Optimal Uncertainty Quantification

Or: Computing Optimal Bounds on Imprecise Probabilities of (Physically Motivated) Imprecise Events

Tim Sullivan
tjs@caltech.edu

California Institute of Technology

EPSRC Symposium on Probability
Workshop on Multiscale Systems: Theory and Applications
University of Warwick, U.K.
12–16 December 2011

Overview

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.

CALTECH PSAAP

National Nuclear Security Administration

Introduction

What is Uncertainty Quantification (UQ)?

Motivating Problems for Optimal UQ

The Basic Idea

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 and are not probabilistic in nature — are nasty.

Motivational UQ Problems (1)

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 κ:

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

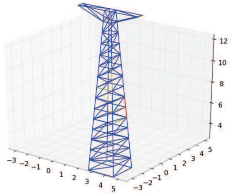
For a given point $x_0 \in U$ and threshold pressure $u_0 \in \mathbb{R}$, ...

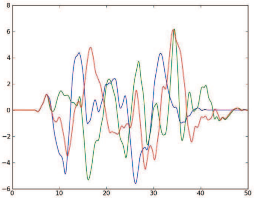
- Given κ and f, is it true that $u(x_0) \geq u_0$?
- What is $\mathbb{P}[u(x_0) \geq u_0]$ if the probability distribution $\mathbb{P}$ associated to random κ, f and boundary conditions is known?
- What if $\mathbb{P}$ is only partially known? What if the space of possibilities for $\mathbb{P}$ is infinite-dimensional?
- How do the answers depend upon the features of κ across various scales? Does the microstructure even matter at all?

Motivational UQ Problems (2)

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?





Motivational UQ Problems (3)

Imperfectly-Known Response

Consider a metric space $\mathcal{X}$ and a 1-Lipschitz function $G\colon \mathcal{X} \to \mathbb{R}$. Given a measurable event $E \subseteq \mathbb{R}$, ...

- For some given $x \in \mathcal{X}$, is $G(x) \in E$?
- When X is distributed according to some given $\mathbb{P} \in \mathcal{P}(\mathcal{X})$, what is $\mathbb{P}[G(X) \in E]$?
- What if $\mathbb{P}$ is incompletely specified? What if, in addition, G is incompletely specified, *e.g.* because it is known only on some $\mathcal{O} \subseteq \mathcal{X}$?
- **N.B.** If G is not uniquely specified, then neither is the set

$$\{x \in \mathcal{X} \mid G(x) \in E\}.$$

Common Themes — Motivation for OUQ

- 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 both simpler and less trivial than you might expect.

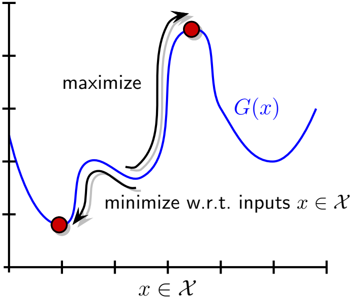


Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only *partially* known? (*I.e.* non-parametric epistemic uncertainty.)

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 both simpler and less trivial than you might expect.

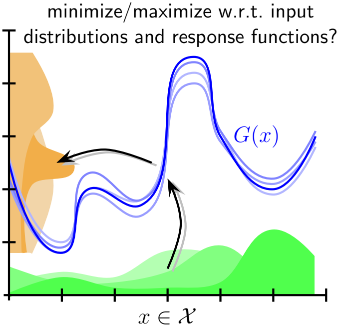


Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only *partially* known? (*I.e.* non-parametric epistemic uncertainty.)

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 $G \colon \mathcal{X} \to \mathcal{Y}$ 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

$$\mathbb{E}_{X \sim \mathbb{P}}[q(X, G(X))]$$

  when the reality $(G, \mathbb{P})$ is only imperfectly known.
- The key step in the Optimal Uncertainty Quantification approach is to specify a feasible set of admissible scenarios (g, μ) that could be $(G, \mathbb{P})$ according to the available information:

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} (g \colon \mathcal{X} \to \mathcal{Y}, \mu \in \mathcal{P}(\mathcal{X})) \text{ is consistent with} \\ \text{all given information about the real system } (G, \mathbb{P}) \\ (\textit{e.g.} \text{ legacy data, first principles, expert judgement}) \end{array} \right\}.$$

- $\mathcal{A}$ encodes everything that we know about the "reality" $(G, \mathbb{P})$.
- *A priori*, all we know about reality is that $(G, \mathbb{P}) \in \mathcal{A}$; we have no idea exactly which (g, μ) in $\mathcal{A}$ is actually $(G, \mathbb{P})$. No $(g, \mu) \in \mathcal{A}$ is "more likely" or "less likely" to be $(G, \mathbb{P})$ than any other.

Formulation of OUQ Problems

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} (g \colon \mathcal{X} \to \mathcal{Y}, \mu \in \mathcal{P}(\mathcal{X})) \text{ is consistent with} \\ \text{all given information about the real system } (G, \mathbb{P}) \\ (\textit{e.g.} \text{ legacy data, first principles, expert judgement}) \end{array} \right\}.$$

- Optimal bounds on the quantity of interest $\mathbb{E}_{X \sim \mathbb{P}}[q(X, G(X))]$ (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 $(g, \mu) \in \mathcal{A}$:

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

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 fixed event E (*i.e.* the response function $g = G$ is fixed and known), $\mathcal{L}(\mathcal{A})$ 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, and in the decision analysis world, distributionally robust decision analysis / optimization.
- 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 of measures $\mathcal{A}$.

