

# A MATHEMATICAL PERSPECTIVE ON THE CERTIFICATION AND DESIGN OF PHYSICAL SYSTEMS IN THE PRESENCE OF UNCERTAINTY

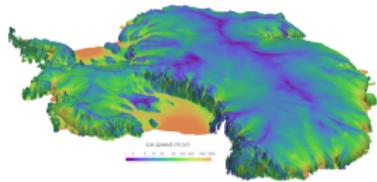
John D. Jakeman

Australian National University Mathematical Sciences  
Institute Colloquium

# MISSION STATEMENT

**Research Objective:** Use theoretically motivated mathematical and statistical algorithms, to quantify the impact of uncertainty in data-based and numerical models of complex systems.

“All models are wrong; the practical question is how wrong do they have to be to not be useful” [BD86]



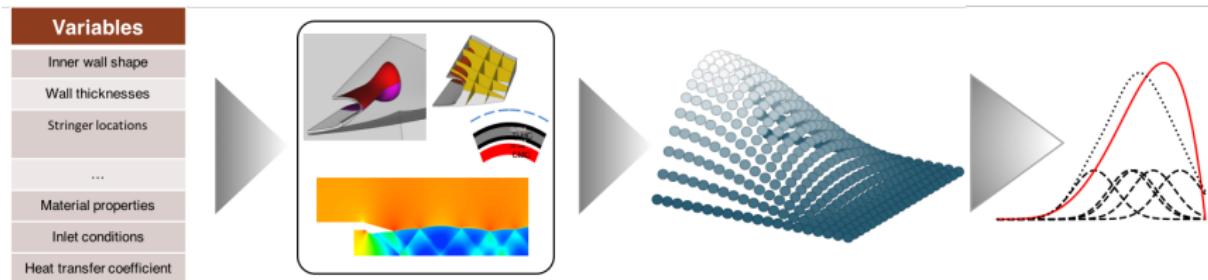
## CHALLENGES

**Cost:** High-fidelity simulation is expensive. Must compute statistics from limited number of samples (simulations).

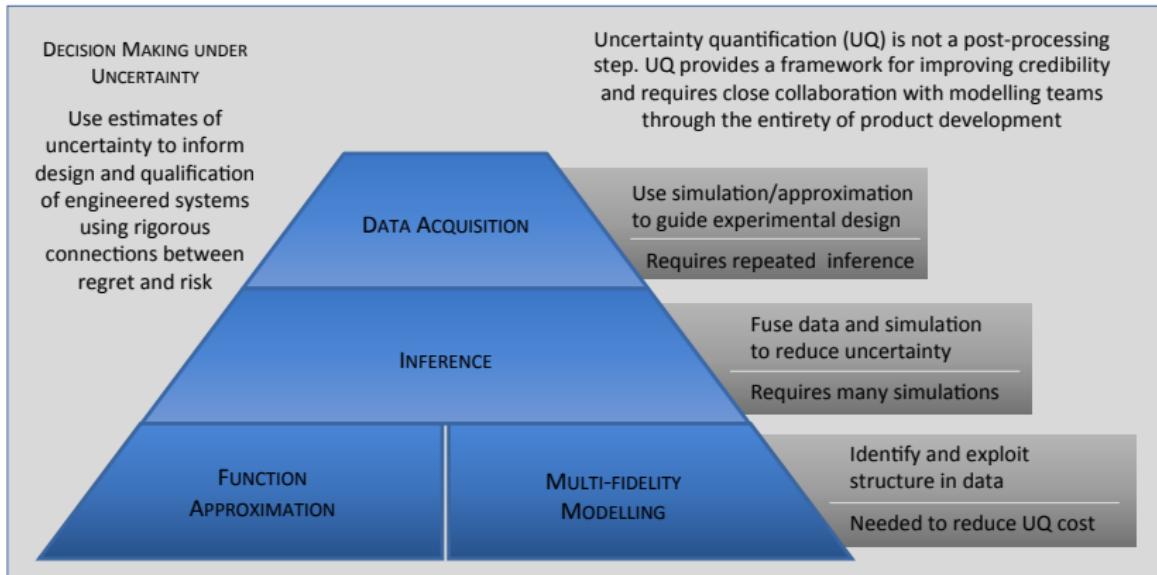
**High-Dimensionality:** Computational cost is often amplified as number of uncertainties increases.

**Inference:** Prior estimates of uncertainty can be overly conservative. Need to condition probabilistic estimates on observed data.

**Data Acquisition:** Determine experiments which are most informative.



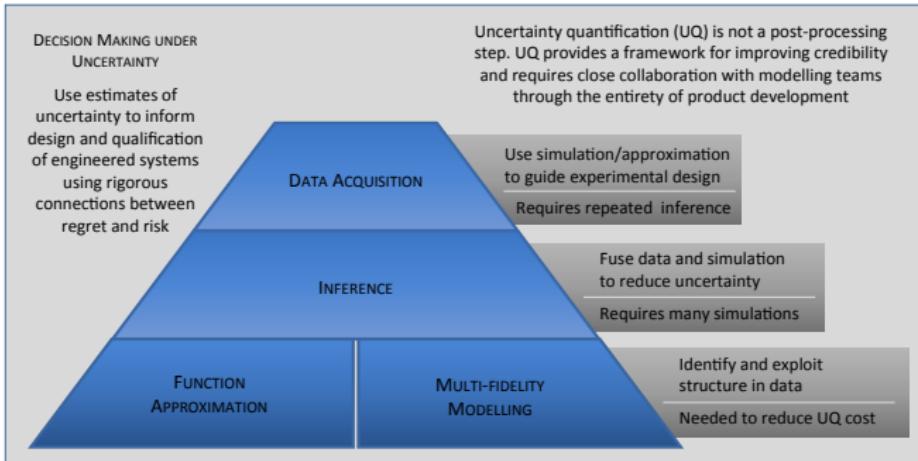
# VISION



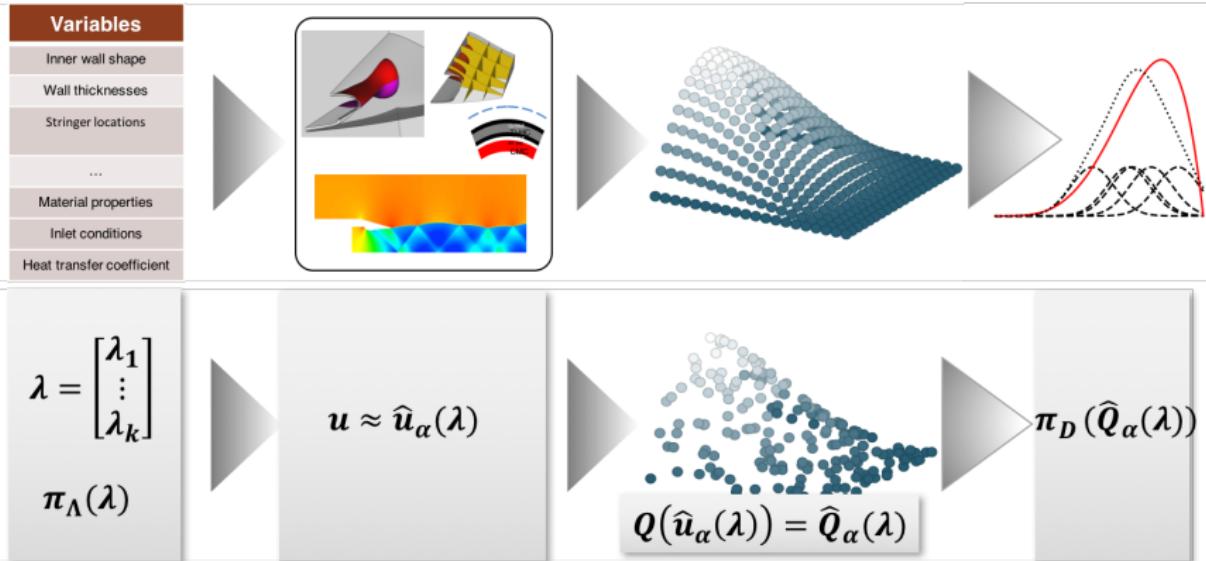
## TABLE OF CONTENTS

1. Function Approximation
2. Multi-fidelity Modelling
3. Inverse Uncertainty Quantification
4. Data Acquisition
5. Design and Decision Making Under Uncertainty

# FUNCTION APPROXIMATION



# FORWARD UNCERTAINTY QUANTIFICATION FUNCTION APPROXIMATION



Must compute statistics from limited number of samples (simulations)  
Computational cost is amplified as number of uncertainties increases

# MULTIVARIATE POLYNOMIAL INTERPOLATION

## TENSOR-PRODUCT INTERPOLATION

Define a set of univariate points

$$\mathcal{Z}_{m_{\beta_i}}^d = (\lambda_d^{(1)}, \dots, \lambda_d^{(m_{\beta_d})}), \quad d = 1, \dots, k$$

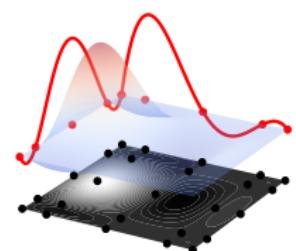
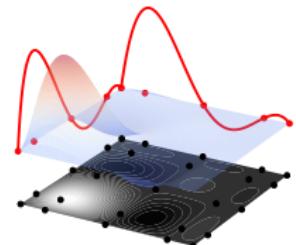
Univariate Lagrange polynomials

$$\phi_{d,j}(\lambda_d) = \prod_{i=1, i \neq j}^{m_{\beta_d}} \frac{\lambda_d - \lambda_d^{(i)}}{\lambda_d^{(j)} - \lambda_d^{(i)}},$$

Multivariate interpolant is given by

$$\hat{Q}_{\alpha, \beta}(\boldsymbol{\lambda}) = \sum_{j \leq \beta} \hat{Q}_{\alpha}(\boldsymbol{\lambda}^{(j)}) \prod_{d=1}^k \phi_{i,j_d}(\lambda_d).$$

$$\left\| \hat{Q}_{\alpha} - \hat{Q}_{\alpha, \beta} \right\|_{L^{\infty}(\Lambda)} \leq C_{k,r} N_{\beta}^{-r/k}$$

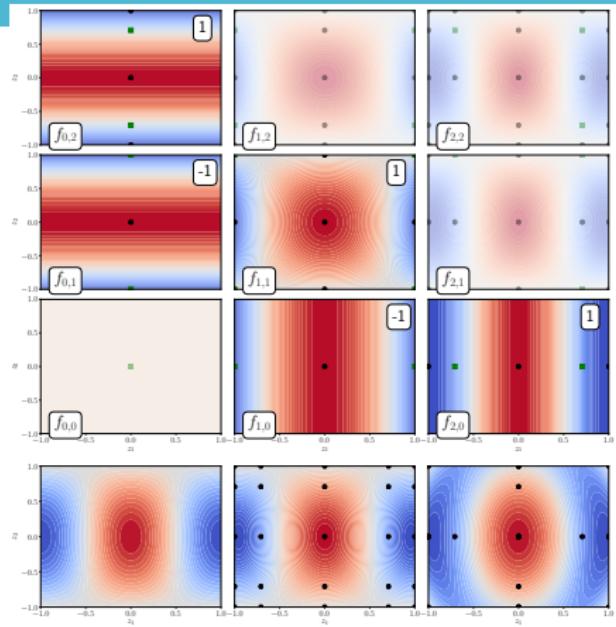


# DELAYING THE CURSE OF DIMENSIONALITY SPARSE GRID INTERPOLATION

$$\hat{Q}_{\alpha, \mathcal{I}}(\boldsymbol{\lambda}) = \sum_{\beta \in \mathcal{I}} c_{\beta} \hat{Q}_{\alpha, \beta}(\boldsymbol{\lambda})$$

$$c_{\beta} = (-1)^{n - \|\beta\|_1} \binom{k-1}{n - \|\beta\|_1}.$$

$$\mathcal{I}(n) = \{\beta \mid (\max(0, n-1) \leq \|\beta\|_1 \leq n-k-2\}$$



Theorem [BNR00]

$$\left\| \hat{Q}_{\alpha} - \hat{Q}_{\alpha, \mathcal{I}(n)} \right\|_{L^{\infty}(\Lambda)} \leq C_{k,r} N_n^{-r} (\log N)^{(r+2)(k-1)+1}$$

## SAMPLING WEIGHTED SPACES ONE-DIMENSIONAL INTERPOLATION

Given the best  $N$ -th order polynomial approximation  $p^*$  of a function  $f$  the error of the interpolant  $f_N$  based upon a set of  $N + 1$  random variable realisations  $\mathcal{Z}_{N+1}$  can be bounded by

$$\|\hat{Q}_\alpha - \hat{Q}_{\alpha,N}(\mathcal{Z})\|_\infty \leq (\Lambda_N(\mathcal{Z}) + 1) \|\hat{Q}_\alpha - p^*\|_\infty$$

where the Lebesgue constant for the grid  $\mathcal{Z}$  is given by

$$\Lambda_N(\mathcal{Z}) = \max_{\lambda \in [a,b]} \sum_{n=1}^N |\phi_n(\lambda)|$$

Best polynomial approximation error cannot be reduced. We can however reduce error by minimizing Lebesgue constant.

