Optimal Uncertainty Quantification

Tim Sullivan
tjs@caltech.edu

California Institute of Technology

APAM Colloquium
Columbia University, New York, U.S.A.
8 December 2011
Overview

1 Introduction
2 Optimal Uncertainty Quantification (OUQ)
3 Optimal Concentration Inequalities and PDEs
4 OUQ Using Legacy Data
5 OUQ for Sesmic Safety Certification
6 Conclusions

Joint work with M. McKerns, M. Ortiz, H. Owhadi (Caltech); C. Scovel (LANL); F. Theil (U. Warwick, UK); and D. Meyer (ex-T.U. München, Germany).

Portions of this work were supported by the U. S. Department of Energy National Nuclear Security Administration under Award Number DE-FC52-08NA28613 through the California Institute of Technology's ASC/PSAAP Center for the Predictive Modeling and Simulation of High Energy Density Dynamic Response of Materials.
Introduction

What is Uncertainty Quantification (UQ)?

Motivation for Optimal UQ:
Some Typical UQ Objectives and Complications
What is Uncertainty Quantification?

In rough terms, **Uncertainty Quantification** (UQ) means

- reasoning under uncertainty about physically-motivated problems
- rigorously quantifying the uncertainties involved
- using mathematical, probabilistic and computational tools.

The conventional wisdom about uncertainties is that

- **aleatoric uncertainties** — which stem from the operation of random chance and can be treated using the methods of probability theory — are nice, and

- **epistemic uncertainties** — which stem from lack of knowledge — are nasty.
Typical UQ Objectives / Problems

What do the following problems have in common?

**Seismic Safety**

- Will a given structure collapse under a given earthquake ground motion?
- What is the probability of collapse under earthquakes that are randomly distributed according to some known probability distribution?
- What if that probability distribution is only partially known? What if it is known, not up to a few real parameters, but only up to an infinite-dimensional family?
Typical UQ Objectives / Problems

What do the following problems have in common?

**Random PDEs — Pressure and Transport in Porous Media**

Consider the following PDE for a pressure field $u$ on $U \subseteq \mathbb{R}^n$ in a medium with porosity described by $\kappa$:

$$-\nabla \cdot (\kappa(x)\nabla u(x)) = f(x), + \text{ boundary conditions.}$$

For a given point $x_0 \in U$ and threshold pressure $u_0 \in \mathbb{R}$,

- Is it true that $u(x_0) \geq u_0$?
- What is $\mathbb{P}[u(x_0) \geq u_0]$ if the probability distribution $\mathbb{P}$ associated to random $\kappa$, $f$ and boundary conditions is known?
- What if $\mathbb{P}$ is only partially known? Again, what if the space of possibilities for $\mathbb{P}$ is infinite-dimensional?
- How do the answers depend upon the features of $\kappa$ across various scales?
Typical UQ Objectives / Problems

What do the following problems have in common?

**Partially-Specified Quantum System**

- Given the (classical) state of a system at time $t = 0$, describe the state at time $t = T$.

- Given the initial state as a wave-function $\psi_0$, describe the state at time $t = T$, i.e. solve the Schrödinger equation

  \[ i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V\psi \]

  on the interval $[0, T]$, with initial condition $\psi(0) = \psi_0$.

- What if $\psi_0$ is incompletely specified? What if $V$ and $m$ are also unknown?
Typical UQ Objectives / Problems

What do the following problems have in common?

**Experimental Design**

Given a choice of one of a number of very expensive experiments to run to gain information about some quantity of interest, which one should you choose if

- the possible outcomes of the candidate experiments are believed to be random with known distribution?
- the possible outcomes’ distributions are unknown, or partially known?

**Other Problems...**

- Prediction and Extrapolation
- Verification and Validation
- ...
Why Optimal UQ?

- Such problems are relatively simple to address if the probability distributions, response functions, & c. are perfectly known, or if the uncertainties are finite-dimensional parametric uncertainties.
- Methods for dealing with them usually depend upon the validity of specific assumptions for their applicability or efficiency. E.g.
  - \{Quasi-, Markov Chain\} Monte Carlo. Need to know the distribution and be able to draw many samples from it.
  - Stochastic Collocation Methods. Need to pick a distribution for the expansion, and require that the randomness and response function have good spectral properties w.r.t. that basis.
- However, in reality, these objects are usually unknown, or incompletely known, and the uncertainties are infinite-dimensional in nature.

The Fear

Even with nice assumptions, probabilistic calculations are harder and more involved than deterministic ones, so infinite-dimensional families of probabilistic problems sound like they would be nearly impossible.
The Idea of Optimal Uncertainty Quantification

If In Doubt, Optimize!

- To obtain robust bounds on output uncertainties given parametric input uncertainties, just optimize w.r.t. those uncertain parameters.
- The **OUQ framework** is the extension of this idea to the infinite-dimensional regime of incompletely specified probability distributions and response functions.
- And, surprisingly, the answers are simpler than you might expect.

Figure: Optimizing $G(x)$ over $x \in X$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only partially known? (i.e. non-parametric epistemic uncertainty.)
The Idea of Optimal Uncertainty Quantification

**If In Doubt, Optimize!**

- To obtain robust bounds on output uncertainties given parametric input uncertainties, just optimize w.r.t. those uncertain parameters.
- The OUQ framework is the extension of this idea to the infinite-dimensional regime of incompletely specified probability distributions and response functions.
- And, surprisingly, the answers are simpler than you might expect.

Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only partially known? (i.e. non-parametric epistemic uncertainty.)
Optimal Uncertainty Quantification

The Problem: Optimal Bounds

OUQ: Formulation, Reduction and Implementation
Problem Setting

The Challenge in General Terms

- Give **optimal bounds** on some quantity of interest $\mathbb{E}_{X \sim \mathbb{P}}[q(X, G(X))]$, which depends on some response function $G : \mathcal{X} \rightarrow \mathcal{Y}$ with $\mathbb{P}$-distributed inputs $X$ in $\mathcal{X}$, given only **incomplete information** about the pair $(G, \mathbb{P})$.
- Archetypical example: to bound $\mathbb{P}[G(X) \leq 0]$, where the event $[G(X) \leq 0]$ corresponds to failure of some kind.

Why Optimality?

- We seek bounds that are both rigorous and optimal in order to be most informative in a decision-making context.
- The bound

$$0 \leq \mathbb{P}[G(X) \leq 0] \leq 1$$

is rigorous, but usually not optimal, and hardly informative!
Formulation of OUQ Problems

- We want to know about the quantity of interest
  \[ \mathbb{E}_{X \sim P} [q(X, G(X))] \]
  when the reality \((G, P)\) is only imperfectly known.
- The key step in the Optimal Uncertainty Quantification approach is to specify a feasible set of admissible scenarios \((g, \mu)\) that could be \((G, P)\) according to the available information:
  \[
  \mathcal{A} := \left\{ (g, \mu) : \begin{array}{l}
  (g : \mathcal{X} \to \mathcal{Y}, \mu \in \mathcal{P}(\mathcal{X})) \text{ is consistent with all given information about the real system } (G, P) \\
  \text{(e.g. legacy data, first principles, expert judgement)}
  \end{array} \right\}.
  \]
- \(\mathcal{A}\) encodes everything that we know about the “reality” \((G, P)\).
- \textit{A priori}, all we know about reality is that \((G, P) \in \mathcal{A}\); we have no idea exactly which \((g, \mu)\) in \(\mathcal{A}\) is actually \((G, P)\). No \((g, \mu) \in \mathcal{A}\) is “more likely” or “less likely” to be \((G, P)\) than any other.
Formulation of OUQ Problems

\[ \mathcal{A} := \left\{ (g, \mu) \mid (g : \mathcal{X} \to \mathbb{R}, \mu \in \mathcal{P}(\mathcal{X})) \text{ is consistent with all given information about the real system } (G, \mathcal{P}) \text{ (e.g. legacy data, first principles, expert judgement)} \right\}. \]

Optimal bounds on the quantity of interest \( \mathbb{E}_{X \sim \mathcal{P}}[q(X, G(X))] \) (optimal w.r.t. the information encoded in \( \mathcal{A} \)) are found by minimizing/maximizing \( \mathbb{E}_{X \sim \mu}[q(X, g(X))] \) over all admissible scenarios \( (g, \mu) \in \mathcal{A} \):

\[ \mathcal{L}(\mathcal{A}) \leq \mathbb{E}_{X \sim \mathcal{P}}[q(X, G(X))] \leq \mathcal{U}(\mathcal{A}), \]

where \( \mathcal{L}(\mathcal{A}) \) and \( \mathcal{U}(\mathcal{A}) \) are defined by the minimization and maximization problems

\[ \mathcal{L}(\mathcal{A}) := \inf_{(g, \mu) \in \mathcal{A}} \mathbb{E}_{X \sim \mu}[q(X, g(X))], \]

\[ \mathcal{U}(\mathcal{A}) := \sup_{(g, \mu) \in \mathcal{A}} \mathbb{E}_{X \sim \mu}[q(X, g(X))]. \]
OUQ in Context

- When the quantity of interest is the probability of some event $E$, $\mathcal{L}(\mathcal{A})$ and $\mathcal{U}(\mathcal{A})$ are the optimal lower and upper probabilities of $E$ w.r.t. the information encoded in $\mathcal{A}$.
- Notions of imprecise probability have a long history stretching back to Boole (1854) and Keynes (1921), with more recent and comprehensive foundations laid out by Kuznetsov (1991), Walley (1991), and Weichselberger (2000).
- In the Bayesian world, such approaches are sometimes known as robust Bayesian inference.
- The idea is an old one, but computability has always been the major hurdle: lots of effort has been spent on representation theorems for various classes $\mathcal{A}$. 
Dimensional Reduction

- A priori, OUQ problems are infinite-dimensional, non-convex, highly-constrained, global optimization problems.
- However, they can be reduced to equivalent finite-dimensional problems in which the optimization is over the extremal scenarios of $\mathcal{A}$.
- The dimension of the reduced problem is proportional to the number of probabilistic inequalities that describe $\mathcal{A}$.

Figure: Just as a linear program finds its extreme value at the extremal points of a convex domain in $\mathbb{R}^n$, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.
Reduction of OUQ Problems — Theorem

**Theorem (Reduction for moment and independence constraints)**

Suppose that $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$ is a product of Radon spaces. Let

$$A := \left\{ (g, \mu) \mid \begin{array}{l}
g: \mathcal{X} \to \mathbb{R} \text{ is measurable, } \mu = \mu_1 \otimes \cdots \otimes \mu_K \in \bigotimes_{k=1}^K \mathcal{P}(\mathcal{X}_k); \\
\langle \text{any conditions on } g \text{ alone}\rangle; \text{ and, for each } g, \\
\text{for some measurable functions } \varphi_i: \mathcal{X} \to \mathbb{R} \text{ and } \varphi_i^{(k)}: \mathcal{X}_k \to \mathbb{R}, \\
\mathbb{E}_{X \sim \mu} [\varphi_i(X)] \leq 0 \text{ for } i = 1, \ldots, n_0, \\
\mathbb{E}_{X_k \sim \mu_k} [\varphi_i^{(k)}(X_k)] \leq 0 \text{ for } i = 1, \ldots, n_k \text{ and } k = 1, \ldots, K
\end{array} \right\} \subseteq A.$$

