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
**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.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Inner wall shape</th>
<th>Wall thicknesses</th>
<th>Stringer locations</th>
<th>...</th>
<th>Material properties</th>
<th>Inlet conditions</th>
<th>Heat transfer coefficient</th>
</tr>
</thead>
</table>
Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.

Use simulation/approximation to guide experimental design

Requires repeated inference

Fuse data and simulation to reduce uncertainty

Requires many simulations

Identify and exploit structure in data

Needed to reduce UQ cost

Decision Making under Uncertainty

Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk

Data Acquisition

Inference

Function Approximation

Multi-Fidelity Modelling

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

Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.

- Use simulation/approximation to guide experimental design
- Requires repeated inference
- Fuse data and simulation to reduce uncertainty
- Requires many simulations
- Identify and exploit structure in data
- Needed to reduce UQ cost

Decision Making Under Uncertainty
Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk.

Data Acquisition
Use simulation/approximation to guide experimental design.

Inference
Requires repeated inference.

Function Approximation

Multi-fidelity Modelling
Needed to reduce UQ cost.
Must compute statistics from limited number of samples (simulations) Computational cost is amplified as number of uncertainties increases
Define a set of univariate points
\[ Z_{m_{βd}}^{d} = (λ_{d}^{(1)},...,λ_{d}^{(m_{βd})}), \ d = 1,...,k \]

Univariate Lagrange polynomials
\[ φ_{d,j}(λ_{d}) = \prod_{i=1, i ≠ j}^{m_{βd}} \frac{λ_{d} - λ_{d}^{(i)}}{λ_{d}^{(j)} - λ_{d}^{(i)}} \]

Multivariate interpolant is given by
\[ \hat{Q}_{α,β}(λ) = \sum_{j ≤ β} \hat{Q}_{α}(λ^{(j)}) \prod_{d=1}^{k} φ_{i,j_{d}}(λ_{d}). \]

\[ \| Q - \hat{Q}_{l} \|_{L^∞(Λ)} ≤ C_{k,r}N_{l}^{-r/k} \]
\[ \hat{Q}_{\alpha, \mathcal{I}}(\lambda) = \sum_{\beta \in \mathcal{I}} c_{\beta} \hat{Q}_{\alpha, \beta}(\lambda) \]

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

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

**Theorem [BNR00]**

\[ \left\| Q - \hat{Q}_{\mathcal{I}(n)} \right\|_{L^\infty(\Lambda)} \leq C_{k,r} N_n^{-r} (\log N)^{(r+2)(k-1)+1} \]
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 $Z_{N+1}$ can be bounded by

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

where the Lebesgue constant for the grid $Z$ is given by

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

Best polynomial approximation error cannot be reduced. We can however reduce error by minimizing Lebesgue constant.
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 \( \mathcal{Z}_N = \{\lambda^{(1)}, \ldots, \lambda^{(N)}\} \) of nodes ensures an at-worst linear Lebesgue constant growth:

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

Fekete nodes usually provide an excellent set of interpolation nodes.
Fekete points maximise determinant but not nested. Use Leja instead

$$\lambda_{N+1} = \arg\max_{\lambda \in I} |\det V_N(\lambda)|$$

**Idea**

Extend Leja sequences for interpolation in $\pi^i_\Lambda$-weighted probability spaces

Let $W_{m,m} = \sqrt{\pi^i_\Lambda(\lambda^{(m)})}$

$$\lambda_{N+1} = \arg\max_{\lambda \in I} |\det W(\lambda)V_N(\lambda)|$$

**Theorem [NJ14]**

Weighted Leja sequences are asymptotically Fekete.
Sampling Weighted Spaces
Leja Sequences

Fekete points maximise determinant but not nested. Use Leja instead.

\[ \lambda_{N+1} = \underset{\lambda \in \mathcal{I}}{\text{argmax}} |\det V_N(\lambda)| \]

Idea: Extend Leja sequences for interpolation in \( \pi_i \Lambda \)-weighted probability spaces.

\[ W_{m,m} = \sqrt{\pi_i \Lambda(\lambda(m))} \]

Theorem [NJ14]: Weighted Leja sequences are asymptotically Fekete.

