Optimal Uncertainty Quantification

Tim Sullivan
tjs@caltech.edu

California Institute of Technology
http://www.psaap.caltech.edu/

PSAAP V&V/UQ Workshop
Stanford University, California, U.S.A.
16–18 August 2011

Overview

1. Optimal Uncertainty Quantification
2. Simple Example: OUQ with Legacy Data
3. Large-Scale Example: Seismic Safety
4. Conclusions / Outlook

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 United States 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.

Optimal Uncertainty Quantification

Overview of the Philosophy and Fundamental Results

Optimization-Driven UQ

Bounds Mean Optimizations!

- Conventional worst/best-case design is an optimization problem over possible design and operation parameters:

$$\min_{x \in \mathcal{X}} G(x), \qquad \max_{x \in \mathcal{X}} G(x).$$

- Insufficient to make statements about *e.g.* probabilities of events.
- We want to handle generic information about the probability distributions and response functions, which are in general incompletely specified.

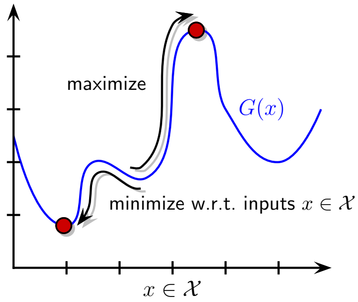


Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only *partially* known? (*I.e.* non-parametric epistemic uncertainty.)

Optimization-Driven UQ

Bounds Mean Optimizations!

- Conventional worst/best-case design is an optimization problem over possible design and operation parameters:

$$\min_{x \in \mathcal{X}} G(x), \qquad \max_{x \in \mathcal{X}} G(x).$$

- Insufficient to make statements about *e.g.* probabilities of events.
- We want to handle generic information about the probability distributions and response functions, which are in general incompletely specified.

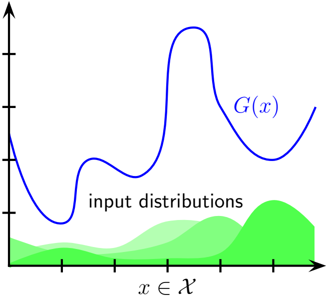


Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only *partially* known? (*I.e.* non-parametric epistemic uncertainty.)

Optimization-Driven UQ

Bounds Mean Optimizations!

- Conventional worst/best-case design is an optimization problem over possible design and operation parameters:

$$\min_{x \in \mathcal{X}} G(x), \qquad \max_{x \in \mathcal{X}} G(x).$$

- Insufficient to make statements about *e.g.* probabilities of events.
- We want to handle generic information about the probability distributions and response functions, which are in general incompletely specified.

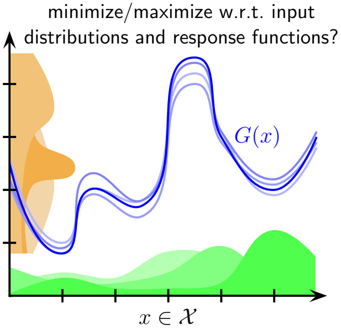


Figure: Optimizing $G(x)$ over $x \in \mathcal{X}$ yields deterministic worst- and best-case outcomes. What if the distribution of the inputs is only *partially* known? (*I.e.* non-parametric epistemic uncertainty.)

Optimal Uncertainty Quantification (OUQ)

- OUQ is a mathematical formulation of UQ that places information at the centre of the problem — items of information are viewed as constraints.
- Particularly suited the regime of high-consequence decision-making with incomplete information.
- Naturally generalizes classical interval analysis and optimization-based UQ methods to the probabilistic regime.
- Basic idea: pick a quantity of interest and optimize (minimize/maximize) with respect to the scenarios compatible with your current state of knowledge.

UQ Problems

- Reliability
- Certification
- Verification
- Validation
- Extrapolation
- Prediction
- Sensitivity
- Model Reduction
- ...

Owhadi *& al.* (2010)
http://arxiv.org/abs/1009.0679

OUQ Paradigm

- Abstract system $G\colon \mathcal{X} \to \mathbb{R}$ with random inputs X with probability distribution $\mathbb{P} \in \mathcal{P}(\mathcal{X})$ — but the pair $(G, \mathbb{P})$ is imperfectly known!
- Quantity of interest $\mathbb{E}[q_G]$, *e.g.* the mean $\mathbb{E}[G]$, or the probability of failure $\mathbb{P}[G \leq 0] \equiv \mathbb{E}[\mathbb{1}[G \leq 0]]$.
- Feasible set of admissible scenarios that could be the reality $(G, \mathbb{P})$:

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

- Optimal bounds on $\mathbb{E}[q_G]$ given the information encoded in $\mathcal{A}$ are found by minimizing/maximizing $\mathbb{E}_\mu[q_g]$ over $(g, \mu) \in \mathcal{A}$:

$$\inf q \leq \inf_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_g] \leq \mathbb{E}[q_G] \leq \sup_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_g] \leq \sup q.$$

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 $\mathcal{A}$.

