Optimal Uncertainty Quantification

Healthy Conservatism in the Face of Epistemic–Aleatoric Uncertainty, and Steps Towards the Computation of Optimal Statistical Estimators

Tim Sullivan

University of Warwick

Stochastic and Statistical Models at the Interface of Modern Industry and Mathematical Sciences

Isaac Newton Institute, Cambridge, U.K

27 March 2013

Credits

Joint work with

- Michael McKerns, Lan Huong Nguyen, Michael Ortiz, and Houman Owhadi (Caltech)
- Clint Scovel (ex-Los Alamos National Laboratory, now Caltech)
- Dominik Meyer (ex-T. U. München)
- Lauren Rast and Vinod Tewary (NIST)
- Florian Theil (Warwick)

Portions of this work were supported by

- the U. S. Department of Energy NNSA under Award No. 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; and
- the Air Force Office of Scientific Research under Grant No. FA9550-12-1-0389.

CALTECH PSAAP

National Nuclear Security Administration

Overview

Overview

Introduction

applied mathematics

probability & statistics

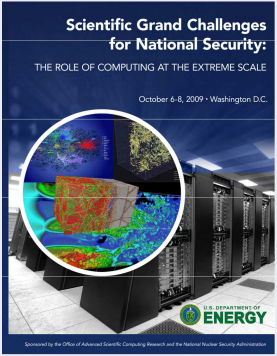
uncertainty quantification

scientific computation

applied sciences

- "UQ is the end-to-end study of the reliability of scientific inferences."
- UQ is naturally about information flow.
- Ideally, the computed relationships between pieces of information should be as sharp as possible.

Grand Challenges



multiphysics modelling
nuclear physics
materials science
chemistry
science of nonproliferation
uncertainty quantification

Prototypical UQ Problem: Reliability Certification

- $G_0 : \mathcal{X} \to \mathcal{Y}$ 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

$$\mathbb{P}_{\mu_0}\big[G_0(X) \in \mathcal{F}\big] \equiv \underbrace{\mathbb{E}_{\mu_0}\big[\mathbb{1}\big[G_0(X) \in \mathcal{F}\big]\big]}_{\substack{\text{“just” an integral} \\ \text{to be evaluated} \\ \text{— directly?} \\ \text{— by MC?} \\ \text{— by gPC?}}},$$

  or at least to know that it is acceptably small (or unacceptably large!).
- **Problem:** In practical applications, one does not know the Universe's G_0 and μ_0 exactly!

Other Quantities of Interest

- For some quantity of interest (measurable function) $q \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, we want to know

$$\underbrace{\mathbb{E}_{\mu_0}\big[q(X, G_0(X))\big]}_{\substack{\text{"just" an integral} \\ \text{to be evaluated} \\ \text{— directly?} \\ \text{— by MC?} \\ \text{— by gPC?}}},$$

or at least to know that it is acceptably small (or unacceptably large!).

- For example:
  - failure probability: $q(x, y) = \mathbb{1}[y \in \mathcal{F}]$,
  - mean performance: $q(x, y) = y$,
  - variance about a nominal output value: $q(x, y) = |y - y_0|^2$.
- Our interest lies in understanding $\mathbb{E}_{\mu_0}\big[q(X, G_0(X))\big]$ when G_0 and μ_0 are only imperfectly known (i.e. epistemic uncertainty), and to obtain bounds that are optimal with respect to the known information.

Overview

Optimal UQ

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

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

- 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).
  - Particularly in high-consequence settings, it makes sense to adopt a posture of healthy conservatism and determine the best and worst outcomes consistent with the information encoded in $\mathcal{A}$.
- Dialogue between UQ practitioners and the domain experts is essential in formulating — and revising — $\mathcal{A}$.

Optimal UQ

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

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

$$\underline{Q}(\mathcal{A}) \leq \mathbb{E}_{\mu_0}[q(X, G_0(X))] \leq \overline{Q}(\mathcal{A}),$$

  where $\underline{Q}(\mathcal{A})$ and $\overline{Q}(\mathcal{A})$ are defined by the optimization problems

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

$$\overline{Q}(\mathcal{A}) := \sup_{(g,\mu)\in\mathcal{A}} \mathbb{E}_{\mu}[q(X, g(X))].$$

- Cf. generalized Chebyshev inequalities in decision analysis (**Smith** (1995)), imprecise probability (**Boole** (1854)), distributionally robust optimization, robust Bayesian inference (surv. **Berger** (1984)).