Distribution function

\[ F_Z \]

Leja points

Arcsine distribution

Contracted distribution function

\[ F_{Z/\tilde{p}} \]

Leja points

Limit distribution
Most literature focuses on convergence of statistics of sequences of random variables. But convergence in distribution does not imply convergence almost surely.
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

\[
\left\| \pi_{\mathcal{D}}^Q(q) - \pi_{\mathcal{D}}^{Q_n}(q) \right\|_{L^\infty(\mathcal{D})} < \epsilon, \quad \left\| \pi_{\mathcal{D}}^Q(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right\|_{L^\infty(\Lambda)} < \epsilon.
\]
Corollary [BJW18c]

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

$$\left\| \pi_{Q}(Q(\lambda)) - \hat{\pi}_{Q}^{n}(Q_{n}(\lambda)) \right\|_{L^{\infty}(\Lambda)} \leq C \left( \left( \frac{\log M}{M} \right)^{\frac{s}{2s+m}} + C_{k,r} N^{-r} (\log N)^{(r+2)(k-1)+1} \right).$$
### 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]
Multi-fidelity Modelling

Decision Making under Uncertainty
- Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk

Data Acquisition
- Use simulation/approximation to guide experimental design
- Requires repeated inference

Inference
- Fuse data and simulation to reduce uncertainty
- Requires many simulations

Function Approximation

Multi-fidelity Modelling
- Identify and exploit structure in data
- Needed to reduce UQ cost

Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.
Multi-fidelity modelling leverages simulations of lower-fidelity models of reduced cost to increase the tractability of sampling/approximating a high-fidelity model

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

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, sparse grids can be naturally extended to multi-fidelity context [HANTT16, dBR17, JEGG18]
The MC estimate of the mean

\[ \hat{Q} = N^{-1} \sum_{n=1}^{N} Q(\lambda^{(i)}) \]

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

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

\[ \hat{Q}^{CV} = \hat{Q} + \alpha \left( \hat{Q}_1 - \mu_1 \right) \]

Given \( r \) samples of \( Q \) and \( r_1 \) of \( Q_1 \), variance in \( \hat{Q}^{CV} \) is

\[ \mathbb{V}[\hat{Q}^{CV}] = (1 - \frac{r_1 - r}{r_1r} \rho^2) \mathbb{V}[\hat{Q}] \]

where \( \rho \) correlation between \( Q \) and \( Q_1 \). 

For multiple models the CV estimator is

\[
\hat{Q}^{CV} = \hat{Q} + \sum_{i=1}^{M} \alpha \left( \hat{Q}_i - \hat{\mu}_i \right) = \hat{Q} + \alpha \Delta
\]

Theorem [GGEJ18]

The Optimal CV weights are

\[
\alpha^* = \arg\min_{\alpha} \gamma^*(\alpha) = -\text{Cov} \left[ \Delta, \Delta \right]^{-1} \text{Cov} \left[ \Delta, \hat{Q} \right]
\]

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

\[
\mathbb{E} [Q_L] = \mathbb{E} [Q_0] + \sum_{\ell=1}^{L} \mathbb{E} [Q_\ell - Q_{\ell-1}]
\]

Theorem [GGEJ18]

Regardless of number of models \( M \), variance of MLMC satisfies

\[
\text{Var} [\hat{Q}^{\text{MLMC}}] < (1 - \rho_{\text{max}}^2) \text{Var} [\hat{Q}]
\]

\( \rho_{\text{max}} \) is max correlation of \( Q \) with \( Q_i \)
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.
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]
Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.

- **Data Acquisition**: Requires many simulations.
  - Identify and exploit structure in data
  - Needed to reduce UQ cost

- **Inference**: Requires repeated inference.
  - Fuse data and simulation to reduce uncertainty

- **Function Approximation** and **Multi-fidelity Modelling**

- **Decision Making under Uncertainty**: Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk.

Use simulation/approximation to guide experimental design.
**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].
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?
Theorem [BJW18b]

The consistent updated density is

\[
\pi^u_{\Lambda}(\lambda) = \pi^i_{\Lambda}(\lambda) \frac{\pi^\text{obs}_D(Q(\lambda))}{\pi^Q_D(Q(\lambda))}.
\]

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 \( \pi^Q_D(q) \).
Example

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

\[
\begin{align*}
\lambda_1 x_1^2 + x_2^2 &= 1, \\
x_1^2 - \lambda_2 x_2^2 &= 1
\end{align*}
\]

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

Initial

Push-forward of Initial
Example
A Simple Non-linear System

Initial

Push-forward of Initial

Observed
Example
A Simple Non-linear System

Initial Push-forward of Initial Observed Updated
We can use rejection sampling to draw samples from the updated density.