## UNIVARIATE INTERPOLATION FEKETE NODES

Let  $\{\psi_n\}_{n=1}^N$  be a polynomial basis. Fekete nodes maximise the determinant of the Vandermonde like Matrix  $V$

$$\mathcal{Z}_N = \operatorname{argmax}_{\mathcal{Z}} |\det V(\mathcal{Z})| \quad V(\mathcal{Z})_{m,n} = \psi_n(\lambda^{(m)})$$

Such a set  $\mathcal{Z}_N = \{\lambda^{(1)}, \dots, \lambda^{(N)}\}$  of nodes ensures an at-worst linear Lebesgue constant growth:

$$\Lambda(\mathcal{Z}_N) = \max_{\mathbf{x}} \sum_{n=1}^N |\psi_n(\lambda)| \leq \sum_{n=1}^N \max_{\lambda} \left| \frac{\det V(\{\mathcal{Z}_N, \lambda\})}{\det V(\mathcal{Z}_N)} \right| \leq N$$

Fekete nodes usually provide an excellent set of interpolation nodes.

# SAMPLING WEIGHTED SPACES LEJA SEQUENCES

Fekete points maximise determinant but not nested. Use Leja instead

$$\lambda_{N+1} = \operatorname{argmax}_{\lambda \in I} |\det V_N(\lambda)|$$

## Idea

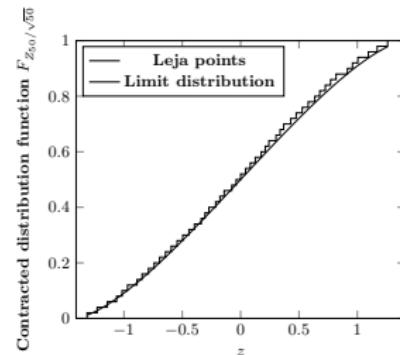
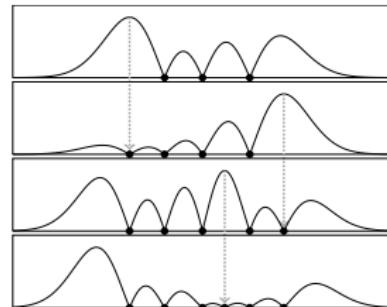
Extend Leja sequences for interpolation in  $\pi_{\Lambda}^i$ -weighted probability spaces

$$\text{Let } W_{m,m} = \sqrt{\pi_{\Lambda}^i(\lambda^{(m)})}$$

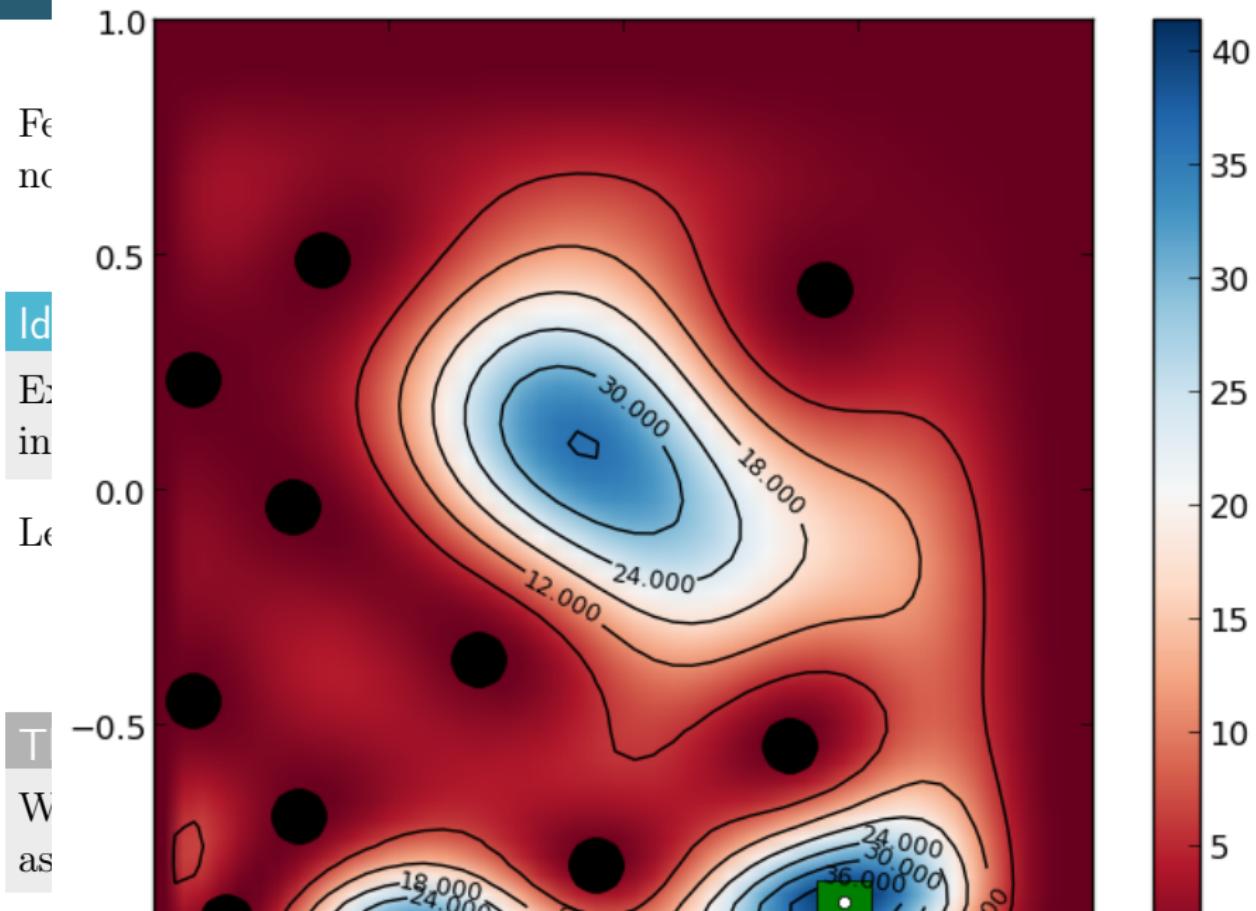
$$\lambda_{N+1} = \operatorname{argmax}_{\lambda \in I} |\det W(\lambda) V_N(\lambda)|$$

## Theorem [NJ14]

Weighted Leja sequences are asymptotically Fekete.

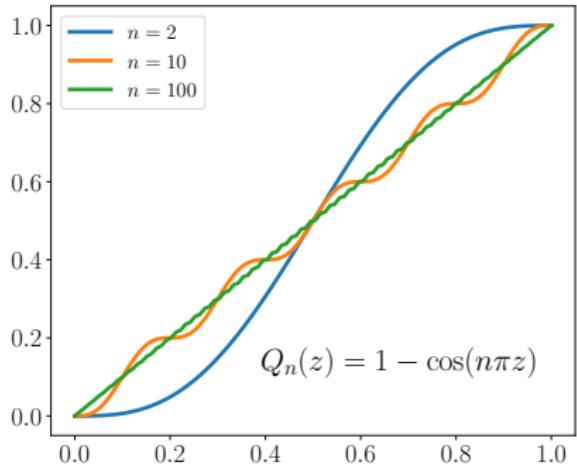
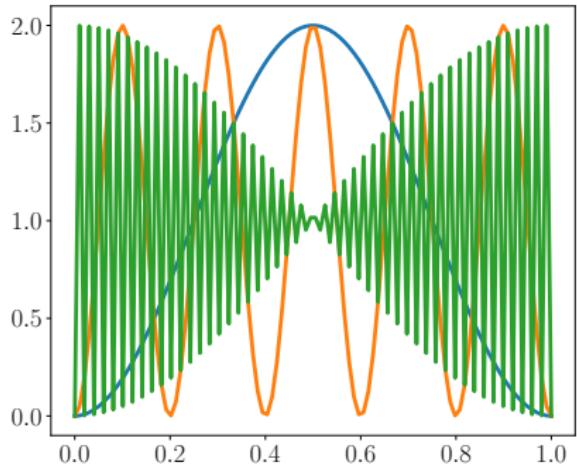


# SAMPLING WEIGHTED SPACES LEJA SEQUENCES



## DENSITY ESTIMATION

Most literature focuses on convergence of statistics of sequences of random variables



But convergence in distribution does not imply convergence almost surely.

# DENSITY ESTIMATION

## CONVERGENCE OF DENSITIES

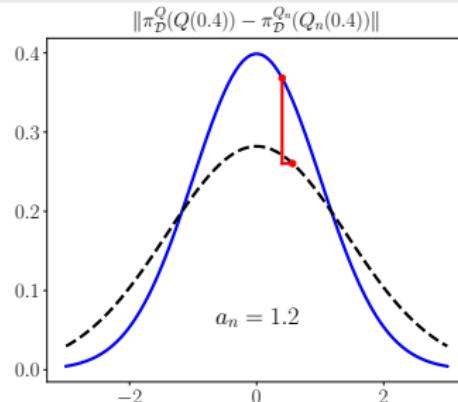
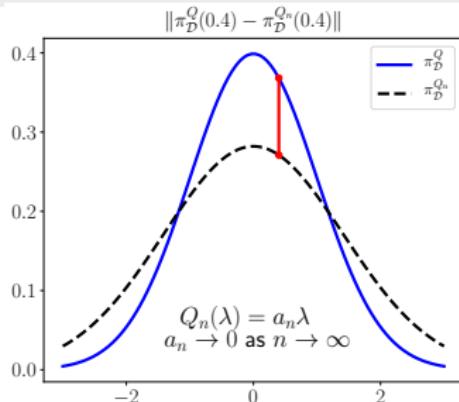
### Theorem [BJW18c]

Let  $Q_n(\lambda)$  be sequence of approximations s.t.

$$\forall \delta > 0 \exists N \text{ s.t. } n > N \Rightarrow \|Q_n(\lambda) - Q(\lambda)\|_{L^\infty(\Lambda)} < \delta.$$

For any  $\epsilon$ , there exists  $N$  s.t.  $n > N$  implies

$$\left\| \pi_{\mathcal{D}}^Q(q) - \pi_{\mathcal{D}}^{Q_n}(q) \right\|_{L^\infty(\mathcal{D})} < \epsilon, \quad \left\| \pi_{\mathcal{D}}^Q(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right\|_{L^\infty(\Lambda)} < \epsilon.$$

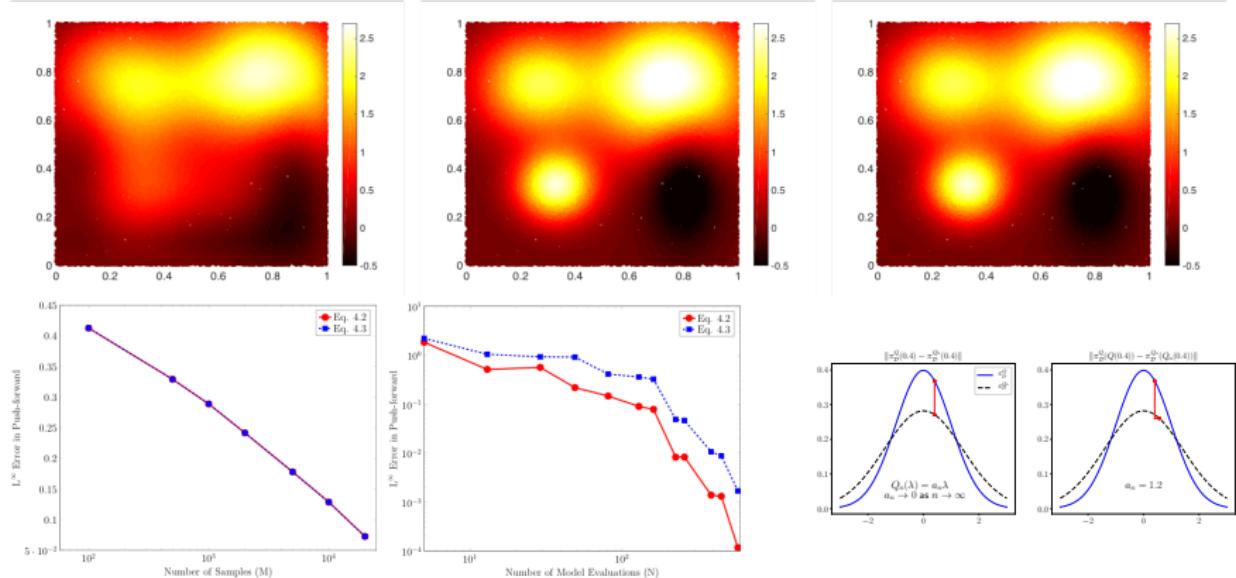


# DENSITY ESTIMATION SAMPLING ON SPARSE GRIDS

Corollary [BJW18c]

Error in push-forward using isotropic level- $n$  CC sparse grid satisfies

$$\|\pi_{\mathcal{D}}^Q(Q(\lambda)) - \hat{\pi}_{\mathcal{D}}^{Q,n}(Q_n(\lambda))\|_{L^\infty(\Lambda)} \leq C \left( \left( \frac{\log M}{M} \right)^{\frac{s}{2s+m}} + C_{k,r} N_n^{-r} (\log N)^{(r+2)(k-1)+1} \right).$$



# FUNCTION APPROXIMATION SUMMARY

## Goal

Build approximations from limited simulation data

## Challenge

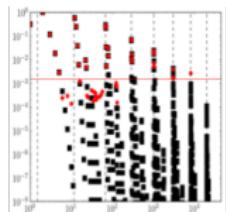
Growth of samples required can grow exponentially with dimension (curse of dimensionality)

## Solution

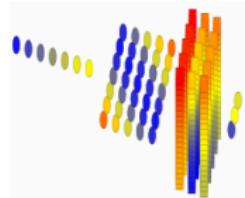
- Exploit structure in function
- Sample in regions of high-probability whilst maximizing conditioning

## Methods

- Sparse grids (smoothness) [JR13, NJ14]
- Compressive sensing (sparsity) [JES15, JNZ17]
- Low-rank decompositions (separability) [GJ18]