Reduction of OUQ Problems — LP Analogy

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 $\mathcal{A}$.

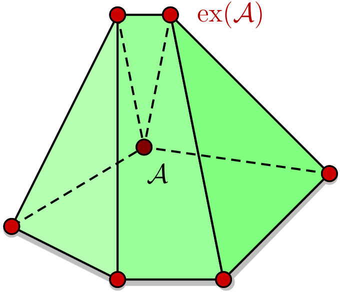


Figure: Just as a linear program finds its extreme value at the extremal points of a convex domain in $\mathbb{R}^n$, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

Reduction of OUQ Problems — Theorem

Theorem (Reduction for moment and independence constraints)

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

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

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

Then

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

$$\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_\Delta) \text{ and } \mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_\Delta).$$

Reduction of OUQ Problems — Sketch Proof

Proof.

- First consider $K = 1$, and fix $g\colon \mathcal{X} \to \mathbb{R}$.
- By definition, since $\mathcal{X}$ is a Radon space, all probability measures on $\mathcal{X}$ are inner regular, and so the set $\mathrm{ex}(\mathcal{A}_\Phi)$ of extreme points of

$$\mathcal{A}_\Phi := \left\{ \mu \in \mathcal{P}(\mathcal{X}) \,\middle|\, \mathbb{E}_{X\sim\mu}[\varphi_1(X)] \leq 0, \dots, \mathbb{E}_{X\sim\mu}[\varphi_n(X)] \right\}$$

  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 in the sense of **Winkler** (1988) — it satisfies a barycentric Choquet-type formula — and so its extreme values over $\mathcal{A}_\Phi$ and $\mathrm{ex}(\mathcal{A}_\Phi)$ 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, allowing an error of ε/K in each marginal. □

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 μ 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.

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,

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

$$\mathcal{U}(\mathcal{A}_{\mathsf{Mrkv}}) := \sup_{\mu \in \mathcal{A}} \mu[X \geq t] \leq \tfrac{m}{t}.$$

- In fact, $\mathcal{U}(\mathcal{A}_{\mathsf{Mrkv}}) = \frac{m}{t}$, and the probability distribution μ that attains this extreme value is

$$\mu = \left(1 - \tfrac{m}{t}\right)\delta_0 + \tfrac{m}{t}\delta_t.$$

McDiarmid's Inequality

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

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

with componentwise oscillations/global sensitivities defined by

$$\operatorname{osc}_k(g) := \sup \left\{ |g(x) - g(x')| \,\middle|\, \begin{array}{c} x, x' \in \mathcal{X}_1 \times \dots \times \mathcal{X}_K, \\ x_i = x'_i \text{ for } i \neq k \end{array} \right\}.$$

Theorem (McDiarmid's Inequality, 1988)

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

Optimal McDiarmid — Non-Propagation

Theorem

For $K = 1$,

$$\mathcal{U}(\mathcal{A}_{McD}) = \begin{cases} 0, & \text{if } D_1 \leq m, \\ 1 - \dfrac{m}{D_1}, & \text{if } 0 \leq m \leq D_1. \end{cases}$$

For $K = 2$,

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

There are similar explicit formulae for $K = 3$ (involving roots of cubic polynomials) and higher K.

Optimal McDiarmid — Non-Propagation

Theorem

For $K = 2$,

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

- 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.

Corollary for Multiscale Systems

In the presence of uncertainty about input probability distributions and the input-output relationship, there can be screening effects and information can fail to propagate.

Example: Random/Multiscale PDEs

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

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

  with appropriate boundary conditions.
- When the probability distribution $\mathbb{P}$ of κ 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:

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

- The OUQ-McDiarmid example can be applied in two ways here: the relative effects of κ and f; and the relative effects of micro and macro features of κ.

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

$$\operatorname*{ess\,inf}_U K > 0, \quad F \geq 0, \quad \int_U F(x)\,\mathrm{d}x > 0,$$

let

$$\mathcal{A} := \left\{ \mu \;\middle|\; \begin{array}{c} \text{under } \mu\text{, the fields } \kappa \text{ and } f \text{ are independent and, } \mu\text{-a.s.} \\ K(x) \leq \kappa(x) \leq e^{D_1} K(x), \\ F(x) \leq f(x) \leq e^{D_2} F(x) \end{array} \right\}.$$

Theorem

$\mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$. 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 $\min\{D_1, D_2\}$.

Example: Random/Multiscale PDEs

Setting II: Independent Porosity Micro- and Macrostructure

Given $D_1, D_2 \geq 0$, and fields $K_1, K_2 \colon U \to \mathbb{R}$ 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

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

Theorem

$\mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$. 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 $\min\{D_1, D_2\}$.

