Optimal Uncertainty Quantification and (Non-)Propagation of Uncertainties Across Scales

Tim Sullivan
tjs@caltech.edu

California Institute of Technology

SIAM Conference on Uncertainty Quantification
Predictive Fidelity in Multiscale Simulations
Raleigh, North Carolina, U.S.A.
1–5 April 2012

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).

1. Introduction
    - Prototypical UQ Problem
    - Formulation and Reduction of Optimal UQ Problems
2. Examples of OUQ and (Non-)Propagation
    - Optimal Concentration Inequalities
    - OUQ with Legacy Data
    - Seismic Safety
3. Closing Remarks
    - Closing Remarks

Portions of this work were supported by the U. S. Department of Energy NNSA 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.

Overview

Prototypical UQ Problem: Reliability Certification

- $G_0 \colon \mathcal{X} \to \mathcal{Y}$ is a system of interest, with random inputs X distributed according to a probability measure μ_0 on $\mathcal{X}$.
- For some subset $\mathcal{F} \subseteq \mathcal{Y}$, the event $[G_0(X) \in \mathcal{F}]$ constitutes failure; we want to know the probability of failure

$$\mathbb{P}_{\mu_0}\big[G_0(X) \in \mathcal{F}\big] \equiv \mathbb{E}_{\mu_0}\Big[\underbrace{\mathbb{1}\big[G_0(X) \in \mathcal{F}\big]}_{\text{q.o.i.}}\Big],$$

  or at least to know that it is acceptably small (or unacceptably large!).
- Our interest lies in understanding $\mathbb{P}_{\mu_0}\big[G_0(X) \in \mathcal{F}\big]$ when G_0 and μ_0 are only imperfectly known, and to obtain bounds that are optimal with respect to the known information.
- Our approach is to treat this as an optimization problem over all (g, μ) that could be (G_0, μ_0).

Overview

Optimal UQ

with **M. McKerns**, **H. Owhadi**, **M. Ortiz** (Caltech), **C. Scovel** (LANL)

- The key step in the Optimal Uncertainty Quantification approach is to specify a feasible set of admissible scenarios (g, μ) that could be (G_0, μ_0) according to the available information:

$$\mathcal{A} = \left\{ (g, \mu) \,\middle|\, \begin{array}{c} (g, \mu) \text{ is consistent with the current} \\ \text{information about (\textit{i.e.} could be) } (G_0, \mu_0) \\ \text{(\textit{e.g.} legacy data, first principles, expert judgement)} \end{array} \right\}.$$

- $\mathcal{A}$ encodes everything that we know about the "reality" (G_0, μ_0).
- *A priori*, all we know about reality is that $(G_0, \mu_0) \in \mathcal{A}$; we have no idea exactly which (g, μ) in $\mathcal{A}$ is actually (G_0, μ_0). No $(g, \mu) \in \mathcal{A}$ is "more likely" or "less likely" to be (G_0, μ_0) than any other.

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns & M. Ortiz. "Optimal Uncertainty Quantification." Submitted to *SIAM Review*. arXiv:1009.0679

Optimal UQ

$$\mathcal{A} = \left\{ (g, \mu) \,\middle|\, \begin{array}{c} (g, \mu) \text{ is consistent with the current} \\ \text{information about (\textit{i.e.} could be) } (G_0, \mu_0) \end{array} \right\}$$

- Optimal bounds on the quantity of interest $\mathbb{P}_{\mu_0}[G_0(X) \in \mathcal{F}]$ (optimal w.r.t. the information encoded in $\mathcal{A}$) are found by minimizing/ maximizing $\mathbb{P}_{\mu}[g(X) \in \mathcal{F}]$ over all admissible scenarios $(g, \mu) \in \mathcal{A}$:

$$\mathcal{L}(\mathcal{A}) \leq \mathbb{P}_{\mu_0}[G_0(X) \in \mathcal{F}] \leq \mathcal{U}(\mathcal{A}),$$

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

$$\mathcal{L}(\mathcal{A}) := \inf_{(g,\mu)\in\mathcal{A}} \mathbb{P}_{\mu}[g(X) \in \mathcal{F}],$$

$$\mathcal{U}(\mathcal{A}) := \sup_{(g,\mu)\in\mathcal{A}} \mathbb{P}_{\mu}[g(X) \in \mathcal{F}].$$

- *Cf.* imprecise probability (**Boole** (1854)), distributionally robust optimization, robust Bayesian inference (surv. **Berger** (1984)).