Let

$$A_{\Delta} := \left\{ (g, \mu) \in A \mid \mu_k \text{ is a convex combination of at most } N_k := 1 + n_0 + n_k \text{ Dirac measures on } \mathcal{X}_k \right\} \subseteq A.$$

Then

$$\dim(A_{\Delta}) \leq \sum_{k=1}^K N_k (1 + \dim(\mathcal{X}_k)) + \prod_{k=1}^K N_k - K,$$

and

$$\mathcal{L}(A) = \mathcal{L}(A_{\Delta}) \text{ and } \mathcal{U}(A) = \mathcal{U}(A_{\Delta}).$$
Reduction of OUQ Problems — Sketch Proof

**Proof.**

- First consider $K = 1$, and fix $g : \mathcal{X} \to \mathbb{R}$.
- Since $\mathcal{X}$ is a Radon space (i.e. “nice”), all probability measures on $\mathcal{X}$ are inner regular, and so the set $\text{ex}(A_\Phi)$ of extreme points of
  
  $$A_\Phi := \{ \mu \in \mathcal{P}(\mathcal{X}) \mid \mathbb{E}_{X \sim \mu}[\varphi_1(X)] \leq 0, \ldots, \mathbb{E}_{X \sim \mu}[\varphi_n(X)] \}$$

  consists of the convex combinations of at most $1 + n$ Dirac masses.
- The map $\mu \mapsto \mathbb{E}_{X \sim \mu}[q(X, g(X))]$ is measure affine (i.e. “nice”), therefore its extreme values over $A_\Phi$ and $\text{ex}(A_\Phi)$ are the same.
- Now vary $g$ — still the same number of Dirac masses regardless of $g$.
- For $K > 1$, apply the previous argument componentwise using Fubini’s theorem.
Reduction of OUQ Problems — Interpretation

The reduction theorem tells us two very important things. It says that, from the perspective of bounding a chosen quantity of interest,

- reasonably general infinite-dimensional feasible sets $\mathcal{A}$ are equivalent to finite-dimensional subsets $\mathcal{A}_\Delta$ — and so we can numerically optimize over that finite-dimensional set; and
- the probability measures in $\mathcal{A}_\Delta$ are very simple (products of finite convex combinations of Dirac point masses), so integration against a measure $\mu$ in $\mathcal{A}_\Delta$ is easy — no need to worry about e.g. MCMC integration against a “general” measure.

Depending on the specific structure of $\mathcal{A}$, there are often additional layers of reduction theorems. E.g. in the McDiarmid example later on, a theorem enables us to “forget” the coordinates in the input spaces.
Numerical Solution of Reduced OUQ Problems

- The finite-dimensional problems $\mathcal{L}(A_{\Delta})$ and $\mathcal{U}(A_{\Delta})$ can be solved numerically.
  - Easily swappable strategies for optimization, population generation, enforcement of constraints, termination criteria.
  - Most of the examples that follow were done using Differential Evolution, which mixes local gradient-based methods with global genetic algorithms.
  - Manages optimizations on scales ranging from the small (seconds-long on a laptop) to the large (days on dozens-of-cores clusters).
Examples I

Optimal Concentration Inequalities: Parameter (In)Sensitivity

OUQ and Random/Multiscale PDEs
Classical Example: Markov’s Inequality

**Theorem (Markov’s Inequality)**

For any non-negative random variable $X$ with given mean $\mathbb{E}[X] = m \geq 0$, for any $t \geq m$,

$$\mathbb{P}[X \geq t] \leq \frac{m}{t}.$$

- Or, in OUQ terms,

  $$A_{\text{Mrkv}} := \{\mu \in \mathcal{P}([0, +\infty)) \mid \mathbb{E}_{X \sim \mu}[X] = m\},$$

  $$\mathcal{U}(A_{\text{Mrkv}}) := \sup_{\mu \in A} \mu[X \geq t] \leq \frac{m}{t}.$$

- In fact, $\mathcal{U}(A_{\text{Mrkv}}) = \frac{m}{t}$, and the probability distribution $\mu$ that attains this extreme value is

  $$\mu = \left(1 - \frac{m}{t}\right)\delta_0 + \frac{m}{t}\delta_t.$$
McDiarmid’s Inequality

Consider the admissible set corresponding to the assumptions of McDiarmid’s inequality (a.k.a. the bounded differences inequality):

$$\mathcal{A}_{\text{McD}} = \left\{ (g, \mu) \mid \begin{array}{l}
g: \mathcal{X}_1 \times \cdots \times \mathcal{X}_K \to \mathbb{R}, \\
\mu = \bigotimes_{k=1}^K \mu_k, \text{ (i.e. } X_1, \ldots, X_K \text{ independent)} \\
\mathbb{E}_{X \sim \mu} [g(X)] \geq m \geq 0, \\
\text{osc}_k(g) \leq D_k \text{ for each } k \in \{1, \ldots, K\} \\
\end{array} \right\},$$

with componentwise oscillations/global sensitivities defined by

$$\text{osc}_k(g) := \sup \left\{ |g(x) - g(x')| \mid x, x' \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_K, \\
x_i = x'_i \text{ for } i \neq k \right\}.$$

**Theorem (McDiarmid’s Inequality, 1988)**

$$\mathcal{U}(\mathcal{A}_{\text{McD}}) := \sup_{(g, \mu) \in \mathcal{A}_{\text{McD}}} \mu[g(X) \leq 0] \leq \exp \left( -\frac{2m^2}{\sum_{k=1}^K D_k^2} \right)$$
Optimal McDiarmid — Non-Propagation

**Theorem**

For $K = 1$,

$$
U(A_{McD}) = \begin{cases} 
0, & \text{if } D_1 \leq m, \\
1 - \frac{m}{D_1}, & \text{if } 0 \leq m \leq D_1.
\end{cases}
$$