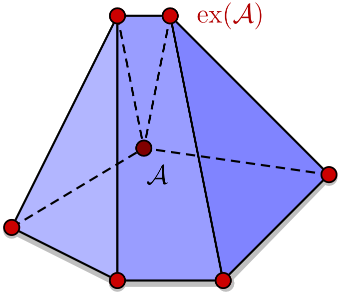


Figure: Just as a linear program finds its extreme value at the extremal points of a convex domain in $\mathbb{R}^n$, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

Reduction Theorem: Moment Constraints

Theorem

For fixed measurable functions $\varphi_i \colon \mathcal{X} \to \mathbb{R}$, *let*

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} g \colon \mathcal{X} \to \mathbb{R} \textit{ is measurable}, \\ \mu \in \mathcal{P}(\mathcal{X}), \\ \langle \textit{some conditions on } g \textit{ alone} \rangle, \\ \mathbb{E}_\mu[\varphi_1] \leq 0, \, \dots, \, \mathbb{E}_\mu[\varphi_{n'}] \leq 0 \end{array} \right\},$$

$$\mathcal{A}_\Delta := \left\{ (g, \mu) \in \mathcal{A} \,\middle|\, \begin{array}{c} \mu = \sum_{i=0}^{n'} \alpha_i \delta_{x_i} \textit{ for some} \\ x_i \in \mathcal{X}, \, \alpha_i \geq 0, \, \sum_{i=0}^{n'} \alpha_i = 1 \end{array} \right\} \subseteq \mathcal{A}.$$

Then

$$\inf_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_g] = \inf_{(g,\mu)\in\mathcal{A}_\Delta} \mathbb{E}_\mu[q_g] \textit{ and } \sup_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_g] = \sup_{(g,\mu)\in\mathcal{A}_\Delta} \mathbb{E}_\mu[q_g].$$

Reduction Theorem: Independence Constraints

Theorem

For fixed measurable functions φ_i^k *and* φ_i, *let*

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} g \colon \mathcal{X} = \prod_{k=1}^{K} \mathcal{X}_k \to \mathbb{R} \text{ is measurable,} \\ \mu = \mu_1 \otimes \cdots \otimes \mu_K \in \bigotimes_{k=1}^{K} \mathcal{P}(\mathcal{X}_k), \\ \langle \text{some conditions on } g \text{ alone} \rangle, \\ \mathbb{E}_\mu[\varphi_1] \leq 0, \dots, \mathbb{E}_\mu[\varphi_{n'}] \leq 0, \\ \mathbb{E}_{\mu_k}[\varphi_1^k] \leq 0, \dots, \mathbb{E}_{\mu_k}[\varphi_{n_k}^k] \leq 0 \end{array} \right\}$$

$$\mathcal{A}_\Delta := \left\{ (g, \mu) \in \mathcal{A} \,\middle|\, \begin{array}{c} \mu_k = \sum_{i=0}^{n_k + n'} \alpha_i^k \delta_{x_i^k} \text{ for some} \\ x_i^k \in \mathcal{X}_k,\ \alpha_i^k \geq 0,\ \sum_{i=0}^{n'} \alpha_i = 1 \end{array} \right\} \subseteq \mathcal{A}.$$

Then

$$\inf_{(g,\mu) \in \mathcal{A}} \mathbb{E}_\mu[q_g] = \inf_{(g,\mu) \in \mathcal{A}_\Delta} \mathbb{E}_\mu[q_g] \text{ and } \sup_{(g,\mu) \in \mathcal{A}} \mathbb{E}_\mu[q_g] = \sup_{(g,\mu) \in \mathcal{A}_\Delta} \mathbb{E}_\mu[q_g].$$

OUQ with Legacy Data

Small-Scale Example of OUQ in Action, and Some of the Common Variations / Complications

The Legacy UQ (Certification) Challenge

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

General Challenge

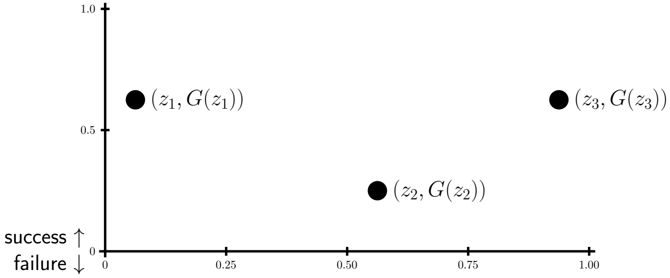
To determine if a system of interest will "fail" only with acceptably small probability, given observations of the system response on some subset $\mathcal{O}$ of the parameter space $\mathcal{X}$ and nowhere else.

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 $\mathcal{O} \subseteq [0, 1]$?

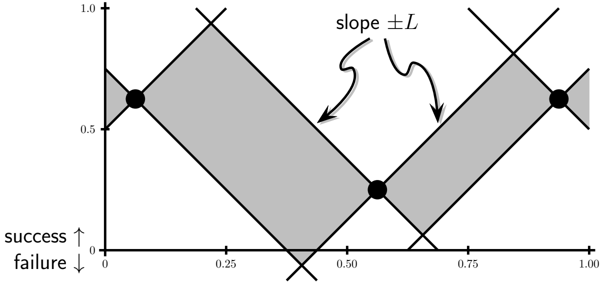


Sharpest Possible Answer...