Optimal Hoeffding and the Effects of Nonlinearity

- Similarly, one can consider the admissible set $\mathcal{A}_{\text{Hfd}}$ that corresponds to the assumptions of Hoeffding's inequality, which bounds deviation probabilities of sums of independent bounded random variables:

$$\mathcal{A}_{\text{Hfd}} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} g \colon \mathbb{R}^K \to \mathbb{R} \text{ given by} \\ g(x_1, \dots, x_K) := x_1 + \dots + x_K, \\ \mu = \mu_1 \otimes \dots \otimes \mu_K \text{ supported on a cube of} \\ \text{side lengths } D_1, \dots, D_K, \text{ and } \mathbb{E}_{X \sim \mu}[g(X)] \geq m \geq 0 \end{array} \right\}.$$

- Hoeffding's inequality is the bound

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

- Interestingly, $\mathcal{U}(\mathcal{A}_{\text{McD}}) = \mathcal{U}(\mathcal{A}_{\text{Hfd}})$ for $K = 1$ and $K = 2$, but $\mathcal{U}(\mathcal{A}_{\text{McD}}) \geq \mathcal{U}(\mathcal{A}_{\text{Hfd}})$ for $K = 3$, and the inequality can be strict.

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

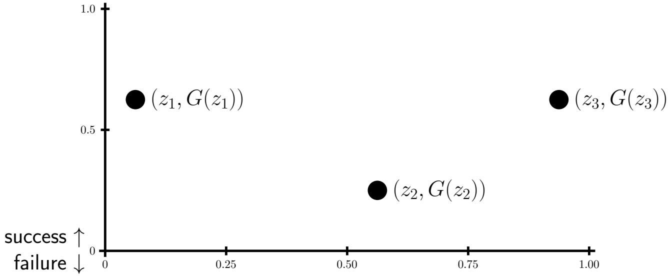
To determine if a system of interest will "fail" only with acceptably small probability, given observations of the system response on some subset $\mathcal{O}$ of the parameter space $\mathcal{X}$ and nowhere else.

Illustrative Example

To bound $\mathbb{P}[G(X) \leq 0]$, where $G \colon [0,1] \to \mathbb{R}$ 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 $\mathcal{O} \subseteq [0, 1]$?

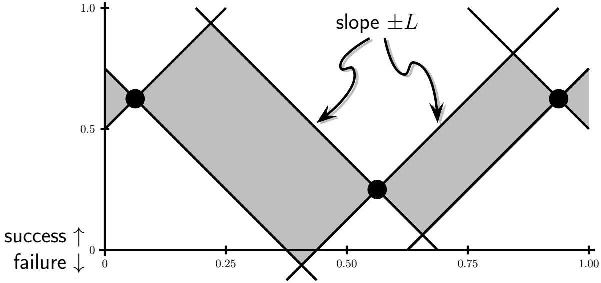


Sharpest Possible Answer...

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 $|G(x) - G(x')| \leq L|x - x'|$?

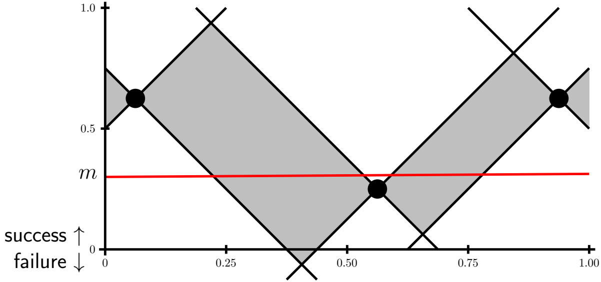


Sharpest Possible Answer...

...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 $\mathbb{E}[G(X)] \geq m$?

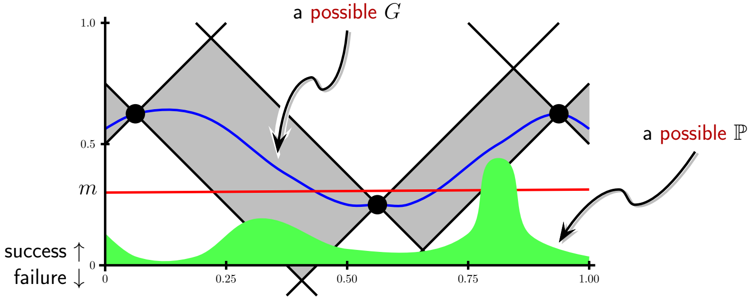


Sharpest Possible Answer...

...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 $\mathbb{E}[G(X)] \geq m$?



Sharpest Possible Answer...

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

Problem Formulation

- For $k \in \{1, \dots, K\}$, metric spaces $(\mathcal{X}_k, d_k)$ and independent $\mathcal{X}_k$-valued random variables X_k.
- Fix constants $L_1, \dots, L_K > 0$ and endow $\mathcal{X} := \mathcal{X}_1 \times \dots \times \mathcal{X}_K$ with the metric

$$d_L(x, x') := \sum_{k=1}^{K} L_k d_k(x_k, x'_k).$$

Knowledge to Encode in $\mathcal{A}$

- G is L-Lipschitz (*i.e.* has Lipschitz constant 1 w.r.t. the metric d_L);
- Observed data: the restriction $G|_{\mathcal{O}}$ of the real response function $G \colon \mathcal{X} \to \mathbb{R}$ to some subset $\mathcal{O} \subseteq \mathcal{X}$;
- Pairwise independence: $X_k \perp\!\!\!\perp X_\ell$ for $k \neq \ell$;
- Mean constraint: $\mathbb{E}_{X \sim \mathbb{P}}[G(X)] \geq m$.

