Sampling, Polynomial Chaos and I SAND2017-7297C r Train Multilevel/Multifidelity Strategies for Forward UQ

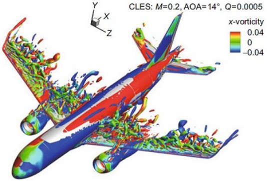
Gianluca Geraci, Alex A. Gorodetsky, John J. Jakeman and Michael S. Eldred

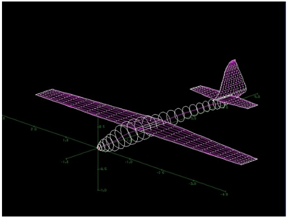
Sandia National Laboratories, Albuquerque

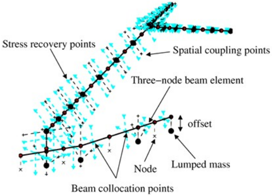
2017 SIAM Annual Meeting
13th July 2017

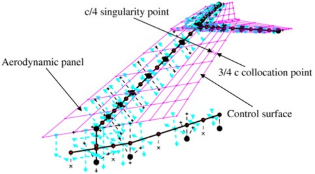
Multifidelity in UQ
Motivation

- Hierarchies of models are ubiquitous in engineering practice
- For centuries we relied on simplified models, then computers arrived...
- Can low-fidelity models still find a place in nowadays computational analysis? Perhaps in UQ...









A little bit more context
Discretization Vs Model Fidelity

Multi-fidelity: several description 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...)

Common features:

- Increasing the model level/fidelity the quality of the solution improves (numerical solution closer to the truth)
- Increasing the level/fidelity the numerical cost also increases

Even if it's always possible to mix discretization levels and model fidelities, exploiting their particular structure can be more advantageous...

Plan of the talk

- Motivation
- (hints of the) Theory
- Numerical examples
- Conclusion

Uncertainty Quantification
Forward Propagation – Why Sampling Methods?

UQ context at a glance:

- High-dimensionality, non-linearity and discontinuities
- Rich physics and many 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
- Regularity or structures of the solution can be also leveraged to compress its representation on high-dimensional spaces

Monte Carlo simulation
Introducing the spatial discretization

Problem statement: We are interested in the expected value of $Q_M = \mathcal{G}(\mathbf{X}_M)$ where
- M is (related to) the number of spatial degrees of freedom
- $\mathbb{E}[Q_M] \xrightarrow{M\to\infty} \mathbb{E}[Q]$ for some RV $Q : \Omega \to \mathbb{R}$

Monte Carlo:

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

two sources of error:
- Sampling error: replacing the expected value by a (finite) sample average
- Spatial discretization: finite resolution implies $Q_M \approx Q$

Looking at the Mean Square Error:

$$\mathbb{E}\left[(\hat{Q}^{MC}_{M,N} - \mathbb{E}[Q])^2\right] = N^{-1}\mathbb{V}ar(Q_M) + (\mathbb{E}[Q_M - Q])^2$$

Accurate estimation ⇒ **Large number** of samples at high (spatial) resolution

Control variate
Pivotal role

A Control Variate MC estimator (function G with $\mathbb{E}[G]$ known)

$$\hat{Q}_N^{MCCV} = \hat{Q}_N^{MC} - \beta\left(\hat{G}_N^{MC} - \mathbb{E}[G]\right)$$

Properties:

- Unbiased, *i.e.* $\mathbb{E}\left[\hat{Q}_N^{MCCV}\right] = \mathbb{E}\left[\hat{Q}_N^{MC}\right]$
- $\underset{\beta}{\operatorname{argmin}}\, \mathbb{V}ar\left(\hat{Q}_N^{MCCV}\right) \rightarrow \beta = -\rho \dfrac{\mathbb{V}ar^{1/2}(Q)}{\mathbb{V}ar^{1/2}(G)}$
- Pearson's $\rho = \dfrac{\mathrm{Cov}(Q, G)}{\mathbb{V}ar^{1/2}(Q)\, \mathbb{V}ar^{1/2}(G)}$ where $|\rho| < 1$

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

Control variate
Pivotal role

A Control Variate MC estimator (function G with $\mathbb{E}[G]$ known)

$$\hat{Q}_N^{MCCV} = \hat{Q}_N^{MC} - \beta \left(\hat{G}_N^{MC} - \mathbb{E}[G] \right)$$

Properties:

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

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

Q: How does the control variate approach enter in our picture?
A: By means of the (geometrical) MLMC and multifidelity strategy

0. Single resolution level
   - Cheap lower fidelity (**Multifidelity**)
1. Applying it recursively
   - Spatial discretization (**Multilevel**)
2. Applying it recursively across resolutions/model forms
   - Spatial discretization and cheap lower fidelity (**Multilevel-Multifidelity**)

Multifidelity

MULTIFIDELITY

PRACTICAL IMPLICATIONS OF UNKNOWN LOW-FIDELITY STATISTICS

Let's modify the high-fidelity QoI, Q_M^{HF}, to decrease its variance

$$\hat{Q}_{M,N}^{\mathrm{HF},CV} = \hat{Q}_{M,N}^{\mathrm{HF}} + \alpha \left(\hat{Q}_{M,N}^{\mathrm{LF}} - \mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \right).$$

Multifidelity

Practical implications of unknown low-fidelity statistics

Let's modify the high-fidelity QoI, Q_M^{HF}, to decrease its variance

$$\hat{Q}_{M,N}^{\mathrm{HF},CV} = \hat{Q}_{M,N}^{\mathrm{HF}} + \alpha \left(\hat{Q}_{M,N}^{\mathrm{LF}} - \mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \right).$$

In practical situations

- the term $\mathbb{E}\left[Q_M^{\mathrm{LF}}\right]$ is unknown (low fidelity $\neq$ analytic function)
- we use an additional and independent set $\Delta^{\mathrm{LF}} = rN^{\mathrm{HF}}$

$$\mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \simeq \frac{1}{(1+r)N^{\mathrm{HF}}} \sum_{i=1}^{(1+r)N^{\mathrm{HF}}} Q_M^{\mathrm{LF},(i)}.$$

Finally the variance is

$$\boxed{\mathbb{V}ar\left(\hat{Q}_{M,N}^{\mathrm{HF},CV} \right) = \mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF}} \right) \left(1 - \frac{r}{1+r} \rho_{HL}^2 \right)}$$

(geometrical) Multilevel

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 $M_0 < M_1 < \dots < M_L \stackrel{\text{def}}{=} M$
- Estimation of $\mathbb{E}[Q_M]$ by means of correction w.r.t. the next lower level

$$Y_\ell \stackrel{\text{def}}{=} Q_{M_\ell} - Q_{M_{\ell-1}} \xrightarrow{\textit{linearity}} \mathbb{E}[Q_M] = \mathbb{E}[Q_{M_0}] + \sum_{\ell=1}^{L} \mathbb{E}\left[Q_{M_\ell} - Q_{M_{\ell-1}}\right] = \sum_{\ell=0}^{L} \mathbb{E}[Y_\ell]$$

