Optimal Uncertainty Quantification

Challenges and Lessons Learned

Tim Sullivan, Mike McKerns, Michael Ortiz, Houman Owhadi
tjs@..., mmckerns@..., ortiz@..., owhadi@caltech.edu
& Clint Scovel
jcs@lanl.gov

California Institute of Technology & Los Alamos National Laboratory

ASME 2010 International Mechanical Engineering Congress & Exposition
Vancuover, British Columbia, Canada
17 November 2010

CALTECH PSAAP

Outline

What is Uncertainty Quantification?

?

What is Uncertainty Quantification?

The Many Faces of Uncertainty Quantification

- *The Elephant in the Room:* there is a growing consensus that UQ is an essential component of objective science.
- *The Blind Men and the Elephant:* unfortunately, as it stands at the moment, UQ has all the hallmarks of an ill-posed problem.

Problems

- Certification
- Extrapolation/Prediction
- Reliability Estimation
- Sensitivity Analysis
- Verification
- Validation
- ...

Methods

- Analysis of Variance
- Bayesian Methods
- Error Bars
- Latin Hypercube Sampling
- (Quasi) Monte Carlo
- Stochastic Collocation
- ...

Optimal Uncertainty Quantification

- We propose a mathematical framework for UQ as an optimization problem, which we call Optimal Uncertainty Quantification (OUQ).
- H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns & M. Ortiz. **"Optimal Uncertainty Quantification"** (2010) Submitted to *SIAM Review*. Preprint at

  http://arxiv.org/pdf/1009.0679v1

- The development and application of OUQ to real, complex problems is a collaborative interdisciplinary effort that requires expertise in
  - applied mathematics, especially probability theory,
  - numerical optimization,
  - (massively) parallel computing,
  - the application area (*e.g.* biology, chemistry, economics, engineering, geoscience, meteorology, physics, . . .).
- This talk will focus on OUQ in the prototypical context of certification.

The Certification Problem

- Suppose that you are interested in a system of interest, $G\colon \mathcal{X} \to \mathbb{R}$, which is a real-valued function of some random inputs $X \in \mathcal{X}$ with probability distribution $\mathbb{P}$ on $\mathcal{X}$.
- Some value $\theta \in \mathbb{R}$ is a *performance threshold*: if $G(X) \leq \theta$, then the system fails; if $G(X) > \theta$, then the system succeeds.
- You want to know the probability of failure

$$\mathbb{P}[G(X) \leq \theta],$$

  or at least to know if it exceeds some maximum acceptable probability of failure ϵ — but you do not know G and $\mathbb{P}$!
- If you have some information about G and $\mathbb{P}$, what are the best rigorous lower and upper bounds that you can give on the probability of failure using that information?
- Optimality is important — the following bounds are true, but useless:

$$0 \leq \mathbb{P}[G(X) \leq \theta] \leq 1.$$

The Importance of Being Optimal: Being overly conservative may lead to economic losses, but being overly optimistic may lead to loss of life, environmental damage *& c.*

Space Shuttle *Columbia*, 1 February 2003

Eyjafjallajökull, Iceland, 27 March 2010

Deepwater Horizon, 21 April 2010

Uncertainty Quantification Methods

So, how can one bound $\mathbb{P}[G(X) \leq \theta]$ given only limited knowledge of G and $\mathbb{P}$?

- Monte Carlo? *Needs many independent $\mathbb{P}$-distributed samples.*
- Stochastic collocation methods? *Fine for SPDEs, if we have a good representation for the randomness and rapid decay of the spectrum.*
- Bayesian inference? *Danger of error propagation in the case of rare events.*
- ...

Each of these methods relies, implicitly or explicitly, on the validity of certain assumptions in order to be applicable or efficient. Those assumptions may not match the information about G and $\mathbb{P}$ — so we place the information about G and $\mathbb{P}$ at the centre of the problem.

What Problem Should You Solve?

Or: What are You Computing, and Why?