Problem Formulation

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

$$\mathcal{A} := \left\{ (g,\mu) \,\middle|\, \begin{array}{c} \mu = \bigotimes_{k=1}^{K} \mu_k \in \bigotimes_{k=1}^{K} \mathcal{P}(\mathcal{X}_k) \subseteq \mathcal{P}(\mathcal{X}), \\ g \colon \mathcal{X} \to \mathbb{R} \text{ is } L\text{-Lipschitz}, \\ g = G \text{ on } \mathcal{O}, \text{ and } \mathbb{E}_{X \sim \mu}[g(X)] \geq m \end{array} \right\}.$$

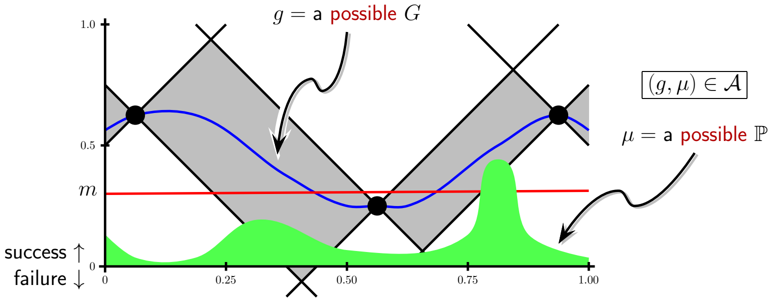
In other words, any (g, μ) for which μ is a product measure and g is L-Lipschitz, agrees with the legacy data, and has the right mean under μ could be $(G, \mathbb{P})$. The reduced admissible set, over which the quantity of interest has the same extreme values, is

$$\mathcal{A}_\Delta := \left\{ (g,\mu) \,\middle|\, \begin{array}{c} \mu = \bigotimes_{k=1}^{K} \mu_k \in \bigotimes_{k=1}^{K} \mathcal{P}(\mathcal{X}_k) \subseteq \mathcal{P}(\mathcal{X}), \\ \text{for some } x_0, x_1 \in \mathcal{X} \text{ and } p \in [0,1]^K, \\ \mu_k = p_k \delta_{x_0^k} + (1 - p_k) \delta_{x_1^k}, \\ g \colon \mathcal{O} \cup \mathcal{C}(x_0, x_1) \to \mathbb{R} \text{ is } L\text{-Lipschitz}, \\ g = G \text{ on } \mathcal{O}, \text{ and } \mathbb{E}_{X \sim \mu}[g(X)] \geq m \end{array} \right\}.$$

The Reduced Problem ($K = 1$)

The original problem entails optimizing over an infinite-dimensional collection of (g, μ) 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.

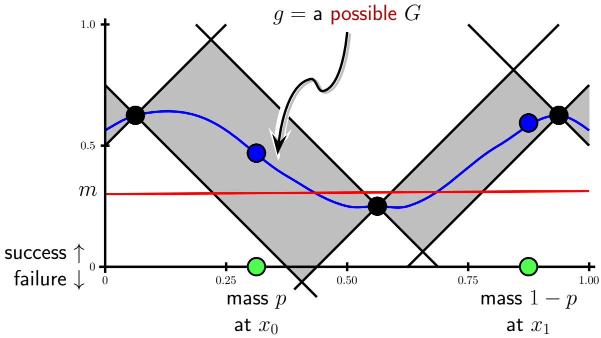
infinite-dimensional problem $\rightsquigarrow$ equivalent 5-dimensional problem!



The Reduced Problem ($K = 1$)

The original problem entails optimizing over an infinite-dimensional collection of (g, μ) 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.

infinite-dimensional problem $\rightsquigarrow$ equivalent 5-dimensional problem!

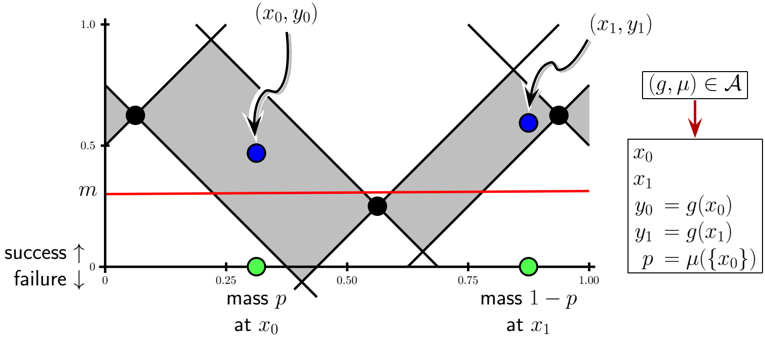


$(g, \mu) \in \mathcal{A}$

The Reduced Problem ($K = 1$)

The original problem entails optimizing over an infinite-dimensional collection of (g, μ) 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.

infinite-dimensional problem $\rightsquigarrow$ equivalent 5-dimensional problem!