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 $\mathcal{A}$.

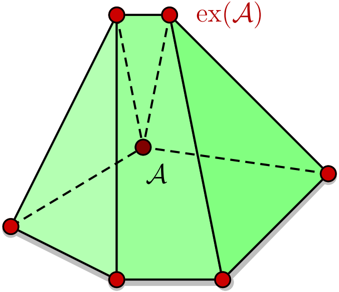


Figure : A linear functional on a convex domain in $\mathbb{R}^n$ finds its extreme value at the extremal points of the domain; similarly, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

*But see e.g. **Bertsimas & Popescu** (2005) and **Smith** (1995) for convex special cases.

Reduction of OUQ Problems — Heuristic

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 $\mathcal{X}_k$.

- To make this heuristic rigorous, we restrict attention to Radon spaces, "nice" spaces on which every Borel probability measure is inner regular. (Polish $\implies$ Radon)
- Our theorem builds on now-classical results by **von Weizsäcker & Winkler** (1980) and **Winkler** (1988) characterizing the extremal measures in moment classes, and "nice" linear/affine functionals on such classes.
- Important point: the extremal measures of a moment class

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

  are the discrete measures that have support on at most $n + 1$ distinct points of $\mathcal{X}$, which we denote by $\Delta_n(\mathcal{X})$.

Reduction of OUQ Problems — Theorem

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 $\mathcal{X}_k$.

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

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

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

Then $$\underline{Q}(\mathcal{A}) = \underline{Q}(\mathcal{A}_\Delta) \text{ and } \overline{Q}(\mathcal{A}) = \overline{Q}(\mathcal{A}_\Delta).$$

Reduction of OUQ Problems — Consequence

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 $\mathcal{X}_k$.

- Computation of the OUQ bounds $\underline{Q}(\mathcal{A})$ and $\overline{Q}(\mathcal{A})$ is equivalent to finite-dimensional problems in which the optimization variables are
  - the positions of the support points $\boldsymbol{x_i} \in \mathcal{X}$ of the discrete measure μ;
  - the weights $w_{\boldsymbol{i}} \in [0, 1]$ of the points $\boldsymbol{x_i}$; and
  - the response values $y_{\boldsymbol{i}} \in \mathcal{Y}$ corresponding to $g(\boldsymbol{x_i})$.

  with objective function

$$\sum_{\boldsymbol{i}=(0,\dots,0)}^{(N_1,\dots,N_K)} w_{\boldsymbol{i}}\, q(\boldsymbol{x_i}, y_{\boldsymbol{i}})$$

  and similar finite sums for the constraints.
- $\implies$ Implementation in the general-purpose open-source Mystic optimization framework, written in Python.

Optimal Concentration Inequalities

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

Example: Chebyshev's Inequality in OUQ Form

$$\mathcal{A}_{\mathsf{Ch}} := \left\{ \mu \in \mathcal{P}(\mathbb{R}) \,\middle|\, \mathbb{E}_\mu[X] = 0 \text{ and } \mathbb{E}_\mu[X^2] \leq \sigma^2 \right\}$$

$$\overline{P}(\mathcal{A}_{\mathsf{Ch}}) := \sup_{\mu \in \mathcal{A}_{\mathsf{Ch}}} \mathbb{P}_\mu[|X| \geq t] = \frac{\sigma^2}{t^2}.$$

Optimal Concentration Inequalities

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

Example: Chebyshev's Inequality in OUQ Form

$$\mathcal{A}_{\mathsf{Ch}} := \left\{ \mu \in \mathcal{P}(\mathbb{R}) \,\middle|\, \mathbb{E}_\mu[X] = 0 \text{ and } \mathbb{E}_\mu[X^2] \leq \sigma^2 \right\}$$

$$\overline{P}(\mathcal{A}_{\mathsf{Ch}}) := \sup_{\mu \in \mathcal{A}_{\mathsf{Ch}}} \mathbb{P}_\mu[|X| \geq t] = \frac{\sigma^2}{t^2}.$$

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.