For $K = 2$,

$$
U(A_{McD}) = \begin{cases} 
0, & \text{if } D_1 + D_2 \leq m, \\
\frac{(D_1 + D_2 - m)^2}{4D_1D_2}, & \text{if } |D_1 - D_2| \leq m \leq D_1 + D_2, \\
1 - \frac{m}{\max\{D_1, D_2\}}, & \text{if } 0 \leq m \leq |D_1 - D_2|.
\end{cases}
$$

There are similar explicit formulae for $K = 3$ (involving roots of cubic polynomials) and higher $K$. 
Optimal McDiarmid — Non-Propagation

Theorem

For $K = 2$,

$$U(A_{McD}) = 1 - \frac{m}{\max\{D_1, D_2\}}, \quad \text{if } 0 \leq m \leq |D_1 - D_2|.$$

- If the “sensitivity gap” $|D_1 - D_2|$ is large enough relative to the performance margin $m$, then $\max\{D_1, D_2\}$ dominates all the uncertainty about $\mathbb{P}[G(X) \leq 0]$.
- The smaller of $D_1$ and $D_2$ could be reduced to zero without improving the worst-case bound on the probability of failure.
- In the presence of uncertainty about input probability distributions and input-output relationship, there can be screening effects and sensitivities can fail to propagate.
Optimal Hoeffding and the Effects of Nonlinearity

Similarly, one can consider the admissible set $A_{Hfd}$ that corresponds to the assumptions of Hoeffding’s inequality, which bounds deviation probabilities of sums of independent bounded random variables:

$$A_{Hfd} := \left\{ (g, \mu) \mid g : \mathbb{R}^K \to \mathbb{R} \text{ given by } g(x_1, \ldots, x_K) := x_1 + \cdots + x_K, \mu = \mu_1 \otimes \cdots \otimes \mu_K \text{ supported on a cube of side lengths } D_1, \ldots, D_K, \text{ and } \mathbb{E}_X \sim \mu [g(X)] \geq m \geq 0 \right\}.$$

Hoeffding’s inequality is the bound

$$\mathcal{U}(A_{Hfd}) \leq \exp \left( -\frac{2m^2}{\sum_{k=1}^{K} D_k^2} \right).$$

Interestingly, $\mathcal{U}(A_{McD}) = \mathcal{U}(A_{Hfd})$ for $K = 1$ and $K = 2$, but $\mathcal{U}(A_{McD}) \geq \mathcal{U}(A_{Hfd})$ for $K = 3$, and the inequality can be strict.
Example: Random PDEs

- Consider the following PDE for a pressure field $u$ on $U \subseteq \mathbb{R}^n$ in a medium with porosity field $\kappa$:

$$-\nabla \cdot (\kappa(x) \nabla u(x)) = f(x),$$

with appropriate boundary conditions.

- When the probability distribution $P$ of $\kappa$ and $f$ is known, such a stochastic PDE is a benchmark application for stochastic expansion methods.

- We seek the least upper bound on the probability that the log-pressure at $x_0 \in U$ exceeds its mean by more than $a$:

$$P[\log u(x_0) \geq \mathbb{E}[\log u(x_0)] + a].$$

- The OUQ-McDiarmid example can be applied in two ways here: the relative effects of $\kappa$ and $f$; and the relative effects of micro and macro features of $\kappa$. 
Example: Random/Multiscale PDEs

### Setting I: Independent Porosity and Source Terms

Given $D_1, D_2 \geq 0$, and fields $K, F \in L^\infty(U)$ with

$$\text{ess inf}_U K > 0, \quad F \geq 0, \quad \int_U F(x) \, dx > 0,$$

let

$$A := \left\{ \mu \mid \text{under } \mu, \text{ the fields } \kappa \text{ and } f \text{ are independent and, } \mu\text{-a.s.} \right. \quad \left. \begin{align*} K(x) &\leq \kappa(x) \leq e^{D_1} K(x), \\ F(x) &\leq f(x) \leq e^{D_2} F(x) \end{align*} \right\}.$$

### Theorem

$\mathcal{U}(A) = \mathcal{U}(A_{\text{McD}})$. *In particular, if $|D_1 - D_2| \geq a$, then the worst-case bound on* $P[\log u(x_0) \geq \mathbb{E}[\log u(x_0)] + a]$ *is independent of* $\min\{D_1, D_2\}$. 
Example: Random/Multiscale PDEs

Setting II: Independent Porosity Micro- and Macrostructure

Given \( D_1, D_2 \geq 0 \), and fields \( K_1, K_2 : U \to \mathbb{R} \) such that \( K_1 \) is smooth and uniformly elliptic in \( U \), and \( K_2 \in L^\infty(U) \) is uniformly elliptic in \( U \) with spatial period \( \delta \ll 1 \), let

\[
A := \begin{cases} 
\mu & \text{under } \mu, \text{ the fields } \kappa_1 \text{ and } \kappa_2 \text{ are independent and, } \mu\text{-a.s.} \\
\kappa = \kappa_1 \kappa_2, \\
\|\nabla \kappa_1\|_{L^\infty} \leq e^{D_1} \|\nabla K_1\|_{L^\infty}, \\
K_1(x) \leq \kappa_1(x) \leq e^{D_1} K_1(x), \\
\kappa_2 \text{ is spatially periodic with period } \delta, \\
K_2(x) \leq \kappa_2(x) \leq e^{D_2} K_2(x) 
\end{cases}
\]

Theorem

\( \mathcal{U}(A) = \mathcal{U}(A_{McD}) \). In particular, if \( |D_1 - D_2| \geq a \), then the worst-case bound on \( \mathbb{P}[\log u(x_0) \geq \mathbb{E}[\log u(x_0)] + a] \) is independent of \( \min\{D_1, D_2\} \).
(Non-)Propagation of Information Across Scales

One can consider hierarchies (directed acyclic graphs) of OUQ modules, representing e.g. a multiscale description of a complex system.