The Reduced Problem ($K \in \mathbb{N}$)

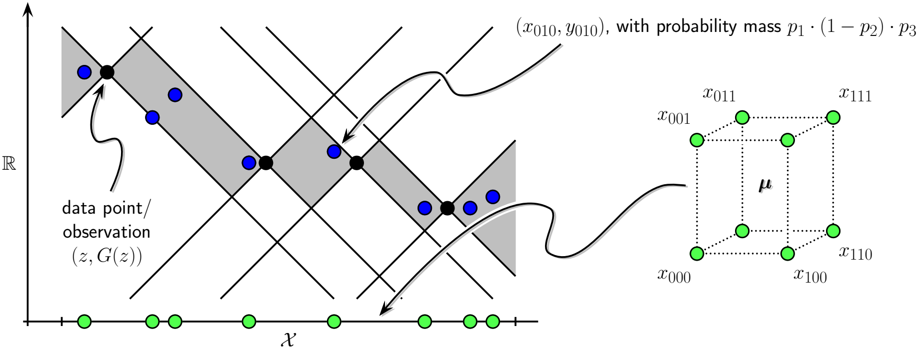


Figure: In the general case, the reduced probability measure μ is supported on the $2 \times 2 \times \cdots \times 2$ discrete (Hamming) cube $\mathcal{C}(x_0, x_1)$ spanned by $x_0, x_1 \in \mathcal{X}$ (the green dots). The blue dots show some feasible values for G over the support of the measure μ. The reduced problem has dimension $3K + 2^K$.

One Data Point

- The case of a single observation can be solved explicitly.
- Suppose that you observe one input-output pair of a function $G\colon [0,1] \to \mathbb{R}$ 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 $\mathbb{E}[G(X)] \geq m$:

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

Critical Data

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:

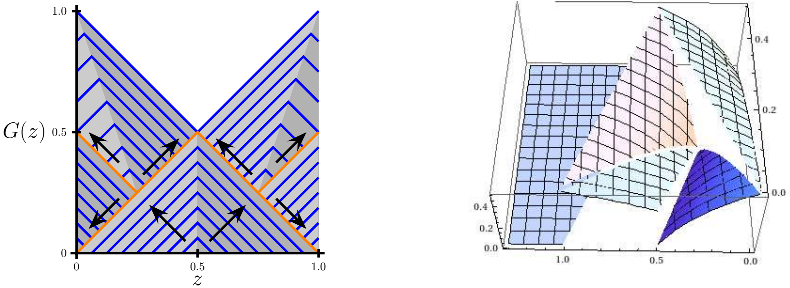
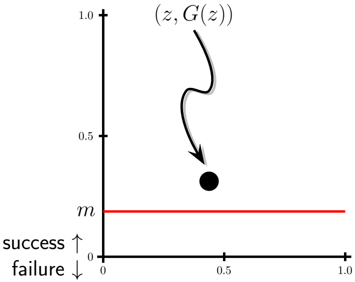
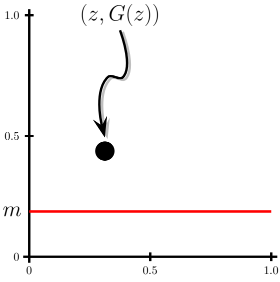


Figure: Schematic contour plot and to-scale surface plot of the least upper bound on the probability of failure, as a function of the observed data point $(z, G(z))$. There are jump discontinuities across the orange lines.

Critical Data



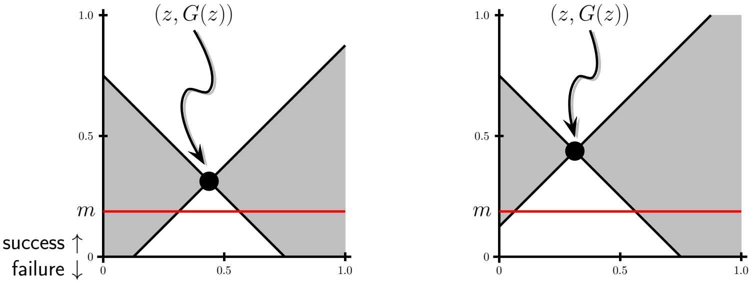
(a) "Subcritical" data point: probability of failure is high.



(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

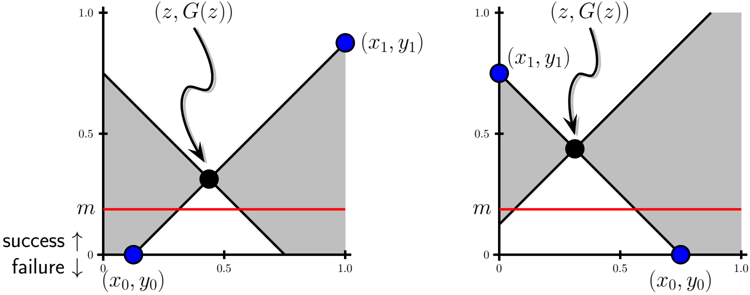


(a) "Subcritical" data point: probability of failure is high.

(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

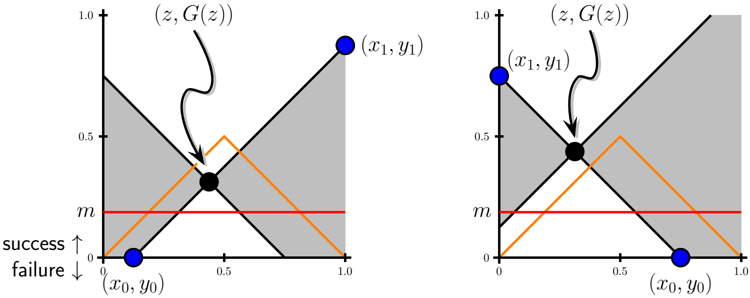


(a) "Subcritical" data point: probability of failure is high.