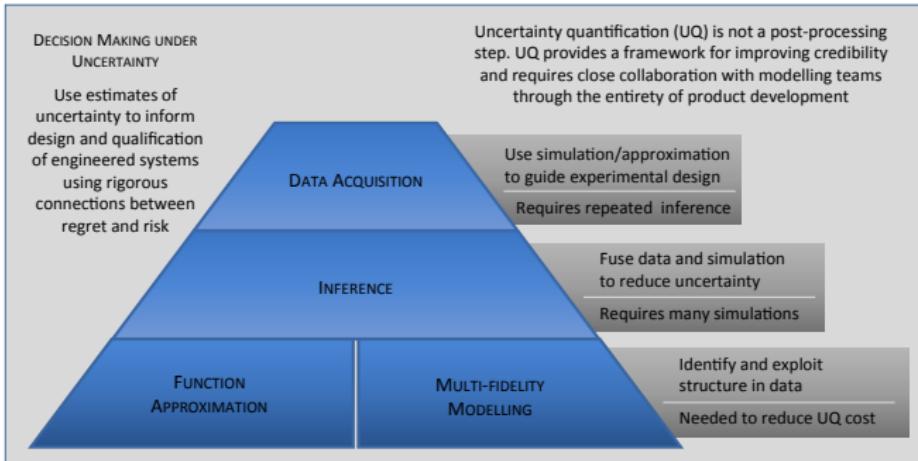


Find most sparse representation



Find low-rank representation

# MULTI-FIDELITY MODELLING



## MULTI-FIDELITY MODELLING BALANCING SOURCES OF ERROR

Multi-fidelity modelling leverages simulations of lower-fidelity models of reduced cost to increase the tractability of sampling/approximating a high-fidelity model

$$\left\| Q - \hat{Q}_{\alpha, \mathcal{I}} \right\|_{L^p(\Lambda)} \leq \underbrace{\left\| Q - \hat{Q}_\alpha \right\|_{L^p(\Lambda)}}_I + \underbrace{\left\| \hat{Q}_\alpha - \hat{Q}_{\alpha, \mathcal{I}} \right\|_{L^p(\Lambda)}}_{II}$$

To minimize simulation cost we should balance physical error (I) with stochastic error (II). I.e. only sample highest fidelity model when stochastic error is smaller than deterministic error [JW15]

If models ensemble forms a hierarchy, i.e.

$$\left\| \hat{Q}_\alpha - Q \right\| \rightarrow 0 \quad \text{as} \quad \max \alpha \rightarrow \infty$$

sparse grids can be naturally extended to multi-fidelity context  
[HANTT16, dBR17, JEGG18]



## CONTROL VARIATE MONTE CARLO FUNDAMENTALS

The MC estimate of the mean

$$\tilde{Q}_{\alpha} = N^{-1} \sum_{n=1}^N \hat{Q}_{\alpha}(\lambda^{(i)})$$

Central Limit Theorem implies error normally distributed with variance  $N^{-1}\mathbb{V}[\hat{Q}_{\alpha}]$ , as  $N \rightarrow \infty$ .

Leverage correlations of low-fidelity models to reduce variance of estimator.

$$\tilde{Q}_{\alpha}^{\text{CV}} = \tilde{Q}_{\alpha} + \gamma \left( \tilde{Q}_{\kappa} - \mu_{\kappa} \right)$$

Given  $r_{\alpha}$  samples of  $\hat{Q}_{\alpha}$  and  $r_{\kappa}$  of  $\hat{Q}_{\kappa}$ , variance in  $\tilde{Q}_{\alpha}^{\text{CV}}$  is

$$\mathbb{V}[\tilde{Q}_{\alpha}^{\text{CV}}] = \left(1 - \frac{r_{\kappa} - r_{\alpha}}{r_{\kappa}r_{\alpha}}\rho^2\right)\mathbb{V}[\tilde{Q}_{\alpha}]$$

where  $\rho$  correlation between  $Q_{\alpha}$  and  $Q_{\kappa}$

## CONTROL VARIATE MONTE CARLO A GENERALISED FRAMEWORK

For multiple models the CV estimator is

$$\tilde{Q}_{\alpha}^{\text{CV}} = \tilde{Q} + \sum_{\alpha \in \mathcal{A}} \gamma_{\alpha} (\tilde{Q}_{\alpha} - \tilde{\mu}_{\alpha}) = \tilde{Q}_{\alpha} + \gamma \vec{\Delta}$$

### Theorem [GGEJ18]

The Optimal CV weights are

$$\gamma^* = -\text{Cov} [\vec{\Delta}, \vec{\Delta}]^{-1} \text{Cov} [\vec{\Delta}, \tilde{Q}]$$

Multi-level MC (MLMC) is a control variate algorithm

$$\mathcal{A} = \{0, \dots, L\}, \gamma = (1, \dots, 1)^T$$

$$\mathbb{E} [\hat{Q}_L] = \mathbb{E} [\hat{Q}_0] + \sum_{\ell=1}^L \mathbb{E} [\hat{Q}_{\ell} - \hat{Q}_{\ell-1}]$$

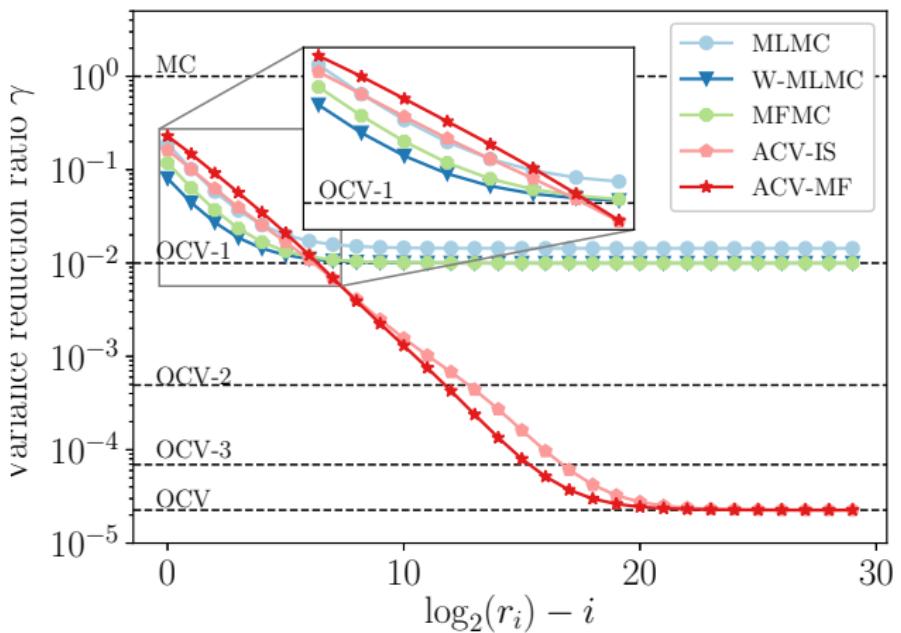
### Theorem [GGEJ18]

Regardless of number of models  $M$ , variance of MLMC satisfies

$$\mathbb{V}[\tilde{Q}^{\text{MLMC}}] < (1 - \rho_{\max}^2) \mathbb{V}[\tilde{Q}]$$

$\rho_{\max}$  is max correlation of  $\hat{Q}_L$  with  $\hat{Q}_i$ ,  $i = 0, \dots, L-1$

Variance reduction for fixed high-fidelity samples of  $Q$  as a function of numbers of samples per level  $r_i(x) = 2^{i+x}$  for 4 low-fidelity models



# MULTIFIDELITY MODELLING SUMMARY

## Goal

Use ensemble of models to reduce errors in statistics

## Challenge

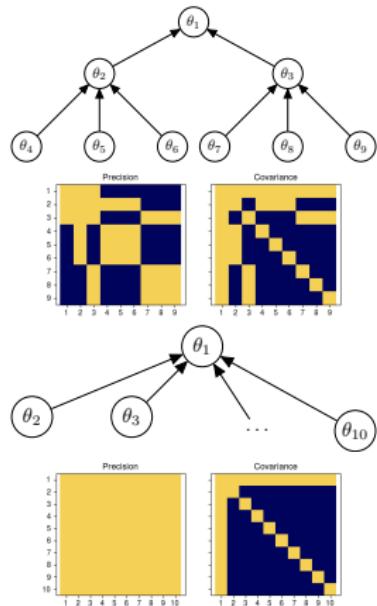
Relationship between models may not be known

## Solution

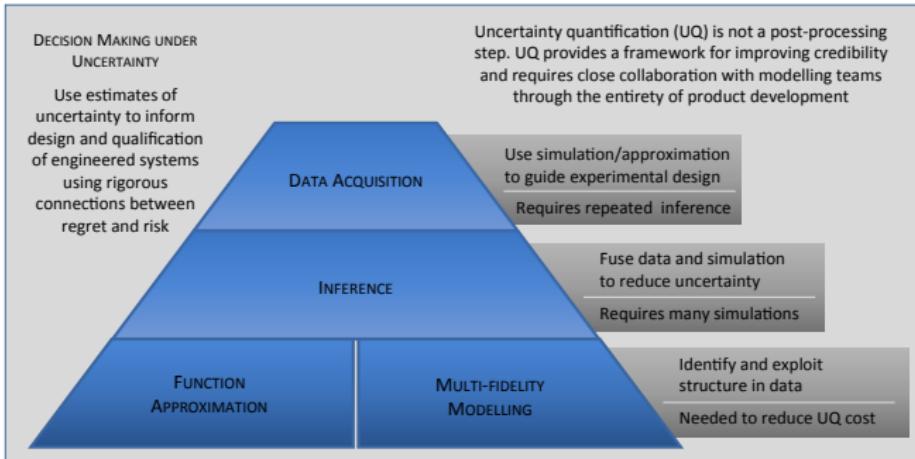
- Learn and exploit relationship between models
- Allocate samples between models to balance deterministic and stochastic errors

## Methods

- Variance reduction methods (MLMC, CVMC)  
[GEGJ18, GGEJ18]
- Multi-index approximation [JEGG18]
- Bayesian network learning [GJGE18]

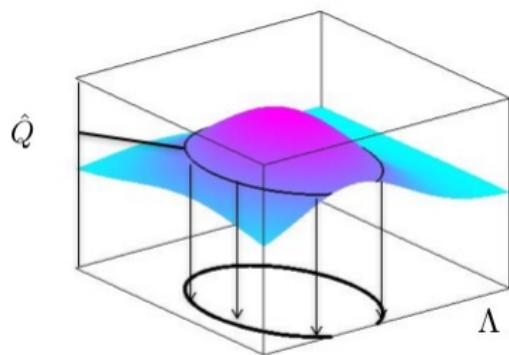


# INVERSE UNCERTAINTY QUANTIFICATION



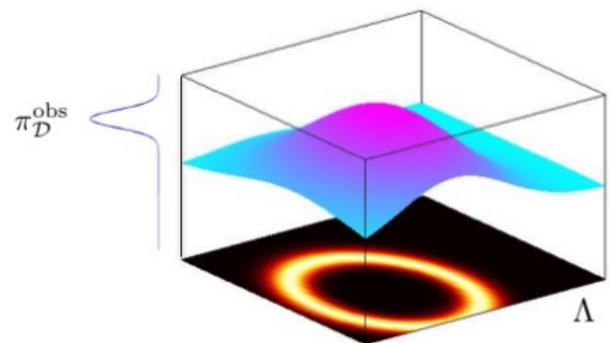
# PARAMETER INFERENCE

## Deterministic Inversion



Find parameter values that produce data. Ill posed must impose regularisation.

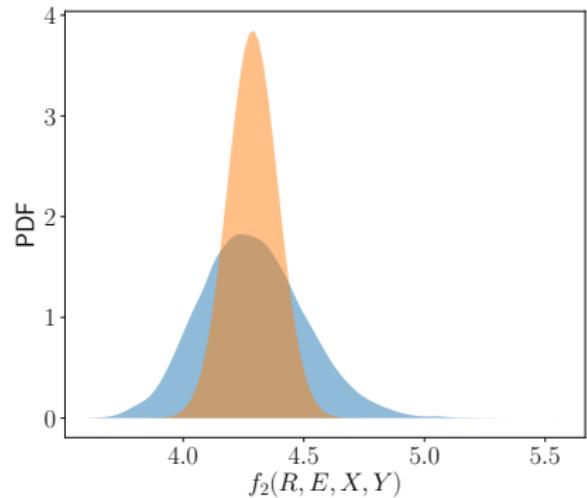
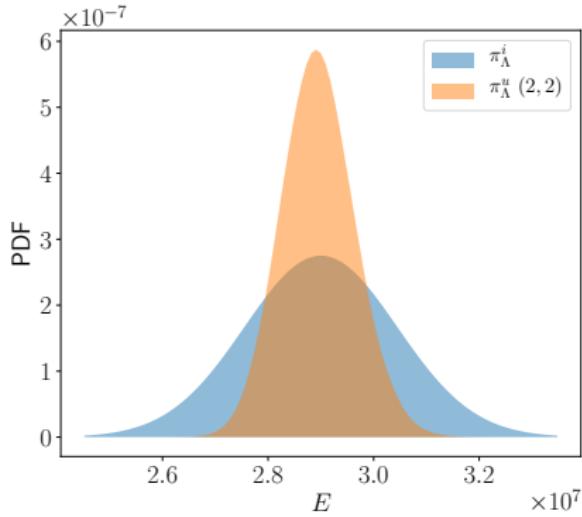
## Stochastic Inversion



Find probability of parameters producing data. Prior distribution is a form of regularisation [Stu10].

## REDUCING UNCERTAINTY USING DATA

We can reduce estimates of uncertainty and improve the performance of design whilst still satisfying constraints.



Can we determine the probability density that when push forward through a model reproduces a given density on the observations?