- You want to know about the probability of failure
$$\mathbb{P}[G(X) \leq \theta]?$$
- You want to know if it's greater than or less than ϵ?
- And you want to do this without ignoring or distorting your existing information set, nor making additional assumptions?
- If you had access to The Ultimate Computer, what problem would you try to solve?
- Worry about computational feasibility later!

Forty-Two?!

Information / Assumptions

- Write down all the information that you have about the system. For example, this information might come from
  - physical laws;
  - expert opinion;
  - experimental data.
- Let $\mathcal{A}$ denote the set of all pairs (f, μ) that are consistent with your information about $(G, \mathbb{P})$:

$$\mathcal{A} \subseteq \left\{ (f, \mu) \,\middle|\, \begin{array}{c} f \colon \mathcal{X} \to \mathbb{R} \text{ is measurable, and} \\ \mu \text{ is a probability measure on } \mathcal{X} \end{array} \right\}.$$

- All you know about reality is that $(G, \mathbb{P}) \in \mathcal{A}$; any $(f, \mu) \in \mathcal{A}$ is an admissible scenario for the unknown reality $(G, \mathbb{P})$.

The Optimal UQ Problem

With this notation, the Optimal UQ Problem is simply to find the greatest lower bound and least upper bound on the probability of failure among all admissible scenarios $(f, \mu) \in \mathcal{A}$. That is, we want to calculate

$$\mathcal{L}(\mathcal{A}) := \inf_{(f,\mu)\in\mathcal{A}} \mu[f \leq \theta]$$

and

$$\mathcal{U}(\mathcal{A}) := \sup_{(f,\mu)\in\mathcal{A}} \mu[f \leq \theta].$$

Any bounds other than these would either be not sharp or not conservative.

Rigorous and Optimal Certification Criteria

Given a maximum acceptable probability of failure $\epsilon \in [0, 1]$, calculation of $\mathcal{L}(\mathcal{A})$ and $\mathcal{U}(\mathcal{A})$ yields unambiguous, rigorous and optimal criteria for certification of the system:

- if $\mathcal{U}(\mathcal{A}) \leq \epsilon$, then the system is safe even in the worst possible case;
- if $\mathcal{L}(\mathcal{A}) > \epsilon$, then the system is unsafe even in the best possible case;
- if $\mathcal{L}(\mathcal{A}) \leq \epsilon < \mathcal{U}(\mathcal{A})$, then there are some admissible scenarios under which the system is safe and others under which it is unsafe: the information encoded in $\mathcal{A}$ is insufficient to rigorously certify the system, and more information must be sought. The system is (temporarily) deemed unsafe due to lack of information.

Reduction of OUQ

- OUQ problems are global, infinite-dimensional, non-convex, highly-constrained (*i.e.* nasty!) optimization problems.
- The non-convexity is a fact of life, but there are powerful reduction theorems that allow a reduction to a search space of very low-dimension.
- Instead of searching over all admissible probability measures μ, we need only to search over those with a very simple "extremal" structure: in the simplest case, these are just finite sums of point masses (Dirac measures) on the input parameter space $\mathcal{X}$.
- That is, we can "pretend" that all the random inputs are discrete random variables and just optimize over the possible values and probabilities that those discrete variables might take.

Reduction of OUQ — Linear Inequalities on Moments

Suppose that the admissible set $\mathcal{A}$ has the following form: all the constraints on the measure μ are linear inequalities on generalized moments. That is, for some given functions $g'_1, \dots, g'_{n'} \colon \mathcal{X} \to \mathbb{R}$,

$$\mathcal{A} = \left\{ (f, \mu) \,\middle|\, \begin{array}{c} f \colon \mathcal{X} \to \mathbb{R} \text{ such that} \\ \langle \text{some conditions on } f \text{ alone} \rangle, \\ \mathbb{E}_\mu[g'_1] \leq 0, \dots, \mathbb{E}_\mu[g'_{n'}] \leq 0 \end{array} \right\}.$$