(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 $\left(1 - \frac{m_+}{y_1}\right)_+$.

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 $6.31\,\text{mm}^2$ and $15.36\,\text{mm}^2$.
- Failure event is $[G(h, \alpha, v) \leq \theta]$, for various values of θ.
- 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

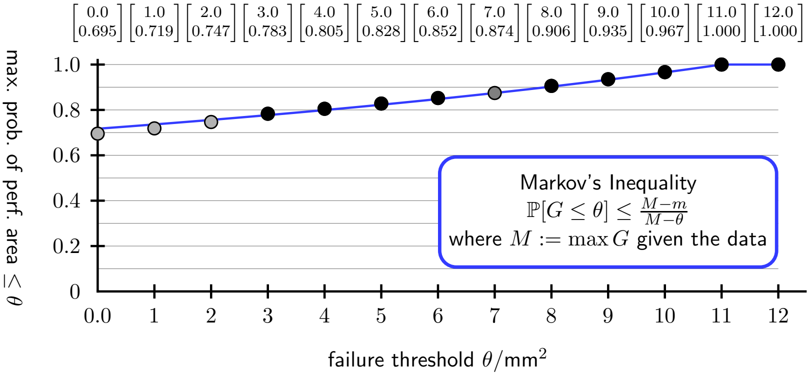


Figure: Maximum probability that perforation area is $\leq \theta$, for various θ, with the data and assumptions of the previous slide, including mean perforation area $\mathbb{E}[G(h, \alpha, v)] \geq m := 11.0\,\text{mm}^2$. Note close agreement of the results with Markov's bound.

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

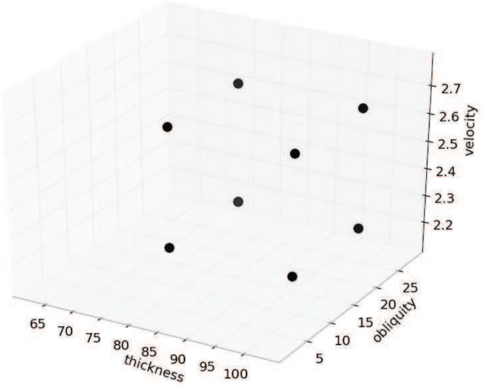
searches over $2 \times 2 \times 2$ product measures are slow and somewhat fragile,

$$\text{searches over } \begin{Bmatrix} 2 \times 1 \times 1 \\ 1 \times 2 \times 1 \\ 1 \times 1 \times 2 \end{Bmatrix} \text{ measures are faster and more robust,}$$

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

- 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.

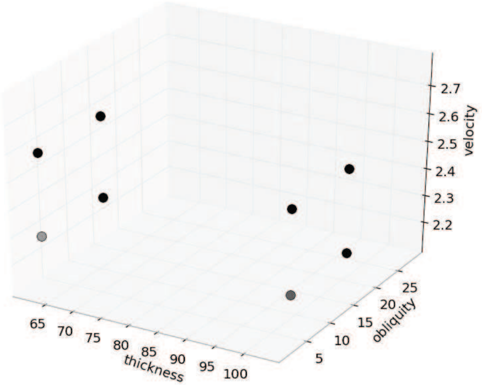
Dimensional Collapse



Iteration 0

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure in another hypervelocity-impact-related setting.

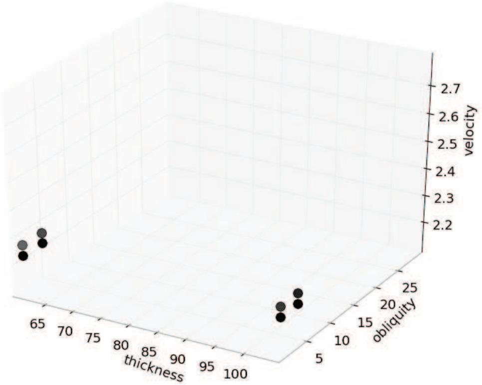
Dimensional Collapse



Iteration 150

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure in another hypervelocity-impact-related setting.

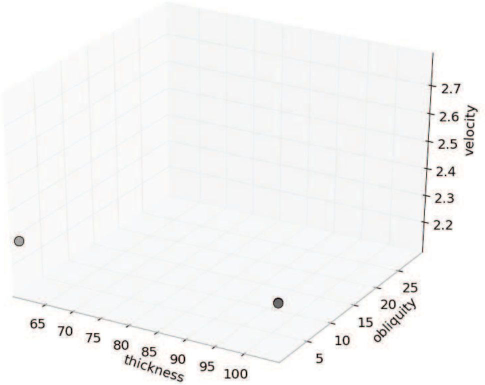
Dimensional Collapse



Iteration 200

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure in another hypervelocity-impact-related setting.

Dimensional Collapse



Iteration 1000

Figure: Collapse of the initial $2 \times 2 \times 2$ product measure to a $2 \times 1 \times 1$ product measure in another hypervelocity-impact-related setting.