# A NEW APPROACH FOR INVERSION PUSH-FORWARD BASED INFERENCE

## Theorem [BJW18b]

The consistent updated density is

$$\pi_{\Lambda}^u(\lambda) = \pi_{\Lambda}^i(\lambda) \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda))}{\pi_{\mathcal{D}}^Q(Q(\lambda))}.$$

## Algorithm: Approximating the Push-forward of the Prior

- 1 Given a set of samples from the prior density:  $\{\lambda_i\}_{i=1}^M$ .
- 2 Evaluate the model and compute the QoIs:  $q_i = Q(\lambda_i)$ .
- 3 Use the set of QoIs and use a standard technique, such as a kernel density method, to estimate  $\pi_{\mathcal{D}}^Q(q)$ .

## EXAMPLE A SIMPLE NONLINEAR SYSTEM

### Example

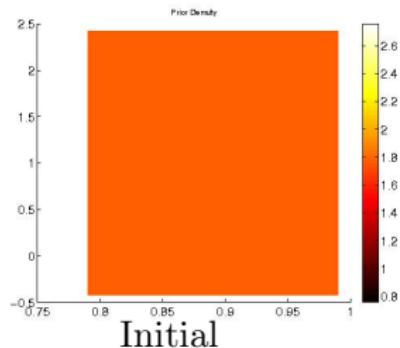
Consider a 2-component nonlinear system of equations with two parameters

$$\begin{aligned}\lambda_1 x_1^2 + x_2^2 &= 1, \\ x_1^2 - \lambda_2 x_2^2 &= 1\end{aligned}$$

- The quantity of interest is the second component:  $q(\lambda) = x_2$ .
- Choose  $\Lambda$  to give interesting variation in QoI.
- Assume that we observe  $q(\lambda) \sim N(0.3, 0.025^2)$ .
- We consider uniform initial density.
- We use 100,000 samples from the initial density and a standard KDE to approximate the push-forward of the that density.

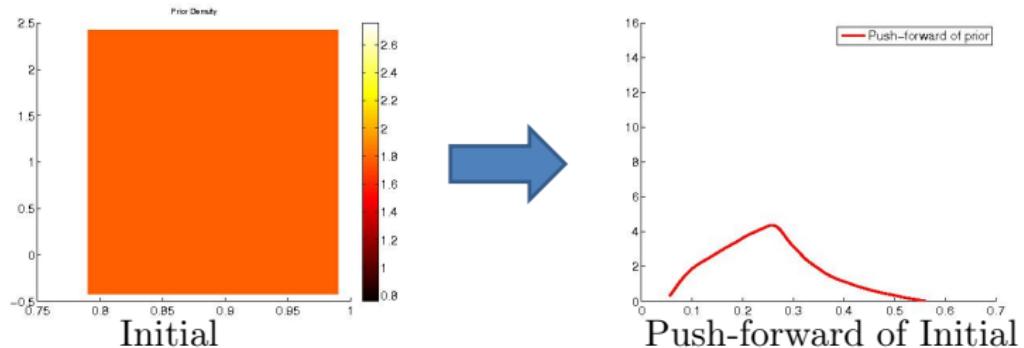
# EXAMPLE

## A SIMPLE NON-LINEAR SYSTEM



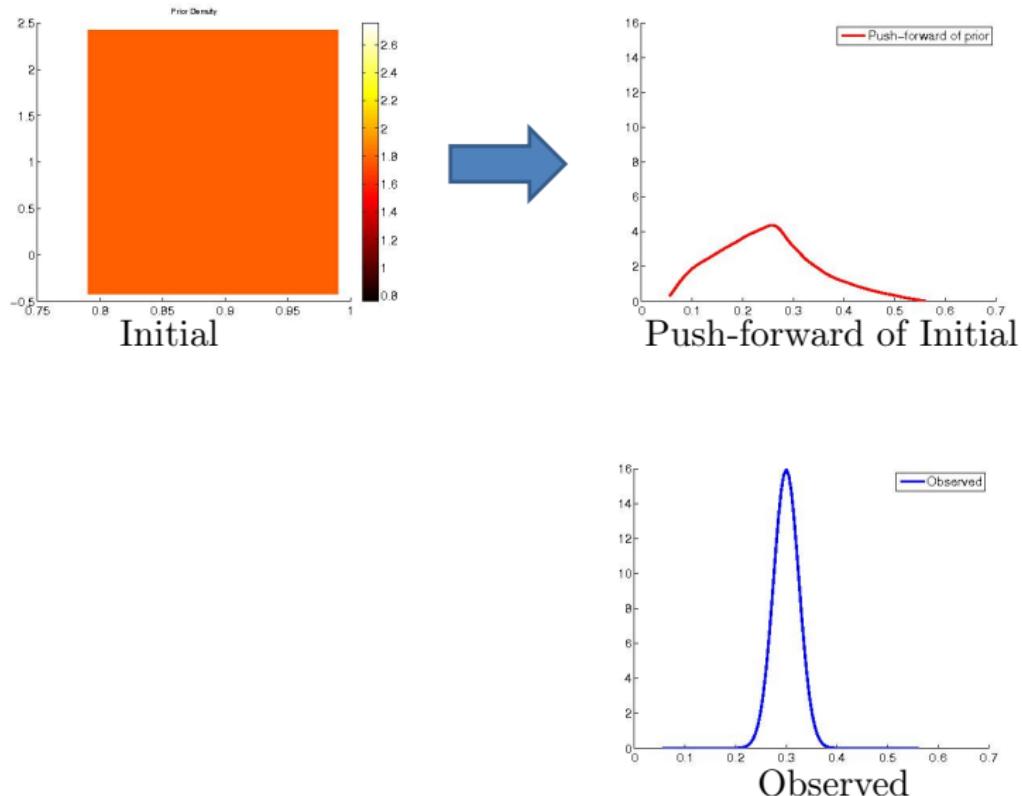
# EXAMPLE

## A SIMPLE NON-LINEAR SYSTEM



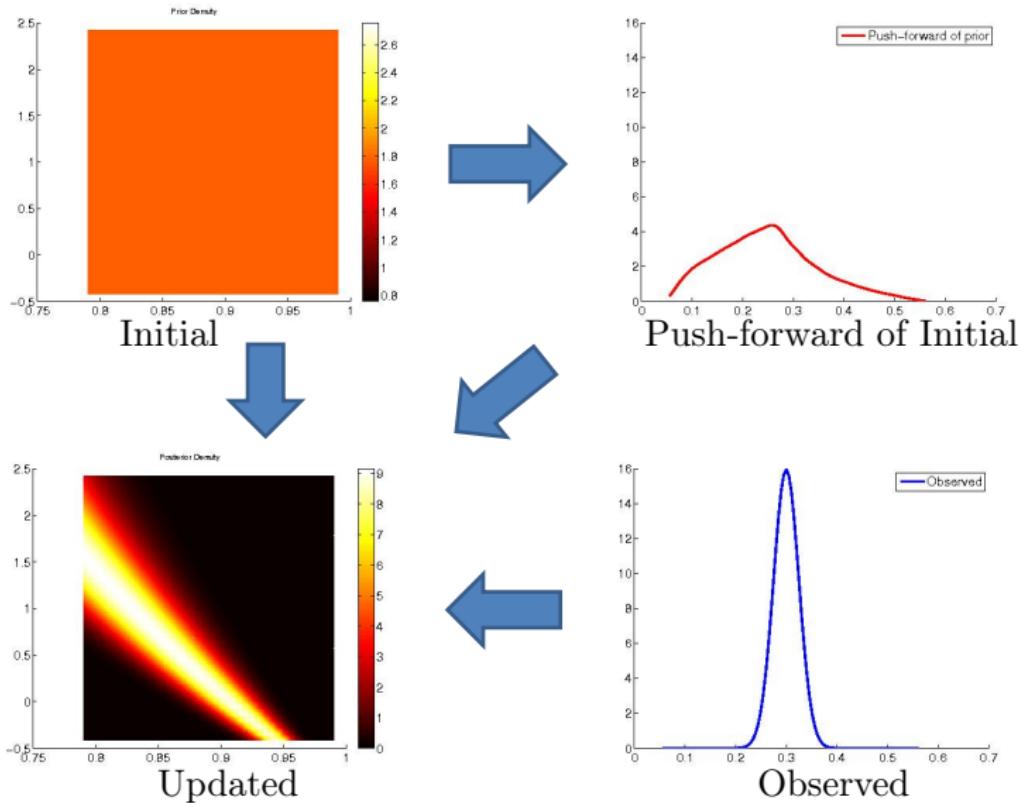
# EXAMPLE

## A SIMPLE NON-LINEAR SYSTEM



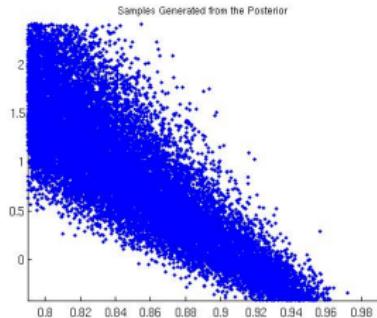
# EXAMPLE

## A SIMPLE NON-LINEAR SYSTEM



## EXAMPLE A SIMPLE NON-LINEAR SYSTEM

We can use rejection sampling to draw samples from the updated density

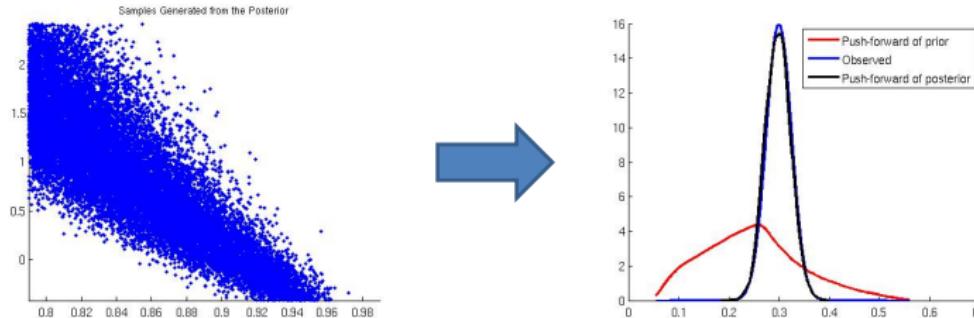


Samples from the updated density

Using rejection sampling for  $\pi_{\Lambda}^u$  using samples from  $\pi_{\Lambda}^i$  is equivalent to sampling  $\pi_{\mathcal{D}}^{\text{obs}}$  using samples from  $\pi_{\mathcal{D}}^Q$ .

## EXAMPLE A SIMPLE NON-LINEAR SYSTEM

Push-forward of samples from updated density matches the observed density



Samples from the updated density

All three densities in  $\mathcal{D}$

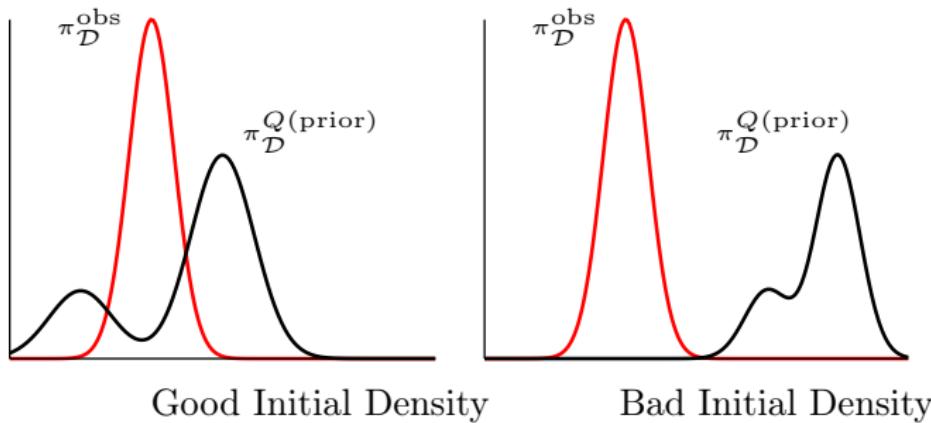
Using rejection sampling for  $\pi_{\Lambda}^u$  using samples from  $\pi_{\Lambda}^i$  is equivalent to sampling  $\pi_{\mathcal{D}}^{\text{obs}}$  using samples from  $\pi_{\mathcal{D}}^Q$ .

## A KEY ASSUMPTION

### Predictability Assumption

We assume that the observed probability measure,  $P_{\mathcal{D}}^{\text{obs}}$ , is absolutely continuous with respect to the push-forward of the initial density,  $P_{\mathcal{D}}^Q$ .

In other words, we assume we can predict everything we observe.

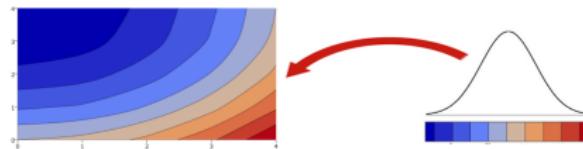


## IMPORTANT PROPERTIES

### Theorem [BJW18b]

The consistent updated density is

$$\pi_{\Lambda}^u(\lambda) = \pi_{\Lambda}^i(\lambda) \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda))}{\pi_{\mathcal{D}}^Q(Q(\lambda))}.$$



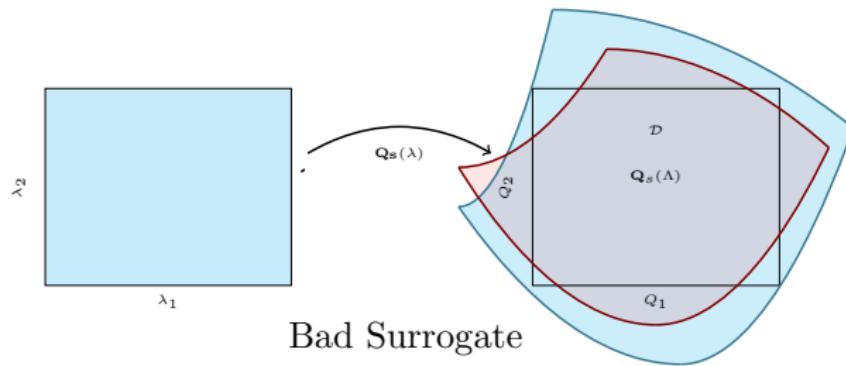
$$P_{\Lambda}(A) = \int_{\mathcal{D}} \left( \int_{A \cap Q^{-1}(q)} dP_q(\lambda) \right) dP_{\mathcal{D}}(q).$$

### Theorem [BJW18b]

Updated density is stable with respect to perturbations in  $\pi_{\mathcal{D}}^{\text{obs}}$  and  $\pi_{\Lambda}^i$ .

