modeling *Journal of Computational Physics*, Under reviw, 2020. `https://arxiv.org/pdf/2008.02672.pdf`
- PyApprox: `https://sandialabs.github.io/pyapprox/index.html`

THANKS!

Acknowledgements

- DARPA Equips Program
- Laboratory Directed Research & Development Funds @ Sandia
- DOE EERE through the A2e program

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 Energys National Nuclear Security Administration under contract DE-NA-0003525.

Backup material

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

(Pseudo-)random generators use **DETERMINISTIC** algorithms to generate only **APPARENTLY RANDOM** numbers

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0,1)$
  - Conversion of the RVs to desired distribution

(Pseudo-)random generators use **DETERMINISTIC** algorithms to generate only **APPARENTLY RANDOM** numbers

Properties for a **good random generator**

- Several statistical tests exist to measure randomness, therefore reliable software has been verified against them
- A long period is needed before the sequence repeats (at least 2^{40} is required)
- A control-based *seed* is provided to skip to an arbitrary point of the sequence (useful in parallel applications)

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

(Pseudo-)random generators use **DETERMINISTIC** algorithms to generate only **APPARENTLY RANDOM** numbers

Properties for a **good random generator**

- Several statistical tests exist to measure randomness, therefore reliable software has been verified against them
- A long period is needed before the sequence repeats (at least 2^{40} is required)
- A control-based *seed* is provided to skip to an arbitrary point of the sequence (useful in parallel applications)

Bottom line...

- do not use your own generator, but use reputable sources
- For instance, Intel Math Kernel Library (MKL) are free

Pre-processing
Variable transformation

- Random generators produce uniform RV $\mathcal{U}(0,1)$, but usually we need other distributions
- Let's assume that the cumulative distribution function F_Ξ for a variable ξ is available
$$F_\Xi(\xi) = P(\Xi \leq \xi)$$
- The random generator produces $U \sim \mathcal{U}(0,1)$, *i.e.* $F_U(u) = u$
- We want to determine the function $g(U)$ which gives $\Xi = g(U)$ with cdf $F_\Xi(\xi)$
- We write the cdf for $F_\Xi(\xi)$
$$F_\Xi(\xi) = \mathrm{P}(\Xi \leq \xi) = \mathrm{P}(g(U) \leq \xi)$$
- We also assume:
- The function g is invertible on its range
- The function g is strictly increasing (only for simplicity)
$$F_\Xi(\xi) = \mathrm{P}(g(U) \leq \xi) = \mathrm{P}(U \leq g^{-1}(\xi)) = F_U(g^{-1}(\xi)) = g^{-1}(\xi)$$
- Finally we can choose $g^{-1}(\xi) = F_\Xi(\xi)$, *i.e.* $\Xi = F_\Xi^{-1}(U)$ in order to get the desired distribution

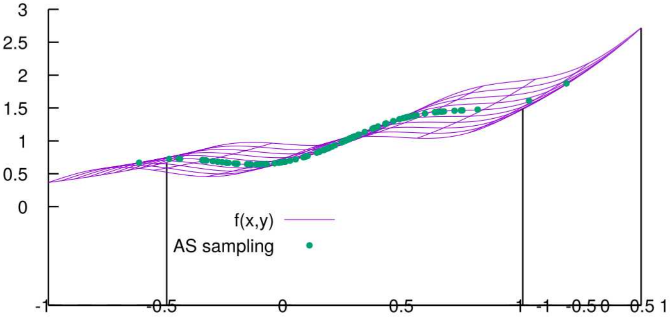
What about the inactive variables?

How do you treat the inactive variables?

$$\mathbf{x} = \mathbf{W_A y} + \mathbf{W_{NA} z}$$

- Given a sample along the Active Variable $\mathbf{y}$, we need to recover $\mathbf{x}$
- This mapping is ill-posed (infinitely many $\mathbf{x}$ exist)
- One possible regularization: conditional expected value of f given $\mathbf{y}$

$$f_{AS}(\mathbf{y}) = \int f(\mathbf{W_A y} + \mathbf{W_{NA} z})\, \rho_{\mathbf{z}|\mathbf{y}}\, \mathrm{d}\,\mathbf{z} \approx f(\mathbf{W_A y} + \mathbf{W_I} \mathbb{E}[\mathbf{z}]) \int \rho_{\mathbf{z}|\mathbf{y}}\, \mathrm{d}\,\mathbf{z} = f(\mathbf{W_A y})$$



Joint Normality: Is this required?