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

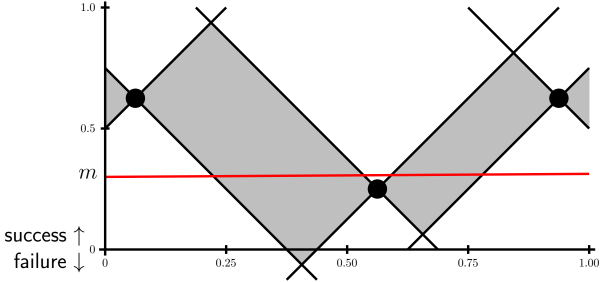


Sharpest Possible Answer...

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

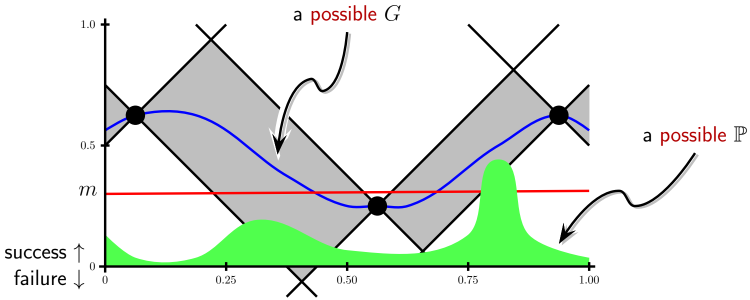


Sharpest Possible Answer. . .

. . . is non-trivial, and can be found using optimization techniques. This is the Optimal UQ viewpoint.

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 optimization techniques. This is the Optimal UQ viewpoint.

Problem Formulation

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

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

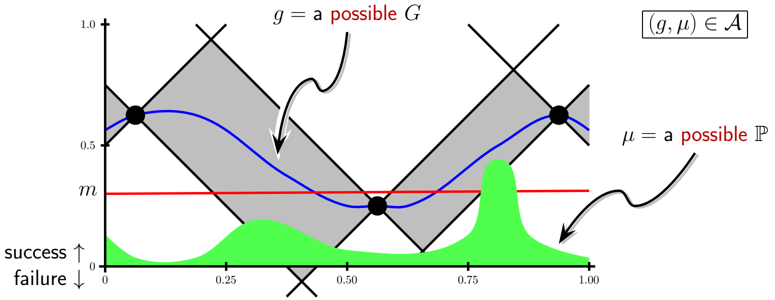
In other words, any (g, μ) for which g is L-Lipschitz, agrees with the legacy data, and has the right mean under μ could be $(G, \mathbb{P})$. The reduced admissible set, over which the quantity of interest has the same extreme values, is

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

The Reduced Problem

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.

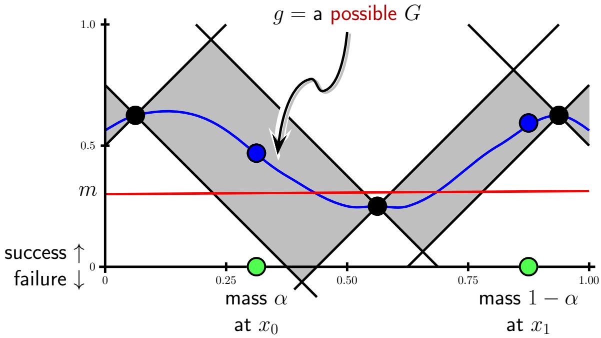
infinite-dimensional problem $\rightsquigarrow$ equivalent 5-dimensional problem!



The Reduced Problem

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.

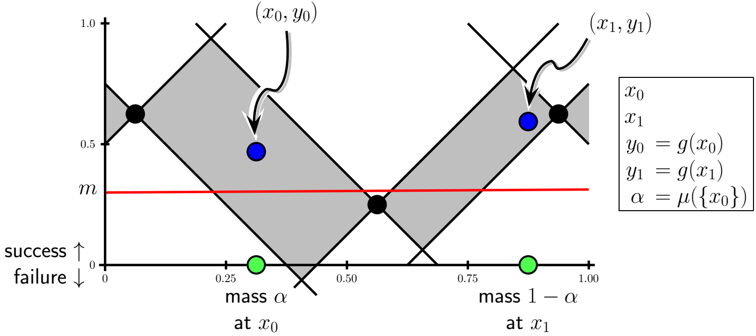
infinite-dimensional problem $\rightsquigarrow$ equivalent 5-dimensional problem!



The Reduced Problem

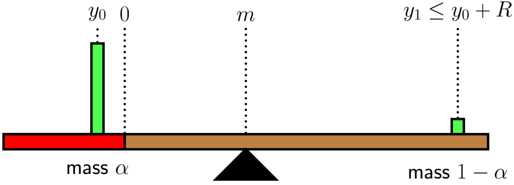
The original problem entails optimizing over an infinite-dimensional collection of (g, μ) that could be $(G, \mathbb{P})$. In the reduced problem, we only have to move around and re-weight two Dirac measures (point masses) and the values of g over those two points.

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



Let's Play Seesaw. . .

Rules. Place two masses α and $1-\alpha$ at positions y_0 and y_1 respectively, such that $|y_0 - y_1| \leq R$ and $\alpha y_0 + (1-\alpha) y_1 \geq m$.
Objective. Maximize the total mass in the failure region $(-\infty, 0]$.



$$\text{Maximum "failure" mass } \alpha_{\max} = \left(1 - \frac{m_+}{R}\right)_+,$$

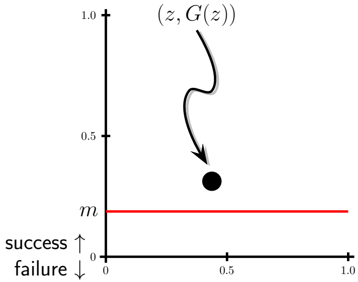
attained by putting mass $\alpha_{\max}$ at $y_0 = 0$, and $1 - \alpha_{\max}$ at $y_1 = R$.