- Multilevel Monte Carlo estimator

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

- The Mean Square Error is

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

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

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)

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

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

Multilevel-Multifidelity (MLMF)

Multilevel-Multifidelity approach
Combination of discretization and model form

- OUTER SHELL – Multi-level

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

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

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

Final properties of the estimator

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

and

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

Multilevel-Multifidelity

Optimal allocation across discretization and model forms

- Target accuracy for the estimator: ε
- Cost per level is now $\mathcal{C}_\ell^{\text{eq}} = \mathcal{C}_\ell^{\text{HF}} + \mathcal{C}_\ell^{\text{LF}} (1 + r_\ell)$
- the (constrained) optimization problem is

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

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

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 ($N_\ell^{\text{LF}} = N_\ell^{\text{HF}} + \Delta_\ell^{\text{LF}} = N_\ell^{\text{HF}}(1 + r_\ell)$)

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

Sparse PC regression

Polynomial Chaos
Basis selection and expansion

Polynomial Chaos methods represent a function $f(\boldsymbol{\xi}) \in L^2(p(\boldsymbol{\xi}))$ as an expansion of orthogonal polynomials

$$f(\boldsymbol{\xi}) \approx \hat{f}(\boldsymbol{\xi}) = \sum_{k=1}^{P} \beta_k \Psi_k(\boldsymbol{\xi}), \quad \boldsymbol{\xi} = (\xi_1, ..., \xi_d).$$

where $\Psi_k(\boldsymbol{\xi}) = \phi_{\alpha_1}(\xi_1) \dots \phi_{\alpha_d}(\xi_d)$ are tensor product of orthonormal polynomials which are orthogonal to $p(\boldsymbol{\xi})$.

A truncation needs to be chosen. For instance, a total degree basis can be selected as

$$\mathcal{A} = \{||\alpha|| \leq n_0\} \quad \text{where} \quad \text{card}(\mathcal{A}) = P = \frac{(n_0 + d)!}{n_0! d!}$$

This basis grown exponentially with the dimension d

Polynomial Chaos

Sparse representation and OMP

- A very flexible (and common) approach to find the coefficients β_k is the regression
- Regression-based PC methods solve the linear system

$$\boldsymbol{\Psi}\boldsymbol{\beta} = \mathbf{b}, \quad \text{where} \quad \mathbb{R}^N \ni \mathbf{b} = \left\{ f(\boldsymbol{\xi}^{(1)}, \dots, f(\boldsymbol{\xi}^{(N)}) \right\}^{\mathrm{T}}$$

- Due to the exponential growth of the basis, $\boldsymbol{\Psi} \in \mathbb{R}^{N \times P}$, very often the system is under-determined, *i.e.* $N << P$
- In the presence of under-determined systems minimizing the residual w.r.t. the ℓ_2 norm typically produces poor solutions
- Compressed sensing methods have been demonstrated to be superior in this situation. These methods try to identify the coefficients β_k with the largest megnitude and enforce as many elements as possible to be zero

Some compressed sampling approaches are

- Basis Pursuit
- Basis Pursuit DeNoising
- Orthogonal Matching Pursuit (OMP)
- Least Angle Regression (LARS)

In particular we use in this numerical investigation OMP:

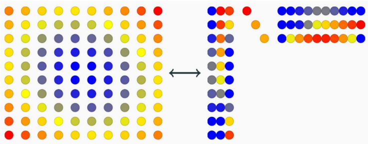
$$\boldsymbol{\beta} = \operatorname{argmin} ||\boldsymbol{\beta}||_{\ell_0} \quad s.t. \quad ||\boldsymbol{\Psi}\boldsymbol{\beta} - \mathbf{b}|| \leq \varepsilon$$

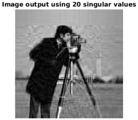
Function Train regression

Functional Tensor Train (in a nutshell)

Main idea

- **MAIN GOAL:** we would like to represent a **function in a tensor product basis**...
- ...tensor product basis has p^d unknowns
- A viable approach is to seek for a low-rank representation of the coefficient tensor
- In 2D optimal low-rank decomposition is the SVD...
- ...in high dimensions optimal low-rank decomposition does not exist
- We will use the tensor-train decomposition to obtain $\mathcal{O}(dnr^2)$ unknowns





Model format: tensor-train

From Tensor-Train (Oseledets, 2010) to the continuous Functional Tensor-Train (Gorodetsky *et al*, 2015)

- TT decomposition provides compression multiway arrays
  - Existence of best approximation guaranteed
  - Storage scales linearly with dimension and polynomially with rank
- TT-ranks are related to the ranks of *reshapings* of a tensor

$$r_k \leq \text{rank} f(\underbrace{i_1, \dots, i_k}_{\boldsymbol{i}_{\leq k}}; \underbrace{i_{k+1}, \dots, i_d}_{\boldsymbol{i}_{>k}})$$

- Approximate multivariate functions instead of multiway arrays
- Adapt to local and global structure
- Efficient, flexible, and adaptive approximation format
- Evaluation through products of matrix-valued functions

$$\begin{aligned} f(x_1, x_2, \dots, x_d) &= \sum_{i_0=1}^{r_0} \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} f_1^{(i_0 i_1)}(x_1) f_2^{(i_1 i_2)}(x_2) \dots f_d^{(i_{d-1} i_d)}(x_d) \\ &= \mathcal{F}_1(x_1)\mathcal{F}_2(x_2)\dots\mathcal{F}_d(x_d) \end{aligned}$$

$$\underbrace{\begin{bmatrix} f_k^{(11)}(x_k) & \cdot & f_k^{(1 r_k)}(x_k) \\ \vdots & \ddots & \vdots \\ f_k^{(r_{k-1} 1)}(x_k) & \cdot & f_k^{(r_{k-1} r_k)}(x_k) \end{bmatrix}}_{\mathcal{F}_k(x_k)}$$

Advantages of the FT format
Tensor-Train Vs Functional Tensor-Train

$$f(x_1, x_2, \dots, x_d) = \mathcal{F}_1(x_1)\mathcal{F}_k(x_k)\dots\mathcal{F}_d(x_d) \qquad \underbrace{\begin{bmatrix} f_k^{(11)}(x_k) & \cdot & f_k^{(1r_k)}(x_k) \\ \vdots & \ddots & \vdots \\ f_k^{(r_{k-1}1)}(x_k) & \cdot & f_k^{(r_{k-1}r_k)}(x_k) \end{bmatrix}}_{\mathcal{F}_k(x_k)}$$