Non linear transformation embedded in the black-box approach

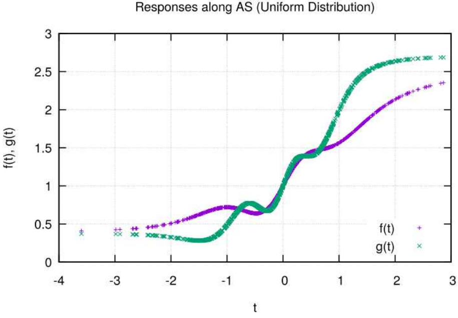
Q: Is the assumption of **joint-normality** on the input space **of the model** required?
A: No, a normal distribution is used only for the AS mapping in order to obtain a shared space between models

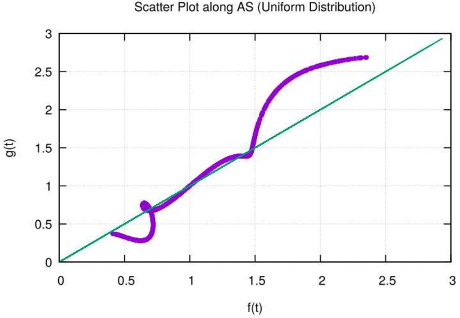
Let's assume, for example $x_i \sim \mathcal{U}(-1,1)$ and $\boldsymbol{\omega}_i \sim \mathcal{N}(0,1)$, we can define (*i.e.* Rosenblatt, Nataf, *etc.*) a non linear function $\mathbf{x} = h(\boldsymbol{\omega})$ such that

$$\boldsymbol{\omega} \rightarrow \boxed{\boxed{h(\omega)} \rightarrow \mathbf{x} \rightarrow \boxed{f(\mathbf{x})}} \rightarrow Q, \quad \text{where} \quad x_i = h(\omega_i) = \operatorname{erf}\left(\frac{\omega_i}{\sqrt{2}}\right)$$

From an AS perspective, only $\boldsymbol{\omega}$ exists (however, for each $\boldsymbol{\omega}$ we can obtain $\mathbf{x}$)

$$\boldsymbol{\omega} = \mathbf{W_A y} + \mathbf{W_{NA} z} \approx \mathbf{W_A t}$$





Joint Normality: Is this required?
Non linear transformation embedded in the black-box approach

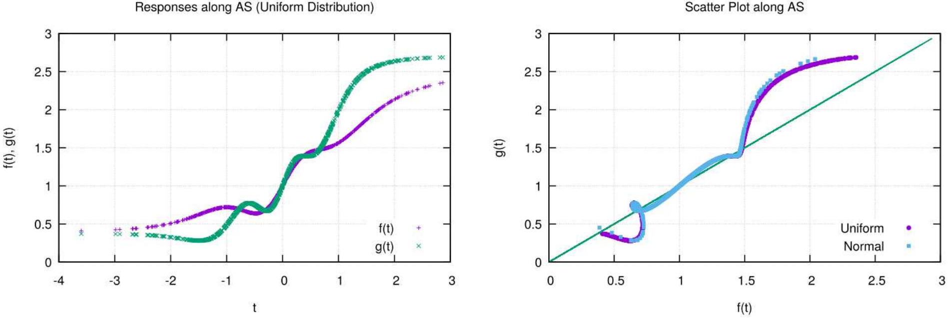
Q: Is the assumption of **joint-normality** on the input space **of the model** required?
A: No, a normal distribution is used only for the AS mapping in order to obtain a shared space between models

Let's assume, for example $x_i \sim \mathcal{U}(-1,1)$ and $\boldsymbol{\omega}_i \sim \mathcal{N}(0,1)$, we can define (*i.e.* Rosenblatt, Nataf, *etc.*) a non linear function $\mathbf{x} = h(\boldsymbol{\omega})$ such that

$$\boldsymbol{\omega} \rightarrow \boxed{\boxed{h(\omega)} \rightarrow \mathbf{x} \rightarrow \boxed{f(\mathbf{x})}} \rightarrow Q, \quad \text{where} \quad x_i = h(\omega_i) = \text{erf}\left(\frac{\omega_i}{\sqrt{2}}\right)$$

From an AS perspective, only $\boldsymbol{\omega}$ exists (however, for each $\boldsymbol{\omega}$ we can obtain $\mathbf{x}$)

$$\boldsymbol{\omega} = \mathbf{W_A y} + \mathbf{W_{NA} z} \approx \mathbf{W_A t}$$



Dissimilar parameterization
Additional input variable for the high-fidelity model

$$f(x,y,z) = \exp(0.7x + 0.3y) + 0.15\sin(2\pi x) + \mathbf{0.75z^3}, \quad \text{where} \quad z \sim \mathcal{N}(0, 1/3)$$

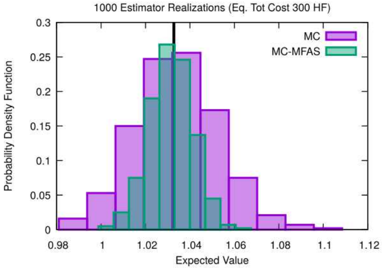


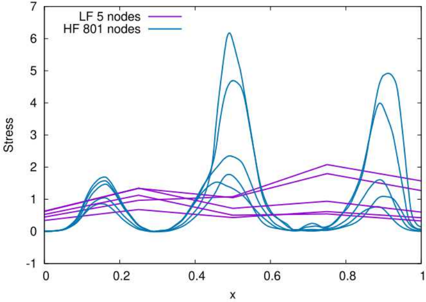
FIGURE: Normalized histograms for 1000 realizations in the case of dissimilar parametrization.