One Data Point

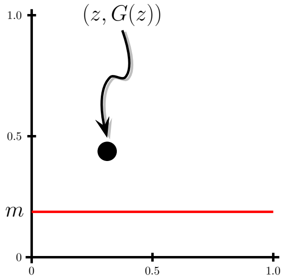
- The case of a single observation can be solved explicitly.
- Suppose that you observe one input-output pair of a function $G\colon [0,1] \to \mathbb{R}$ with Lipschitz constant L.
- You know $(z, G(z))$ — assume that $z \in [0, \frac{1}{2}]$ and $G(z) > 0$.
- Four cases for the least upper bound on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

$$\mathbb{P}[G(X) \leq 0] \leq \begin{cases} \left(1 - \frac{m_+}{L-(Lz-G(z))}\right)_+, & \text{if } G(z) \leq Lz, \\ \left(1 - \frac{m_+}{L-(Lz+G(z))}\right)_+, & \text{if } Lz < G(z) \leq L\left|\frac{1}{2} - z\right|, \\ \left(1 - \frac{2m_+}{L+(G(z)-Lz)}\right)_+, & \text{if } L\left|\frac{1}{2} - z\right| < G(z) \leq L|1 - 3z|, \\ \left(1 - \frac{m_+}{Lz+G(z)}\right)_+, & \text{if } G(z) > L\max\{z, 1 - 3z\}. \end{cases}$$

Critical Data



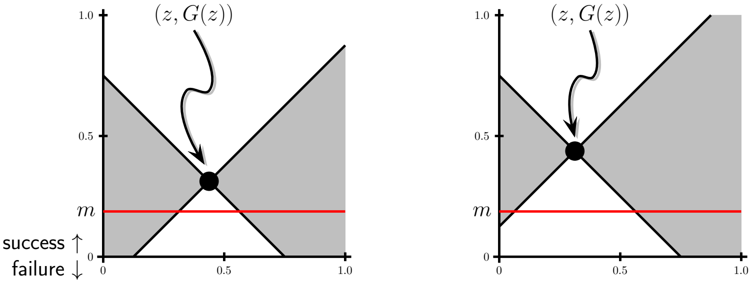
(a) "Subcritical" data point: probability of failure is high.



(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

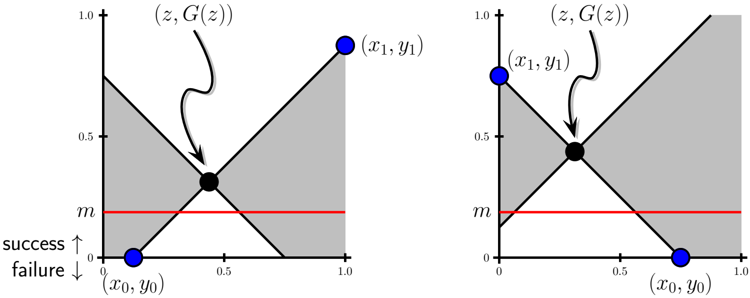


(a) "Subcritical" data point: probability of failure is high.

(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

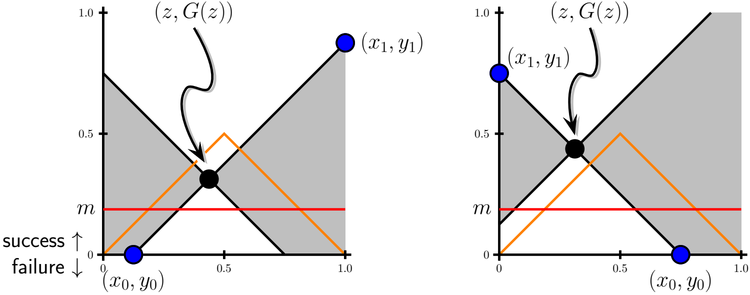


(a) “Subcritical” data point: probability of failure is high.

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

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:

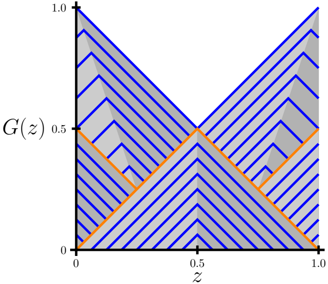


Figure: A schematic diagram of the level sets of the least upper bound on the probability of failure, as a function of the observed data point $(z, G(z))$. There are jump discontinuities across the orange lines.

Redundant and Non-Binding Data

- Now consider a set of observations $\mathcal{O} = \{z_1, \dots, z_N\}$.
- Which data points $(z_n, G(z_n))$ contribute non-trivial constraints, and actually determine the set of feasible (x_0, y_0) and (x_1, y_1)? (*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 $z_1, z_2 \in [0, 1]$:

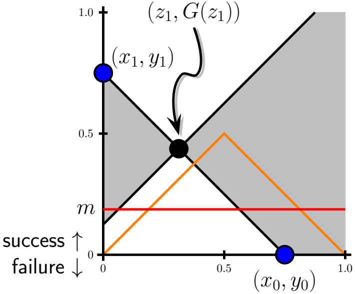


Figure: The extremizer for the problem with data point $(z_1, G(z_1))$ is feasible with respect to the new data point $(z_2, G(z_2))$, so the two problems have the same extreme value. The new data point is a relevant but non-binding data point.