# UTILIZING SURROGATE MODELS

What happens when we use a surrogate model to compute the push-forward

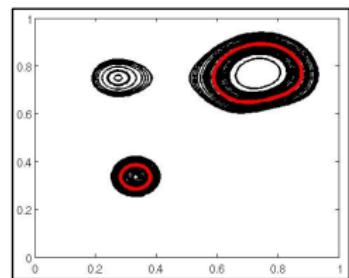
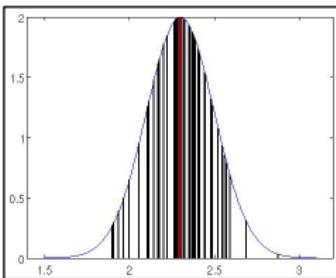
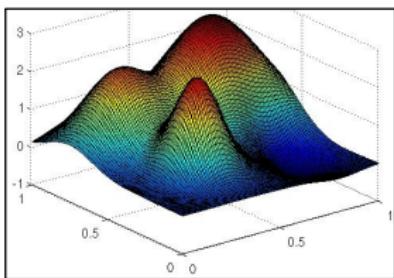


Theorem [BJW18c]

The error in the updated density using an isotropic level- $n$  CC sparse grid satisfies

$$\|\pi_{\Lambda}^u(\lambda) - \hat{\pi}_{\Lambda}^{u,n}(\lambda)\|_{L^1(\Lambda)} \leq C \left( \left( \frac{\log M}{M} \right)^{\frac{s}{2s+m}} + C_{k,r} N_n^{-r} (\log N)^{(r+2)(k-1)+1} \right).$$

## EXAMPLE GAUSSIAN PEAKS



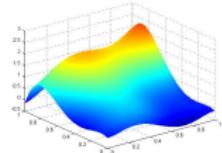
Gaussian peaks function Density on observations Contours in  $\Lambda$

- We let  $\Lambda = [0, 1]^2$  and consider a sum of Gaussian peaks.
- The initial density is uniform over  $\Lambda$ .
- The goal is to investigate how the accuracy of the surrogate model affects the updated density.

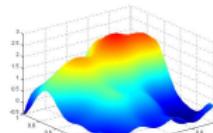
# EXAMPLE GAUSSIAN PEAKS

Level 2 (17 pts)

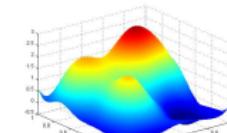
$$Q_s(\lambda)$$



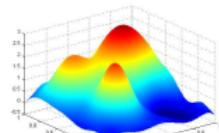
Level 3 (49 pts)



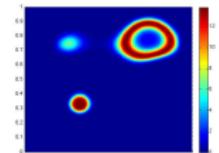
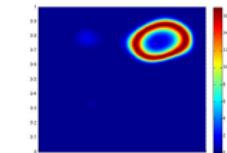
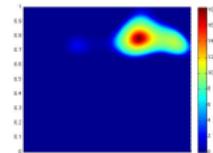
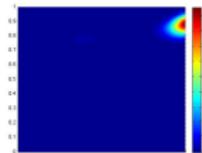
Level 4 (97 pts)



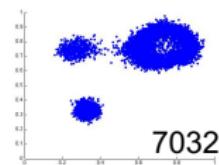
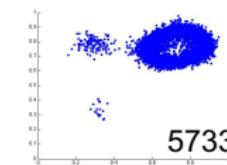
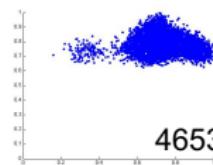
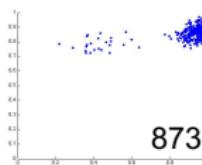
Level 5 (161 pts)



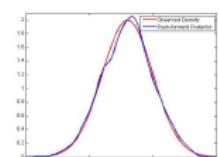
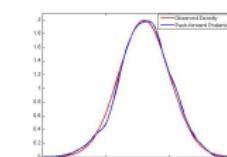
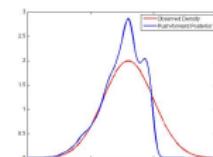
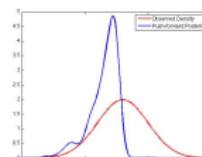
$$\pi_{\Lambda}^{\text{post}}(\lambda)$$



Samples  
from  
posterior



$$\pi_{\mathcal{D}}^{\text{obs}}(q) \text{ and } \pi_{\mathcal{D}}^{Q(\text{post})}(q)$$



$$\int_{\Lambda} \pi_{\Lambda}^{\text{post}}(\lambda) d\mu_{\Lambda}$$

$$0.4789$$

$$0.8704$$

$$0.9787$$

$$0.9825$$

# INVERSE UNCERTAINTY QUANTIFICATION SUMMARY

## Goal

Condition estimates of uncertainty on experimental data

## Challenge

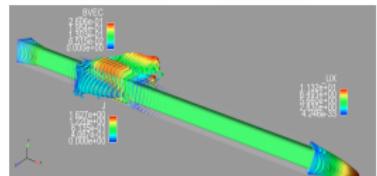
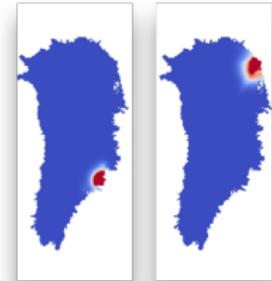
Develop new formulations and reduce sample complexity

## Solution

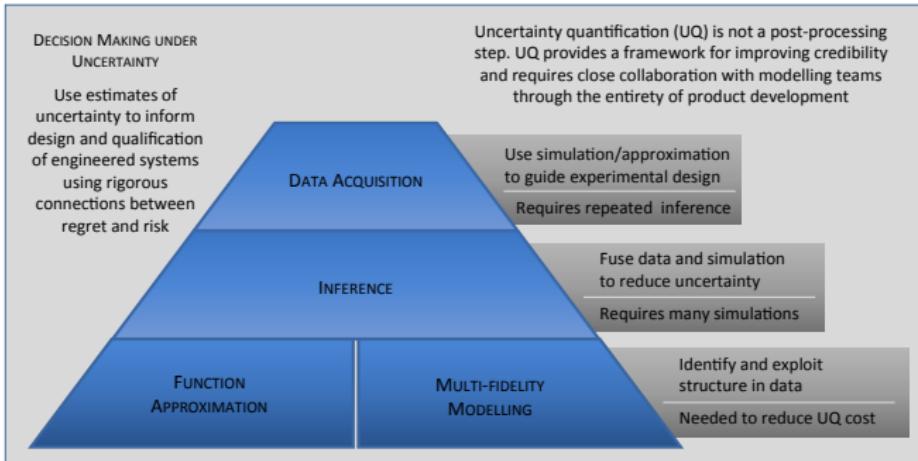
- Find input measure whose push-forward matches observed density
- Reduce inverse problem to one forward solve

## Methods

- Bayesian inference
- Push-forward based inference  
[BJW18a, BJW18c]

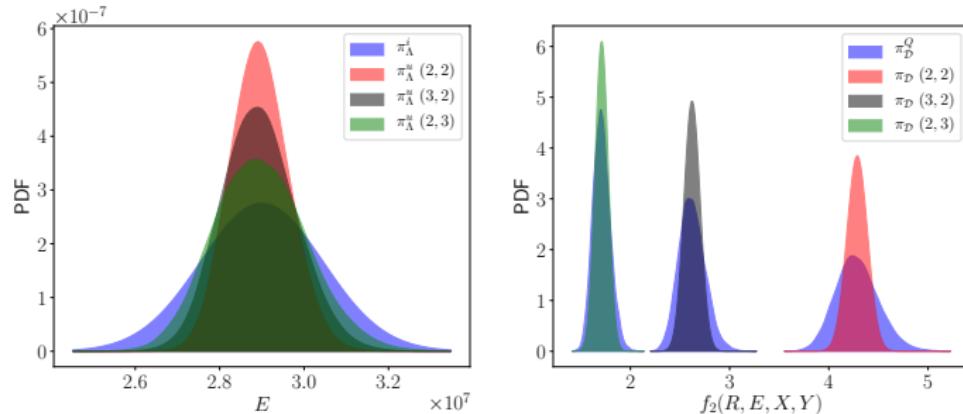


# DATA ACQUISITION



# OPTIMAL EXPERIMENTAL DESIGN (OED)

Data are not equally informative



How does one select the maximise information gain whilst minimizing cost of experimentation.

Use measure of change in uncertainty

$$KL(\pi_{\Lambda}^i : \pi_{\Lambda}^u) := \int_{\Lambda} \pi_{\Lambda}^u \log \left( \frac{\pi_{\Lambda}^u}{\pi_{\Lambda}^i} \right) d\mu_{\Lambda}.$$

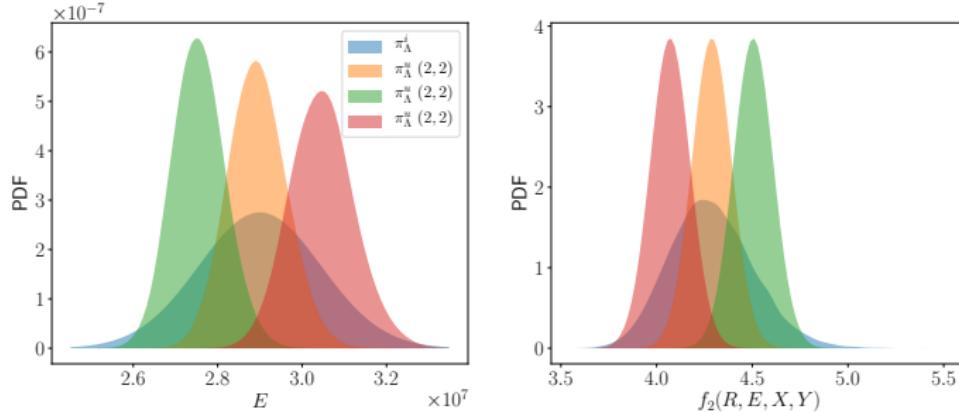
## INFORMATION THEORETIC DESIGN

OED must select a design before experimental data become available.

In the absence of data, we use the simulation model to quantify the information gain of a given experimental design for all possible realisations of data for that design.

Let  $\mathcal{O}$  denote the space of densities that may be observed in reality

$$\mathcal{O} = \left\{ \hat{N}(\mathbb{E}[\pi_{\mathcal{D}}^Q] + \tau \mathbb{V}[\pi_{\mathcal{D}}^Q]^{1/2}, \sigma^2) : \tau \in \{-1, 0, 1\} \right\},$$



## Expected Information Gain

$$EIG(Q) := \int_{\mathcal{D}} KL(d) \pi_{\mathcal{D}}^Q(d) d\mu_{\mathcal{D}}.$$

Given samples from push-forward:

$q^{(j)} = Q(\lambda^{(j)})$  compute

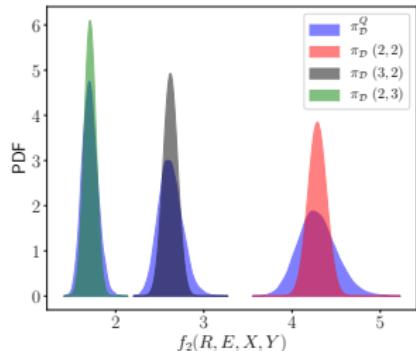
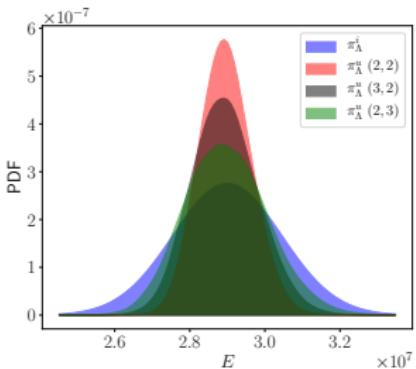
$$I_Q(\tau) \approx \frac{1}{N} \sum_{i=1}^N \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda^{(i)}))}{\pi_{\mathcal{D}}^Q(Q(\lambda^{(i)}))} \log \left( \frac{\pi_{\mathcal{D}}^{\text{obs}}(Q(\lambda^{(i)}))}{\pi_{\mathcal{D}}^Q(Q(\lambda^{(i)}))} \right)$$

## OED definition

Let  $Q^z \in \mathcal{Q}$  be a specific design in space of all possible designs, then OED solves

$$Q^{\text{opt}} := \arg \max_{Q^z \in \mathcal{Q}} E(I_{Q^z}).$$

We would choose (2,2)



# DATA ACQUISITION APPLICATIONS

## Direct field acoustic testing (DFAT)



# DATA ACQUISITION SUMMARY

## Goal

Determine data that maximizes reduction in uncertainty whilst minimizing cost

## Challenge

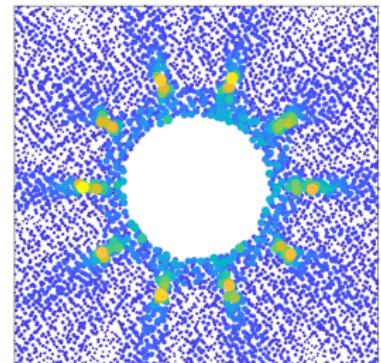
Requires solving many inverse problems

## Solution

- Use push-forward based inference - only one forward problem required
- Use gradient based optimization