In this case we used 2 active directions for the HF and 1 for the LF

Non-linear elastic waves propagation – Hyperbolic CLAWs 1D

Non-linear elasticity problem
Can we enhance the correlation for this problem as well?

Let's consider an 'extreme' scenario (within the previous test problem)



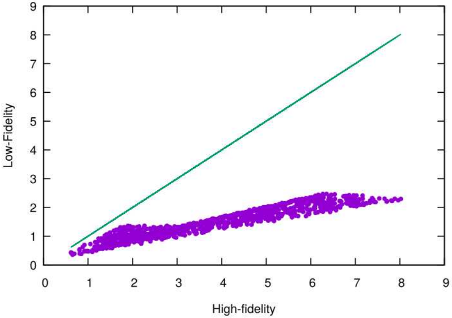
	N_x	N_t	Δ_t
Low-fidelity	5	50	36×10^{-4}
High-fidelity	801	600	30×10^{-5}

Table: HF to LF Cost ratio ~ 2800

- We compute the AS without the gradient (we use a linear regression)
- We use 40 HF samples for our estimator
- We perform 250 repetitions

Non-linear elasticity problem

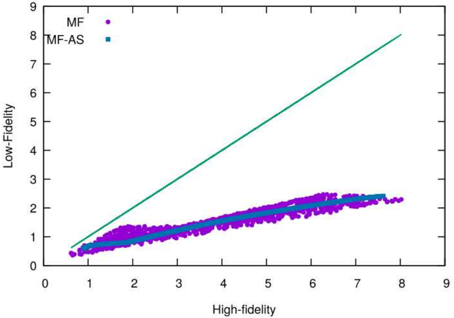
Can we enhance the correlation for this problem as well?

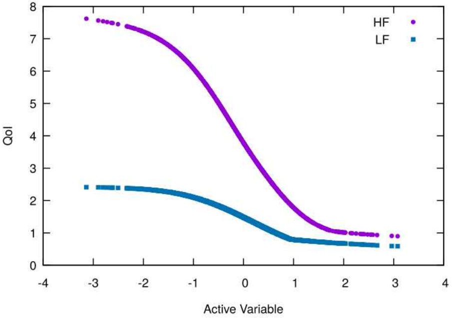


Active Direction Agnostic sampling: $\rho^2 = 0.89$

Non-linear elasticity problem

Can we enhance the correlation for this problem as well?





Active Direction Agnostic sampling: $\rho^2 = 0.89$

Active Direction Aware sampling:
$\rho^2 = 0.99$

Lid- and Buoyancy-driven Cavity Flow
Alleviating the cost of AS estimation

- The cost of the pilot samples accounted to $30 \times 1 + 30 \times 0.001 = 30.03$ HF (coming from HF mainly in this case)
- Can we re-use the HF samples without discarding them?

1. Pilot samples are generated in the physical space (30 as done before)
2. The LF samples are discarded
3. The HF pilot samples are projected onto the active direction
4. LF samples are generated at the Active Variables locations of the HF
5. Correlation is estimated and the oversampling is computed (always on the active variables)
6. The MF estimator is evaluated

- Items (1-6) are **repeated 300 times** and the estimated mean are reported

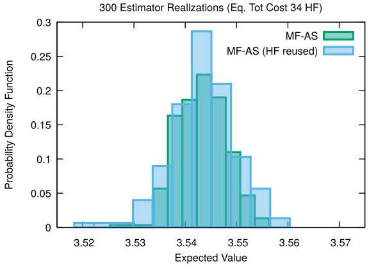


FIGURE: Probability density function for the estimators MF-AS computed with 300 independent realizations with and without reusing the HF samples.

Lid- and Buoyancy-driven Cavity Flow

Projecting onto the Active Variables from the pilot realizations

- By reusing the HF samples, we need to handle samples that have not been generated along the active variables
- Due to the nature of the mapping (inactive variables) this projection will exhibit a noisy behavior
- A very simple approach to improve this step is to perform a regression over the active variables

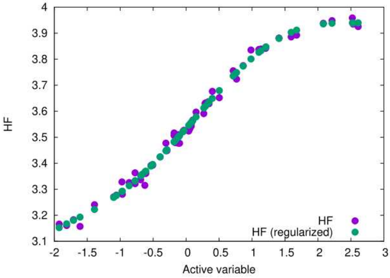


FIGURE: High-fidelity realizations for 40 pilot samples projected on to the active variable space with and without regularization.

