Optimal Uncertainty Quantification and Sensitivity Analysis with Incomplete Information

Tim Sullivan
tjs@caltech.edu

California Institute of Technology

INFORMS 2011 Annual Meeting
Decision Support Models and Sensitivity Analysis
Charlotte, North Carolina, U.S.A.
13–16 November 2011

Overview

1. Introduction
   - Problem Description
   - Optimal Uncertainty Quantification (OUQ)
2. Examples of OUQ
   - Optimal Concentration Inequalities
   - Seismic Safety
   - Knowledge Acquisition / Experimental Design
3. 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.

CALTECH PSAAP

National Nuclear Security Administration

Introduction

The Problem: Optimal Bounds

Optimal Uncertainty Quantification: Formulation, Reduction and Implementation

Problem Setting

Challenge

- 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

- 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:

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

- $\mathcal{A}$ encodes everything that we know about the "reality" $(G, \mathbb{P})$.
- *A priori*, all we know about $(G, \mathbb{P})$ 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

- 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:

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

- 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

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

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 $\mathcal{A}$.

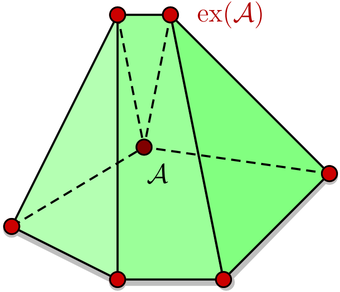


Figure: Just as a linear program finds its extreme value at the extremal points of a convex domain in $\mathbb{R}^n$, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

Reduction of OUQ Problems — Theorem

Theorem (Reduction for moment and independence constraints)

For fixed measurable functions $\varphi_i \colon \mathcal{X} \to \mathbb{R}$ *and* $\varphi_i^{(k)} \colon \mathcal{X}_k \to \mathbb{R}$, *let*

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} g \colon \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_K \to \mathbb{R} \text{ is measurable,} \\ \mu = \mu_1 \otimes \cdots \otimes \mu_K \in \bigotimes_{k=1}^K \mathcal{P}(\mathcal{X}_k), \\ \langle \text{any conditions on } g \text{ alone} \rangle, \\ \mathbb{E}_{X \sim \mu}\big[\varphi_i(X)\big] \leq 0 \text{ for } i = 1, \dots, n_0, \\ \mathbb{E}_{X_k \sim \mu_k}\big[\varphi_i^{(k)}(X_k)\big] \leq 0 \text{ for } i = 1, \dots, n_k \text{ and } k = 1, \dots, K \end{array} \right\}$$

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

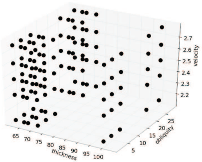
Then

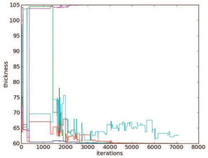
$$\dim(\mathcal{A}_\Delta) \leq \sum_{k=1}^K N_k (1 + \dim(\mathcal{X}_k)) + \prod_{k=1}^K N_k - K,$$

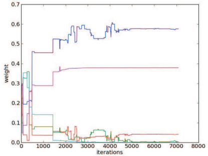
$$\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_\Delta) \text{ and } \mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_\Delta).$$

Reduced OUQ Problems

- The finite-dimensional problems $\mathcal{L}(\mathcal{A}_\Delta)$ and $\mathcal{U}(\mathcal{A}_\Delta)$ can be solved numerically.
- Current tool of choice: *mystic*, a Python-based open-source optimization framework.
  - Easily swappable strategies for optimization, population generation, enforcement of constraints, termination criteria.
  - Manages optimizations on scales ranging from the small (second-long on a laptop) to the large (days on dozens-of-cores clusters).
- Depending on the specific structure of $\mathcal{A}$, there are additional layers of reduction theorems. *E.g.* in the McDiarmid example that follows, a theorem enables us to "forget" the coordinates in the input spaces.