Examples of Redundant and Non-Binding Data

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

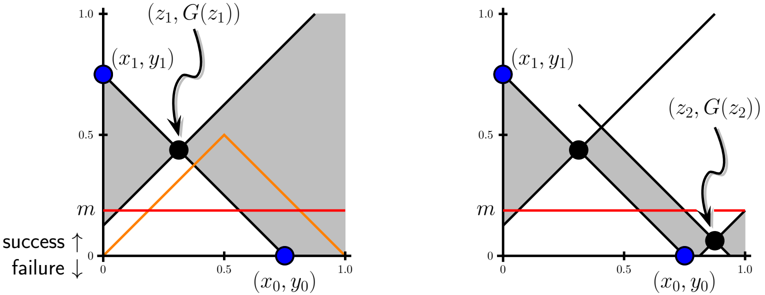


Figure: The extremizer for the problem with data point $(z_1, G(z_1))$ is feasible with respect to the new data point $(z_2, G(z_2))$, so the two problems have the same extreme value. The new data point is a relevant but non-binding data point.

Incorporating Observational Uncertainties

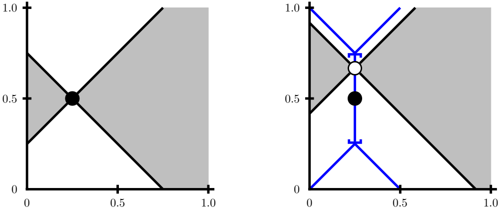
- Suppose that we are not entirely sure about the observed data points $(z, G(z))$ — there is observational uncertainty.
- For example, suppose that $z \in \mathcal{X}$ is observed perfectly, but $G(z)$ is only observed to an error of $\pm\delta$: we observe $(z, \tilde{G}(z))$, knowing that the true value $G(z)$ at z satisfies $|G(z) - \tilde{G}(z)| \leq \delta$.
- Corresponding admissible sets:

$$\mathcal{A} := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} g \colon [0,1] \to \mathbb{R} \text{ is } L\text{-Lipschitz}, \\ \mu \text{ a probability measure on } [0,1], \\ |g(z) - \tilde{G}(z)| \leq \delta \text{ for all } z \in \mathcal{O}, \text{ and } \mathbb{E}_\mu[g(X)] \geq m \end{array} \right\},$$

$$\mathcal{A}_\Delta := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} x_0, x_1 \in [0,1], g \colon \mathcal{O} \cup \{x_0, x_1\} \to \mathbb{R} \\ \mu = \alpha\delta_{x_0} + (1-\alpha)\delta_{x_1}, \\ |g(x) - g(x')| \leq L|x - x'| \text{ for all } x, x' \in \mathcal{O} \cup \{x_0, x_1\}, \\ |g(z) - \tilde{G}(z)| \leq \delta \text{ for all } z \in \mathcal{O}, \\ \alpha g(x_0) + (1-\alpha) g(x_1) \geq m \end{array} \right\}.$$

Incorporating Observational Uncertainties

- The less we know about the observational errors (or the larger δ is), the larger the set of admissible scenarios $\mathcal{A}$, and hence the wider the bounds on the quantity of interest.
- The "true" data points enter as new optimization variables, constrained to lie close to the observed data points.



(a) no observational uncertainty

(b) observational uncertainty of $\pm\delta$

Figure: Feasible sets without and with observational uncertainty.

Bounds Using (Validated) Models

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

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

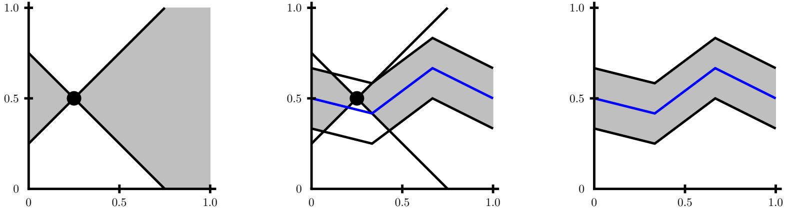
- Corresponding admissible sets:

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

$$\mathcal{A}_\Delta := \left\{ (g, \mu) \,\middle|\, \begin{array}{c} x_0, x_1 \in [0,1], g\colon \mathcal{O} \cup \{x_0, x_1\} \to \mathbb{R} \\ \mu = \alpha\delta_{x_0} + (1-\alpha)\delta_{x_1}, \\ |g(x) - g(x')| \leq L|x - x'| \text{ for all } x, x' \in \mathcal{O} \cup \{x_0, x_1\}, \\ |g(x) - F(x)| \leq C_V \text{ for all } x \in \mathcal{O} \cup \{x_0, x_1\}, \\ |g(x) - G(z)| \leq L|x - z| \text{ for all } z \in \mathcal{O}, \\ \alpha g(x_0) + (1-\alpha) g(x_1) \geq m \end{array} \right\}.$$

Bounds Using (Validated) Models

- The knowledge that F is a quantitatively validated model for G restricts the set $\mathcal{A}$ of admissible scenarios, and hence sharpens the bounds on the quantity of interest.
- The resulting bounds are tighter than those arrived at by either legacy data $G|_{\mathcal{O}}$ or the model F alone.