Lid- and Buoyancy-driven Cavity Flow

Projecting onto the Active Variables from the pilot realizations

- By reusing the HF samples, we need to handle samples that have not been generated along the active variables
- Due to the nature of the mapping (inactive variables) this projection will exhibit a noisy behavior
- A very simple approach to improve this step is to perform a regression over the active variables

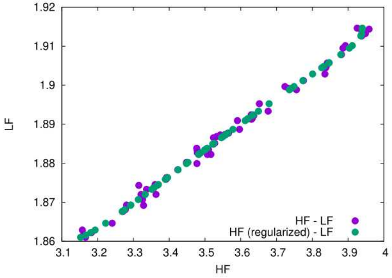


FIGURE: High-fidelity realizations for 40 pilot samples projected on to the active variable space with and without regularization.

Lid- and Buoyancy-driven Cavity Flow

Can I re-use also the LF pilot samples?

- We can conceptually apply the same strategy for the LF samples, however there is an additional **challenge**...
- ...we **do not have a common sample set to estimate the correlation** along the active variables
- In order to compute the correlation before evaluating the additional LF samples we use the PC expansion (analytical expression)
- Once the correlation is evaluated and the LF oversampling is defined the initial LF set might be fully re-used
- We can now perform MF-AS (re)starting from legacy dataset

1 30 pilot samples extracted from a dataset of 500 evaluations (LF and HF are consistent)

2 300 repetitions of the estimator with full re-use of both HF and LF

- **NOTE:** there is a non-zero probability of using the same evaluation multiple time (for different estimator realizations)

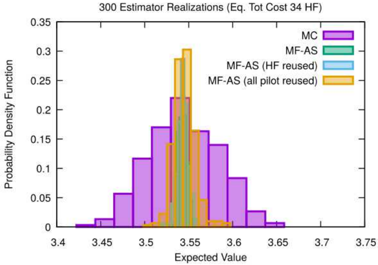


FIGURE: Probability density function for the estimators MF-AS computed with 300 independent realizations with and without reusing the pilot samples.

Overview of recent developments in surrogate-based MF UQ

Surrogate-based MF UQ
Motivation

Why do we want to use surrogate-based UQ if we already have sampling-based MF approaches?

- **Sampling methods are very robust** and often the only viable solution for UQ studies of high-dimensional, noisy and possibly discontinuous problems...
- ...however **many applications** (especially their QoIs) are much **more regular than one might expect** *a priori*
- In these circumstances, **surrogate-based approach offer a huge advantage in term of their convergence rate**

A recent example:

- DARPA SEQUOIA – aero-thermo-structural design of a nozzle (RANS+FEM): the QoIs where **reasonably well behaved** and lower order (at least along the active direction(s))
- DARPA SCRAMJET – supersonic combustion (LES): the **QoIs were very noisy** (additional error contribution coming from unconverged statistics)

We currently continue the development in both areas to cover **different needs for different applications**

The two main building blocks
Non-intrusive PC and SC

- **Polynomial Chaos:** Spectral projection using orthogonal polynomial basis

$$\hat{f} = \sum_{k=0}^{P+1} \beta_k \Psi_k$$

- **Stochastic Collocation:** Form interpolants for known coefficients

Notes:

- Common tools are regression, tensor/sparse quadrature, etc.

Seminal idea
Decreasing 'complexity' for the discrepancy function

- The concept of **multifidelity** has been known/exploited in the optimization community for decades
- One of the first applications of this concept in UQ:

  Ng and **Eldred**. *Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation*. In 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2012.

The **main idea** is quite simple and effective: Can you use a **LF model to capture most of the response** and use only **fewer HF evaluations to correct** it?

$$Q_{HF} = \exp -0.05\xi^2 \cos 0.5\xi - \mathbf{0.5} \exp -\mathbf{0.02}(\xi - \mathbf{5})^{\mathbf{2}}$$
$$Q_{LF} = \exp -0.05\xi^2 \cos 0.5\xi$$

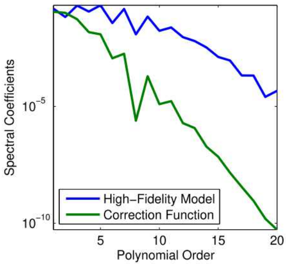


FIGURE: Spectral content

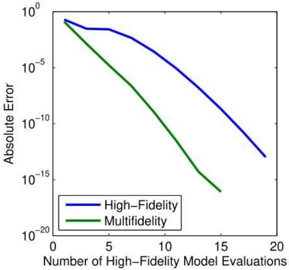


FIGURE: Error (Mean)

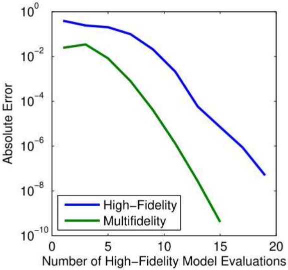


FIGURE: Error (Mean)

'Complexity' of a function
Order, sparsity, low-rank structure...