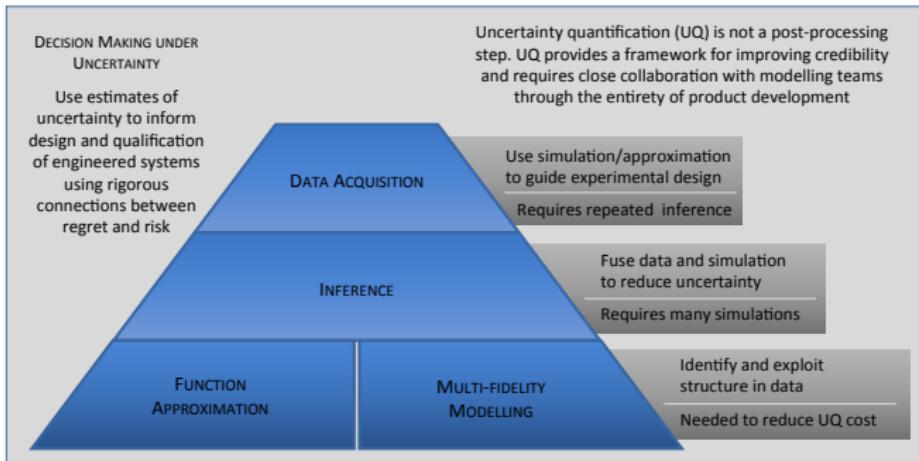
## Methods

- OED using SVD of Jacobians [BJP<sup>+</sup>18]
- OED using Push-forward based inference [WWJ17]

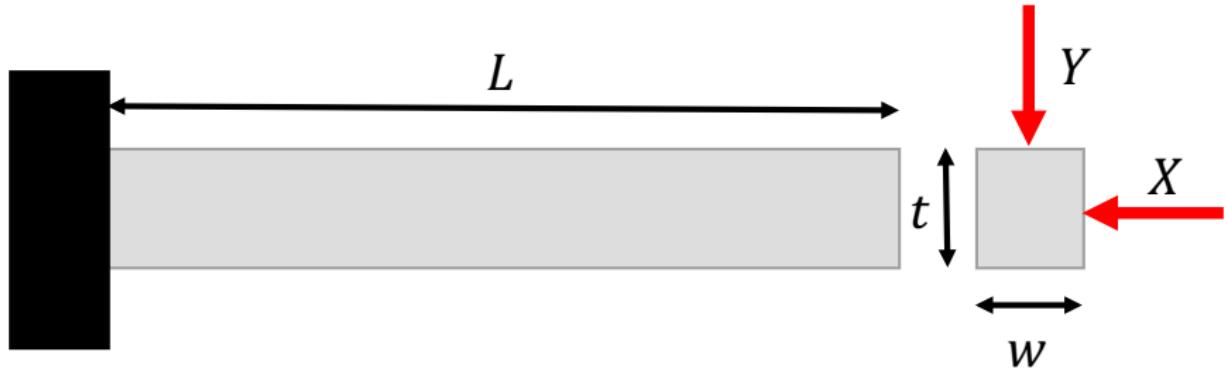


# DESIGN AND DECISION MAKING UNDER UNCERTAINTY

---



## A SIMPLE MOTIVATING EXAMPLE CANTILEVER BEAM



$$f_1(\boldsymbol{\lambda}) = 1 - \frac{6L}{Rwt} \left( \frac{X}{w} + \frac{Y}{t} \right) \geq 0 \quad f_2(\boldsymbol{\lambda}) = 1 - \frac{4L^3}{2.2535Ewt} \sqrt{\frac{X^2}{w^4} + \frac{Y^2}{t^4}} \geq 0$$

Uncertainty	Symbol	Prior
Yield stress	$R$	$N(40000, 2000)$
Young's modulus	$E$	$N(2.9e7, 1.45e6)$
Horizontal load	$X$	$N(500, 100)$
Vertical Load	$Y$	$N(1000, 100)$

# DESIGN UNDER UNCERTAINTY

## Deterministic Design

$$\underset{w,t}{\operatorname{argmin}} \ wt$$

$$f_1(\lambda) \geq 0$$

$$f_2(\lambda) \geq 0$$

$$1 \leq w \leq 4 \quad 1 \leq t \leq 4$$

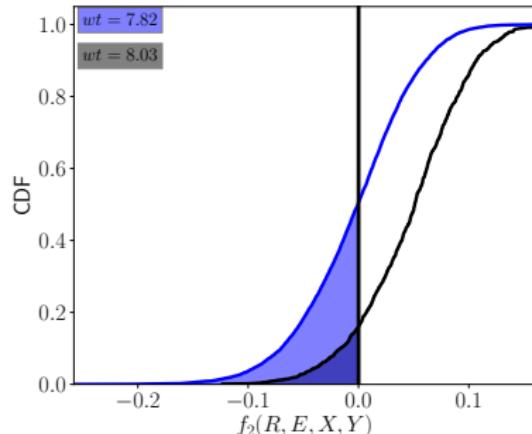
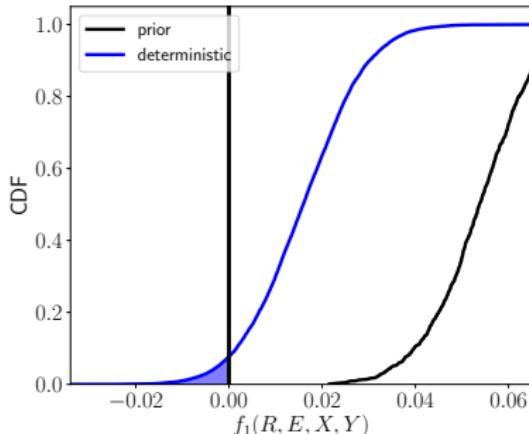
## Design under uncertainty

$$\underset{w,t}{\operatorname{argmin}} \ wt$$

$$P(f_1(\lambda) \leq 0) \leq \delta_1$$

$$P(f_2(\lambda) \leq 0) \leq \delta_2$$

$$1 \leq w \leq 4 \quad 1 \leq t \leq 4$$



# PUTTING IT ALL TOGETHER CANTILEVER BEAM

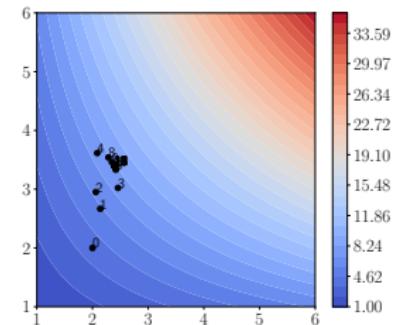
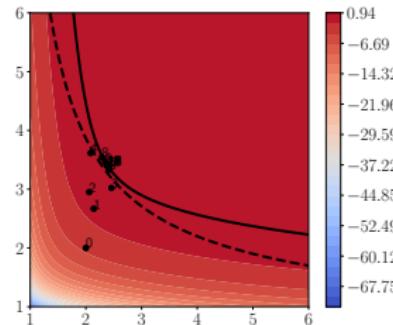
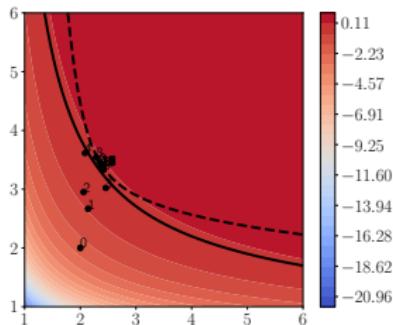
## Design under uncertainty

$$\underset{w,t}{\operatorname{argmin}} \ wt$$

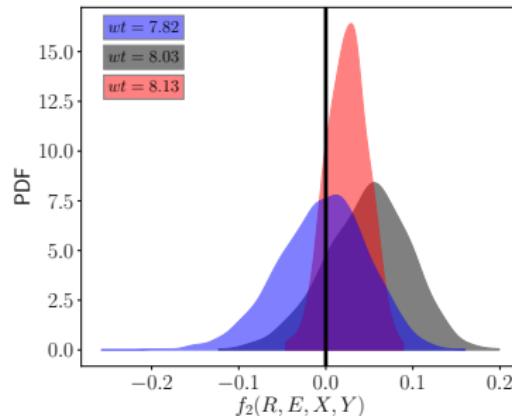
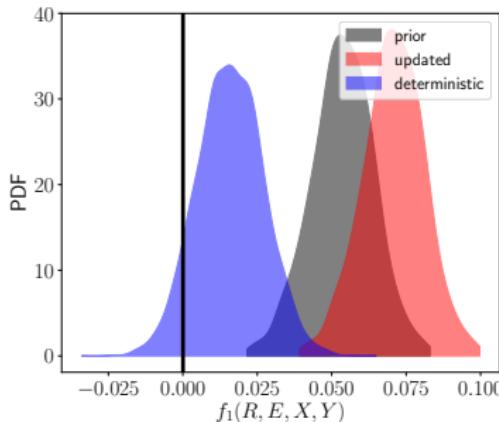
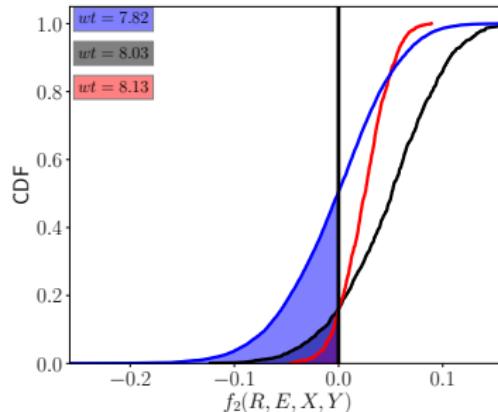
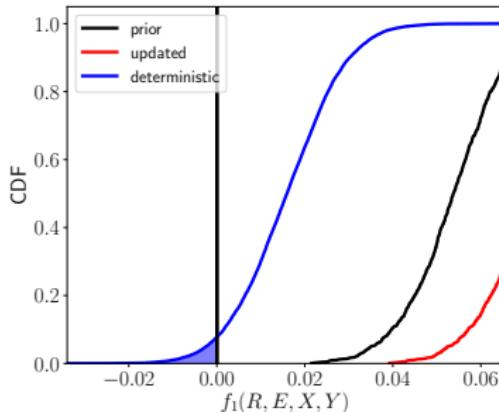
$$P(f_1(\lambda) \leq 0) \leq \delta_1$$

$$P(f_2(\lambda) \leq 0) \leq \delta_2$$

$$1 \leq w \leq 4 \quad 1 \leq t \leq 4$$



# PUTTING IT ALL TOGETHER CANTILEVER BEAM



## DESIGN UNDER UNCERTAINTY APPLICATION

Minimise weight subject to thrust and thermal and structural failure constraints



# UNCERTAINTY QUANTIFICATION QUESTIONS

## Function Approximation:

- Are sparse grids suitable for all classes of functions?
- What other ways can one delay the curse of dimensionality?
- How can one handle dependent probability measures?

## Inference:

- What happens as the variance of the observed density is reduced?
- How does this approach scale with the number of parameters and data?
- Is all data equally informative?

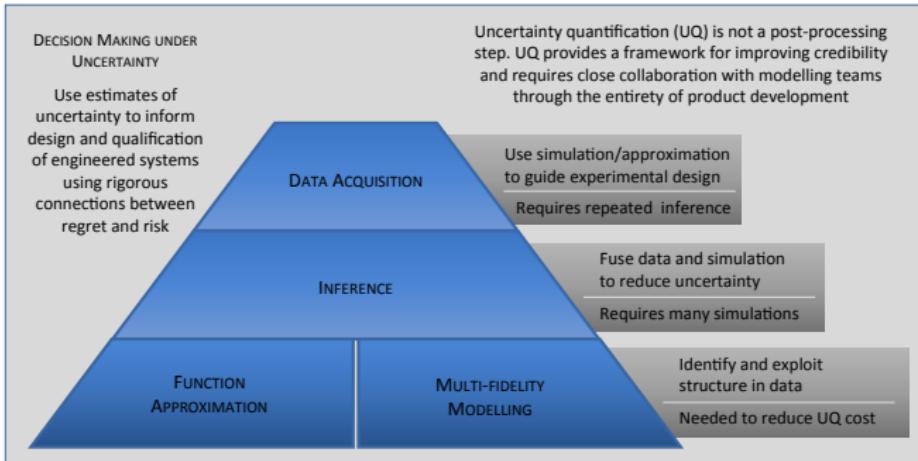
## Multi-fidelity Modelling:

- How do multi-fidelity sparse grids work?
- What if parameter-to-input map is not smooth?
- Can other types of model relationships be exploited?
- What if models have different parameterisations?

## Data Acquisition:

- What if we want to select more than one experiment?
- Expected information gain minimises average performance. Can we be more conservative?