Examples of OUQ in Action

McDiarmid's Inequality: Parameter (In)Sensitivity

Large-Scale Example: Seismic Safety Certification

Improving the Bounds: Optimal Knowledge Acquisition / Experimental Design

McDiarmid's Inequality

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

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

with componentwise oscillations/global sensitivities defined by

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

Note that saying "$(G, \mathbb{P}) \in \mathcal{A}_{\mathsf{McD}}$" specifies neither G nor $\mathbb{P}$ exactly. As usual, we want to know the worst-case probability of failure

$$\mathcal{U}(\mathcal{A}_{\mathsf{McD}}) := \sup_{(g,\mu) \in \mathcal{A}_{\mathsf{McD}}} \mu[g(X) \leq 0]$$

McDiarmid's Inequality

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

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

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 $K = 2$,

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

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

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.

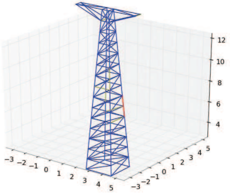
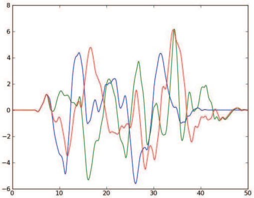
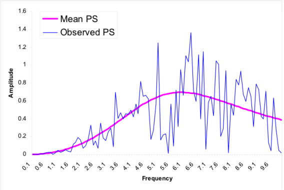


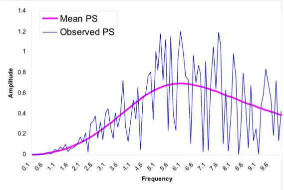
Figure: A 198-member steel truss electrical tower.

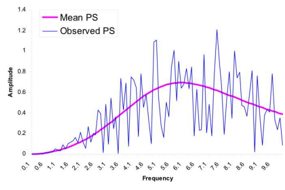


Frequency Domain Formulation

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

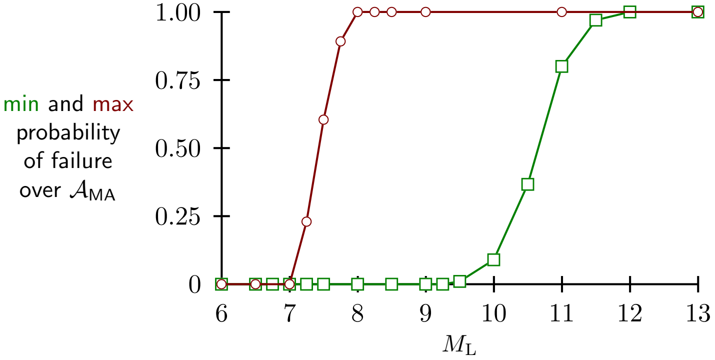
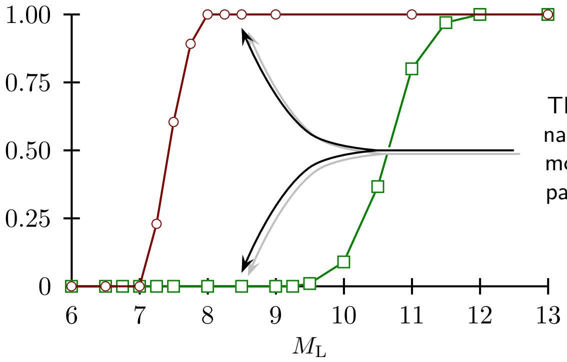


Figure: The minimum and maximum probability of failure as a function of Richter magnitude M_{L}, where the power spectrum is constrained to have mean equal to the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. 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 $\mathcal{A}_{\mathrm{MA}}$



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

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

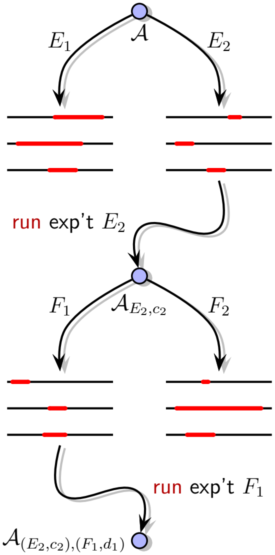
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



Conclusions

- Optimal UQ is a general framework for the sharp propagation of information/uncertainties. It can assist in decision-making under uncertainty by identifying key vulnerabilities in and assumptions about the system, and what new information would be most informative.
- Dimensional reduction theorems make what is mathematically *The Right Thing To Do* into a computationally tractable approach — small problems can be done in minutes on a laptop, larger ones in hours/days on clusters.
- Future work: connections between OUQ and (robust) Bayesian inference — (families of) priors and posteriors on $\mathcal{A}$?

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

Open-source optimization framework: dev.danse.us/trac/mystic