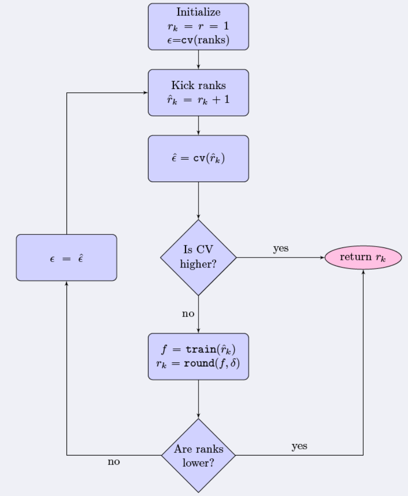
- Nonlinear parameterizations: can capture local effects
- Sparse representations:
  - Example: Rank-2 additive function
  $$f(x_1, \dots, x_d) = f_1(x_1) + \dots + f_d(x_d)$$
  - FT format:
  $$f(x_1, x_2, \dots, x_d) = [f_1(x_1) \ 1] \begin{bmatrix} 1 & 0 \\ f_2(x_2) & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 \\ f_d(x_d) \end{bmatrix}$$
  - TT storage requirement: $4p(d-2)$ floats
  - FT storage requirement: $d(p+3)-4$ floats
  - In the limit, FT requires almost 4 times less storage
  - Difference is more striking for higher order interactions

Rank adaptation

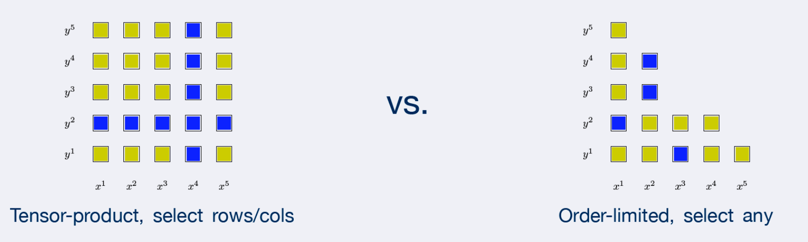
On-line procedure to select the optimal rank

- Use FT rounding and CV
- Increase ranks until either
  - Rounding lowers all ranks
    → data not informative enough
  - CV error increases
    → avoid overfitting
- Rounding threshold is a parameter
  - Similar to regularization
  - Relation to other approaches?

- Train model through a LS or other supervised learning objective function e.g., huber loss, hinge loss, etc.
- We have developed ALS and all-at-once optimization algorithms
- Have used BFGS as well as stochastic gradient descent
- In the result that follow will use All-at-once optimization with least squares loss



Low-rank vs. sparse comparisons



Pros:

1. No dimensional dependent enumeration
2. Linear scaling with dimension
3. Captures high-order effects

Cons:

1. Requires rank estimation

Pros:

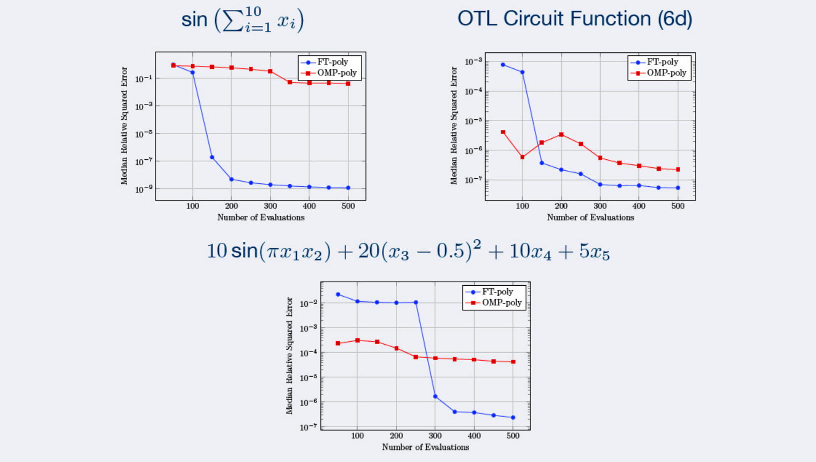
1. Adapted to decaying spectrums
2. Mature and robust algorithms

Cons:

1. Enumeration of basis functions
2. Curse Of Dimensionality for fixed order

Functional Tensor Train and Polynomial Chaos
Sparsity and low-rank

Test functions – OMP Vs FT



- Motivation
- (hints of the) Theory
- **Numerical examples**
- Conclusion

Numerical investigation

Description of the approach

- Set of several test functions available in literature (multifidelity UQ)
- Array of 1000 repetitions with the following features
  - Number of coarse evaluations $\{200, 400, 600\}$
  - Number of Δ evaluations $\{10, 20, 40, 80, 160\}$
  - Polynomial degree ranging from 0 to 6 with step of 1
- Dataset problems (fixed set of realizations across levels/resolutions)
  - Aero-thermo-structural nozzle analysis
  - Cardiovascular problem

Main steps of the analysis

1. All the polynomial orders are compared and the one corresponding to the smaller estimator variability is selected
2. The effect of the number of Δ evaluations is studied
3. Expected values distributions are compared, for MLMC, MLPCE and MLFT.

Additional outcome

- The variance of each estimator can be studied for each separate level
- The variance decays per level can be studied for all the estimators

Synthetic Problems

Synthetic problems

Function definition and investigation strategy

- **Currin** *et al.* (1988) – Exponential function

$$f(\boldsymbol{\xi}) = \left[1 - \exp\left(-\frac{1}{2\xi_2}\right)\right] \frac{2300\xi_1^3 + 1900\xi_1^2 + 2092\xi_1 + 60}{100\xi^3 + 500\xi_1^2 + 4\xi_1 + 20}$$

$$f_{low}(\boldsymbol{\xi}) = \frac{1}{4}\left[f(\xi_1 + 0.05, \xi_2 + 0.05) + f(\xi_1 + 0.05, \max(0, \xi_2 - 0.05)\right] + \frac{1}{4}\left[f(\xi_1 - 0.05, \xi_2 + 0.05) + f(\xi_1 - 0.05, \max(0, \xi_2 - 0.05)\right]$$

- **Park** (1991) – **F1**

$$f(\boldsymbol{\xi}) = \frac{\xi_1}{2}\left[\sqrt{1 + \left(\xi_2 + \xi_3^2\right)\frac{\xi_4}{\xi_1^2}}\right] + (\xi_1 + 3\xi_4)\exp(1 + \sin(\xi_3))$$

$$f_{low}(\boldsymbol{\xi}) = \left[1 + \frac{\sin(\xi_1)}{10}\right] f(\boldsymbol{\xi}) - 2\xi_1 + \xi_2^2 + \xi_3^2 + 0.5$$