(a) data alone (b) data and model (c) model alone

Figure: Feasible sets given data (observations), or a model, or both.

Ancillary Measurements

- Suppose that instead of having a model F for G, we have a model F' for some ancillary quantity G'. Can F' be used to make statements about G?
- Yes, provided we know (approximately) how G and G' are related. *E.g.*, if

$$G(x) = \mathcal{T}(G'(x))$$

  for some transformation $\mathcal{T}$, then just apply the previous slide to the model F defined in terms of $\mathcal{T}$ and F' by

$$F(x) := \mathcal{T}(F'(x)).$$

- $\mathcal{T}$ could express something very direct and deterministic, like a Fourier transform, or something more statistical, like a covariance or correlation.

(Non-)Propagation of Information through Hierarchies

One can consider hierarchies (directed acyclic graphs) of OUQ modules:

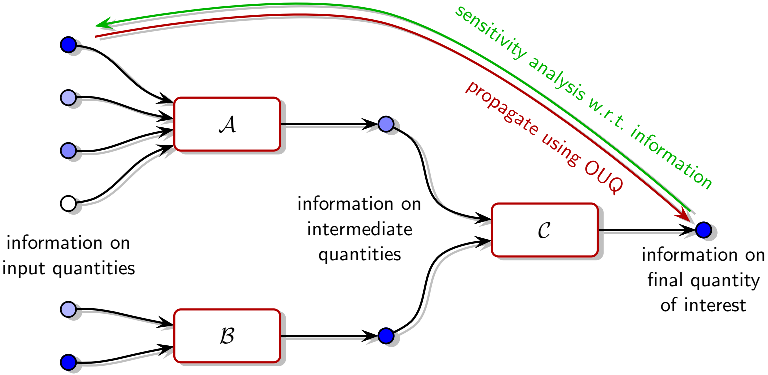


Figure: Because OUQ is a *sharp* information propagation scheme, the results of sensitivity analysis ("inverse OUQ") give non-trivial insights into the roles of the various pieces of input information. Some inputs may even be irrelevant!

OUQ-Driven Experimental Planning

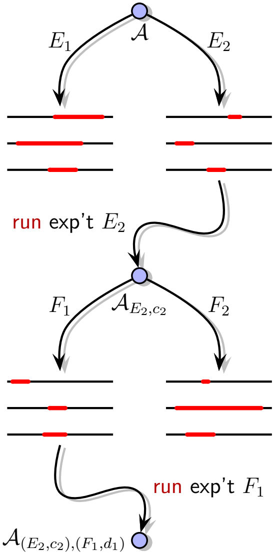
- Range of prediction for q given $\mathcal{A}$:

$$\mathcal{R}(q|\mathcal{A}) := \sup_{(g,p)\in\mathcal{A}} \mathbb{E}_p[q_g] - \inf_{(g,p)\in\mathcal{A}} \mathbb{E}_p[q_g],$$

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

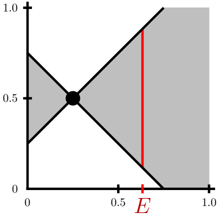
- Let $\mathcal{A}_{E,c}$ denote those scenarios in $\mathcal{A}$ that are consistent with getting outcome c from some experiment E.
- The optimal next experiment E^* satisfies a minimax criterion, *i.e.* E^* is the most predictive even in its least predictive outcome:

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

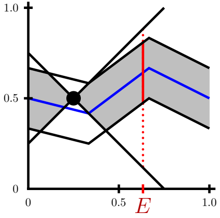


OUQ-Driven Experimental Planning with Models

As before, a validated model F is an advantage: it restricts the possibilities for G and hence reduces the time to discovery (*i.e.* an acceptably narrow range of prediction for q).



(a) data alone



(b) data and model

Figure: Upon adding to the legacy data $G|_{\mathcal{O}}$ the validated model F, the range of possible outcomes for $G(E)$ shrinks, and so does the range of prediction $\sup_c \mathcal{R}(q|\mathcal{A}_{E,c})$ for the quantity of interest.

OUQ for Seismic Safety

High-Dimensional Example of OUQ in Action

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.
- There are two sources of uncertainty on which we perform OUQ:
  - the source term of the earthquake,
  - the transfer function from source to the truss site.

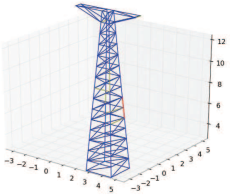


Figure: A 198-member steel truss electrical tower.

Problem Formulation

- We know the system reponse function G in this case: it is the mapping from source term s and transfer function ψ, to ground acceleration at the truss site,

$$\ddot{u}_0(t) = (\psi \star s)(t);$$

through the (linear) dynamics of the structure,

$$M\ddot{v} + C\dot{v} + Kv = f - MT\ddot{u}_0;$$

to

$$\min_{\text{members } i} \left(S_i - \sup_{t \geq 0} |Y_i(t)| \right),$$

which is positive iff the structure survives.