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 $\mathcal{A}$.

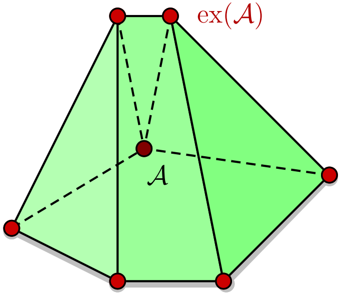


Figure: A linear functional on a convex domain in $\mathbb{R}^n$ finds its extreme value at the extremal points of the domain; similarly, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

Reduction of OUQ Problems — Theorem

Theorem (Generalized moment and indep. constraints)

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

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

$$\mathcal{A}_{\Delta} := \left\{ (g, \mu) \in \mathcal{A} \,\middle|\, \begin{array}{c} \mu_k \in \Delta_{N_k}(\mathcal{X}_k) \\ \text{where } N_k := n_0 + n_k \end{array} \right\} \subseteq \mathcal{A}.$$

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

Heuristic

If you have N_k pieces of information relevant to the random variable X_k, then just pretend that X_k takes at most $N_k + 1$ values in $\mathcal{X}_k$.

Reduction of OUQ Problems — Comments

- The reduction theorem is very general: no compactness required, just some (weak) regularity properties:
  - the spaces involved just need to be Radon;
  - the functions involved must be integrable.

  We generalize results of **Karr** (1983), which required $\mathcal{X}$ to be compact and the functions to be bounded and continuous.
- The proof involves Choquet theory on the probability simplex $\mathcal{P}(\mathcal{X})$ and on $\bigotimes_{k=1}^{K} \mathcal{P}(\mathcal{X}_k)$, and uses results of **Winkler** (1988) and **von Weizsäcker & Winkler** (1979/80) on the extreme points of generalized moment sets and extreme values of measure-affine functions (functions that satisfy barycentric Choquet formulae).
- Similar arguments can be applied for other admissible classes $\mathcal{A}$, but generalized moment classes have computationally nice extreme points.

Overview

Optimal Concentration Inequalities

Classical inequalities of probability theory can be seen as OUQ statements:

Example: Markov's Inequality in OUQ Form

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

$$\mathcal{U}(\mathcal{A}_{\mathsf{Mrkv}}) := \sup_{\mu \in \mathcal{A}_{\mathsf{Mrkv}}} \mathbb{P}_\mu[X \leq 0] = \frac{M - m}{M}.$$

How about other deviation/concentration-of-measure inequalities?

- McDiarmid's inequality: deviations from the mean of bounded-differences functions of independent random variables.
- Hoeffding's inequality: deviations from the mean of sums of independent random variables.
- Samuels' conjecture: deviations of sums of non-negative independent random variables with given means.

Consider

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

with componentwise oscillations/global sensitivities defined by

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

Theorem (McDiarmid's Inequality, 1988)

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

Optimal McDiarmid and Screening Effects

Theorem (Optimal McDiarmid Inequality)

For $K = 1$,

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

For $K = 2$, *if* $D_1 \geq D_2 \geq 0$,

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

There are similar explicit formulae for $K = 3$ (which involves solving auxiliary cubic polynomials), $K = 4$ (quartics), and so on.

Optimal McDiarmid and Screening Effects

Theorem (Optimal McDiarmid Inequality)

For $K = 2$, *if* $D_1 \geq D_2 \geq 0$,

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

- There is an explicitly-identified regime in which the worst-case bound on the probability of failure is controlled only by m and D_1.
- In this regime, the statement that $\operatorname{osc}_2(g) \leq D_2$ carries no information,
  - not in the sense that it contains zero bits,
  - but that it is a non-binding constraint: its inclusion/removal does not change the extreme value of the OUQ problem.

Optimal Hoeffding and the Effects of Nonlinearity

- Similarly, one can consider $\mathcal{A}_{\mathsf{Hfd}}$ "$\subseteq$" $\mathcal{A}_{\mathsf{McD}}$ corresponding to the assumptions of Hoeffding's inequality, which bounds deviation probabilities of sums of independent bounded random variables:

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

- Hoeffding's inequality is the bound

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

- Interestingly, $\mathcal{U}(\mathcal{A}_{\mathsf{Hfd}}) = \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$ for $K = 1$ and $K = 2$, but $\mathcal{U}(\mathcal{A}_{\mathsf{Hfd}}) \leq \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$ for $K = 3$, and the inequality can be strict. Thus, sometimes linearity is binding information, sometimes not.