- **Park** (1991) – **F2**

$$f(\boldsymbol{\xi}) = \frac{2}{3}\exp(\xi_1 + \xi_2) - \xi_4 \sin(\xi_3) + \xi_3$$

$$f_{low}(\boldsymbol{\xi}) = 1.2 f(\boldsymbol{\xi}) - 1$$

- **Short Column** (Eldred 2012 and Berchier 2016)

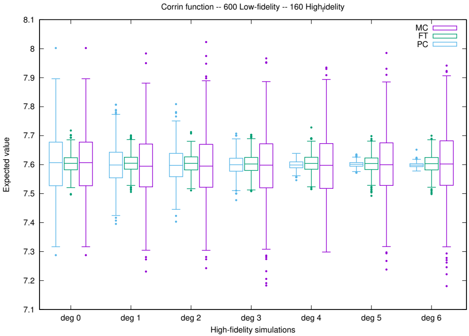
$$f(\boldsymbol{\xi}) = 1 - \frac{4M}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

$$f_{low}(\boldsymbol{\xi}) = 1 - \frac{4P}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

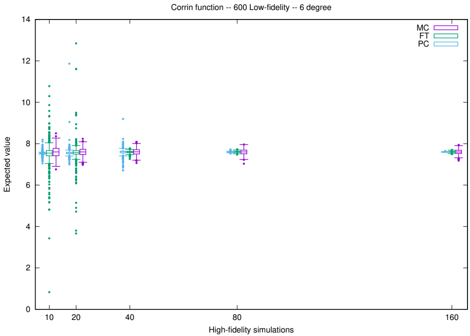
Synthetic problems
Currin – degree and high-fidelity effect

CURRIN

$N_{low} = 600$, $N_{high} = 160$



$N_{low} = 600$, **degree=6**



$$f(\boldsymbol{\xi}) = \left[1 - \exp\left(-\frac{1}{2\xi_2}\right)\right] \frac{2300\xi_1^3 + 1900\xi_1^2 + 2092\xi_1 + 60}{100\xi^3 + 500\xi_1^2 + 4\xi_1 + 20}$$

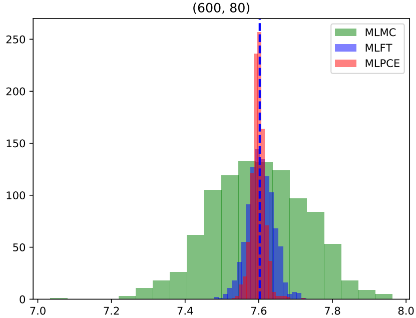
$$f_{low}(\boldsymbol{\xi}) = \frac{1}{4}\left[f(\xi_1 + 0.05, \xi_2 + 0.05) + f(\xi_1 + 0.05, \max(0, \xi_2 - 0.05)\right] + \frac{1}{4}\left[f(\xi_1 - 0.05, \xi_2 + 0.05) + f(\xi_1 - 0.05, \max(0, \xi_2 - 0.05)\right]$$

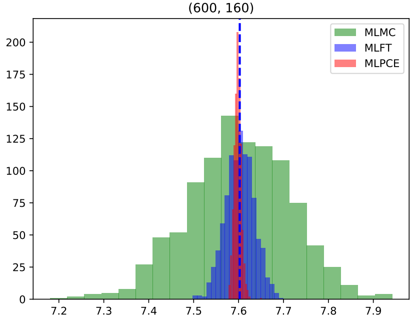
Synthetic problems
Currin – Expected values distribution

CURRIN

$N_{low} = 600$, $N_{high} = 80$, **degree = 6**

$N_{low} = 600$, N_{high}**=160**, **degree = 6**





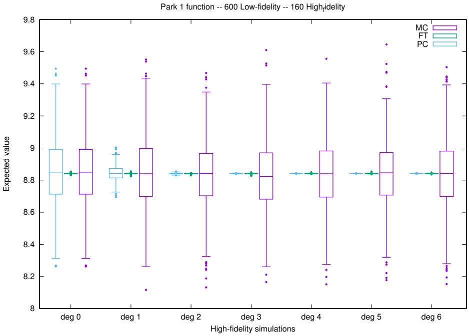
$$f(\boldsymbol{\xi}) = \left[1 - \exp\left(-\frac{1}{2\xi_2}\right)\right] \frac{2300\xi_1^3 + 1900\xi_1^2 + 2092\xi_1 + 60}{100\xi^3 + 500\xi_1^2 + 4\xi_1 + 20}$$

$$f_{low}(\boldsymbol{\xi}) = \frac{1}{4}\left[f(\xi_1 + 0.05, \xi_2 + 0.05) + f(\xi_1 + 0.05, \max(0, \xi_2 - 0.05)\right]$$
$$+ \frac{1}{4}\left[f(\xi_1 - 0.05, \xi_2 + 0.05) + f(\xi_1 - 0.05, \max(0, \xi_2 - 0.05)\right]$$

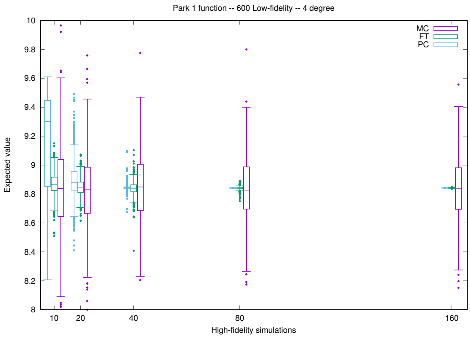
SYNTHETIC PROBLEMS
PARK 1 – DEGREE AND HIGH-FIDELITY EFFECT

PARK 1

N_{low} = **600**, N_{high}**=160**



N_{low} = 600, **degree=4**



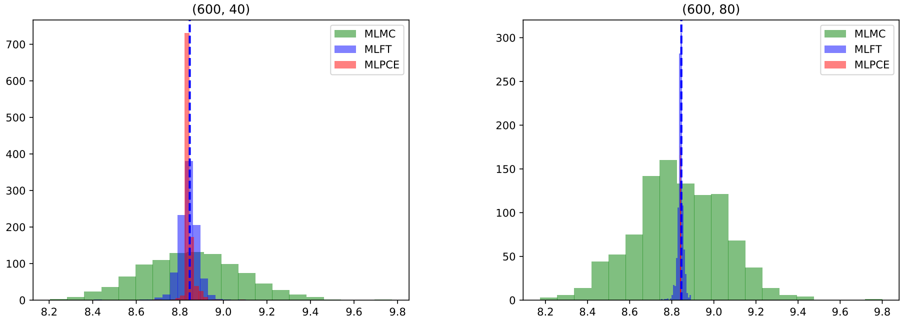
$$f(\boldsymbol{\xi}) = \frac{\xi_1}{2}\left[\sqrt{1 + \left(\xi_2 + \xi_3^2\right)\frac{\xi_4}{\xi_1^2}}\right] + \left(\xi_1 + 3\xi_4\right)\exp\left(1 + \sin\left(\xi_3\right)\right)$$

$$f_{low}(\boldsymbol{\xi}) = \left[1 + \frac{\sin\left(\xi_1\right)}{10}\right] f(\boldsymbol{\xi}) - 2\xi_1 + \xi_2^2 + \xi_3^2 + 0.5$$