Figure: Because OUQ is a sharp information propagation scheme, the results of sensitivity analysis ("inverse OUQ") give non-trivial insights into the roles of the various pieces of input information. Some inputs may even be irrelevant!
Examples II

OUQ Using Legacy Data

Redundant and Non-Binding Data
The Legacy UQ (Certification) Challenge

Another illustrative and accessible example of OUQ in action is furnished by the problem of **UQ with legacy data**.

**General Challenge**
To determine if a system of interest will “fail” only with acceptably small probability, given observations of the system response on some subset $\mathcal{O}$ of the parameter space $\mathcal{X}$ and nowhere else.

**Illustrative Example**
To bound $\mathbb{P}[G(X) \leq 0]$, where $G: [0, 1] \rightarrow \mathbb{R}$ is a function known only on some subset $\mathcal{O} \subseteq [0, 1]$, and the probability distribution $\mathbb{P}$ of $X$ on $[0, 1]$ is also only partially known.
The Effect of Information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of $G$ on $\mathcal{O} \subseteq [0, 1]$?

Sharpest Possible Answer...

With so little information, the only rigorous bounds that can be given are the trivial ones: $0 \leq \mathbb{P}[G(X) \leq 0] \leq 1$. 
The Effect of Information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of $G$ on $\mathcal{O} \subseteq [0, 1]$, and that $|G(x) - G(x')| \leq L|x - x'|$?

Sharpest Possible Answer...

...we might discover that $\mathbb{P}[G(X) \leq 0] = 0$ or $= 1$, but otherwise no improvement on the trivial bound $0 \leq \mathbb{P}[G(X) \leq 0] \leq 1$. 
The Effect of Information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of $G$ on $\mathcal{O} \subseteq [0, 1]$, that $|G(x) - G(x')| \leq L|x - x'|$, and that $\mathbb{E}[G(X)] \geq m$?

Sharpest Possible Answer... is non-trivial, and can be found using the optimization techniques of the OUQ framework.
The Effect of Information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of $G$ on $\mathcal{O} \subseteq [0, 1]$, that $|G(x) - G(x')| \leq L|x - x'|$, and that $\mathbb{E}[G(X)] \geq m$?

![Diagram showing a possible $G$ and $\mathbb{P}$, with success and failure axes.]  

Sharpest Possible Answer... 

...is non-trivial, and can be found using the optimization techniques of the OUQ framework.
Problem Formulation

What is the admissible set $\mathcal{A}$ in this case?

$$\mathcal{A} := \left\{ (g, \mu) \mid \begin{array}{c} \mu \text{ a probability measure on } [0, 1], \\ g : [0, 1] \to \mathbb{R} \text{ is } L\text{-Lipschitz}, \\ g = G \text{ on } \mathcal{O}, \text{ and } \mathbb{E}_\mu[g(X)] \geq m \end{array} \right\}.$$  

In other words, any $(g, \mu)$ for which $g$ is $L$-Lipschitz, agrees with the legacy data, and has the right mean under $\mu$ could be $(G, \mathbb{P})$. The reduced admissible set, over which the quantity of interest has the same extreme values, is

$$\mathcal{A}_\Delta := \left\{ (g, \mu) \mid \begin{array}{c} \mu \text{ a probability measure on } [0, 1], \\ \mu = p\delta_{x_0} + (1-p)\delta_{x_1} \text{ for some } p, x_0, x_1 \in [0, 1], \\ g : \mathcal{O} \cup \{x_0, x_1\} \to \mathbb{R} \text{ is } L\text{-Lipschitz}, \\ g = G \text{ on } \mathcal{O}, \text{ and } \mathbb{E}_\mu[g(X)] \geq m \end{array} \right\}.$$
The Reduced Problem

The original problem entails optimizing over an infinite-dimensional collection of \((g, \mu)\) that could be \((G, \mathbb{P})\). In the reduced problem, we only have to move around and re-weight two Dirac measures (point masses) and the values of \(g\) over those two points.

\[
\text{infinite-dimensional problem} \rightsquigarrow \text{equivalent 5-dimensional problem!}
\]
The Reduced Problem

The original problem entails optimizing over an infinite-dimensional collection of \((g, \mu)\) that could be \((G, \mathbb{P})\). In the reduced problem, we only have to move around and re-weight two Dirac measures (point masses) and the values of \(g\) over those two points.

\[
\text{infinite-dimensional problem} \sim \text{equivalent 5-dimensional problem!}
\]
The Reduced Problem

The original problem entails optimizing over an infinite-dimensional collection of \((g, \mu)\) that could be \((G, \mathbb{P})\). In the reduced problem, we only have to move around and re-weight two Dirac measures (point masses) and the values of \(g\) over those two points.

\[
\text{infinite-dimensional problem} \quad \leadsto \quad \text{equivalent 5-dimensional problem!}
\]

\[
\begin{align*}
(x_0, y_0) & \quad \text{mass } p \quad \text{at } x_0 \\
(x_1, y_1) & \quad \text{mass } 1 - p \quad \text{at } x_1
\end{align*}
\]

\[
\begin{align*}
(g, \mu) & \in \mathcal{A} \\
x_0 \\
x_1 \\
y_0 & = g(x_0) \\
y_1 & = g(x_1) \\
p & = \mu(\{x_0\})
\end{align*}
\]
One Data Point