McDiarmid's Inequality

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

with componentwise oscillations/global sensitivities defined by

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

Theorem (McDiarmid's Inequality, 1988)

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

Optimal McDiarmid and Screening Effects

Theorem (Optimal McDiarmid for $K = 1, 2$)

For $K = 1$,

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

For $K = 2$,

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

In the highlighted case, $\min\{D_1, D_2\}$ carries no information — not in the sense of 0 bits, but the sense of being a non-binding constraint.

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

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

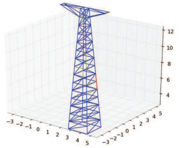
- Interestingly, $\overline{P}(\mathcal{A}_{\mathsf{Hfd}}) = \overline{P}(\mathcal{A}_{\mathsf{McD}})$ for $K = 1$ and $K = 2$, but $\overline{P}(\mathcal{A}_{\mathsf{Hfd}}) \leq \overline{P}(\mathcal{A}_{\mathsf{McD}})$ for $K = 3$, and the inequality can be strict. Thus, sometimes linearity is binding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

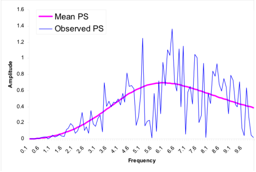
Seismic Safety Certification

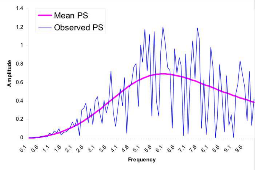
- Consider the survivability of a truss structure under an random earthquake of known intensity drawn from an incompletely specified probability distribution.
- Consider a random ground motion u, with the constraint that the mean power spectrum is the Matsuda–Asano shape function s_{MA}:

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

- Such shape functions are a common tool in the seismological community, but usually u is generated by filtering white noise through s.
- We used 200 3d Fourier modes, leading to a 1200-dimensional OUQ problem.







Numerical Vulnerability Curves (CDF Envelopes)

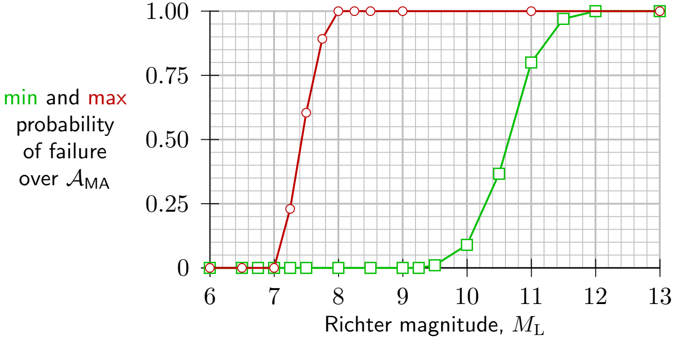


Figure : The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required $\approx$ 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 Vulnerability Curves (CDF Envelopes)

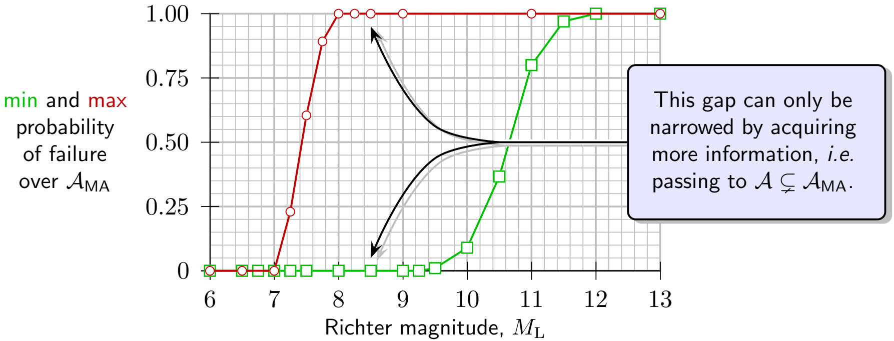


Figure : The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required $\approx$ 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 Vulnerability Curves (CDF Envelopes)