# ANSWERS TO SOME QUESTIONS

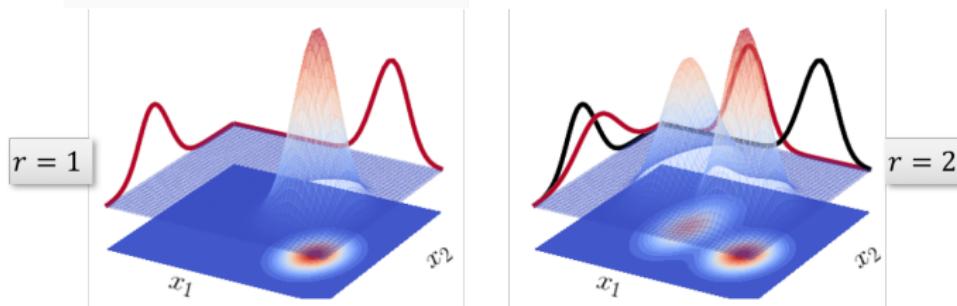
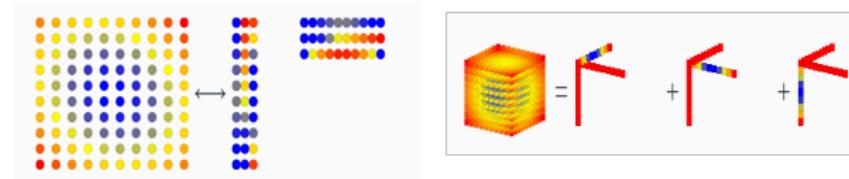


# DELAYING THE CURSE OF DIMENSIONALITY

## LOW-RANK TENSOR DECOMPOSITIONS

The canonical tensor decomposition represents a tensor as the sum of outer product of  $d$  vectors

$$A = \sum_{i=1}^r v_1 \circ \cdots \circ v_d$$



Number of samples required grows quadratically with rank  $r$  and linearly with dimension  $d$  and number of univariate bases  $p$  [GJ18]

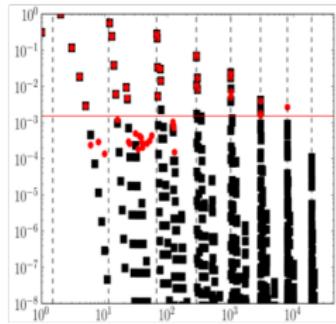
# DELAYING THE CURSE OF DIMENSIONALITY

## SPARSE APPROXIMATION

Approximate function with small number of nonzero terms  $f_{\Lambda}(\lambda) = \sum_{i \in I} \alpha_i \phi_i(\lambda)$

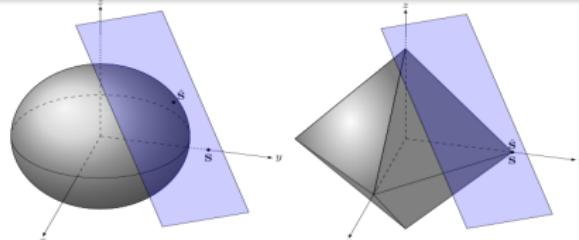
$$s = \#\{i \mid |\alpha_i| > \delta\}$$

$l_0$  -minimization (**NP HARD**)

$$\min_{\alpha} \|\alpha\|_0 \text{ s.t. } \|f - f_{\Lambda}\|_2 \leq \epsilon$$


$l_1$  -minimization (**Finds sparse solution under certain conditions**)

$$\min_{\alpha} \|\alpha\|_1 \text{ s.t. } \|f - f_{\Lambda}\|_2 \leq \epsilon$$



If a function is sparse the number of samples required to compute the coefficients only grows **linearly** with dimension

Sampling strategies for weighted probability spaces are needed [JNZ17]

### QUESTIONS

---

- How do multifidelity sparse grids work?
- What if parameter-to-input map is not smooth?
- Can other types of model relationships be exploited?
- What if models have different parameterisations?

# MULTIFIDELITY APPROXIMATION NATURALLY EXTENDING SPARSE GRIDS

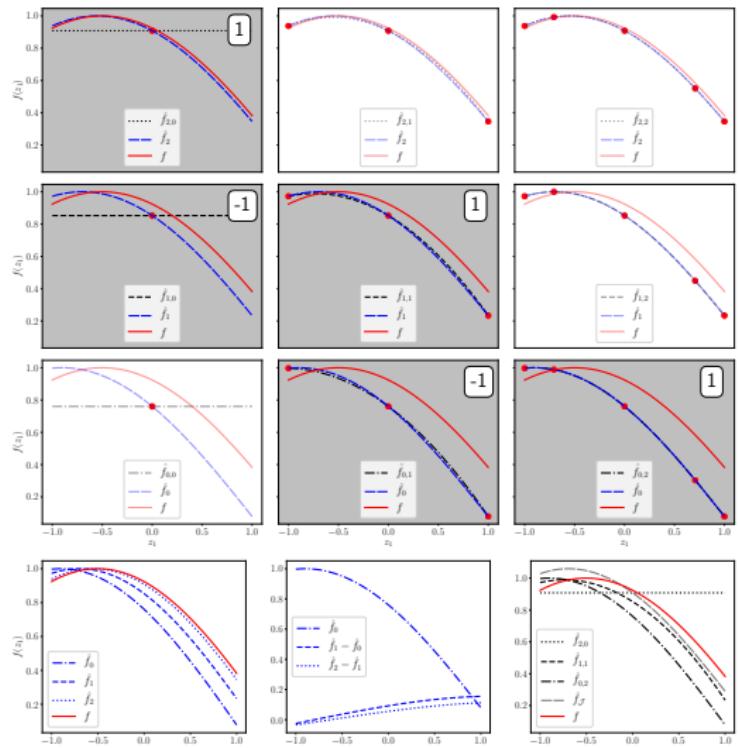
Sparse grids can be used for approximation and solving PDEs [LHH<sup>+</sup>13].

We can combine these tasks in a natural way

Use most accurate interpolation on coarsest mesh level and least accurate interpolation on finest level

Effectiveness of method depends of rate of discrepancy decay vs approximation error decay

$$\hat{Q}_{\mathcal{J}}(\boldsymbol{\lambda}) = \sum_{[\alpha, \beta] \in \mathcal{J}} c_{\alpha, \beta} \hat{Q}_{\alpha, \beta}(\boldsymbol{\lambda})$$



## MULTI-INDEX APPROXIMATION ERROR

Isotropic approximation not well suited for multi-index approximation.

Work needed to add a new tensor-product interpolant and resulting reduction in error are given by

$$\Delta E_{\alpha,\beta} = \left\| \hat{Q}_{\mathcal{J} \cup \{[\alpha,\beta]\}} - \hat{Q}_{\mathcal{J}} \right\| \quad \Delta W_{\alpha,\beta} = \left| \text{Work}[\hat{Q}_{\mathcal{J} \cup \{[\alpha,\beta]\}}] - \text{Work}[\hat{Q}_{\mathcal{J}}] \right|$$

Work required to build sparse grid and resulting approximation error are

$$\text{Work}[\hat{Q}_{\mathcal{J}}] = \sum_{[\alpha,\beta] \in \mathcal{J}} \Delta W_{\alpha,\beta} \quad \left\| Q - \hat{Q}_{\mathcal{J}} \right\| = \sum_{[\alpha,\beta] \notin \mathcal{J}} \Delta E_{\alpha,\beta}$$

Must choose index set to minimise error and work

## MULTI-INDEX APPROXIMATION ALLOCATING WORK

Find quasi-optimal index set as a binary knapsack problem  
[HANTT16, dBR17, NE12]

$$\max \sum_{[\beta] \in \mathbb{N}_0^{n_\beta}} \Delta E_{\alpha, \beta} \delta_{\alpha, \beta} \quad \text{s.t.} \quad \sum_{[\beta] \in \mathbb{N}_0^{n_\beta}} \Delta W_{\alpha, \beta} \delta_{\alpha, \beta} \leq W_{\max}, \quad \delta_{\alpha, \beta} \in \{0, 1\}$$

The resulting grid depends on the space of functions that  $f_\alpha$  belongs to. Assumptions on number of continuous mixed derivatives leads to isotropic sparse grids

$$\mathcal{J} = \left\{ [\alpha, \beta] \in \mathbb{N}_0^{n_\alpha + n_\beta} \mid \frac{\Delta E_{\alpha, \beta}}{\Delta W_{\alpha, \beta}} > \epsilon \right\},$$

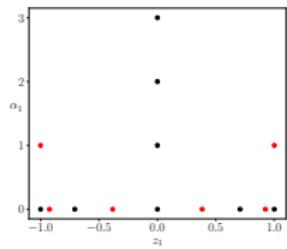
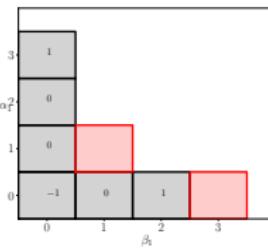
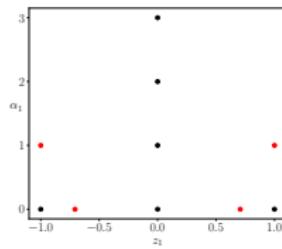
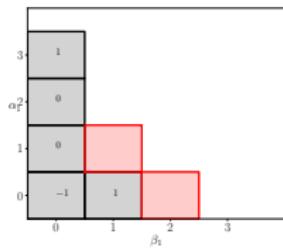
Theorem [NTT16]

If  $Q$  is Hilbert valued function and

$$\left( \sum_{\alpha, \beta} \left( \frac{\Delta E_{\alpha, \beta}}{\Delta W_{\alpha, \beta}} \right)^p \Delta W_{\alpha, \beta} \right)^{\frac{1}{p}} = C(p) < \infty \text{ then}$$

$$\left\| Q - \hat{Q}_{\mathcal{J}} \right\|_{L^2(\Lambda)} \leq C(p) \text{Work} [\hat{Q}_{\mathcal{J}}]^{1 - \frac{1}{p}} \text{ for } 0 < p \leq 1$$

Constructing quasi-optimal index set is intractable so we greedy adaptivity to construct an approximate solution [Heg03, GG03]



# MULTI-INDEX APPROXIMATION ANALYSIS OF AIRCRAFT NOZZLE

Unmanned combat vehicle aircraft demonstrator, capable of carrier take-off and landing

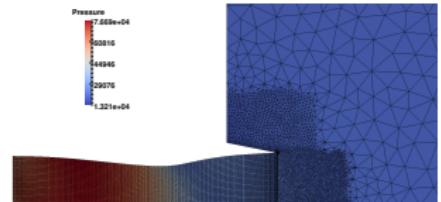
Complex nozzle shape integrated into aft end of vehicle Advanced materials and significant heat environment and thermal management issues

Uncertainties in all areas of multi-physics problem



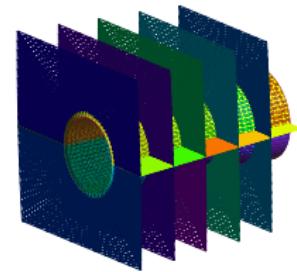
# MULTI-INDEX APPROXIMATION ANALYSIS OF AIRCRAFT NOZZLE

Steady fluid analysis: 2D Axissymmetric Euler



Thermal analysis : Conduction & convection modeled

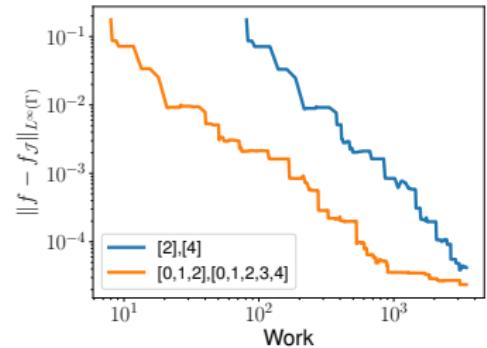
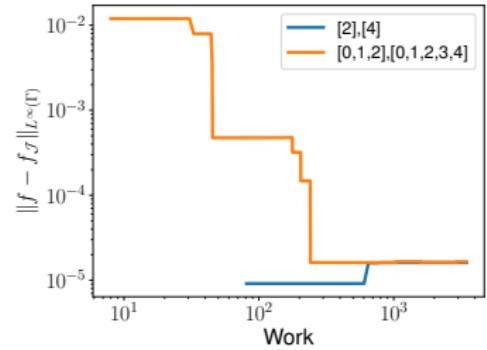
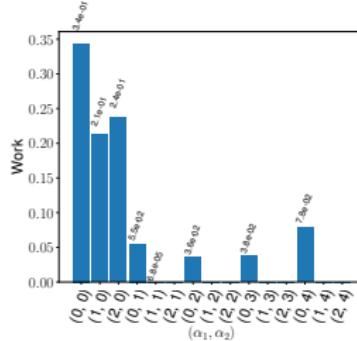
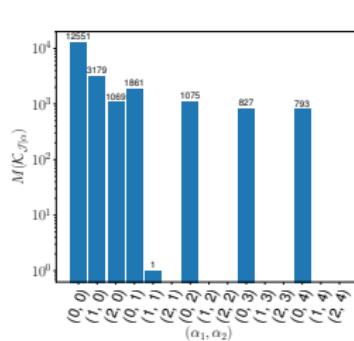
Structural analysis: Pressure and temperature-induced forces in load layers, only temperature-induced forces in thermal layer



# MULTI-INDEX APPROXIMATION ANALYSIS OF AIRCRAFT NOZZLE

We can obtain an almost optimal order of magnitude reduction in cost by sampling coarse meshes more often than fine meshes

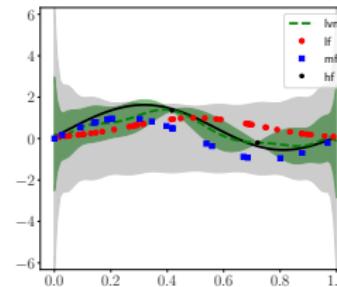
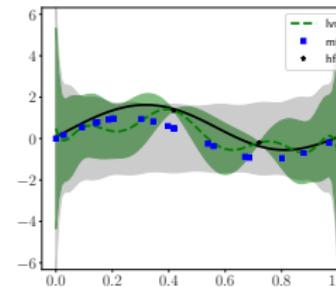
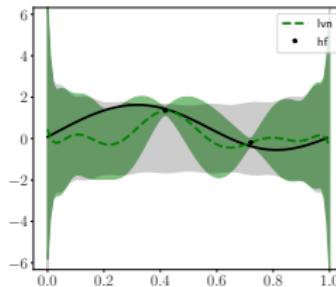
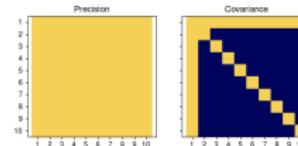
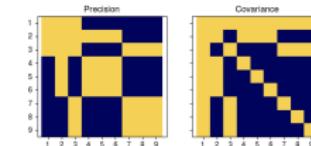
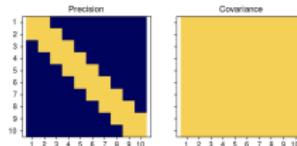
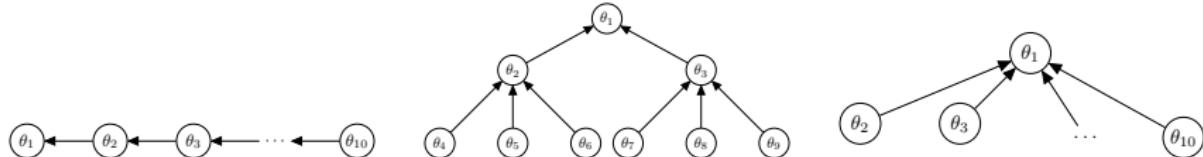
$\alpha_1 \backslash \alpha_2$	0	1	2	3	4
0	36.3	39.3	44.2	61.5	131.2
1	88.5	90.5	95.7	113.1	184.2
2	293.6	297.4	300.0	317.5	384.5



# MULTIFIDELITY MODELLING EXPLOITING DIFFERENT MODEL RELATIONSHIPS

Use Bayesian networks to efficiently compute Bayesian regression basis coefficients. Bayesian generalisation of MLMV and CVMC

Represent each model with polynomial basis with coefficients  $\theta_i$ .  
Estimate high-fidelity  $\theta$  using graph covariance on all models  $\theta$ .