Synthetic problems
Park 1 – Expected values distribution

PARK 1

N_{low} = 600, N_{high}=40, degree = 4 N_{low} = 600, N_{high}=80, degree = 4



$$f(\boldsymbol{\xi}) = \frac{\xi_1}{2}\left[\sqrt{1+\left(\xi_2+\xi_3^2\right)\frac{\xi_4}{\xi_1^2}}\right] + (\xi_1 + 3\xi_4)\exp\left(1+\sin\left(\xi_3\right)\right)$$

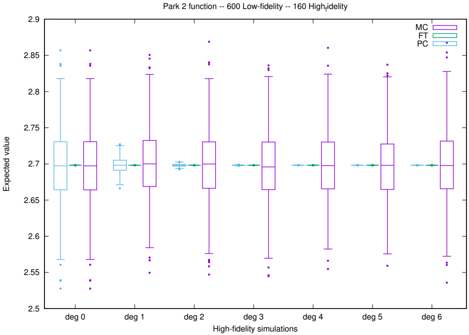
$$f_{low}(\boldsymbol{\xi}) = \left[1 + \frac{\sin\left(\xi_1\right)}{10}\right] f(\boldsymbol{\xi}) - 2\xi_1 + \xi_2^2 + \xi_3^2 + 0.5$$

Synthetic problems

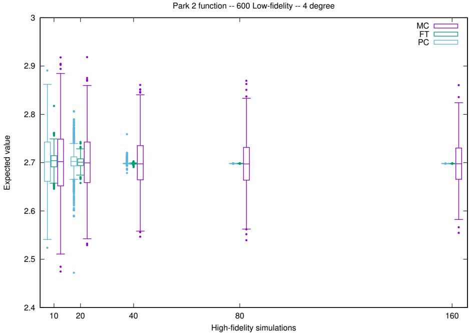
Park 2 – degree and high-fidelity effect

PARK 2

N_{low} = **600**, N_{high}=**160**



N_{low} = 600, **degree=4**



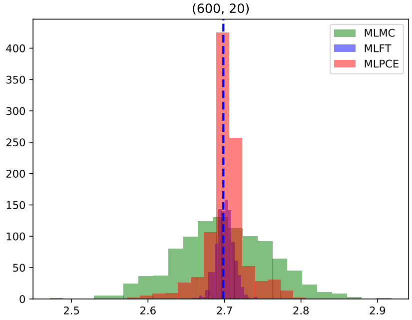
$$f(\boldsymbol{\xi}) = \frac{2}{3} \exp(\xi_1 + \xi_2) - \xi_4 \sin(\xi_3) + \xi_3$$

$$f_{low}(\boldsymbol{\xi}) = 1.2 f(\boldsymbol{\xi}) - 1$$

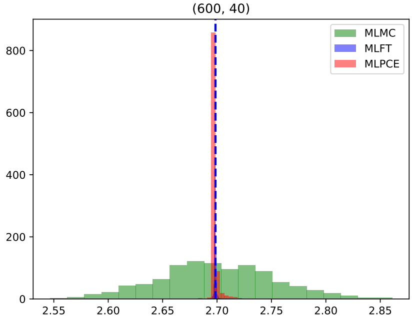
Synthetic problems
Park 2 – Expected values distribution

PARK 2

N_{low} = 600, N_{high}=20, degree = 4



N_{low} = 600, N_{high}=40, degree = 4



$$f(\boldsymbol{\xi}) = \frac{2}{3}\exp(\xi_1 + \xi_2) - \xi_4 \sin(\xi_3) + \xi_3$$

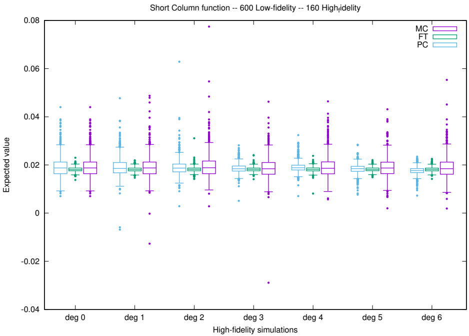
$$f_{low}(\boldsymbol{\xi}) = 1.2 f(\boldsymbol{\xi}) - 1$$

Synthetic problems

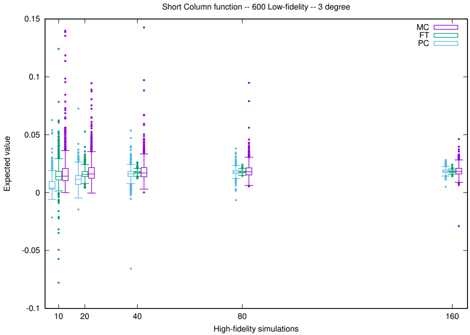
Short Column – degree and high-fidelity effect

SHORT COLUMN

$N_{low} = 600$, N_{high}=160



$N_{low} = 600$, degree=3



$$f(\boldsymbol{\xi}) = 1 - \frac{4M}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

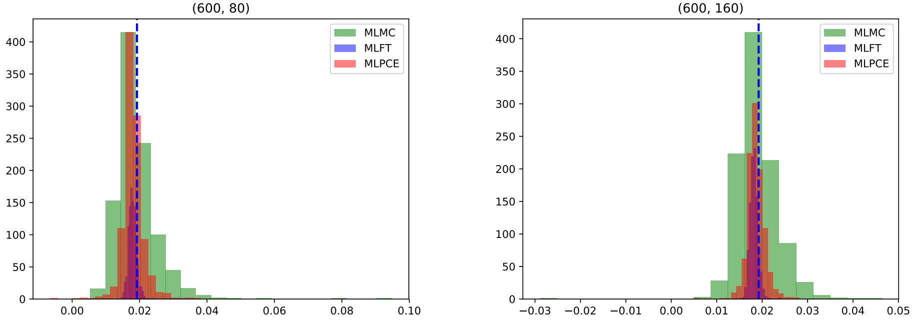
$$f_{low}(\boldsymbol{\xi}) = 1 - \frac{4P}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

SYNTHETIC PROBLEMS

SHORT COLUMN – EXPECTED VALUES DISTRIBUTION

SHORT COLUMN

N_{low} = 600, N_{high}=80, degree = 3 N_{low} = 600, N_{high}=160, degree = 3



$$f(\boldsymbol{\xi}) = 1 - \frac{4M}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

$$f_{low}(\boldsymbol{\xi}) = 1 - \frac{4P}{bh^2Y} - \left(\frac{P}{bhY}\right)^2$$