The original idea was based on the following assumptions:

- the LF model is able to capture the high frequencies of the response
- only the low-order terms are included in the discrepancy term $\rightarrow$ **few evaluations of the discrepancy are needed** to build the response for the discrepancy

In many **practical applications**:

- the **LF model only capture low-order effects**
- however the discrepancy term can have a structure that we can still exploit

Two possible structures that we can exploit are:

- **Sparsity** $\rightarrow$ Compressed sensing: orthogonal matching pursuit (OMP), basis pursuit denoising (BPDN), least angle regression (LARS), least absolute selection and shrinkage operator (LASSO)...
- **Low-rank** $\rightarrow$ Functional Tensor-Train decomposition (TT)

Exploiting favorable function's structures
Three main strategies

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)

Exploiting favorable function's structures

Three main strategies

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)
2. Exploiting **Restricted Isometry Property** (RIP)

Exploiting favorable function's structures

Three main strategies

In order we have tried **several approaches**:

1. **Optimal resources allocation** (direct extension of MLMC concepts to surrogates)
2. Exploiting **Restricted Isometry Property** (RIP)
3. **Greedy Multilevel Refinement**

Exploiting favorable function's structures
Strategy 1: Extending the MLMC sampling approach to surrogates

Main idea: Two parameters can be added to parametrize the variance of the recovered discrepancy term

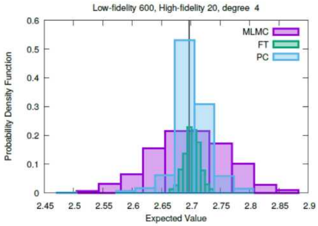
$$\mathbb{V}ar\left[\hat{Y}_\ell\right] = \frac{\mathbb{V}ar\left[Y_\ell\right]}{\gamma N^{\mathbf{k}}} \rightarrow N_\ell = \sqrt[k]{\frac{\sum_{q=0}^{L} \sqrt[k+1]{\mathbb{V}ar\left[Y_q\right] \mathcal{C}_q^k}}{\gamma \varepsilon^2/2}} \sqrt[k+1]{\frac{\mathbb{V}ar\left[Y_\ell\right]}{\mathcal{C}_\ell}}$$

Notes:

- γ and k can be obtained as by-product of the k-fold cross-validation process
- this approach can be extended to level-dependent parameters, *i.e.* γ_ℓ and k_ℓ (slightly different closed form solution)

Findings:

- **Abrupt transition** in both sparse and low-rank recovery **does not allow to efficiently estimate the parameters** and exploit the faster convergence



(a) $N_{low} = 600$, $N_{high} = 20$ and $deg = 4$ (b) $N_{low} = 600$, $N_{high} = 40$ and $deg = 4$

EXPLOITING FAVORABLE FUNCTION'S STRUCTURES

STRATEGY 2: RESTRICTED ISOMETRY PROPERTY (RIP) FROM *Jakeman, Narayan, Zhou, 2016*

Main idea: Address/Avoid abrupt transition by **ensuring enough samples** for accurate recovery

$$\text{RIP}: \quad N_\ell \geq s_\ell L_\ell \log^3(s_\ell) \log(C_\ell)$$

where

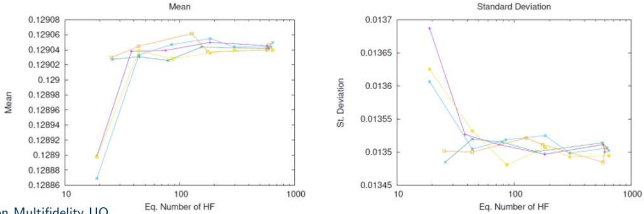
- s_ℓ is the sparsity, *i.e.* number of non-zero coefficients
- L_ℓ is the mutual coherence, *i.e.* if a_i are the normalized ($a_i^T a_i = 1$) columns of the matrix A then $L = \max \left| a_i^T a_j \right|$ for $i \neq j$
- C_ℓ is the cardinality of the dictionary

Algorithm:

- Start with pilot sample to estimate sparsity at each level ℓ
- Number of samples is increased to allow the recovery

Findings:

- **RIP is quite conservative** and it is likely to overshoot so it is necessary to add a constraint on the profile $\rightarrow$ very difficult to handle the feedback



Exploiting favorable function's structures
Strategy 3: Greedy Multilevel refinement

Main issues discovered with strategy #1 and #2 are:

- **Difficult to estimate** a trend
- **Difficult to handle** the allocation strategy in order to avoid overshoot in term on number of samples