Theorem (General reduction theorem)

If $\mathcal{X}$ is a Suslin space, then $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_\Delta)$ and $\mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_\Delta)$, where

$$\mathcal{A}_\Delta = \left\{ (f, \mu) \in \mathcal{A} \,\middle|\, \begin{array}{c} \mu \textit{ is a sum of at most } n' + 1 \\ \textit{weighted Dirac measures on } \mathcal{X} \end{array} \right\}.$$

Reduction of OUQ — Independent Inputs

Similarly, if we have K independent inputs, *i.e.* $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$ and

$$\mathcal{A} = \left\{ (f, \mu) \,\middle|\, \begin{array}{c} f \colon \mathcal{X} \to \mathbb{R} \text{ such that} \\ \langle \text{some conditions on } f \text{ alone} \rangle, \\ \mu = \mu_1 \otimes \cdots \otimes \mu_K, \\ \mathbb{E}_\mu[g'_1] \leq 0, \dots, \mathbb{E}_\mu[g'_{n'}] \leq 0, \\ \mathbb{E}_{\mu_k}[g_i^k] \leq 0 \text{ for } i = 1, \dots, n_k \text{ and } k = 1, \dots, K \end{array} \right\}.$$

Theorem (Reduction for independent input parameters)

If $\mathcal{X}_1, \dots, \mathcal{X}_K$ are Suslin spaces, then $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_\Delta)$ and $\mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_\Delta)$, where

$$\mathcal{A}_\Delta = \left\{ (f, \mu) \in \mathcal{A} \,\middle|\, \begin{array}{c} \mu_k \text{ is a sum of at most } n' + n_k + 1 \\ \text{weighted Dirac measures on } \mathcal{X}_k \end{array} \right\}.$$

McDiarmid's Inequality

- McDiarmid's inequality says that if $\mathbb{E}[G(X)] \geq m$ and the maximum oscillation of G with respect to changes of its k^{th} argument is at most D_k, then

$$\mathbb{P}[G(X) \leq \theta] \leq \exp\left(-2\frac{(m-\theta)_+^2}{\sum_{k=1}^K D_k^2}\right).$$

- That is, for

$$\mathcal{A}_{\mathsf{McD}} := \left\{ (f,\mu) \,\middle|\, \begin{array}{c} f\colon \mathcal{X} \to \mathbb{R}, \\ \mathcal{D}_k[f] := \sup |f(x^1,\dots,x^k,\dots,x^K) - \\ \qquad - f(x^1,\dots,\widetilde{x}^k,\dots,x^K)| \leq D_k, \\ \mu = \mu_1 \otimes \dots \otimes \mu_K \text{ on } \mathcal{X}, \\ \mathbb{E}_\mu[f] \geq m \end{array} \right\},$$

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

Optimal McDiarmid Inequality

- The reduction theorems mentioned before, along with other reduction theorems that eliminate dependency upon the coordinate positions in the parameter space $\mathcal{X}$, yield finite-dimensional problems that can be solved exactly to give optimal concentration inequalities with the same assumptions as McDiarmid's inequality.
- Write $a := (m - \theta)_+$ for the mean performance margin.

Optimal McDiarmid, $K = 1$

$$\mathcal{U}(\mathcal{A}_{\mathsf{McD}}) = \left(1 - \frac{a}{D_1}\right)_+$$

Optimal McDiarmid Inequality

Optimal McDiarmid, $K = 2$

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

- By a combinatorial induction procedure, this can be entended to give an optimal inequality given the assumptions of McDiarmid's inequality for $K \in \mathbb{N}$ input random variables.
- Note that not all parameter sensitivities are created equal! If the "sensitivity gap" between the largest parameter sensitivity D_1 and the second-largest one D_2 is big enough, then all the output uncertainty is controlled by D_1 and the performance margin $a := (m - \theta)_+$.