Synthetic problems
(Partial) Summary

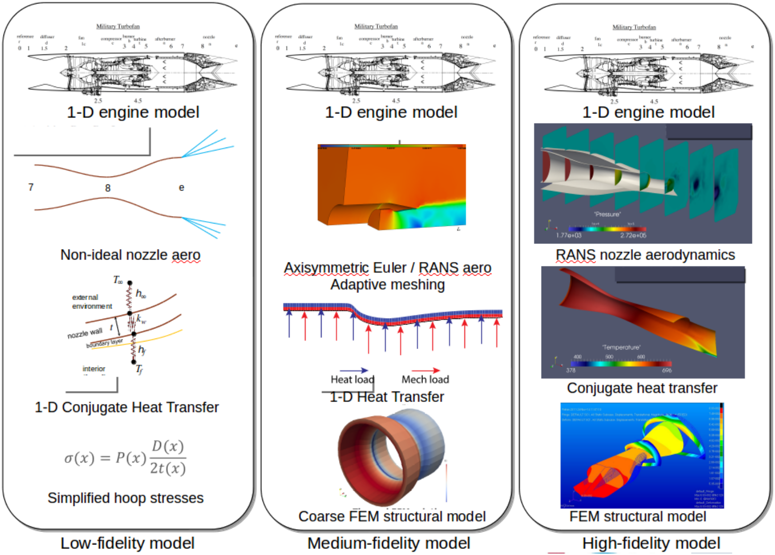
- Main goal is to collect evidence regarding the behaviors of the different approaches for different problems
- We do not want to select the 'best method', we know we will need all of them
- Non smooth transition is evident for both FT and PC
- Transition related to singular values (FT) or sparsity (PC)
- MC more 'reliable' for low samples allocations
- ML/FT after the transition converge to the exact results (expected for smooth problems)

Application Problems

Nozzle flow – Aero-Thermo-Structure interaction

Aero-Thermo-Structural analysis

Nozzle thrust – Computational setting



Aero-Thermo-Structural analysis
UQ Case description

- 2D RANS model realizations with SU2
- Thermo-Structural FEM solver
- 102 uniform random parameters for representing the manufacture uncertainties

	Relative Cost
Coarse	0.38
Δ	1.0

Table: Computational cost per realization.

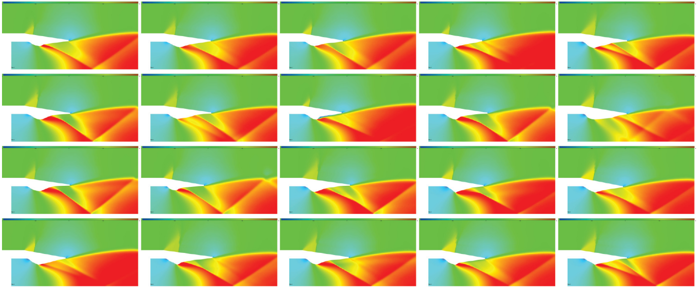
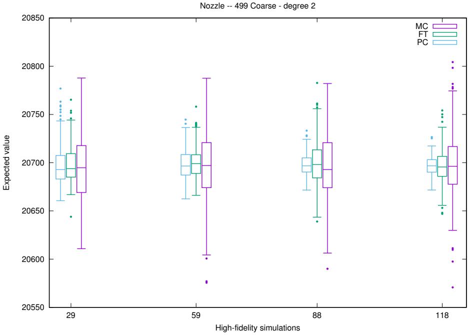
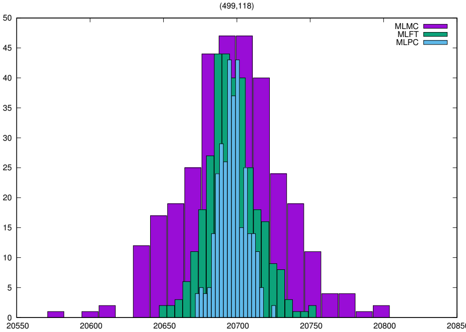


Figure: Example of Euler computations for different nozzle geometries.

Application problems
Nozzle – degree and high-fidelity effect



$N_{Coarse} = 499$, **degree=2**

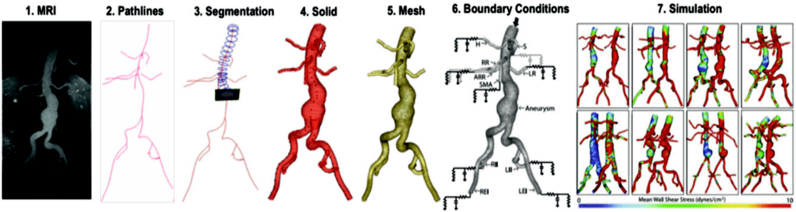


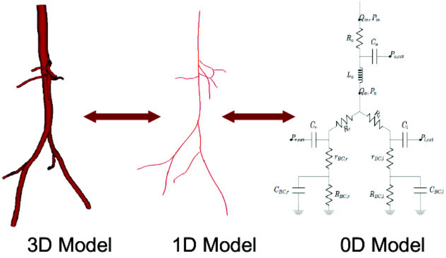
$N_{Coarse} = 499$, $N_{High} = 118$, **degree=2**

Cardiovascular flow – Flow/Structure interaction

Cardiovascular Flow introduction

Courtesy of C. Fleeter (Stanford), Prof. D. Schiavazzi (Notre Dame) and Prof. A. Mardsen (Stanford)



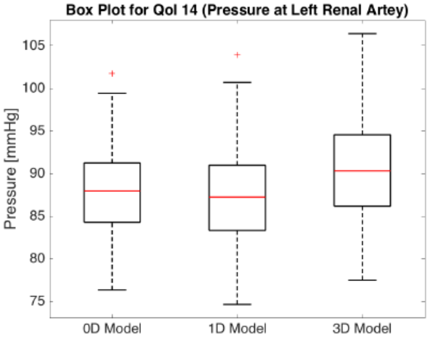


Solver	Cost (1 simulation)	Effective Cost (No. 3D Simulations)
3D	96 hr	1
1D	11.67 min	2E-3
0D	5 sec	1.45E-5

Cardiovascular Flow introduction
Computational setting and UQ setup

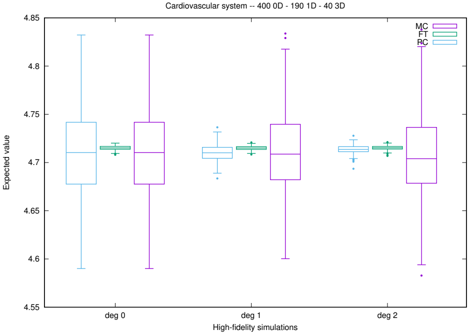
- We considered 9 uncertain BC parameters (*i.e.* resistances)
- Steady inlet flow (5 L/min)
- 20 QoIs:
    - Flows and pressures at the branches outlets
    - Min and Max wall shear stress