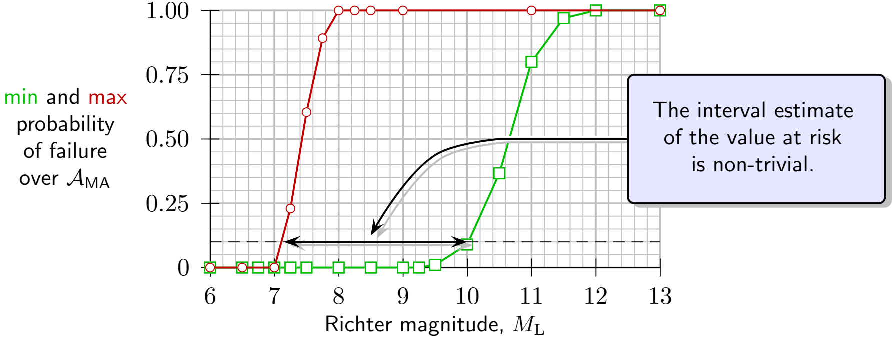
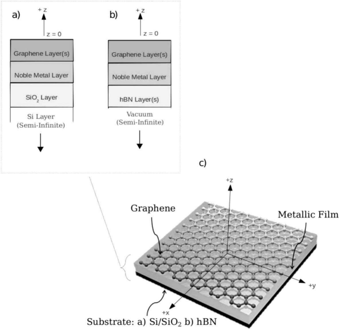


Figure : The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980 Livermore earthquake. Each data point required ≈ 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.

Other Completed or In-Progress Applications

- Hypervelocity impact (e.g. micrometeorites) given legacy data.
- Hypervelocity impact with a detailed multi-physics mechanical model.
- Optimal control of magnetically induced localized hyperthermia for the non-invasive treatment of brain tumours.
- Design of graphene + noble metal sandwich structures for light weight and low loss plasmonics applications.



Skip Details

OUQ with Legacy Data

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

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

- Note that $\mathcal{O}$ need not be statistically representative.
- Simple examples of "smooth enough" modulus of continuity include Lipschitz constants or Hölder conditions.
- Mathematically interesting interactions between the measure-theoretic constraints and the metric geometry of the space $\mathcal{X}$, e.g. the fact that any Lipschitz function on the support of a discrete measure $\mu \in \Delta_n(\mathcal{X})$ can be extended to the whole space without changing the Lipschitz constant (**McShane** (1934)).

One Random Parameter, One Data Point

- The case of a single observation in 1d can be solved explicitly.
- Suppose that you have one observation $(z, G_0(z)) \in [0, \frac{1}{2}] \times \mathbb{R}$ of a function $G_0 \colon [0,1] \to \mathbb{R}$ with Lipschitz constant $L \geq 0$.
- Explicit piecewise and discontinuous least upper bound on $\mathbb{P}_{\mu_0}[G_0(X) \leq 0]$ given L, $(z, G_0(z))$, and that $\mathbb{E}_{\mu_0}[G_0(X)] \geq m$:

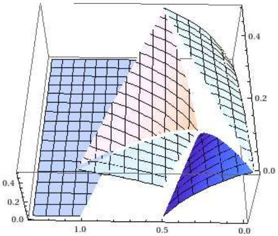


Figure : Surface plot of the least upper bound $\overline{P}$ on $\mathbb{P}_{\mu_0}[G_0(X) \leq 0]$, as a function of the observed data point $(z, G_0(z))$.