Overview

- Another interesting class of admissible function-measure pairs arises in the case of partially observed smooth enough functions, *e.g.*

$$\mathcal{A} = \left\{ (g,\mu) \,\middle|\, \begin{array}{c} g\colon \mathcal{X} \to \mathbb{R} \text{ has prescribed modulus of continuity,} \\ g = G_0 \text{ on } \mathcal{O} \subseteq \mathcal{X}, \text{ (some legacy data)} \\ \mu \in \mathcal{P}(\mathcal{X}),\ \mathbb{E}_\mu[\varphi_i] \leq 0 \text{ for } i = 1, \dots, n \end{array} \right\}$$

- Note that $\mathcal{O}$ need not be statistically representative.
- Simple examples of "smooth enough" modulus of continuity include Lipschitz constants or Hölder conditions.
- Interesting interactions between the measure-theoretic constraints and the metric geometry of the space $\mathcal{X}$; it is essential that any Lipschitz function on the support of a discrete measure $\mu \in \Delta_n(\mathcal{X})$ can be extended to the whole space (**McShane** (1934)).

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi & M. Ortiz.
"Optimal uncertainty quantification for legacy data observations of Lipschitz functions."
Submitted to *Mathematical Modelling and Numerical Analysis*. arXiv:1202.1928

One Random Parameter, One Data Point

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

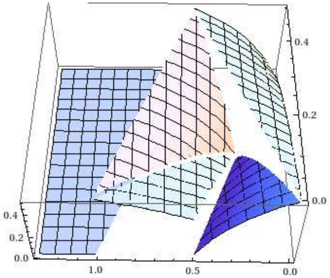


Figure: Surface plot of the least upper bound on the probability of failure, as a function of the observed data point $(z, G(z))$.

3-Parameter Hypervelocity Impact 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):

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

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

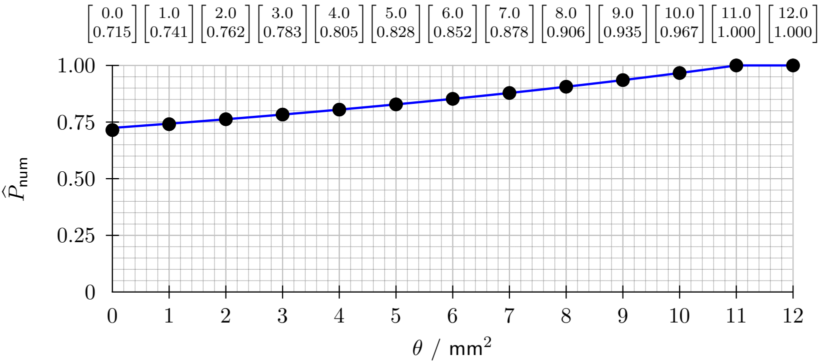


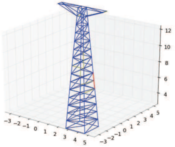
Figure: Maximum probability that perforation area is $\leq \theta$, for various θ, with the data and assumptions of the previous slide, including mean perforation area $\mathbb{E}[G(h, \alpha, v)] \geq m := 11.0\,\mathrm{mm}^2$. For $\theta \geq 2\,\mathrm{mm}^2$, the results are within 10^{-6} of Markov's bound, which indicates that 2 binding data points are those that constrain the maximum of the response function; the other 30 are non-binding.

Overview

Seismic Safety Certification

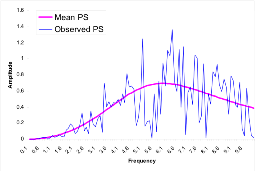
⟶ **H. Owhadi** @ 18:00, Wed. 4 Apr., MS61, in State B

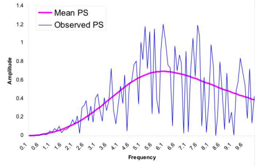
- Consider the survivability of a truss structure under an random earthquake of known intensity drawn from an incompletely specified probability distribution.