## Example

Consider the linear problem,

$$q = A\lambda, \quad \lambda \in \mathbb{R}^k, \quad q \in \mathbb{R}^p, \quad A \in \mathbb{R}^{p \times k}$$

where we assume  $p \leq k$  and  $A$  is rank- $p$ .

- If we assume Gaussian prior and a Gaussian noise model, then the statistical Bayesian posterior is given by:

$$\pi_{\Lambda, \text{post}}(\lambda) \sim \exp \left( - \left( \underbrace{\frac{1}{2} \|C_q^{-1/2}(q - \tilde{q})\|_2^2}_{\text{Data mismatch}} + \underbrace{\frac{1}{2} \|C_x^{-1/2}(\lambda - \bar{\lambda})\|_2^2}_{\text{Tikhonov regularisation}} \right) \right)$$

- MAP point is identical to Tikhonov-regularised solution to a deterministic optimisation problem.

# PUSH-FORWARD BASED INFERENCE

## CONNECTIONS WITH DETERMINISTIC OPTIMISATION

- Set  $C_A = AC_\lambda A^T$  and  $\bar{q} = A\bar{\lambda}$ .
- The consistent Bayesian updated density is given by,

$$\pi_{\Lambda}^u(\lambda) \sim \exp \left( - \left( \underbrace{\frac{1}{2} \|C_q^{-1/2}(q - \tilde{q})\|_2^2}_{\text{Data mismatch}} + \underbrace{\frac{1}{2} \|C_\lambda^{-1/2}(\lambda - \bar{\lambda})\|_2^2}_{\text{Tikhonov regularisation}} - \underbrace{\frac{1}{2} \|C_A^{-1/2}(q - \bar{q})\|_2^2}_{\text{Un-regularisation}} \right) \right)$$

- We can rewrite the regularisation terms as,

$$\frac{1}{2} \|C_\lambda^{-1/2}(\lambda - \bar{\lambda})\|_2^2 - \frac{1}{2} \|C_A^{-1/2}(q - \bar{q})\|_2^2 = \frac{1}{2} (\lambda - \bar{\lambda})^T \underbrace{(C_\lambda^{-1} - A^T(AC_\lambda A^T)^{-1}A)}_R (\lambda - \bar{\lambda})$$

- If  $A$  is rank- $k$  and invertible  $\Rightarrow R = 0$ .
  - $\Rightarrow$  Regularisation gets “turned off” if there is a unique solution.
- What happens if  $p < k$ ? (less observations than parameters)

## Example

Consider the following linear map,

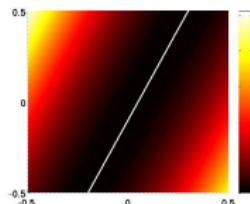
$$A = \begin{bmatrix} 2 & -1 \end{bmatrix},$$

with

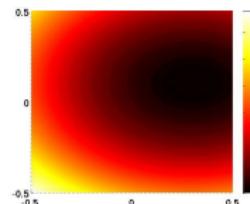
$$\bar{\lambda} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}, \quad C_\lambda = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 0.25 \end{bmatrix}, \quad \tilde{q} = [0.1], \quad C_q = [0.25].$$

# PUSH-FORWARD BASED INFERENCE CONNECTIONS WITH DETERMINISTIC OPTIMISATION

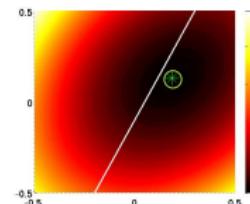
Tikhonov regularisation:



Data mismatch

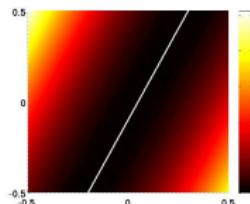


Tikhonov regularisation

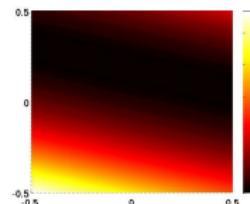


Optimisation function

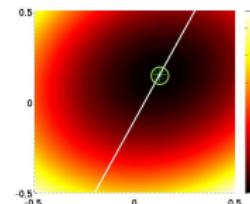
Modified Tikhonov regularisation:



Data mismatch



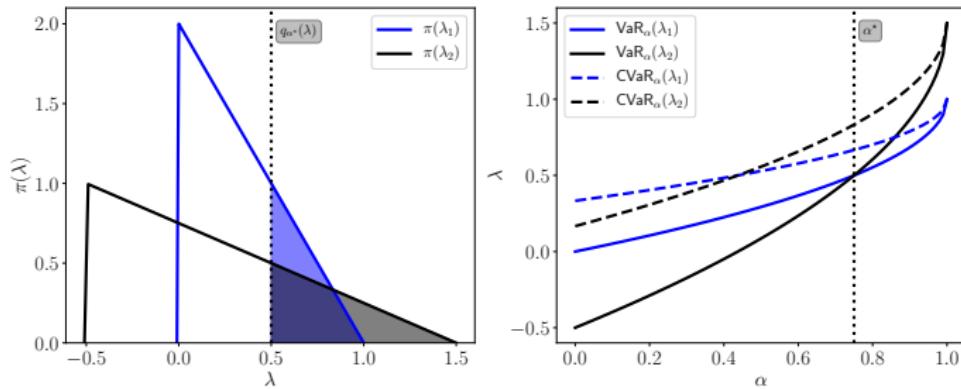
Modified regularisation



Modified optimisation function

## STAKEHOLDER DEFINED UTILITY

What if we do not want to be correct on average but care about avoiding certain outcomes.



$$VaR_\delta(\lambda) = \inf\{\lambda \in \Lambda \mid F_\lambda(\lambda) \geq \delta\}$$

$$CVaR_\alpha(\lambda) = \frac{1}{1-\alpha} \int_\delta^1 VaR_\delta(\lambda) \, d\delta$$

## ACKNOWLEDGMENTS

- Troy Butler (University of Colorado, Denver)
- Mike Eldred (SNL)
- Gianluca Geraci (SNL)
- Alex Gorodetsky (University of Michigan)
- Drew Kouri (SNL)
- John Lewis (SNL)
- Akil Narayan (University of Utah)
- Scott Walsh (University of Colorado, Denver)
- Tim Walsh (SNL)
- Tim Wildey (SNL)

John Jakeman's work was funded partially funded by DARPA EQUIPS.

## REFERENCES

-  George E P Box and Norman R Draper, Empirical model-building and response surface, John Wiley & Sons, Inc., New York, NY, USA, 1986.
-  T. Butler, J.D. Jakeman, M. Pisolov, S. Walsh, and T. Wildey, A new approach to optimal experimental design using singular values of jacobians for non-linear observable maps, Computer Methods in Applied Mechanics and Engineering (2018), Submitted, arXiv:1811.04988.
-  T. Butler, J. Jakeman, and T. Wildey, Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems, SIAM Journal on Scientific Computing 40 (2018), no. 2, A984–A1011.
-  T. Butler, J. Jakeman, and T. Wildey, Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems, SIAM Journal on Scientific Computing 40 (2018), no. 2, A984–A1011.
-  T. Butler, J.D Jakeman, and T. Wildey, Convergence of probability densities using approximate models for forward and inverse problems in uncertainty quantification, SIAM Journal on Scientific Computing 40 (2018), no. 5, A3523–A3548.
-  V. Barthelmann, E. Novak, and K. Ritter, High dimensional polynomial interpolation on sparse grids, Advances in Computational Mathematics 12 (2000), 273–288.
-  Jouke H. S. de Baar and Stephen G. Roberts, Multifidelity sparse-grid-based uncertainty quantification for the hokkaido nansei-oki tsunami, Pure and Applied Geophysics 174 (2017), no. 8, 3107–3121.

## REFERENCES

-  G. Geraci, M.S. Eldred, A. Gorodetsky, and D. Jakeman, J, Leveraging active subspaces for efficient multifidelity uncertainty quantification, 6th European Conference on Computational Mechanics and 7th European Conference on Computational Fluid Dynamics (Glasgow, Scotland, UK), 2018, accepted.
-  T. Gerstner and M. Griebel, Dimension-adaptive tensor-product quadrature, *Computing* 71 (2003), no. 1, 65–87.
-  A. Gorodetsky, G. Geraci, M. Eldred, and J.D. Jakeman, A generalized framework for approximate control variates, *Statistics and Computing* (2018), Submitted.
-  Alex A. Gorodetsky and John D. Jakeman, Gradient-based optimization for regression in the functional tensor-train format, *Journal of Computational Physics* 374 (2018), 1219 – 1238.
-  A. Gorodetsky, J.D. Jakeman, G. Geraci, and M. Eldred, Multifidelity modeling with Bayesian network learning, In preparation.
-  A. Haji-Ali, F. Nobile, L. Tamellini, and R. Tempone, Multi-index stochastic collocation for random pdes, *Computer Methods in Applied Mechanics and Engineering* 306 (2016), 95 – 122.
-  M. Hegland, Adaptive sparse grids, Proc. of 10th Computational Techniques and Applications Conference CTAC-2001 (K. Burrage and Roger B. Sidje, eds.), vol. 44, April 2003, pp. C335–C353.
-  J.D. Jakeman, M. Eldred, G. Geraci, and A. Gorodetsky, Adaptive multi-index collocation and sensitivity analysis, *International Journal for Numerical Methods in Engineering* (2018), Submitted.

## REFERENCES

-  J.D. Jakeman, M.S. Eldred, and K. Sargsyan, Enhancing  $\ell_1$ -minimization estimates of polynomial chaos expansions using basis selection, *Journal of Computational Physics* 289 (2015), no. 0, 18 – 34.
-  J.D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, and D. Pflüger, Polynomial chaos expansions for dependent random variables, *Computer Methods in Applied Mechanics and Engineering* (2018), Submitted.
-  J.D. Jakeman and A. Narayan, Generation and application of multivariate polynomial quadrature rules, *Computer Methods in Applied Mechanics and Engineering* 338 (2018), 134 – 161.
-  J.D. Jakeman, A. Narayan, and T. Zhou, A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions, *SIAM Journal on Scientific Computing* 39 (2017), no. 3, A1114–A1144.
-  J.D. Jakeman and S.G. Roberts, Local and dimension adaptive stochastic collocation for uncertainty quantification, *Sparse Grids and Applications* (J. Garcke and M. Griebel, eds.), Lecture Notes in Computational Science and Engineering, vol. 88, Springer Berlin Heidelberg, 2013, pp. 181–203 (English).
-  J.D. Jakeman and T. Wildey, Enhancing adaptive sparse grid approximations and improving refinement strategies using adjoint-based a posteriori error estimates, *Journal of Computational Physics* 280 (2015), no. 0, 54 – 71.

## REFERENCES

-  J.W. Larson, M. Hegland, B. Harding, S. Roberts, L. Stals, A.P. Rendell, P. Strazdins, M.M. Ali, C. Kowitz, R. Nobes, J. Southern, N. Wilson, M. Li, and Y. Oishi, Fault-tolerant grid-based solvers: Combining concepts from sparse grids and mapreduce, *Procedia Computer Science* 18 (2013), 130 – 139, 2013 International Conference on Computational Science.
-  L. Ng and M. Eldred, Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation, *Proceedings of 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference* (Honolulu, Hawaii), 2012.
-  A. Narayan and J.D. Jakeman, Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation, *SIAM Journal on Scientific Computing* 36 (2014), no. 6, A2952–A2983.
-  F. Nobile, L. Tamellini, and R. Tempone, Convergence of quasi-optimal sparse-grid approximation of hilbert-space-valued functions: application to random elliptic pdes, *Numerische Mathematik* 134 (2016), no. 2, 343–388.
-  A. M. Stuart, Inverse problems: A bayesian perspective, *Acta Numerica* 19 (2010), 451–559.
-  S Walsh, T. Wildey, and J.D. Jakeman, A consistent Bayesian formulation for stochastic inverse problems based on push-forward measures, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 4 (2017), no. 1, 011005–011005–19.