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 $G_0(h, \alpha, v) =$ 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 $[G_0(h, \alpha, v) \leq \theta]$, for various values of θ.
- Constrain the mean perf. area: $\mathbb{E}_{\mu_0}[G_0(h, \alpha, v)] \geq m := 11.0\,\text{mm}^2$.
- Modified Lipschitz constraint (multi-valued data):

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

$$|y - y'| \leq \sum_{k=1}^{3} L_k |x_k - x'_k| + 1.0\,\text{mm}^2.$$

3-Parameter Hypervelocity Impact Example: Results

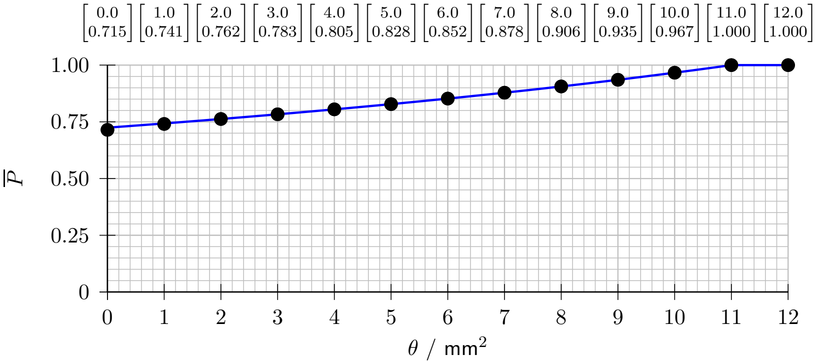


Figure : Maximum probability that perforation area is $\leq \theta$, for various θ, with the data and assumptions of the previous slide, including mean perforation area $\mathbb{E}[G_0(h, \alpha, v)] \geq 11.0\,\text{mm}^2$. For $\theta \geq 2\,\text{mm}^2$, the results are within 10^{-6} of Markov's bound, which indicates that 2 binding data points are those that constrain the maximum of the response function; the other 30 are non-binding.

Models and Neighbourhoods

- One can consider feasible sets in which the constraints on g are of the form $d(g, F) \leq C$ for some model function F.

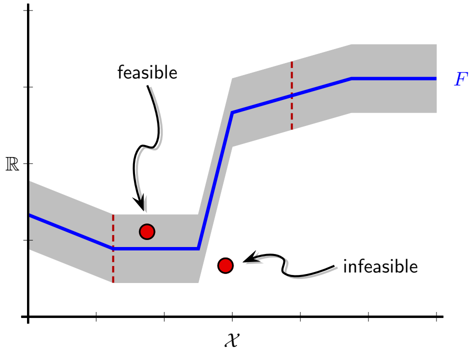


Figure : Assuming that reality G_0 is uniformly close to the model F means assuming that the model has approximately the right cliffs in exactly the right places; Hausdorff (graphical) closeness is a much looser assumption.

Models and Neighbourhoods

- One can consider feasible sets in which the constraints on g are of the form $d(g, F) \leq C$ for some model function F.

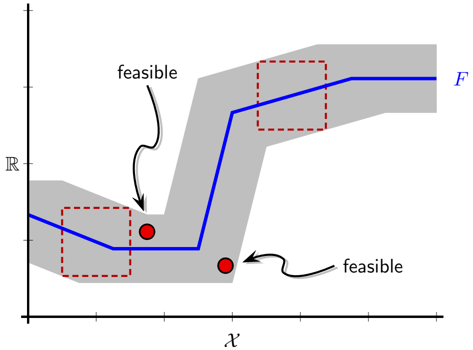


Figure : Assuming that reality G_0 is uniformly close to the model F means assuming that the model has approximately the right cliffs in exactly the right places; Hausdorff (graphical) closeness is a much looser assumption.

Dimensional Collapse and Acceleration

- Often, the solutions of OUQ problems have lower dimension than the reduction theorems might suggest.
- As in the earlier McDiarmid example, the structure of the solutions indicates the “key players” in the UQ problem.

Dimensional Collapse and Acceleration

- Often, the solutions of OUQ problems have lower dimension than the reduction theorems might suggest.
- As in the earlier McDiarmid example, the structure of the solutions indicates the “key players” in the UQ problem.
- For example, the product probability measure on (h, α, v) that maximizes the probability of non-perforation in the previous impact example, given the mean perforation area (i.e. 1 constraint), has support on $2 \times 1 \times 1$ points, not all the available $2 \times 2 \times 2$ points.
- In the course of the calculation, observe two kinds of “collapses”:
  - support points collide (distance between them tends to zero);
  - probability masses of support points decay to zero.
- CLPS is a module for the implementation of OUQ in Mystic that
  - numerically detects these phenomena at runtime;
  - pauses the optimizer and returns collapse metadata;
  - restarts the calculation with the observed collapses as new constraints $\Longrightarrow$ faster exploration of a lower-dimensional search space.