Samples from the updated density

Using rejection sampling for $\pi^u_\Lambda$ using samples from $\pi^i_\Lambda$ is equivalent to sampling $\pi^{\text{obs}}_D$ using samples from $\pi^Q_D$. 
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 $\pi^u_\Lambda$ using samples from $\pi^i_\Lambda$ is equivalent to sampling $\pi^{obs}_D$ using samples from $\pi^Q_D$. 
A Key Assumption

Predictability Assumption

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

In other words, we assume we can predict everything we observe.
**Important Properties**

**Theorem [BJW18b]**

The consistent updated density is

$$
\pi^u_\Lambda(\lambda) = \pi^i_\Lambda(\lambda) \frac{\pi^\text{obs}_D(Q(\lambda))}{\pi^Q_D(Q(\lambda))}.
$$

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

**Theorem [BJW18b]**

Updated density is stable with respect to perturbations in $\pi^\text{obs}_D$ and $\pi^i_\Lambda$. 
What happens when we use a surrogate model to compute the push-forward

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

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

Samples from posterior

$Q_s(\lambda)$

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

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

0.4789 0.8704 0.9787 0.9825
## Inverse Uncertainty Quantification

### Summary

<table>
<thead>
<tr>
<th>Goal</th>
<th>Condition estimates of uncertainty on experimental data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenge</td>
<td>Develop new formulations and reduce sample complexity</td>
</tr>
<tr>
<td>Solution</td>
<td>Find input measure whose push-forward matches observed density</td>
</tr>
<tr>
<td></td>
<td>Reduce inverse problem to one forward solve</td>
</tr>
<tr>
<td>Methods</td>
<td>Bayesian inference</td>
</tr>
<tr>
<td></td>
<td>Push-forward based inference [BJW18a, BJW18c]</td>
</tr>
</tbody>
</table>
**Data Acquisition**

- **Decision Making Under Uncertainty**
  - Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk.

- **Data Acquisition**
  - Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.

- **Inference**
  - Use simulation/approximation to guide experimental design.
  - Requires repeated inference.

- **Function Approximation**
  - Fuse data and simulation to reduce uncertainty.
  - Requires many simulations.

- **Multi-fidelity Modelling**
  - Identify and exploit structure in data.
  - Needed to reduce UQ cost.
Data are not equally informative

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

Use measure of change in uncertainty

$$KL(\pi^i_\Lambda : \pi^u_\Lambda) := \int_{\Lambda} \pi^u_\Lambda \log \left( \frac{\pi^u_\Lambda}{\pi^i_\Lambda} \right) d\mu_\Lambda.$$
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

$$\mathcal{O} = \left\{ \hat{N}(E[\pi^O_D] + \tau \sqrt{\pi^O_D}^{1/2}, \sigma^2) : \tau \in \{-1, 0, 1\} \right\},$$
**Expected Information Gain**

\[
\text{EIG}(Q) := \int_D KL(d) \pi_D^Q(d) d\mu_D.
\]

Given samples from push-forward:

\[
q^{(j)} = Q(\lambda^{(j)}) \text{compute}
\]

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

**OED definition**

Let \(Q^z \in Q\) be a specific design in space of all possible designs, then OED solves

\[
Q^{\text{opt}} := \arg\max_{Q^z \in Q} E(I_{Q^z}).
\]
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+18]
- OED using Push-forward based inference [WWJ17]
Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.

Use simulation/approximation to guide experimental design
Requires repeated inference
Fuse data and simulation to reduce uncertainty
Requires many simulations
Identify and exploit structure in data
Needed to reduce UQ cost

Use estimates of uncertainty to inform design and qualification of engineered systems using rigorous connections between regret and risk

Decision Making Under Uncertainty

Design and Decision Making Under Uncertainty
A Simple Motivating Example
Cantilever Beam

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

<table>
<thead>
<tr>
<th>Uncertainty</th>
<th>Symbol</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield stress</td>
<td>( R )</td>
<td>( N(40000, 2000) )</td>
</tr>
<tr>
<td>Young’s modulus</td>
<td>( E )</td>
<td>( N(2.9e7, 1.45e6) )</td>
</tr>
<tr>
<td>Horizontal load</td>
<td>( X )</td>
<td>( N(500, 100) )</td>
</tr>
<tr>
<td>Vertical Load</td>
<td>( Y )</td>
<td>( N(1000, 100) )</td>
</tr>
</tbody>
</table>
Design under uncertainty

Deterministic Design

\[
\text{argmin}_{w,t} \quad wt \\
\begin{align*}
  f_1(\lambda) &\geq 0 \\
  f_2(\lambda) &\geq 0 \\
  1 \leq w &\leq 4 \\
  1 \leq t &\leq 4
\end{align*}
\]

