Optimal Uncertainty Quantification

Towards a Paradigm Shift in Predictive Science

Tim Sullivan

with B. Li, M. McKerns, D. Meyer, L. H. Nguyen,

M. Ortiz, H. Owhadi, C. Scovel, F. Theil

California Institute of Technology

PSAAP V&V/UQ Workshop

University of Michigan, Ann Arbor, Michigan, U.S.A.

8–10 August 2012

Outline

Portions of this work were supported by

- the U. S. Department of Energy (National Nuclear Security Administration) 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;
- the Air Force Office of Scientific Research under Grant No. FA9550-12-1-0389.

CALTECH PSAAP

National Nuclear Security Administration

Overview

Introduction

applied mathematics

probability & statistics

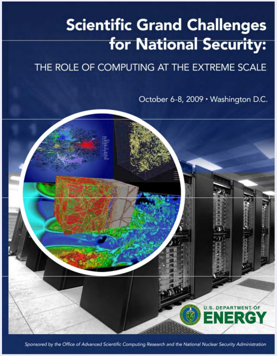
uncertainty quantification

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

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

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

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

Overview

Optimal UQ

- 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:

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

- *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) than any other.
- Dialogue between UQ practitioners and the domain experts is essential in formulating — and revising — $\mathcal{A}$.

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

Optimal UQ

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

- 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 optimization 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)).

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 $\mathcal{A}$.

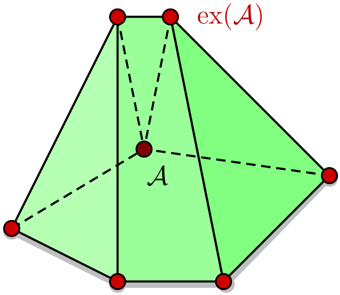


Figure: A linear functional on a convex domain in $\mathbb{R}^n$ finds its extreme value at the extremal points of the domain; similarly, OUQ problems reduce to searches over finite-dimensional families of extremal scenarios.

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

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

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

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

Optimal Concentration Inequalities

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

Example: Markov's Inequality in OUQ Form

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

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

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.

McDiarmid's Inequality

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

with componentwise oscillations/global sensitivities defined by

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

Theorem (McDiarmid's Inequality, 1988)

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

Optimal McDiarmid and Screening Effects

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

For $K = 1$,

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

For $K = 2$,

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

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

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

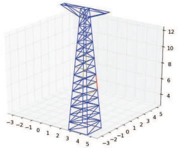
- Interestingly, $\mathcal{U}(\mathcal{A}_{\mathsf{Hfd}}) = \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$ for $K = 1$ and $K = 2$, but $\mathcal{U}(\mathcal{A}_{\mathsf{Hfd}}) \leq \mathcal{U}(\mathcal{A}_{\mathsf{McD}})$ for $K = 3$, and the inequality can be strict. Thus, sometimes linearity is binding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

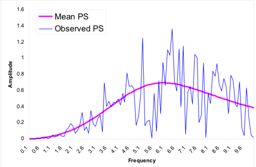
Seismic Safety Certification

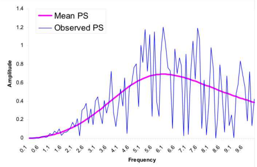
- Consider the survivability of a truss structure under an random earthquake of known intensity drawn from an incompletely specified probability distribution.
- Consider a random ground motion u, with the constraint that the mean power spectrum is the Matsuda–Asano shape function s_{MA}:

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

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







Numerical Results: Vulnerability Curves

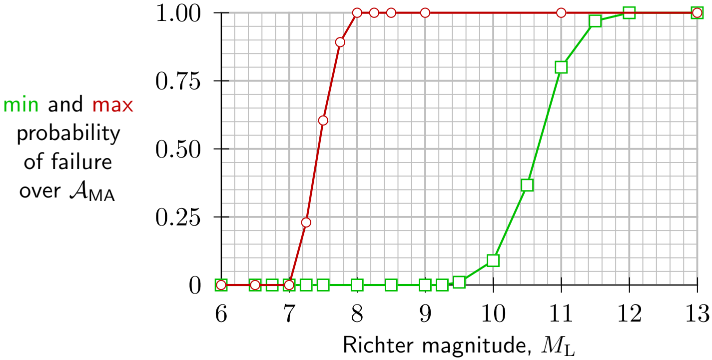


Figure: The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980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