Redundant and Non-Binding Data

- Now consider a set of observations $\mathcal{O} = \{z_1, \dots, z_N\}$, 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 $z_1, z_2 \in [0, 1]$:

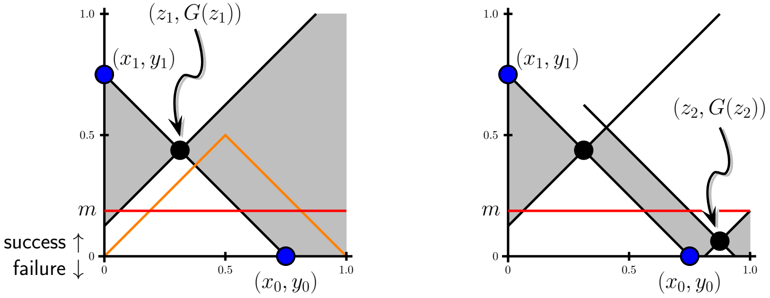


Figure: The extremizer for the problem with data point $(z_1, G(z_1))$ is feasible with respect to the new data point $(z_2, G(z_2))$, so the two problems have the same extreme value. The new data point is a relevant but non-binding data point.

Algorithm for Handling Large Data Sets with Redundancies

Theorem (Sufficient Condition to be Non-Binding)

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

$$|g(x) - G(z)| \leq d_L(x, z) \text{ for each } x \in \operatorname{supp}(\mu), \qquad (*)$$

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

N.B. The feasibility check $(*)$ is a simple algebraic check; it does not require any (potentially slow or expensive) optimizations.

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;
- $\widetilde{\mathcal{O}}_i$ = that data points that are not enforced at iteration i, but are potentially binding.

Sketch Algorithm

1. Initialize with $\mathcal{O}_0 = \varnothing$ and $\widetilde{\mathcal{O}}_0 = \mathcal{O}$.
2. Then, for $i = 1, 2, \dots$
    1. For each $z \in \widetilde{\mathcal{O}}_{i-1}$, 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 $z \in \widetilde{\mathcal{O}}_{i-1}$ with most extreme extreme value of $\mathbb{E}_\mu[q_g]$.
    2. Let $\mathcal{O}_i := \mathcal{O}_{i-1} \cup \{z_*\}$.
    3. Let $\widetilde{\mathcal{O}}_i$ 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 $\widetilde{\mathcal{O}}_i = \varnothing$.

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 $7860\,\mathrm{kg}\cdot\mathrm{m}^{-3}$;
  - Young's modulus 2.1×10^{11} Pa;
  - yield stress 2.5×10^{8} 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.

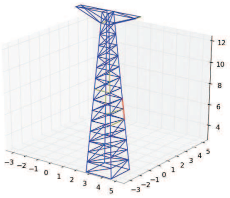
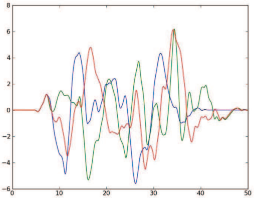
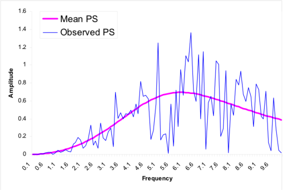


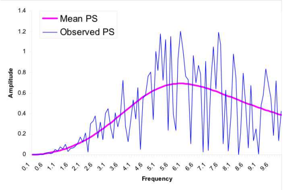
Figure: A 198-member steel truss electrical tower.

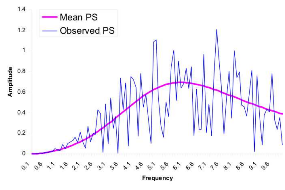


Frequency Domain Formulation

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







Matsuda–Asano shape function (mean power spectrum) with Richter magnitude M_{L} and site-specific natural frequency ω_{g} and damping ξ_{g}:

$$s_{\mathsf{MA}}(\omega) := C_1 e^{C_2 M_{\mathrm{L}}} \frac{\omega_{\mathrm{g}}^2 \omega^2}{(\omega_{\mathrm{g}}^2 - \omega^2)^2 + 4\xi_{\mathrm{g}}^2 \omega_{\mathrm{g}}^2 \omega^2}.$$

Frequency Domain Formulation

$$\mathcal{A}_{\mathsf{MA}} := \left\{ \mu \,\middle|\, \begin{array}{c} \mu \text{ is a prob. dist. on ground motions,} \\ \text{and } \mathbb{E}_{\mu}[\text{power spectrum}] = s_{\mathsf{MA}} \end{array} \right\}$$

- 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 $\mu_{\mathsf{f.w.n.}} \in \mathcal{A}_{\mathsf{MA}}$ — there are *many* others!.
- The collection $\mathcal{A}_{\mathsf{MA}}$ 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

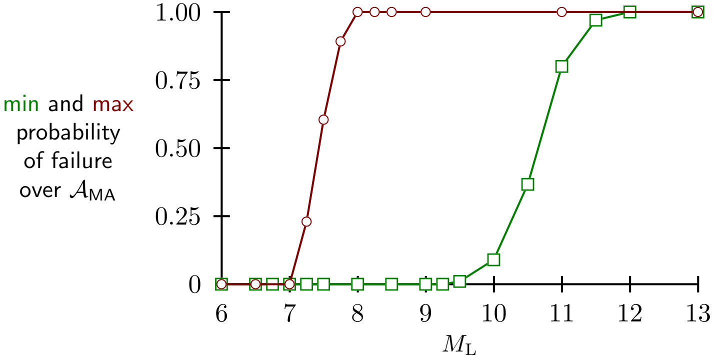
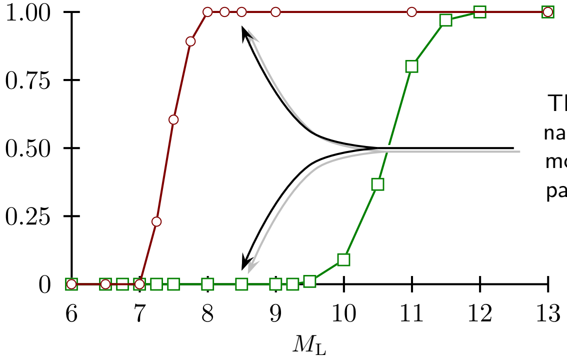