- The case of a single observation can be solved explicitly.
- Suppose that you observe one input-output pair of a function $G: [0, 1] \to \mathbb{R}$ with Lipschitz constant $L$.
- You know $(z, G(z))$ — assume that $z \in [0, \frac{1}{2}]$ and $G(z) > 0$.
- Four cases for the least upper bound on the probability of failure given $L$, $(z, G(z))$, and that $\mathbb{E}[G(X)] \geq m$:

$$U(A) = \begin{cases} 
(1 - \frac{m}{L-(Lz-G(z))})_+ & , \text{if } G(z) \leq Lz, \\
(1 - \frac{m}{L-(Lz+G(z))})_+ & , \text{if } Lz < G(z) \leq L\frac{1}{2} - z, \\
(1 - \frac{2m}{L+(G(z)-Lz})_+ & , \text{if } L\frac{1}{2} - z < G(z) \leq L\frac{1}{2} - 3z, \\
(1 - \frac{m}{Lz+G(z)})_+ & , \text{if } G(z) > L \max\{z, 1 - 3z\}.
\end{cases}$$
Critical Data

(a) “Subcritical” data point: probability of failure is high.

(b) “Supercritical” data point: probability of failure is lower.

Figure: Construction of the least upper bound on $\mathbb{P}[G(X) \leq 0]$ given one observation in two of the four cases. In each case shown, the probability of failure is the probability mass at $x_0$, which is given by $\left(1 - \frac{m_+}{y_1}\right)_+$. 
Critical Data

(a) “Subcritical” data point: probability of failure is high.

(b) “Supercritical” data point: probability of failure is lower.

Figure: Construction of the least upper bound on $\mathbb{P}[G(X) \leq 0]$ given one observation in two of the four cases. In each case shown, the probability of failure is the probability mass at $x_0$, which is given by $\left(1 - \frac{m_+}{y_1}\right)_+$. 
Critical Data

(a) “Subcritical” data point: probability of failure is high.

(b) “Supercritical” data point: probability of failure is lower.

Figure: Construction of the least upper bound on $\mathbb{P}[G(X) \leq 0]$ given one observation in two of the four cases. In each case shown, the probability of failure is the probability mass at $x_0$, which is given by $\left(1 - \frac{m+}{y_1}\right)_+$. 
Critical Data

(a) “Subcritical” data point: probability of failure is high.
(b) “Supercritical” data point: probability of failure is lower.

Figure: Construction of the least upper bound on $\mathbb{P}[G(X) \leq 0]$ given one observation in two of the four cases. In each case shown, the probability of failure is the probability mass at $x_0$, which is given by $\left(1 - \frac{m^+}{y_1}\right)_+$. 
The intuition that “an observation \((z, G(z))\) with \(G(z)\) large \(\implies\) failure is less likely” is more-or-less valid, but in a rather interesting way:

**Figure:** Schematic contour plot and to-scale surface plot of the least upper bound on the probability of failure, as a function of the observed data point \((z, G(z))\). There are jump discontinuities across the orange lines.
Medium-Dimensional Example

- Legacy data = 32 data points (steel-on-aluminium shots A48–A81, less two mis-fires) from summer 2010 at Caltech’s SPHIR facility:

\[ X = (h, \alpha, v) \in \mathcal{X} := [0.062, 0.125] \text{ in} \times [0, 30] \text{ deg} \times [2300, 3200] \text{ m/s}. \]

Output \( G(h, \alpha, v) = \) the induced perforation area in mm\(^2\); the data set contains results between 6.31 mm\(^2\) and 15.36 mm\(^2\).

- Failure event is \( [G(h, \alpha, v) \leq \theta] \), for various values of \( \theta \).
- Constrain the mean perf. area: \( \mathbb{E}[G(h, \alpha, v)] \geq m := 11.0 \text{ mm}^2 \).
- Modified Lipschitz constraint (multi-valued data):

\[
L = \left( \frac{175.0}{\text{in}}, \frac{0.075}{\text{deg}}, \frac{0.1}{\text{m/s}} \right) \text{ mm}^2
\]

\[
|y - y'| \leq \sum_{k=1}^{3} L_k |x_k - x'_k| + 1.0.
\]
Numerical Results

Figure: Maximum probability that perforation area is \( \leq \theta \), for various \( \theta \), with the data and assumptions of the previous slide, including mean perforation area \( \mathbb{E}[G(h, \alpha, v)] \geq m := 11.0 \text{ mm}^2 \). Note close agreement of the results with Markov’s bound.
Dimensional Collapse

- In practice, we do not run the reduced problem (the search over $\mathcal{A}_\Delta$) at full dimensionality.
- *E.g.*, in the previous example, relatively speaking searches over $2 \times 2 \times 2$ product measures are slow and somewhat fragile,

$$\begin{align*}
\{ & 2 \times 1 \times 1 \\
& 1 \times 2 \times 1 \\
& 1 \times 1 \times 2 
\} 
\end{align*}$$

searches over

$$\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_{222}) \leq \mathcal{L}(\mathcal{A}_{112}) \leq \mathcal{U}(\mathcal{A}_{112}) \leq \mathcal{U}(\mathcal{A}_{222}) = \mathcal{U}(\mathcal{A}).$$

- One often sees the higher-dimensional measure “collapsing” as the optimization calculation progresses, and this gives hints as to
  - which lower-dimensional problems to try;
  - the “key uncertainties” in the problem.
Dimensional Collapse

**Figure:** Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure.
Dimensional Collapse

Iteration 150

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure.
Dimensional Collapse

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure.
Dimensional Collapse

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure.
Redundant and Non-Binding Data

- Now consider a set of observations $\mathcal{O} = \{z_1, \ldots, z_N\}$, $N$ large.
- Which data points $(z_n, G(z_n))$ contribute non-trivial constraints, and actually determine the set of feasible $(x_0, x_1, y, p)$? (i.e. which data points are relevant as opposed to being redundant?)
- More importantly, which data points determine the extreme values of the probability of failure? (i.e. which data points are binding as opposed to being non-binding?)
- Not all data points are created equal: we don’t want to solve an optimization problem with $N = 10^6$ constraints if only 42 of them actually matter.
Examples of Redundant and Non-Binding Data