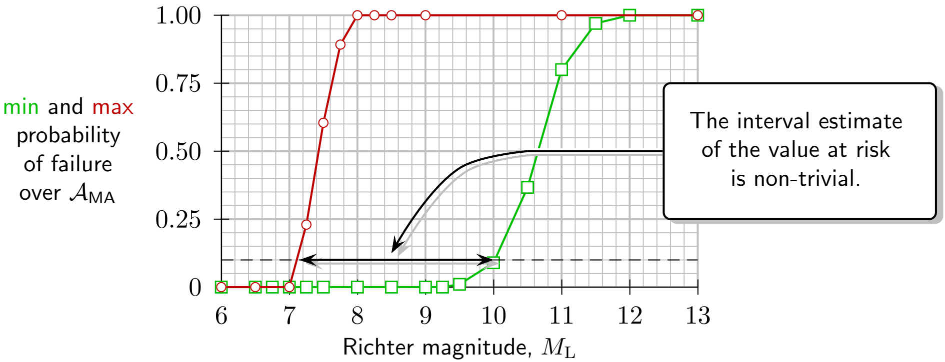


Figure: The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980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

Numerical Results: Vulnerability Curves

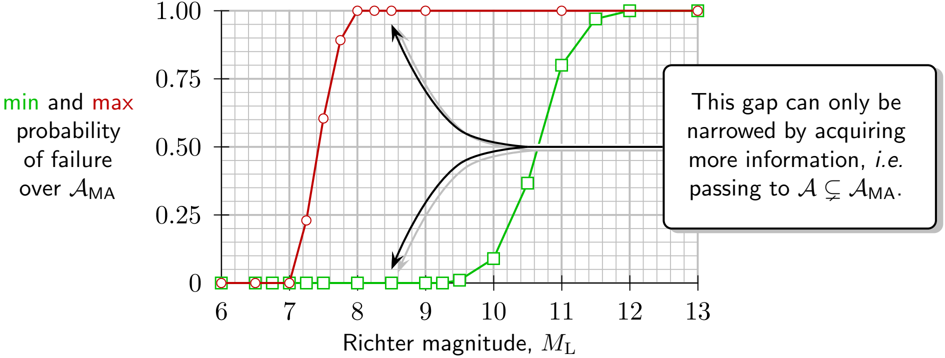


Figure: The minimum and maximum probability of failure as a function of Richter magnitude, M_{L}, where the ground motion u is constrained to have $\mathbb{E}_{\mu}[|\hat{u}|^2] =$ the Matsuda–Asano shape function s_{MA} with natural frequency ω_{g} and natural damping ξ_{g} taken from the 24 Jan. 1980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

Overview

Optimal Knowledge Acquisition

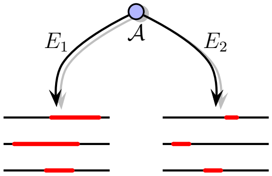
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



Optimal Knowledge Acquisition

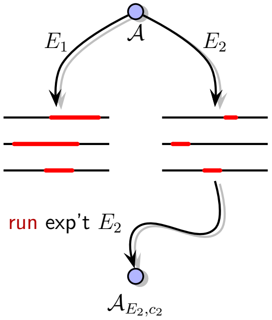
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



Optimal Knowledge Acquisition

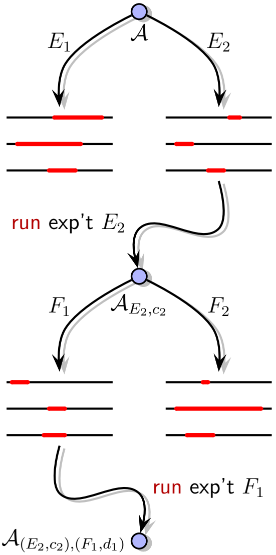
- Range of prediction given $\mathcal{A}$:

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

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

- 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:

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



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.
- In this sense, (O)UQ and extreme-scale scientific computing are natural partners:
  - UQ calculations for complex systems clearly demand large computational resources;
  - but those same resources are expensive — and will probably be non-deterministic! — and so UQ offers a way to perform large calculations on such systems.

"OUQ++": 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.

Prove a 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, ...

Or Compute?

OUQ++ offers a way to compute the optimal statistical estimator for your problem, a 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”.

	I	II
PDEs	Compute tables for special functions, plug them into PDE *ansätze*	Discretize the PDE and compute directly using FE, FD, ...
E.g. **McD**	McDiarmid's inequality $\bar{p} \leq e^{-2m^2/\sum_i D_i^2}$	Optimal McDiarmid inequality, $\bar{p} = \mathcal{U}(\mathcal{A}_{\mathrm{McD}})$
UQ/Stats	Compute tables for statistics and plug them into (theorem-derived) estimators	OUQ++?

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