Numerical Effects of Dimensional Collapse

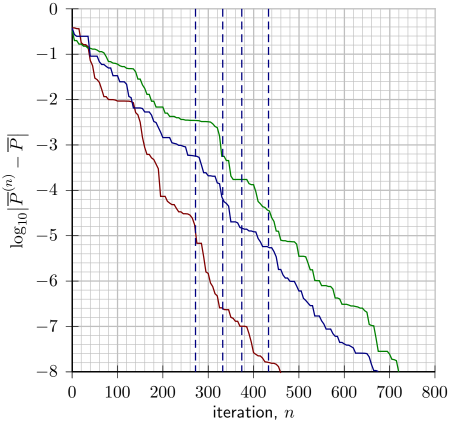


Figure : Semi-log plot showing typical numerical convergence of the impact OUQ problem with dimensionality $2 \times 2 \times 2$ both without (green) and with (blue) CLPS features. The vertical dashed blue lines indicate the occurrence of collapse events. For comparison, the solid red line shows the numerical convergence of a typical run with dimensionality $2 \times 1 \times 1$.

Numerical Effects of Dimensional Collapse

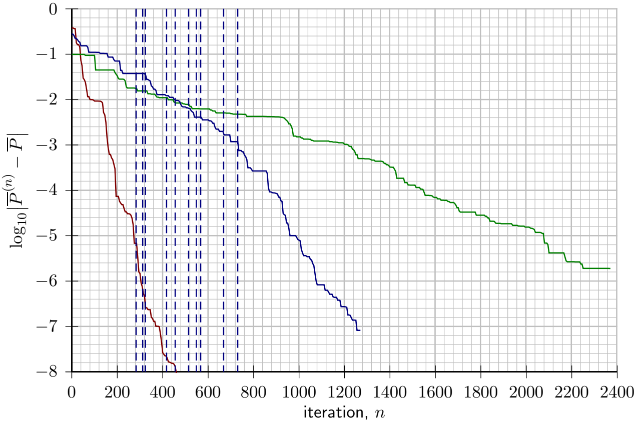


Figure : Semi-log plot showing typical numerical convergence of the impact OUQ problem with dimensionality $4 \times 4 \times 4$ both without (green) and with (blue) CLPS features. The vertical dashed blue lines indicate the occurrence of collapse events. For comparison, the solid red line shows the numerical convergence of a typical run with dimensionality $2 \times 1 \times 1$.

Overview

Optimal Knowledge Acquisition

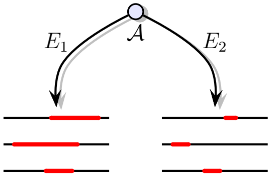
- Range of prediction given $\mathcal{A}$:

$$\mathcal{R}(\mathcal{A}) := \overline{Q}(\mathcal{A}) - \underline{Q}(\mathcal{A}),$$

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

- 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:

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



Optimal Knowledge Acquisition

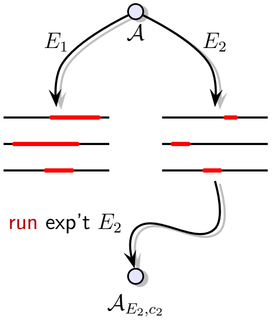
- Range of prediction given $\mathcal{A}$:

$$\mathcal{R}(\mathcal{A}) := \overline{Q}(\mathcal{A}) - \underline{Q}(\mathcal{A}),$$

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

- 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:

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



Optimal Knowledge Acquisition

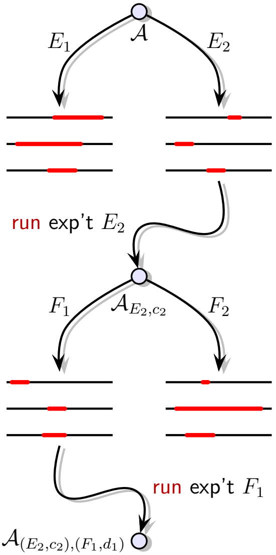
- Range of prediction given $\mathcal{A}$:

$$\mathcal{R}(\mathcal{A}) := \overline{Q}(\mathcal{A}) - \underline{Q}(\mathcal{A}),$$