Selection of the Best Next Experiment

- Suppose that you are offered a choice of running just one very expensive experiment from a collection $E_1, E_2, \dots$: each experiment E_i will measure some functional $\Phi_i(G, \mathbb{P})$ to very high accuracy. *E.g.*

$$
\begin{aligned}
\Phi_1(f,\mu) &:= \mathbb{E}_\mu[f], \\
\Phi_2(f,\mu) &:= \mu[X \in A] \text{ for some set } A \subseteq \mathcal{X}, \\
\Phi_3(f,\mu) &:= \mathcal{D}_1[f], \\
\Phi_4(f,\mu) &:= \mathbb{V}_\mu[f | X \in A].
\end{aligned}
$$

- Which experiment should you run? How can one objectively say that one experiment is "better" or "worse" than another?
- In the Optimal UQ framework, we can assess how predictive or decisive a potential experiment may be in terms of "overlap".

Selection of the Best Next Experiment

- Let $J_{\text{safe},\epsilon}(\Phi_i)$ be the closed interval in $\mathbb{R}$ spanned by the possible values of $\Phi_i(f, \mu)$ among all safe scenarios $(f, \mu) \in \mathcal{A}$, *i.e.* those with $\mu[f \leq \theta] \leq \epsilon$.
- Let $J_{\text{unsafe},\epsilon}(\Phi_i)$ be the closed interval in $\mathbb{R}$ spanned by the possible values of $\Phi_i(f, \mu)$ among all unsafe scenarios $(f, \mu) \in \mathcal{A}$, *i.e.* those with $\mu[f \leq \theta] > \epsilon$.
- Determination of these two intervals means solving four OUQ problems.
- What could you conclude if you were told $\Phi_i(G, \mathbb{P})$?

$$
\begin{aligned}
\Phi_i(G, \mathbb{P}) \in J_{\text{safe},\epsilon}(\Phi_i) \setminus J_{\text{unsafe},\epsilon}(\Phi_i) &\implies \text{system is safe,} \\
\Phi_i(G, \mathbb{P}) \in J_{\text{unsafe},\epsilon}(\Phi_i) \setminus J_{\text{safe},\epsilon}(\Phi_i) &\implies \text{system is unsafe,} \\
\Phi_i(G, \mathbb{P}) \in J_{\text{safe},\epsilon}(\Phi_i) \cap J_{\text{unsafe},\epsilon}(\Phi_i) &\implies \text{cannot decide,} \\
\Phi_i(G, \mathbb{P}) \notin J_{\text{safe},\epsilon}(\Phi_i) \cup J_{\text{unsafe},\epsilon}(\Phi_i) &\implies \text{faulty assumptions!}
\end{aligned}
$$

Selection of the Best Next Experiment

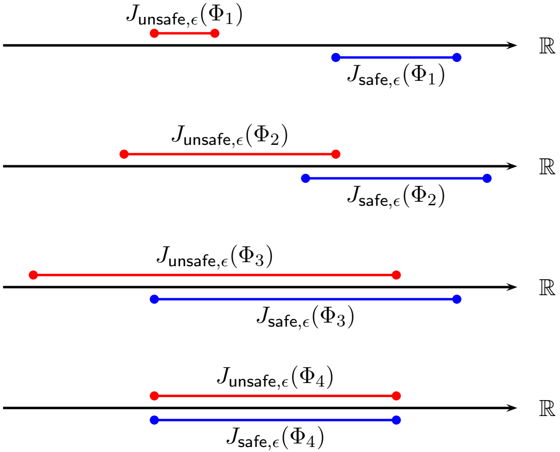


Figure: Outcome intervals for four possible experiments E_1, E_2, E_3 and E_4.

Hypervelocity Impact

Figure: Caltech's Small Particle Hypervelocity Impact Range (SPHIR): a two-stage light gas gun that launches 1–$50\,\mathrm{mg}$ projectiles at speeds of 2–$10\,\mathrm{km}\cdot\mathrm{s}^{-1}$.