Proposed solution: Greedy refinement - compete refinement candidates to **maximize induced change per unit cost**

Algorithm:

- One or more **candidates are generated** per each level
- The **impact of each candidate on the final QoIs statistics is evaluated** and normalized by the relative cost of level increment
- **Greedy selection** of the best candidate
- **Generation of new candidates** for the selected level

Greedy Multilevel refinement
Level candidate generators

- **Uniform refinement:** coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC
  - Regression PCE: least squares / compressed sensing using a fixed sample ratio
- **Anisotropic refinement:** coarse-grained refinement with one expansion order / grid level candidate per model level
  - Tensor / sparse grids: projection PCE and nodal/hierarchical SC
- **Index-set-based refinement:** fine-grained refinement with multiple index set candidates per model level; exponential growth in size of candidate set with dimension.
  - Generalized sparse grids: projection PCE and nodal/hierarchical SC
- **Basis selection:** coarse-grained refinement with a few expansion order frontier advancements per model level
  - Regression PCE

Greedy Multilevel refinement
Test case

Steady-state diffusion

$$-\frac{d}{dx}\left[a(x,\boldsymbol{\xi})\frac{du}{dx}(x,\boldsymbol{\xi})\right] = 10, \quad (x,\boldsymbol{\xi}) \in (0,1) \times I_{\boldsymbol{\xi}},$$

- x is the spatial coordinate
- $\boldsymbol{\xi}$ a vector of independent random input parameters
- $a(x,\boldsymbol{\xi})$
- in our test $d = 9$, *i.e.* $I_{\boldsymbol{\xi}} = [-1,1]^9$ denotes the (random) diffusivity field

Dirichlet boundary conditions are also assumed

$$u(0,\boldsymbol{\xi}) = 0, \quad u(1,\boldsymbol{\xi}) = 0.$$

QoIs defined as the solution u at specified spatial locations: $\bar{x} = 0.05, 0.5, 0.95$. We represent the random diffusivity field a using the following expansion

$$a(x,\boldsymbol{\xi}) = 1 + \sigma \sum_{k=1}^{d} \frac{1}{k^2\pi^2} \cos(2\pi k x)\boldsymbol{\xi}_k$$

Multilevel setup: discretization corresponding to 4, 8, 16, 32 and 64 elements

Greedy Multilevel refinement
Compressed sensing – Statistics

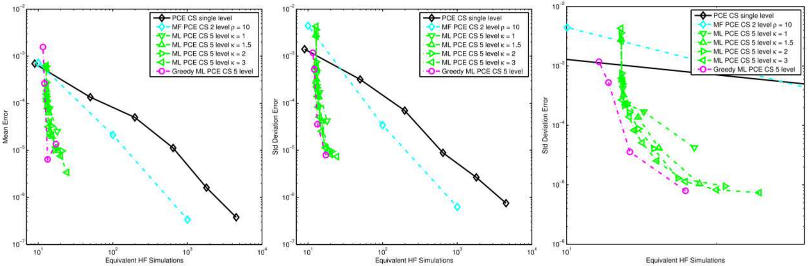


FIGURE: Convergence for greedy multilevel PCE based on compressed sensing. Test problem is steady state diffusion with nine random variables and one, two, or five discretization levels.

Greedy Multilevel refinement

Compressed sensing – Samples allocation

Conv Tol	N_1	N_2	N_3	N_4	N_5
1.e-1	198	9	9	9	9
1.e-2	644	198	9	9	9
1.e-3	1802	644	9	9	9
1.e-4	4505	1802	50	9	9

Table: Final sample profiles for greedy multilevel compressed sensing applied to steady state diffusion (9 random variables, 5 discretization levels).

Notes:

- We impose a collocation ration of 0.9, *i.e.* the system is underdetermined
- The first order correspond to 10 terms, therefore 9 simulations are needed (initialization/pilot)
- The second order correspond to 55 terms, therefore 50 simulations are needed

Greedy Multilevel refinement
Generalized Sparse Grid – Statistics

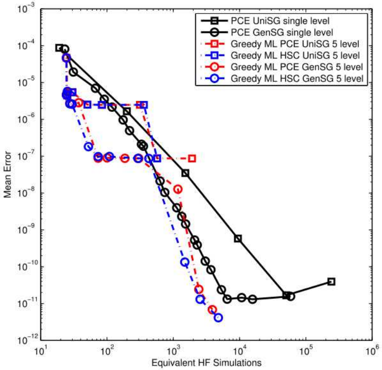


FIGURE: Convergence for greedy multilevel PCE based on (generalized) sparse grids. Test problem is steady state diffusion with nine random variables and one or five discretization levels (solid and dashed lines, respectively).