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

- 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:

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



Optimal Knowledge Acquisition

- The "experiments" E_i of the previous slide could be
  - actual physical experiments on the full system of interest;
  - partial or subsystem experiments;
  - simulations of same.
- Thus, OUQ offers a systematic application of the scientific method to drive experimental and computational campaigns in an optimal goal-oriented fashion.
- Like a good chess player, one could even plan many moves ahead, i.e. plan an optimal experimental campaign — or discover that the experiments are not worth doing at all!
- What are the fundamental properties of this kind of "UQ game"? — **Open question**

Optimal Statistical Estimators

- The natural next step for OUQ is to extend it to make optimal use of random sample data.
- Suppose that you are given some samples $\xi_1, \dots, \xi_n$ of a random variable Ξ and have to use them to estimate some other quantity $Q(\Xi)$, e.g. to fit the coefficients of a model, or to make a prediction.

Paradigm I   Prove a General(ish) Theorem

One can spend a lot of time and effort designing a good statistical estimator or test, and proving its properties, e.g. χ^2 test, BLUE, ...

Paradigm II   Compute for the Circumstances

Compute the optimal statistical estimator for your problem, a schema-specific computed formula into which to plug $\xi_1, \dots, \xi_n$.

Analogy with Early Scientific Computing

- Similarities between developments in the UQ community now and the development of scientific computing in the era of von Neumann &al.
- Transition from "compute a function for general application" to "compute for the specific application".

	Paradigm I	**Paradigm II**
PDEs	Compute tables for special functions, and couple them with PDE ansätze	Discretize the PDE and compute directly using FE, FD, ...
E.g. **McD**	McDiarmid's inequality $\overline{P} \leq e^{-2m^2 / \sum_i D_i^2}$	Optimal McDiarmid-type inequality, $\overline{P}(\mathcal{A}_{\mathrm{McD}})$
UQ/Stats	Compute tables for statistics and plug them into (theorem-derived) estimators	Computation of Optimal Statistical Estimators?

Overview

Conclusions

- 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 w.r.t. that information;
  - get natural notions of information content in optimization-theoretic terms about constraints: active/inactive, binding/non-binding, . . .
- We have theoretical (closed-form pen-and-paper) and real (high-dimensional engineering systems) examples in hand showing these phenomena at work.
- Growing computational resources make large OUQ-type problems increasingly tractable, cf. Bayesian methods in 20th Century.
- Many open questions, especially concerning the inclusion of random sample data, algorithmic properties of OUQ, &c.
- **Interesting times for UQ.** The community is on the verge of transforming UQ/statistical practice much as happened with PDEs post-WWII.

References

- Publications

  - M. Adams, M. Aivazis, L. H. Nguyen, B. Li, P.-H. T. Kamga, M. McKerns, J. Mihaly, M. Ortiz, H. Owhadi, A. J. Rosakis, T. J. Sullivan & J. Tandy. "Optimal uncertainty quantification of hypervelocity impact." In preparation.
  - L. H. Nguyen, T. J. Sullivan, M. McKerns & H. Owhadi. "Dimensional reduction and acceleration of optimal distributionally-robust uncertainty quantification calculations." In preparation.
  - H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns & M. Ortiz. "Optimal Uncertainty Quantification." To appear in *SIAM Review* $\approx$ Summer 2013. arXiv:1009.0679
  - L. Rast, T. J. Sullivan & V. K. Tewary. "Stratified graphene-noble metal systems for low-loss plasmonics applications." To appear in *Phys. Rev. B.* arXiv:1301.5620
  - 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 *Math. Mod. Num. Anal.* arXiv:1202.1928

- Software
  - Mystic (optimization framework): http://dev.danse.us/trac/mystic
  - Pathos (distributed computing): http://dev.danse.us/trac/pathos