Hypervelocity Impact

Figure: Caltech's Small Particle Hypervelocity Impact Range (SPHIR): a two-stage light gas gun that launches 1–$50\,\text{mg}$ projectiles at speeds of 2–$10\,\text{km}\cdot\text{s}^{-1}$.

Hypervelocity Impact: Surrogate Model

Experimentally-derived deterministic surrogate model for the perforation area (in mm^2), with three independent inputs:

- plate thickness $h \in \mathcal{X}_1 := [1.52, 2.67]\,\mathrm{mm} = [60, 105]\,\mathrm{mils}$;
- impact obliquity $\alpha \in \mathcal{X}_2 := [0, \frac{\pi}{6}]$;
- impact speed $v \in \mathcal{X}_3 := [2.1, 2.8]\,\mathrm{km} \cdot \mathrm{s}^{-1}$.

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

The quantity $v_{\mathsf{bl}}(h, \alpha)$ given by

$$v_{\mathsf{bl}}(h, \alpha) := 0.579 \left(\frac{h}{(\cos \alpha)^{0.448}} \right)^{1.400}$$

is called the ballistic limit, the impact speed below which no perforation occurs. The failure event is non-perforation, *i.e.* $[H = 0] \equiv [H \leq 0]$.

Admissible scenarios, $\mathcal{A}$	$\mathcal{U}(\mathcal{A})$	Method
$\mathcal{A}_{\text{McD}}$: independence, oscillation and mean constraints (exact response H not given)	$\leq 66.4\%$ $= 43.7\%$	McD. ineq. Opt. McD.
$\mathcal{A} := \{(f, \mu) \mid f = H \text{ and } \mathbb{E}_\mu[H] \in [5.5, 7.5]\}$	$\overset{\text{num}}{=} 37.9\%$	OUQ
$\mathcal{A} \cap \left\{(f, \mu) \,\middle\vert\, \begin{array}{c}\mu\text{-median velocity} \\ = 2.45\,\text{km} \cdot \text{s}^{-1}\end{array}\right\}$	$\overset{\text{num}}{=} 30.0\%$	OUQ
$\mathcal{A} \cap \left\{(f, \mu) \,\middle\vert\, \mu\text{-median obliquity} = \frac{\pi}{12}\right\}$	$\overset{\text{num}}{=} 36.5\%$	OUQ
$\mathcal{A} \cap \left\{(f, \mu) \,\middle\vert\, \text{obliquity} = \frac{\pi}{6}\ \mu\text{-a.s.}\right\}$	$\overset{\text{num}}{=} 28.0\%$	OUQ

Warning!

It is tempting to say that some of these bounds are “sharper” than others. Except for the first line, every one of these bounds is sharp given the available information. In the case of asymmetric information, think before condemning a bound as “not sharp”.

Numerical Convergence

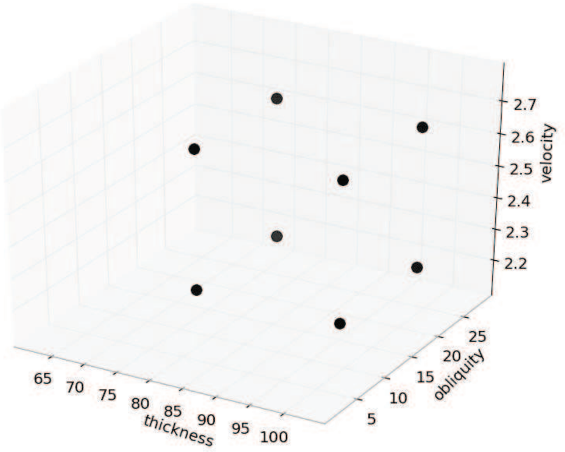


Figure: Support of the $2 \times 2 \times 2$-point measure μ at iteration 0.