Greedy Multilevel refinement
Generalized Sparse Grid – Samples allocation

Conv Tol	N_1	N_2	N_3	N_4	N_5
1.e-2	43	23	19	19	19
1.e-4	211	83	19	19	19
1.e-6	391	271	156	19	19
1.e-8	1359	743	327	59	19
1.e-10	3535	2311	1039	391	19
1.e-12	10319	5783	2783	1343	43
1.e-14	26655	14991	8063	3703	1535

TABLE: Final sample profiles for greedy multilevel refinement applied to steady state diffusion (9 random variables, 5 discretization levels).

Notes:

- All levels incur a minimum $2n + 1 = 19$ evaluation cost due to the initial set of level-one candidate index sets

Greedy Multilevel refinement
CS/GSG – Statistics

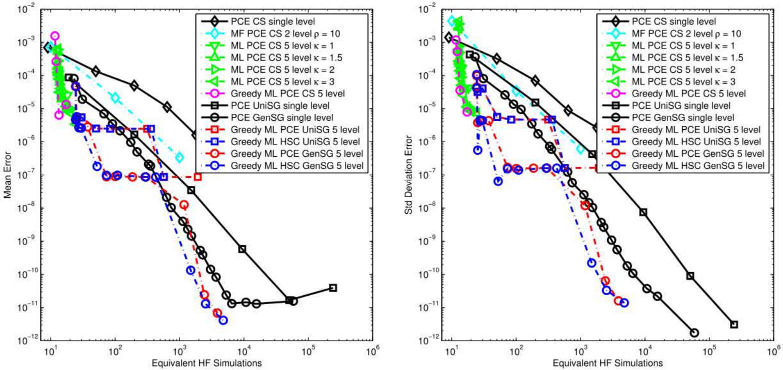


FIGURE: Convergence for greedy multilevel PCE comparing generalized sparse grids and compressed sensing.

Notes:

- The explicit nature of the sparse grid approaches allows for more precise convergence
- The compressed sensing approaches, while supporting sample profiles at the lower end of the cost spectrum, are currently hampered in accuracy by solution of the large implicit systems that are allocated at the coarse level

Bayesian inversion

Generalities on the approach adopted in this work

Bayesian calibration

- Sandia's UQ software **Dakota** (see Dakota Theory Manual for more details)
- **Markov Chain Monte Carlo** for computing a sample-based **posterior distribution**
- We are interested in calibrating the parameters $\boldsymbol{\theta}$
- We assume that a surrogate for the computational model is available for the QoI: $\mathbf{q} = \mathbf{q}(\boldsymbol{\theta})$
- Reference data $\mathbf{d}$ are available

Bayesian inversion
Few details

Bayesian rule

$$f_{\boldsymbol{\Theta}|\boldsymbol{D}}(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\,\mathcal{L}(\boldsymbol{\theta};\boldsymbol{d})}{f_{\boldsymbol{D}}(\boldsymbol{d})}$$

- **Posterior probability** $f_{\boldsymbol{\Theta}|\boldsymbol{D}}(\boldsymbol{\theta}|\boldsymbol{d})$
- Conservative **Prior distribution** $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$
- **Likelihood** $\mathcal{L}(\boldsymbol{\theta};\boldsymbol{d})$
- **Evidence** $f_{\boldsymbol{D}}(\boldsymbol{d})$

If the difference between the model quantity of interest $\mathbf{q}$ and the data $\mathbf{d}$ is Gaussian

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{d}) = \frac{1}{\sqrt{(2\pi)^n\,|\boldsymbol{\Sigma}_{\boldsymbol{d}}|}} \exp\left(-\frac{1}{2}\boldsymbol{r}^T\boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1}\boldsymbol{r}\right),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{d}}$ represents the covariance matrix of the Gaussian data.

NOTES:

- From computational perspective it is more convenient to work with the negative log-likelihood

$$-\log\mathcal{L}(\boldsymbol{\theta};\boldsymbol{d}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{d}}| + \frac{1}{2}\boldsymbol{r}^T\boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1}\boldsymbol{r}$$

- The term $\boldsymbol{r}^T\boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1}\boldsymbol{r}$ is called **Misfit Function**
- Minimizing the **Misfit Function** corresponds to maximizing the **Likelihood**
- Maximizing the Likelihood (MLE) does not in general correspond to the Maximum A posteriori (MAP) point
- Posterior probability is analytically intractable and therefore MCMC is used to approximate it
- We use the QUESO library in Dakota to perform MCMC

Bayesian inversion

Why does having a surrogate help?