Consider the previous one-dimensional example, but now with two observations at $z_1, z_2 \in [0, 1]$:

Figure: The extremizer for the problem with data point $(z_1, G(z_1))$ is feasible with respect to the new data point $(z_2, G(z_2))$, so the two problems have the same extreme value. The new data point is a relevant but non-binding data point.
Algorithm for Handling Large Data Sets with Redundancies

Theorem (Sufficient Condition to be Non-Binding)

Suppose that \((g, \mu) \in A_\Delta\) is an extremizer for the legacy OUQ problem with data set \(O\), and let \(z \in X \setminus O\). If \((g, \mu)\) is feasible with respect to \((z, G(z))\), then the new observation is non-binding. That is, if

\[
|g(x) - G(z)| \leq d_L(x, z) \text{ for each } x \in \text{supp}(\mu),
\]

then the extreme values for the problems with data sets \(O\) and \(O \cup \{z\}\) are the same, and given by \((g, \mu)\).

**N.B.** The feasibility check (*) is a simple algebraic check; it does not require any (potentially slow or expensive) optimizations.
OUQ Using Legacy Data  Redundant and Non-Binding Constraints

Algorithm for Handling Large Data Sets with Redundancies

Work with two subsets of the full set of data points, $\mathcal{O}$:
- $\mathcal{O}_i =$ the data points that are enforced at iteration $i$;
- $\tilde{\mathcal{O}}_i =$ that data points that are not enforced at iteration $i$, but are potentially binding.

**Sketch Algorithm**

1. Initialize with $\mathcal{O}_0 = \emptyset$ and $\tilde{\mathcal{O}}_0 = \mathcal{O}$.
2. Then, for $i = 1, 2, \ldots$
   1. For each $z \in \tilde{\mathcal{O}}_{i-1}$, find the extreme values of $\mathbb{E}_\mu[q_g]$ with respect to the data set $\mathcal{O}_{i-1} \cup \{z\}$; let $z_*$ denote a/the $z \in \tilde{\mathcal{O}}_{i-1}$ with most extreme extreme value of $\mathbb{E}_\mu[q_g]$.
   2. Let $\mathcal{O}_i := \mathcal{O}_{i-1} \cup \{z_*\}$.
   3. Let $\tilde{\mathcal{O}}_i$ consist of those $z \in \mathcal{O} \setminus \mathcal{O}_i$ such that the extremizer for $\mathcal{O}_i$ is infeasible with respect to $z$ (and hence $z$ is possibly binding).
   4. Terminate if $\tilde{\mathcal{O}}_i = \emptyset$. 
Bounds Using (Validated) Models

- Suppose that the real response function $G: \mathcal{X} \rightarrow \mathbb{R}$ has been modelled by $F: \mathcal{X} \rightarrow \mathbb{R}$, which can be exercised at will.
- We need information/assumptions relating $F$ to $G$, e.g.

$$\|G - F\|_\infty := \sup_{x \in \mathcal{X}} |G(x) - F(x)| \leq C_V.$$  

- Under such an assumption, admissible scenarios $(g, \mu) \in A$ must satisfy $\|g - F\|_\infty \leq C_V$.

(a) data alone  

(b) data and model  

(c) model alone
Better Validation Metrics

**Figure**: The uniform neighbourhood (dark grey) of the function $F$ is relatively small where $F$ has a cliff or discontinuity, whereas the Hausdorff graphical neighbourhood (light grey) is relatively large. More precisely, uniformly (resp. Hausdorff) close functions have *approximately* the same-size cliffs/discontinuities in $\mathbb{R}$ at *exactly* (resp. *approximately*) the same places in $\mathcal{X}$. 
Examples III

OUQ for Sesmic Safety Certification

Knowledge Acquisition and Experimental Design
Large-Scale Example: Seismic Safety

- Consider the safety of a truss structure under an earthquake.

- The truss dynamics and material properties are assumed to be known:
  - density $7860 \text{ kg} \cdot \text{m}^{-3}$;
  - Young’s modulus $2.1 \times 10^{11} \text{ Pa}$;
  - yield stress $2.5 \times 10^8 \text{ Pa}$;
  - damping ratio 0.07.

- **Failure** consists of any truss member $i$’s axial strain $Y_i$ exceeding its yield strain $S_i$.

- The uncertainty with respect to which we perform OUQ is the **unknown earthquake ground motion** that the structure will experience.

Figure: A 198-member steel truss electrical tower.
Frequency Domain Formulation

An admissible set $\mathcal{A}$ can be constructed using the common seismological technique of considering the mean power spectrum, which is relatively well understood:

\[ s_{MA}(\omega) := C_1 e^{C_2 M_L} \frac{\omega_g^2 \omega^2}{(\omega_g^2 - \omega^2)^2 + 4\xi_g^2 \omega_g^2 \omega^2}. \]

Matsuda–Asano shape function (mean power spectrum) with Richter magnitude $M_L$ and site-specific natural frequency $\omega_g$ and damping $\xi_g$: 
A \_\text{MA} := \left\{ \mu \mid \mu \text{ is a prob. dist. on ground motions,} \right. \\
\left. \text{and } \mathbb{E}_\mu[\text{power spectrum}] = s_{\text{MA}} \right\}

- The typical approach is to repeatedly sample white noise, then filter those samples through a shape function (such as the Matsuda–Asano one) to generate samples with a “typical” power spectrum, and use the resulting ground motions as tests for the safety of the structure.

- This procedure amounts to sampling from just one possible probability distribution $\mu_{f.w.n.} \in A_{\text{MA}}$ — there are many others!