Numerical Convergence

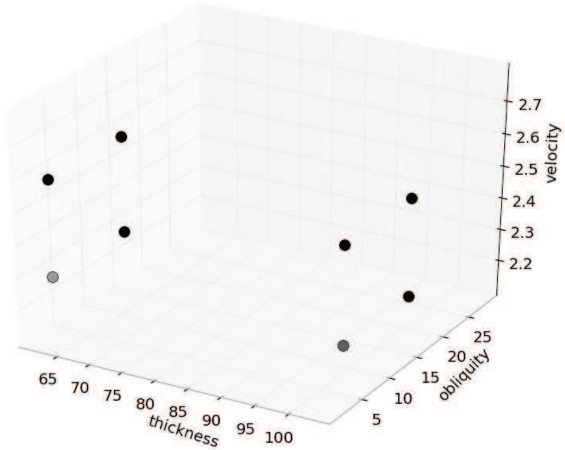


Figure: Support of the $2 \times 2 \times 2$-point measure μ at iteration 150.

Numerical Convergence

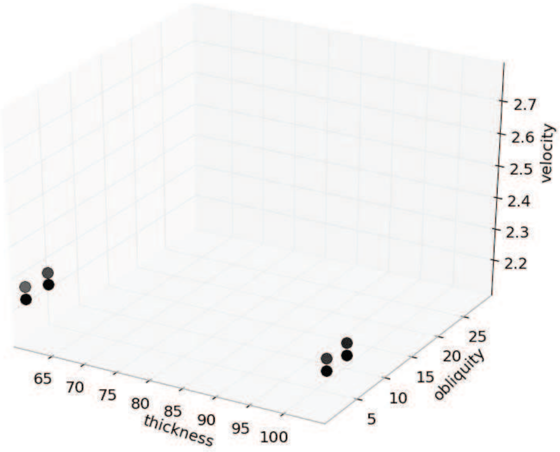


Figure: Support of the $2 \times 2 \times 2$-point measure μ at iteration 200.

Numerical Convergence

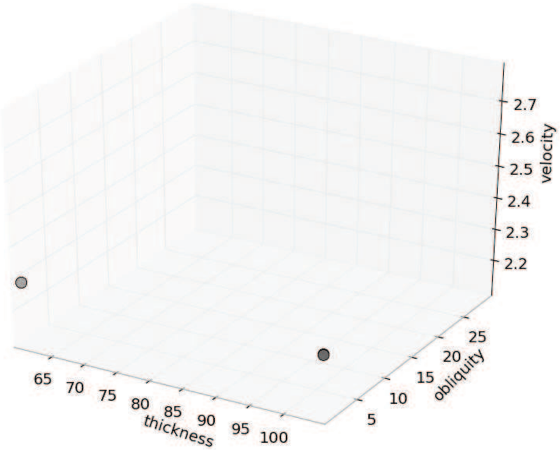
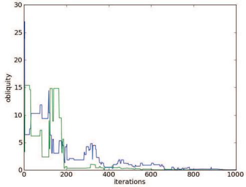
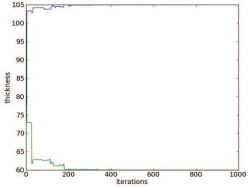


Figure: Support of the $2 \times 2 \times 2$-point measure μ at iteration 1000.

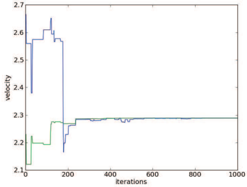
Numerical Convergence



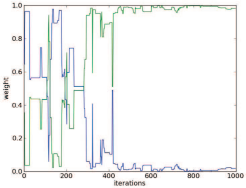
(a) oblquity positions



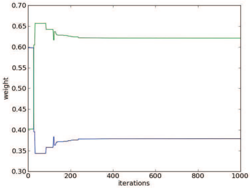
(b) thickness positions



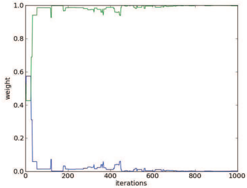
(c) velocity positions



(d) obliquity masses



(e) thickness masses



(f) velocity masses

Comments

- Over these parameter ranges, the oscillations of H are:

$$\mathcal{D}_h[H] = 8.86\,\mathrm{mm}^2, \quad \mathcal{D}_\alpha[H] = 4.17\,\mathrm{mm}^2, \quad \mathcal{D}_v[H] = 7.20\,\mathrm{mm}^2,$$

  so the "screening effects" apply in the optimal McDiarmid inequality.
- In the full OUQ analysis, the measure that maximizes the probability of failure yields important information about the "key players" in the system.
- For given mean perforation area, the worst-case probability of failure is not controlled by the impact velocity or the oblquity, but by the thickness of the plate.
- The locations of the measure μ's support points collapse to
  - the two extremes of the thickness (h) range;
  - the lower extreme of the obliquity (α) range;
  - a single non-trivial value in the velocity (v) range.

Extremizers are Attractors?

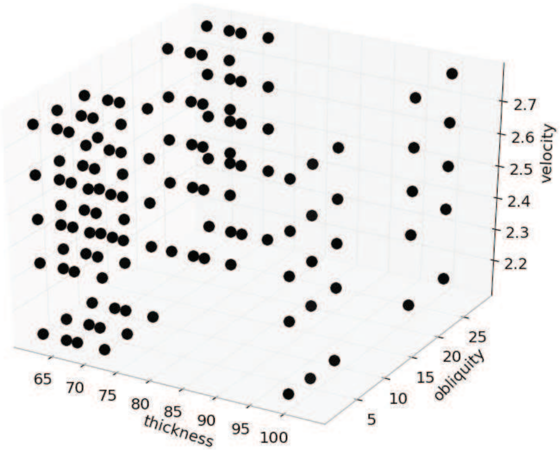


Figure: Support of the $5 \times 5 \times 5$-point measure μ at iteration 0.

Extremizers are Attractors?

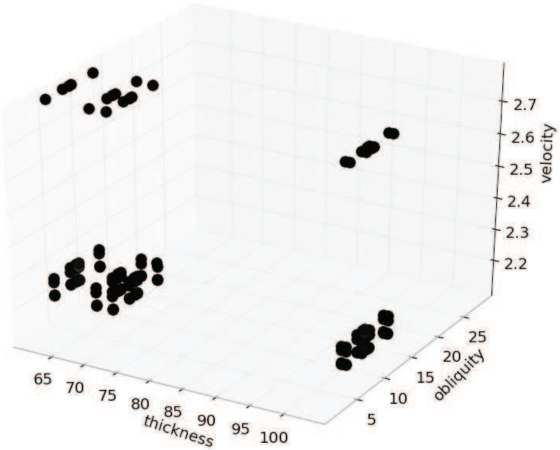


Figure: Support of the $5 \times 5 \times 5$-point measure μ at iteration 1000.

Extremizers are Attractors?

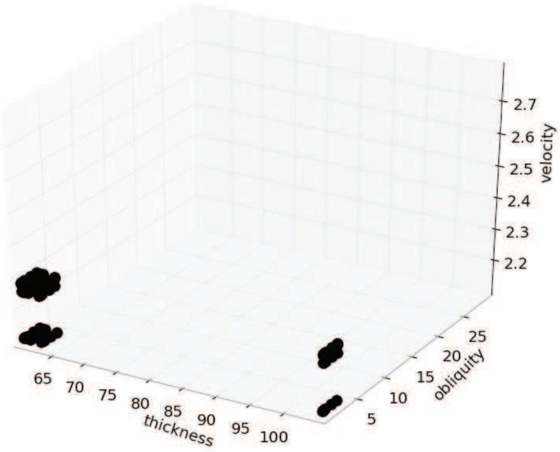


Figure: Support of the $5 \times 5 \times 5$-point measure μ at iteration 3000.

Extremizers are Attractors?