- Consider a random ground motion u, with the constraint that the mean power spectrum is the Matsuda–Asano shape function s_{MA}:

$$\mathbb{E}_{u\sim\mu}\left[|\hat{u}(\omega)|^2\right] = s_{\mathsf{MA}}(\omega) \propto \frac{\omega_{\mathrm{g}}^2 \omega^2 e^{M_{\mathrm{L}}}}{(\omega_{\mathrm{g}}^2 - \omega^2)^2 + 4\xi_{\mathrm{g}}^2\omega_{\mathrm{g}}^2\omega^2}.$$





- Such shape functions are a common tool in the seismological community, but usually u is generated by filtering white noise through s.
- Further development of this approach with **S. Mitchell** and the group of **S. Krishnan**.

Numerical Results: Vulnerability Curves

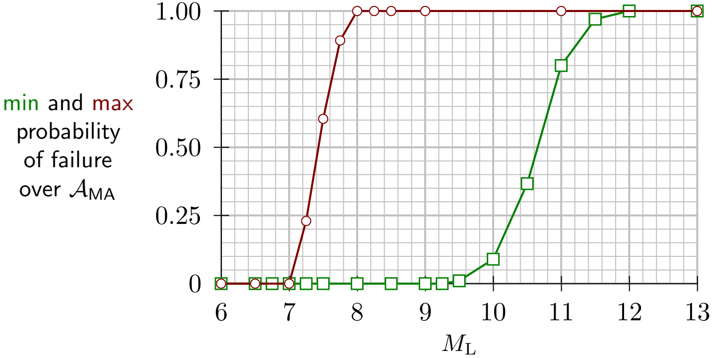


Figure: The minimum and maximum probability of failure as a function of Richter magnitude M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required O(1 day) on 44+44 AMD Opterons (*shc* and *foxtrot* at Caltech). The forward model used 200 Fourier modes for a 3-dimensional ground motion u.

Numerical Results: Vulnerability Curves

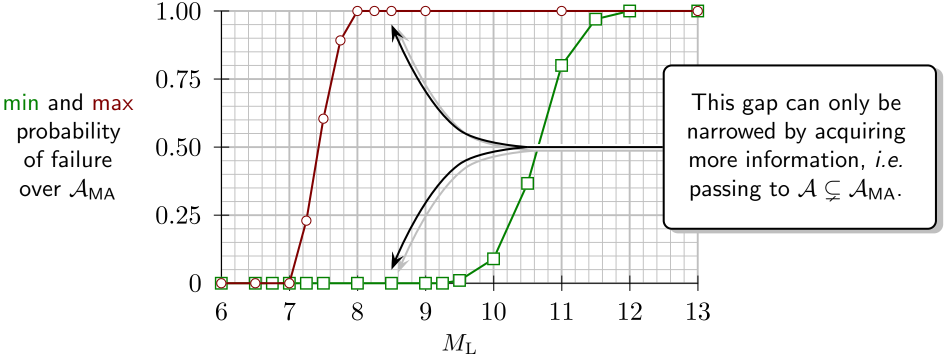


Figure: The minimum and maximum probability of failure as a function of Richter magnitude M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required O(1 day) on 44+44 AMD Opterons (*shc* and *foxtrot* at Caltech). The forward model used 200 Fourier modes for a 3-dimensional ground motion u.

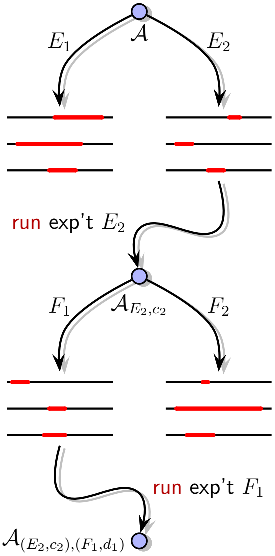
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



Overview

Closing Remarks

- By posing UQ as an optimization problem we
  - place the available information ($\cong$ constraints) about the input uncertainties at the centre of the problem;
  - obtain optimal bounds on output uncertainties with respect to that information;
  - get natural notions of information content in optimization-theoretic terms *re* constraints: active/inactive, binding/non-binding, ...
- We have theoretical and numerical examples in hand showing these phenomena at work.
- We also have 100s-dimensional "real world" examples: see **H. Owhadi**'s talk tomorrow for more details.
- Also a huge number of open questions, especially concerning the inclusion of random sample data, algorithmic properties of OUQ, ...

open-source optimization framework
mystic: A Simple Model-Independent Inversion Framework
dev.danse.us/trac/mystic