Figure: The minimum and maximum probability of failure as a function of Richter magnitude M_{L}, where the power spectrum is constrained to have mean equal to the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake.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

Numerical Results: Vulnerability Curves

min and max probability of failure over $\mathcal{A}_{\mathsf{MA}}$



This gap can only be narrowed by acquiring more information, *i.e.* passing to $\mathcal{A} \subsetneq \mathcal{A}_{\mathsf{MA}}$.

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 ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required O(1 day) on 44+44 AMD Opterons (*shc* and *foxtrot* at Caltech).

Optimal Knowledge Acquisition / Experimental Design

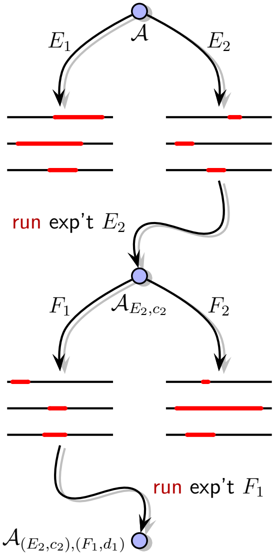
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



Experimental Design — Example

- Consider the fixed response function

$$H(h, \alpha, v) := 10.396 \left(\left(\tfrac{h}{1.778}\right)^{0.476} (\cos\theta)^{1.028} \tanh\left(\tfrac{v}{v_{\mathrm{bl}}} - 1\right)\right)_{+}^{0.468},$$

$$v_{\mathrm{bl}}(h, \theta) := 0.579 \left(\tfrac{h}{(\cos\theta)^{0.448}}\right)^{1.400}.$$

- Given: h, θ and v are independent random variables in the cuboid

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

  and $\mathbb{E}[H(h, \theta, v)] \in [5.5, 7.5]\,\mathrm{mm}^2$. OUQ analysis reveals that the least upper bound on $\mathbb{P}[H(h, \theta, v) = 0]$ is 0.378969. . . (*vs.* 0.038. . . if one just assumes a uniform distribution).
- I offer to tell you (at great expense!) one of

$$\mathbb{E}[h], \quad \mathbb{E}[\theta], \quad \mathbb{E}[v],$$
$$\mathbb{V}[h], \quad \mathbb{V}[\theta], \quad \mathbb{V}[v], \quad \mathbb{V}[H(h, \theta, v)].$$

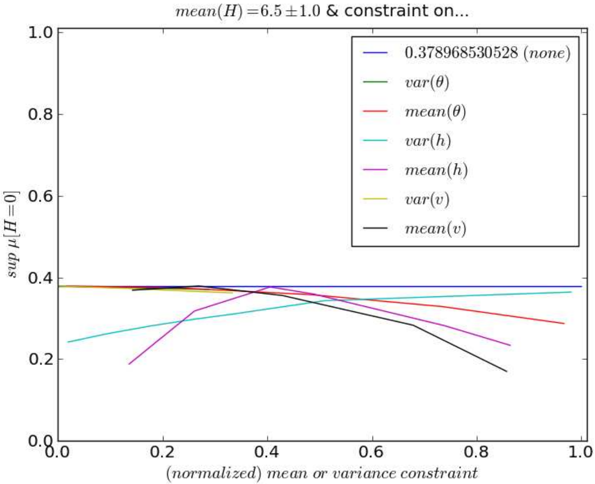


Figure: Learning the variance of h (light blue) would provide the greatest reduction on $\mathbb{P}[H = 0]$ in the minimax sense, although other pieces of information would yield lower upper bounds on $\mathbb{P}[H = 0]$ for particular outcomes.

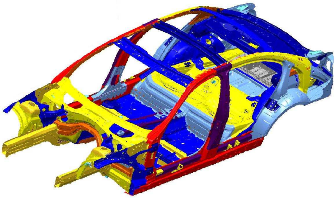
Concluding Remarks

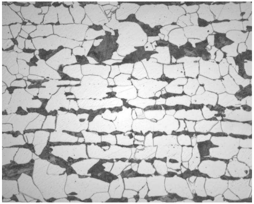
Conclusions

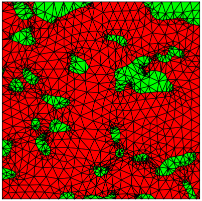
- Optimal UQ is (an opening gambit towards) a general framework for the sharp propagation of information/uncertainties. It can assist in decision-making under uncertainty by
    - forcing the user/client and UQ practitioner to clearly state all assumptions and information;
    - identifying key vulnerabilities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

Links

Preprint: arXiv:1009.0679v2
Under consideration at *SIAM Review*

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