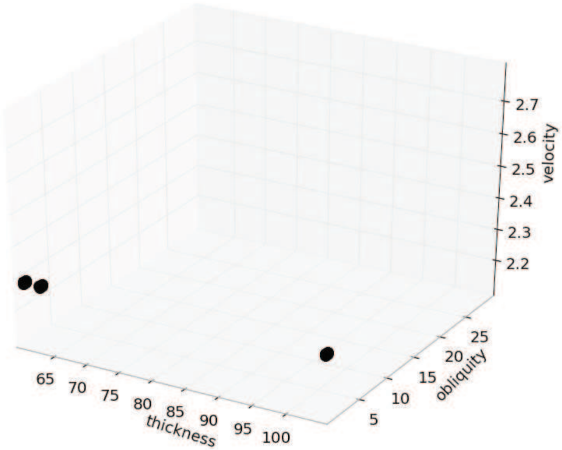


Figure: Support of the $5 \times 5 \times 5$-point measure μ at iteration 7100.

Optimal UQ with Legacy Data

- So far, we have assumed that the response function G can be exercised at will. What if this is not the case?
- What are the sharpest bounds on global parameter sensitivities or the probability of failure given only some fixed legacy data

$$(z_1, G(z_1)), \dots, (z_N, G(z_N))$$

  and some constraints on how much G can vary in between those data points?
- It is essential to have some information on how much G can vary in between the data points, *e.g.* $|\partial G / \partial x^i| \leq L_i$.
- Again, this can be posed as an OUQ problem and the reduction theorems make it a finite-dimensional (albeit non-convex) problem.

Optimal UQ with Legacy Data

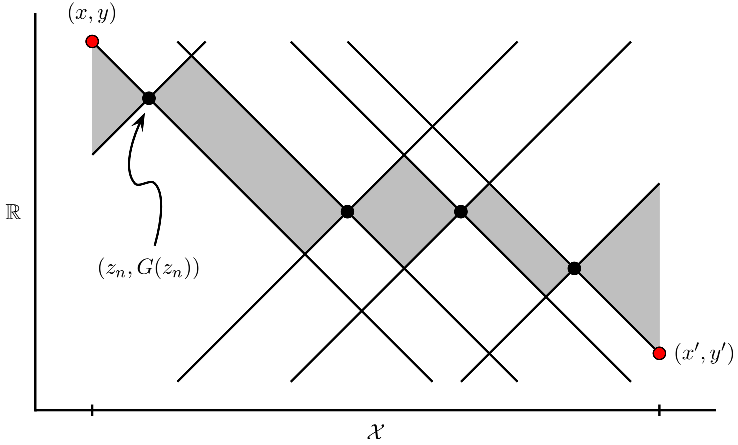


Figure: The legacy OUQ problem amounts to the optimal placement of point masses and function values in a feasible set in $\mathcal{X} \times \mathbb{R}$ that is an intersection of cones through the given data points.

Conclusions

- UQ is an essential component of modern science, with many high-consequence applications. However, there is no established consensus on how to formally pose "the UQ problem", nor a common language in which to communicate and quantitatively compare UQ methods and results.
- OUQ is an opening gambit. OUQ is not just an effort to provide answers, but an effort to well-pose the question: UQ is the challenge of optimally bounding functions of unknown responses and unknown probabilities, given some information about them.
- A key feature is that the OUQ viewpoint explicitly requires the user to explicitly state all the assumptions in operation — once listed, they can be perturbed to see if the answers are robust.
- Although the optimization problems involved are large, in many cases of interest, their dimension can be substantially reduced.

References and Acknowledgements

- H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns & M. Ortiz. **"Optimal Uncertainty Quantification"** (2010) Submitted to *SIAM Review*. Preprint at

  http://arxiv.org/pdf/1009.0679v1

- Optimization calculations performed using *mystic*:

  http://dev.danse.us/trac/mystic

- We gratefully acknowledge portions of this work 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.

  http://www.psaap.caltech.edu/