Design under uncertainty

\[
\text{argmin}_{w,t} \quad wt \\
\begin{align*}
  P(f_1(\lambda) \leq 0) &\leq \delta_1 \\
  P(f_2(\lambda) \leq 0) &\leq \delta_2 \\
  1 \leq w &\leq 4 \\
  1 \leq t &\leq 4
\end{align*}
\]
Design under uncertainty

\[
\text{argmin } wt \\
\text{subject to } w, t \\
P(f_1(\lambda) \leq 0) \leq \delta_1 \\
P(f_2(\lambda) \leq 0) \leq \delta_2 \\
1 \leq w \leq 4 \quad 1 \leq t \leq 4
\]
Putting it all together

Cantilever Beam
Minimise weight subject to thrust and thermal and structural failure constraints
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

Uncertainty quantification (UQ) is not a post-processing step. UQ provides a framework for improving credibility and requires close collaboration with modelling teams through the entirety of product development.
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

$$f_{\Lambda}(\lambda) = \sum_{i \in \Lambda} \alpha_i \phi_i(\lambda)$$

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

$l_0$-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]
Multi-fidelity Modelling

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?
Sparse grids can be used for approximation and solving PDEs [LHH+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.
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

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

Work required to build sparse grid and resulting approximation error are

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

Must choose index set to minimise error and 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

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

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

1

1Top image: Northrop Grumman X-47B UCAS Data-sheet, 2015
Steady fluid analysis: 2D Axissymetric Euler

Thermal analysis: Conduction & convection modeled

Structural analysis: Pressure and temperature-induced forces in load layers, only temperature-induced forces in thermal layer
We can obtain an almost optimal order of magnitude reduction in cost by sampling coarse meshes more often than fine meshes.
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 $\theta$. 

\[ \begin{align*}
\theta_1 & \quad \theta_2 & \quad \theta_3 & \quad \ldots & \quad \theta_{10} \\
\theta_4 & \quad \theta_5 & \quad \theta_6 & \quad \theta_7 & \quad \theta_8 & \quad \theta_9 \\
\theta_1 & \quad \theta_2 & \quad \theta_3 & \quad \theta_5 & \quad \theta_4 & \quad \theta_6 & \quad \theta_8 & \quad \theta_7 & \quad \theta_9
\end{align*} \]

\[ \begin{align*}
\theta_{10} \\
\theta_2 & \quad \theta_3 & \quad \ldots & \quad \theta_{10}
\end{align*} \]
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:

\[ \pi_{\Lambda,\text{post}}(\lambda) \sim \exp \left( - \left( \frac{1}{2} \| C_q^{-1/2} (q - \tilde{q}) \|_2^2 + \frac{1}{2} \| C_x^{-1/2} (\lambda - \overline{\lambda}) \|_2^2 \right) \right) \]

- MAP point is identical to Tikhonov-regularised solution to a deterministic optimisation problem.
Set $C_A = AC\lambda A^T$ and $\bar{q} = A\bar{\lambda}$.

The consistent Bayesian updated density is given by,

$$
\pi^u_A(\lambda) \sim \exp \left( - \left( \frac{1}{2} \|C_q^{-1/2}(q - \bar{q})\|_2^2 + \frac{1}{2} \|C_\lambda^{-1/2}(\lambda - \bar{\lambda})\|_2^2 - \frac{1}{2} \|C_A^{-1/2}(q - \bar{q})\|_2^2 \right) \right)
$$

We can rewrite the regularisation terms as,

$$
\frac{1}{2} \|C_\lambda^{-1/2}(\lambda - \bar{\lambda})\|_2^2 - \frac{1}{2} \|C_A^{-1/2}(q - \bar{q})\|_2^2 = \frac{1}{2}(\lambda - \bar{\lambda})^T \left( C_\lambda^{-1} - A^T (AC\lambda AT)^{-1} A \right)(\lambda - \bar{\lambda})
$$

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:

\[ \text{Data mismatch} + \text{Tikhonov regularisation} = \text{Optimisation function} \]

Modified Tikhonov regularisation:

\[ \text{Data mismatch} + \text{Modified regularisation} = \text{Modified optimisation function} \]
What if we do not want to be correct on average but care about avoiding certain outcomes.

\[ \text{VaR}_\delta(\lambda) = \inf \{ \lambda \in \Lambda \mid F_\lambda(\lambda) \geq \delta \} \]

\[ \text{CVaR}_\alpha(\lambda) = \frac{1}{1 - \alpha} \int_\delta^1 \text{VaR}_\delta(\lambda) \, d\delta \]
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.


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.


A. Gorodetsky, J.D. Jakeman, G. Geraci, and M. Eldred, Multifidelity modeling with Bayesian network learning, In preparation.