Solver	No. Simulations
3D	100
1D	2000
0D	10 000

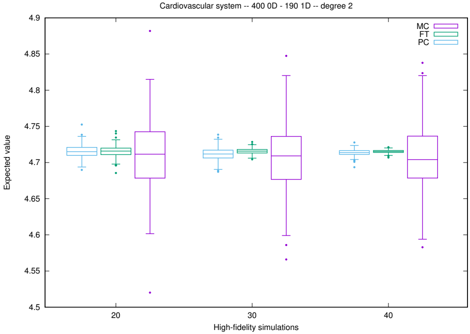


Application problems

Cardiovascular system – degree and high-fidelity effect

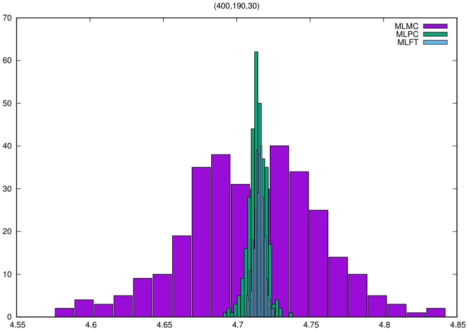


N_{0D} = **400**, N_{1D}=**190**, N_{3D}=**40**

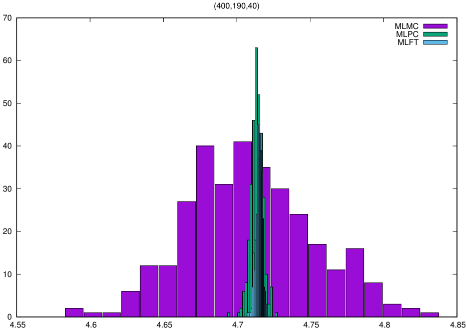


N_{0D} = **400**, N_{1D}=**190**, **degree=2**

Application problems
Cardiovascular system – Expected values distribution



N_{0D} = 400, N_{1D}=190, N_{3D}=30, degree = 2



N_{0D} = 400, N_{1D}=190, N_{3D}=40, degree = 2

Looking forward

Multilevel strategy

Optimal samples allocation for MLMC vs MLPCE/FT

- MLMC samples allocation is obtained by using the following relationship (exact) between the variance of the estimator and the variance of the QoI

$$\mathbb{V}ar\left(\hat{Y}_\ell\right) = \frac{\mathbb{V}ar\left(Y_\ell\right)}{N_\ell}$$

- The final samples allocation is

$$N_\ell = \frac{2}{\varepsilon^2}\left[\sum_{k=0}^{L}\left(\mathbb{V}ar\left(Y_k\right)\mathcal{C}_k\right)^{1/2}\right]\sqrt{\frac{\mathbb{V}ar\left(Y_\ell\right)}{\mathcal{C}_\ell}},$$

Multilevel strategy

Optimal samples allocation for MLMC vs MLPCE/FT

- MLMC samples allocation is obtained by using the following relationship (exact) between the variance of the estimator and the variance of the QoI

$$\mathbb{V}ar\left(\hat{Y}_\ell\right) = \frac{\mathbb{V}ar\left(Y_\ell\right)}{N_\ell}$$

- The final samples allocation is

$$N_\ell = \frac{2}{\varepsilon^2}\left[\sum_{k=0}^{L}\left(\mathbb{V}ar\left(Y_k\right)\mathcal{C}_k\right)^{1/2}\right]\sqrt{\frac{\mathbb{V}ar\left(Y_\ell\right)}{\mathcal{C}_\ell}},$$

- By assuming a similar relationship for PCE in a previous work we added two free parameters γ and κ

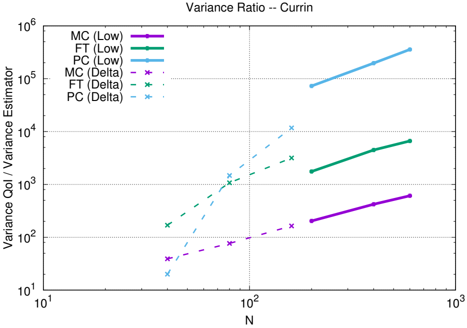
$$\mathbb{V}ar\left(\hat{Y}_\ell\right) = \frac{\mathbb{V}ar\left(Y_\ell\right)}{\gamma N_\ell^k}$$

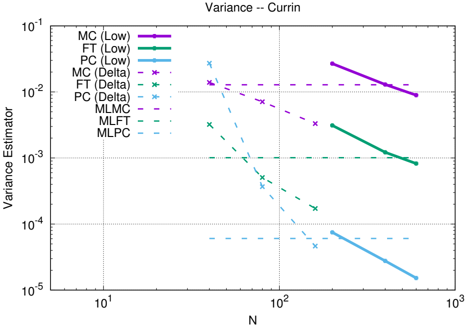
- The optimal samples allocation in this case is

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

Multilevel strategy
Currin – Variance ratio and decay

$$\mathbb{V}ar\left(Y_\ell\right)/\mathbb{V}ar\left(\hat{Y}_\ell\right) \qquad \mathbb{V}ar\left(\hat{Y}_\ell\right)$$



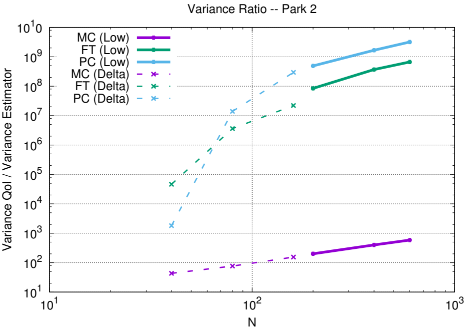


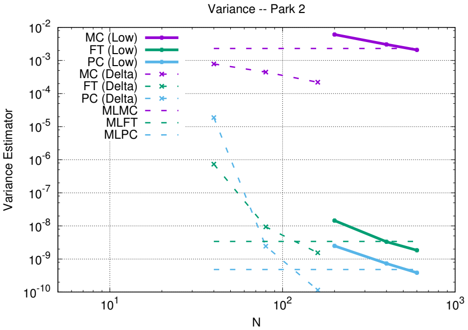
	γ		κ	
	Low	Δ	Low	Δ
MC	0.997	$8.47E-01$	1.005	$1.03E+00$
FT	2.737	$7.97E-02$	1.224	$2.11E+00$
PC	34.040	$1.26E-06$	1.446	$4.59E+00$