- The collection $A_{\text{MA}}$ can be traversed using OUQ. In our example, the optimizer manipulates 200 3-dimensional random Fourier coefficients: the reduced OUQ problem has dimension 600.
Numerical Results: Vulnerability Curves

Figure: The minimum and maximum probability of failure as a function of Richter magnitude $M_L$, where the power spectrum is constrained to have mean equal to the Matsuda–Asano shape function $s_{MA}$ with natural frequency $\omega_g$ and natural damping $\xi_g$ taken from the 24 Jan. 1980 Livermore earthquake. Each data point required $O(1 \text{ day})$ on 44+44 AMD Opterons ($shc$ and $foxtrot$ at Caltech).
Numerical Results: Vulnerability Curves

**Figure:** The minimum and maximum probability of failure as a function of Richter magnitude $M_L$, where the power spectrum is constrained to have mean equal to the Matsuda–Asano shape function $s_{MA}$ with natural frequency $\omega_g$ and natural damping $\xi_g$ taken from the 24 Jan. 1980 Livermore earthquake. Each data point required $O(1 \text{ day})$ on 44+44 AMD Opterons (shc and foxtrot at Caltech).

This gap can only be narrowed by acquiring more information, i.e. passing to $A \subsetneq A_{MA}$. 
Range of prediction given $\mathcal{A}$:

$$\mathcal{R}(\mathcal{A}) := \mathcal{U}(\mathcal{A}) - \mathcal{L}(\mathcal{A}),$$

$\mathcal{R}(\mathcal{A})$ small $\iff$ $\mathcal{A}$ very predictive.

Let $\mathcal{A}_{E,c}$ denote those scenarios in $\mathcal{A}$ that are consistent with getting outcome $c$ from some experiment $E$.

The optimal next experiment $E^*$ solves a minimax problem, i.e. $E^*$ is the most predictive even in its least predictive outcome:

$$E^* \text{ minimizes } E \mapsto \sup_{\text{outcomes } c \text{ of } E} \mathcal{R}(\mathcal{A}_{E,c}).$$
Experimental Design — Example

- Consider the fixed response function

\[ H(h, \alpha, v) := 10.396 \left( \left( \frac{h}{1.778} \right)^{0.476} (\cos \theta)^{1.028} \tanh \left( \frac{v}{v_{bl}} - 1 \right) \right)^{0.468}, \]

\[ v_{bl}(h, \theta) := 0.579 \left( \frac{h}{(\cos \theta)^{0.448}} \right)^{1.400}. \]

- Given: \( h, \theta \) and \( v \) are independent random variables in the cuboid

\[ (h, \alpha, v) \in [1.52, 2.67] \text{ mm} \times [0, \frac{\pi}{6}] \times [2.1, 2.8] \text{ km/s} \]

and \( \mathbb{E}[H(h, \theta, v)] \in [5.5, 7.5] \text{ mm}^2 \). OUQ analysis reveals that the least upper bound on \( \mathbb{P}[H(h, \theta, v) = 0] \) is 0.378969\ldots (vs. 0.038\ldots if one just assumes a uniform distribution).

- I offer to tell you (at great expense!) one of

\[ \mathbb{E}[h], \quad \mathbb{E}[\theta], \quad \mathbb{E}[v], \]

\[ \mathbb{V}[h], \quad \mathbb{V}[\theta], \quad \mathbb{V}[v], \quad \mathbb{V}[H(h, \theta, v)]. \]
Learning the variance of $h$ (light blue) would provide the greatest reduction on $\mathbb{P}[H = 0]$ in the minimax sense, although other pieces of information would yield lower upper bounds on $\mathbb{P}[H = 0]$ for particular outcomes.
Concluding Remarks
Conclusions

- **Optimal UQ** is (an opening gambit towards) a general framework for the sharp propagation of information/uncertainties. It can assist in decision-making under uncertainty by
  - forcing the user/client and UQ practitioner to clearly state all assumptions and information;
  - identifying key vulnerabilities in and assumptions about the system;
  - identifying what new information would be most informative.

- Dimensional reduction theorems make what is mathematically *The Right Thing To Do* into a **computationally tractable approach**.

- Simple situations $\rightarrow$ exact solutions and non-trivial mathematical insights.

- More complicated situations $\rightarrow$ numerical solutions that advance the boundaries of large-scale optimization.

- Some measure of defence against **GIGO**: sharp propagation of uncertainties can help to identify **GI** given **GO**.
Future Directions

- Many further applications of the reduction theorems and the OUQ framework in pure and applied contexts:
  - Work on Samuels’ conjecture (bounds sums of independent random variables of given mean) — with Y. Chen.
  - Further development of the seismic safety applications — with S. Mitchell and the research group of S. Krishnan.
  - Design and prediction of biological reactions — with M. Kennedy.
  - OUQ characterization of the effects of material microstructure morphology in bi-phase steels — with D. Balzani.
Future Directions

- Improvements to be made to the computational implementation of OUQ problems:
  - Exploit problem structure (e.g. multilinearity, partial convexity).
  - Automation of dimensional collapse and reduction.
  - Development of algorithms for identifying redundant or non-binding constraints, or activating a few constraints at a time à la the simplex algorithm — with L. H. Nguyen.

- OUQ with random sample data. Are there well-defined optimal bounds on probabilities when some of the information comes from a few (perhaps corrupted) realizations of random processes?

- Connections between OUQ and Bayesian inference — (families of) priors and posteriors on $\mathcal{A}$? In particular, can one have both robustness (posterior conclusions are stable w.r.t. changes of the prior) and consistency (posterior concentrates around the frequentist truth)?
Conclusions

Links

Under consideration at SIAM Review

Open-source optimization framework: dev.danse.us/trac/mystic
(OUQ tools in the development branch)