- Denote by $\mathcal{S}$ the space of possible source terms s, and by Ψ the space of possible transfer functions ψ: we must describe which probability distributions on $\mathcal{S} \times \Psi$ are going to be admissible.

Why Use OUQ?

- Conventional worst-case design is simply the exercise of finding the minimum value of the system response $G(s, \psi)$ over all $(s, \psi) \in \mathcal{S} \times \Psi$.
- If the structure can fail *anywhere* in $\mathcal{S} \times \Psi$, then the structure is declared to be unsafe.
- This is a sound defensive design approach, but is "far too pessimistic to be practical" (Drenick, 1973): in worst-case design, the (s, ψ) that compromise the truss are "tuned" to the structure — they are, in general, statistically unreasonable earthquakes.
- OUQ, on the other hand, offers a way to explore a broad class of statistically reasonable earthquakes, and to provide uniform rigorous bounds on the probability of failure across such a class.

Problem Formulation — Transfer Function

- Write the transfer function $\psi \in \Psi$ with respect to the piecewise-linear basis of "tent" functions φ_i:

$$\psi(t) := \frac{\sqrt{q}}{\tau'} \sum_{i=1}^{q} c_i \varphi_i(t)$$

  with $q = 20$, $\tau' = 10\,\mathrm{s}$, $c = (c_1, \dots, c_q)$ a random vector of unknown distribution in $[-1, 1]^q$ such that $\sum_{i=1}^{q} c_i^2 \leq 1$, $\sum_{i=1}^{q} c_i = 0$.
- No statistical constraints on c, so we are optimizing the placement of one Dirac mass in the space Ψ, *i.e.* performing conventional worst-case analysis to identify a single worst transfer function.

Problem Formulation — Source Term

- Treat the source term as a succession of boxcar time impulses of random direction and duration:

$$s = \sum_{i=1}^{B} X_i s_i(t)$$

where

  - $X_1, \dots, X_B \in [-a_{\mathrm{max}}, a_{\mathrm{max}}]^3 \subseteq \mathbb{R}^3$ independent with independent components, with mean $\mathbb{E}[X_i] = 0$;
  - step functions $s_i(t) := \mathbb{1}\left[\sum_{j=1}^{i-1} \tau_j \leq t < \sum_{j=1}^{i} \tau_j\right]$ with durations $\tau_1, \dots, \tau_B \in [0, \tau_{\mathrm{max}}]$ independent, with mean $\bar{\tau}_1 \leq \mathbb{E}[\tau_i] \leq \bar{\tau}_2$.
- Use Esteva's (1970) semi-empiricial expressions for earthquakes on firm ground:

$$a_{\mathrm{max}} = \frac{a_0 e^{\lambda M_{\mathrm{L}}}}{(R_0 + R)^2},$$

$$a_0 = 12.3 \times 10^6\,\mathrm{m}^3 \cdot \mathrm{s}^{-2}, \qquad \lambda = 0.8, \qquad R_0 = 25\,\mathrm{km},$$

$$\bar{\tau}_1 = 1\,\mathrm{s}, \qquad \bar{\tau}_2 = 2\,\mathrm{s}, \qquad \tau_{\mathrm{max}} = 6\,\mathrm{s}, \qquad B = 20.$$

Reduced Problem

- Apply the OUQ reduction theorems: we move around
  - one Dirac mass in Ψ, $\implies$ q parameters;
  - two Dirac masses on each of the three components of each X_i, $\implies$ $3 \times 3B$ parameters;
  - two Dirac masses for each duration τ_i $\implies$ $3B$ parameters.
- Reduced OUQ problem is a global optimization problem in dimension

$$12B + q = 260.$$

- A fully-converged maximum-probability-of-failure OUQ calculation takes $O(24\,\text{hrs})$ on parts of Caltech's *shc* and *foxtrot* clusters, totalling 88 AMD Opterons.

Numerical Convergence

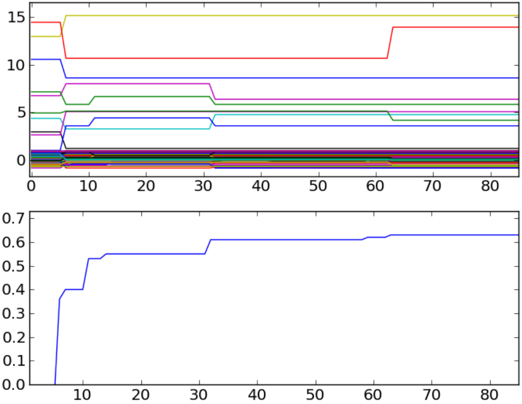
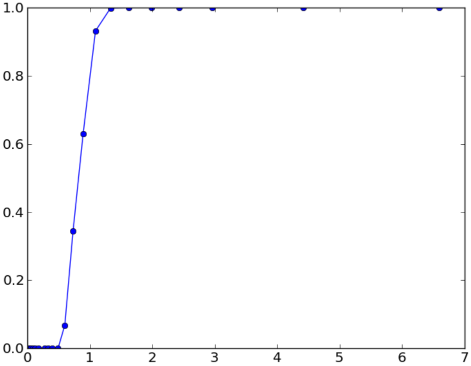


Figure: Numerical convergence of reduced optimization variables (top) and probability of failure (bottom) for Richter magnitude 6.5.