Table: Fitted values for γ and κ *per level*

Multilevel strategy
Park 2 – Variance ratio and decay

$$\mathbb{V}ar\left(Y_\ell\right)/\mathbb{V}ar\left(\hat{Y}_\ell\right) \qquad \mathbb{V}ar\left(\hat{Y}_\ell\right)$$





	γ Low	γ Δ	κ Low	κ Δ
MC	$1.13E+00$	$1.43E+00$	$9.78E-01$	$9.17E-01$
FT	$3.61E+03$	$5.33E-03$	$1.90E+00$	$4.44E+00$
PC	$5.82E+04$	$6.59E-11$	$1.70E+00$	$8.65E+00$

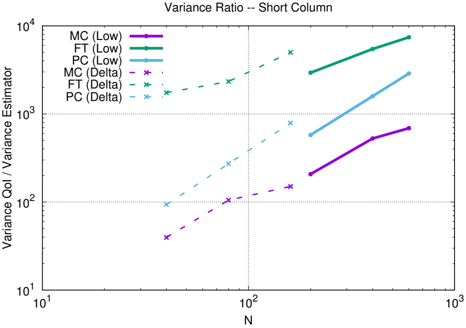
Table: Fitted values for γ and κ *per level*

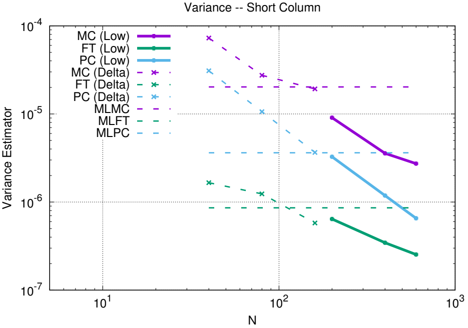
Multilevel strategy

Short Column – Variance ratio and decay

$$\mathbb{V}ar\left(Y_\ell\right)/\mathbb{V}ar\left(\hat{Y}_\ell\right)$$

$$\mathbb{V}ar\left(\hat{Y}_\ell\right)$$





	γ		κ	
	Low	Δ	Low	Δ
MC	0.572	1.275	1.120	0.959
FT	33.024	97.752	0.848	0.759
PC	0.247	0.321	1.463	1.538

Table: Fitted values for γ and κ *per level*

Multilevel strategy

Optimal samples allocation: work in progress

- All the results obtained in this numerical investigation suggest that we should use γ_ℓ and κ_ℓ

$$\mathbb{V}ar\left(\hat{Y}_\ell\right) = \frac{\mathbb{V}ar\left(Y_\ell\right)}{\gamma_\ell {N_\ell}^{\kappa_\ell}}$$

- The optimal samples allocation in this case is

$$N_\ell = \sqrt[\kappa_\ell+1]{\frac{\sum_{q=0}^{L} \frac{\kappa_\ell}{\kappa_q} N_q \mathcal{C}_q}{\gamma_\ell \varepsilon^2/2}} \sqrt[\kappa_\ell+1]{\mathbb{V}ar\left(Y_\ell\right) \mathcal{C}_\ell}$$

- The optimization problem is now more complex and requires non-linear iterations

Plan of the talk

- Motivation
- (hints of the) Theory
- Numerical examples
- **Conclusion**

Concluding remarks
Work still in progress

Summary

- Multifidelity, Multilevel and multilevel-multifidelity sampling estimators
- Key feature: Optimal allocation across all resolutions/models
- Extension of the multilevel/multifidelity idea to PC and FT
- Preliminary set of comparisons between MLMC, MLPC and MLFT

Work in progress

- Optimal allocation for MLMC/MLMF cannot be obtained in close form...
  ... much more challenging to do so for PC and FT
- Iterative procedure for the 'optimization' of degree, rank number of samples at each level
- Variance estimation (for linear regression) is very challenging without cross-validation
- Variance estimation is even more challenging for non-linear regression

Acknowledgements

- Nozzle case: Rick Fenrich, Dr. Victorien Menier and Prof. Juan Alonso (Stanford)
- Cardiovascular case: Casey Fleeter, Prof. Daniele Schiavazzi (ND) and Prof. Alison Mardsen (Stanford)

ML Giles, M.B., Multilevel Monte Carlo path simulation. *Oper. Res.* **56**, 607-617.

MLMF 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)*

MLPCE M.S. Eldred, G. Geraci & J.D. Jakeman, Multilevel Monte-Carlo Hybrids Exploiting Multifidelity Modeling and Sparse Polynomial Chaos, *SIAM Conference on Uncertainty Quantification, 2016*

FT A.A. Gorodetsky, S. Karaman, Y.M. Marzouk, Function-Train: a continuous analogue of the tensor-train decomposition (Submitted). *Available on arXiv:1510.09088v2*

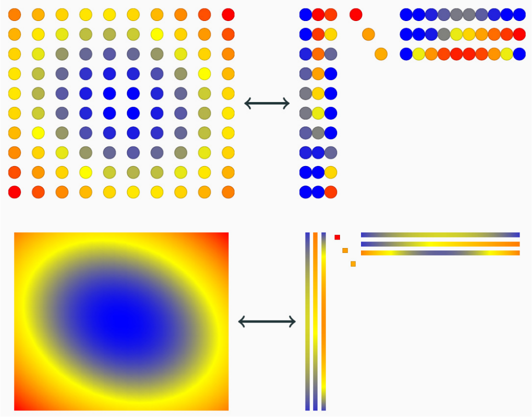
THANKS!

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.

Functional Tensor Train (in a nutshell)

Main idea – Discrete Vs Continuous representations

Example: Compression of a bivariate function

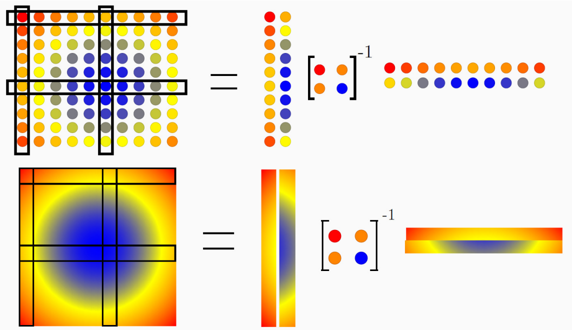


- Singular Value Decomposition (SVD) is a sum of (outer) products of vectors
- Functional SVD (fSVD) is the sum of products of univariate functions

Functional Tensor Train (in a nutshell)

Choosing a more efficient representation

Example: Compression of a bivariate function (CUR/skeleton decomposition)



- SVD is expensive, it requires $\mathcal{O}(N^2)$ evaluations
- SVD can be replaced with another (suboptimal) factorization
- CUR decomposition used columns and rows directly from the original matrix
- Theorem: If a rank r fSVD exists then a rank r CUR also exists