- The computational code can be queried directly, but MCMC requires a very large number of evaluations to converge
- Surrogates can provide:
  - Computing local accurate proposal density (by using Hessian information)
  - Pre-solving for the MAP in order to eliminate the initial burn-in phase

Computing a local accurate proposal density

- The MCMC proposal covariance to be the inverse of the Hessian of the negative log posterior

$$\nabla^2_{\boldsymbol{\theta}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right] = \nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) - \nabla^2_{\boldsymbol{\theta}} \left[\log(\pi_0(\boldsymbol{\theta})) \right]$$

- A standard approximation is the multivariate normal (MVN) distribution with mean centered at the actual point in the chain and prescribed covariance

$$-\nabla^2_{\boldsymbol{\theta}} \left[\log(\pi_0(\boldsymbol{\theta})) \right] = \boldsymbol{\Sigma}_{\mathbf{0}}^{-1} \quad \rightarrow \quad \nabla^2_{\boldsymbol{\theta}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right] = \nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_{\mathbf{0}}^{-1} \left[\log(\pi_0(\boldsymbol{\theta})) \right]$$

- The Heassian of the Misfit Function can be computed through the surrogate model as

$$\nabla^2_{\boldsymbol{\theta}} M(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta})^T \boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta}) + \nabla^2_{\boldsymbol{\theta}} \boldsymbol{q}(\boldsymbol{\theta}) \cdot \left[\boldsymbol{\Sigma}_{\boldsymbol{d}}^{-1} \boldsymbol{r} \right].$$

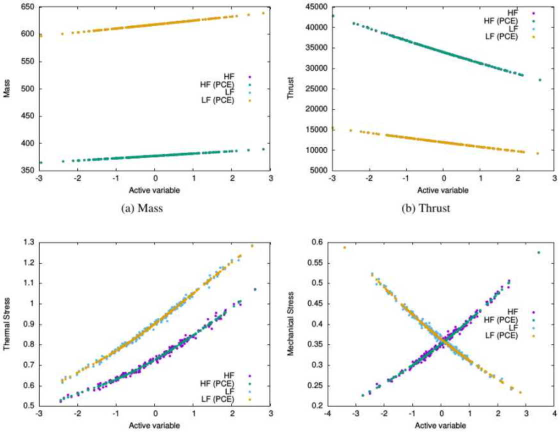
Avoiding the burn-in phase

- When a surrogate is available the burn-in can be avoided by pre-solving for the MAP point using an optimizer to minimize the negative log posterior

$$\boldsymbol{\theta}_{MAP} = \underset{\boldsymbol{\theta}}{\text{argmin}} \left[-\log(\pi_{\text{post}}(\boldsymbol{\theta})) \right]$$

Preliminary results for the Sequoia problem

Scenario 1 – inconsistent parameterization and same sample set



(a) Mass (b) Thrust

(c) Thermal Stresses (d) Mechanical Stresses

FIGURE: QoIs w.r.t. the active variable for the nozzle problem in the case of inconsistent parameterization for both the original data and the PCE regression with respect to the active variable (Scenario 1).

Preliminary results for the Sequoia problem

Scenario 2 – consistent parameterization and independent same sample set

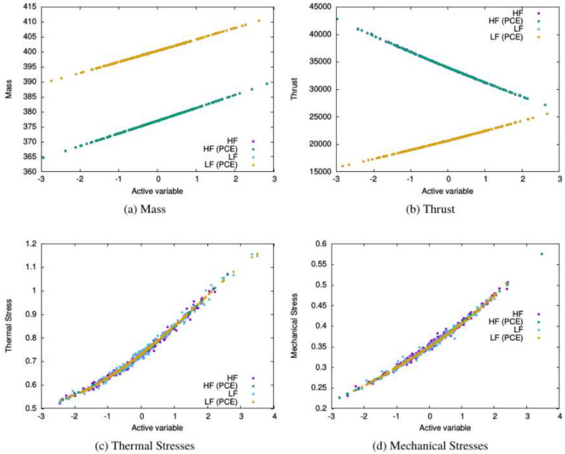


FIGURE: QoIs w.r.t. the active variable for the nozzle problem in the case of inconsistent parameterization for both the original data and the PCE regression with respect to the active variable (Scenario 2).

Bayesian inversion
Posterior distribution

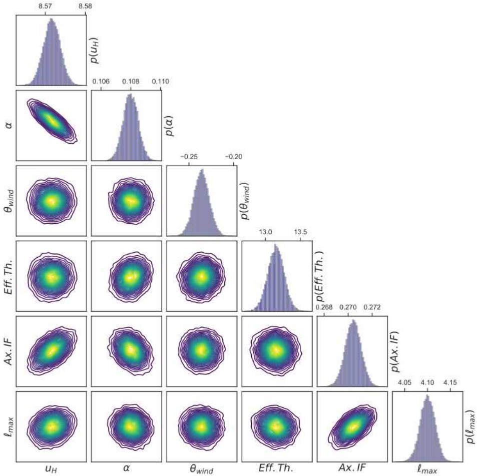


Figure: Visualization of the six-dimensional posterior distribution obtained through emulator-based inference from u data only. Marginal distributions are shown as histograms and pairwise joint distributions are displayed as contour plots.