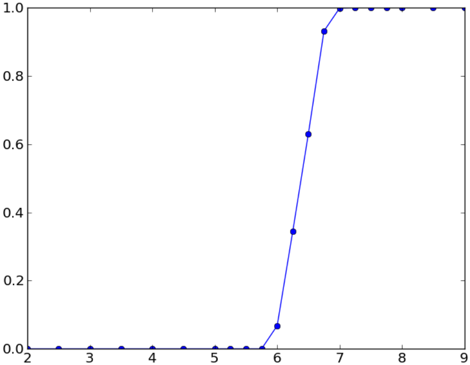
Numerical Results



$a_{\max}/\mathsf{m} \cdot \mathsf{s}^{-2}$

Figure: Maximum probability of failure $\sup_{\mu \in \mathcal{A}} \mu[\text{failure}]$. Note the sharp transition around $a_{\max} \approx 0.9\,\mathsf{m} \cdot \mathsf{s}^{-2}$, *i.e.* $M_{\mathrm{L}} \approx 6.5$.

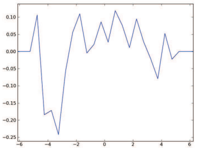
Numerical Results



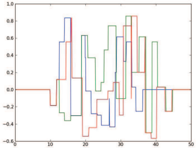
M_{L}

Figure: Maximum probability of failure $\sup_{\mu \in \mathcal{A}} \mu[\text{failure}]$. Note the sharp transition around $a_{\max} \approx 0.9\,\mathrm{m} \cdot \mathrm{s}^{-2}$, *i.e.* $M_{\mathrm{L}} \approx 6.5$.

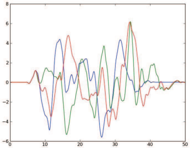
Numerical Results Near Critical Excitation



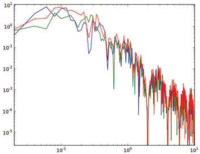
(a) extremal transfer function ψ at Richter $M_L = 6.5$



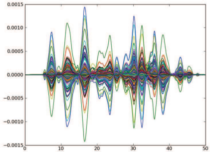
(b) the 3 components of the extremal source function s



(c) the corresponding ground accelerations at the truss site



(d) the corresponding power spectra



(e) time evolution of all member strains

Numerical Results Near Critical Excitation

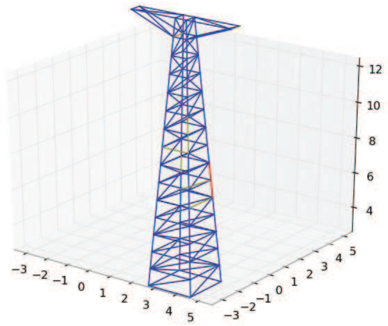
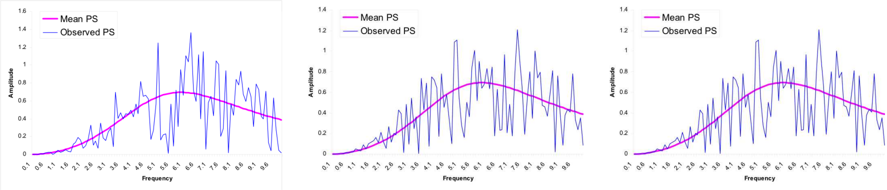


Figure: The members highlighted in red and yellow are the ten weakest members under a Richter magnitude 6.5 earthquake.

Frequency Domain Formulation

An alternative admissible set 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):

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

Frequency Domain Formulation

- The typical approach in the literature is to filter white noise through a shape function (such as the Matsuda–Asano one) to generate a "typical" power spectrum, and use the resulting earthquake as the test for the safety of the structure.
- This procedure amounts to sampling from just one possible probability distribution μ on earthquakes for which the μ-mean power spectrum is s_{MA} — there are many others!.
- The collection of *all* earthquake distributions with s_{MA} as their mean power spectrum can be traversed using OUQ.
- The optimizer manipulates random Fourier coefficients rather than pulse durations, and the reduced problem has dimension $O(600)$.

Numerical Results

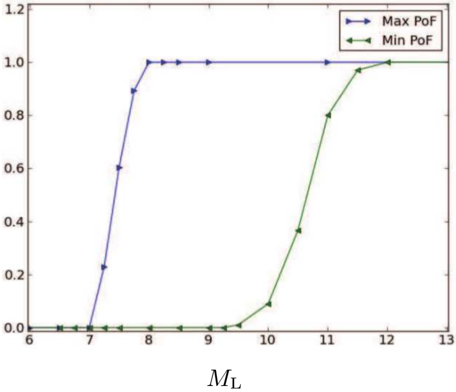


Figure: The minimum and maximum probability of failure as a function of Richter magnitude in the frequency domain formulation, where the power spectrum is constrained to have mean equal to the Matsuda–Asano shape function s_{MA} with natural frequency ω_g and natural damping ξ_g taken from the 24 January 1980 Livermore earthquake.

Conclusions

Conclusions / Outlook

- OUQ is an information propagation scheme.
- In principle, arbitrary input information can be propagated to give optimal bounds on any chosen output quantity of interest.
- Even in simple settings, there are interesting and non-trivial results to be seen.
- The method has been applied to real-world examples, entailing developments in global optimization in hundreds-dimensional parameter spaces.

  $\Longrightarrow$ extreme scale computation (M. Stalzer's talk)