CALTECH PSAAP Predictive Science Academic Alliance Program

NNSA National Nuclear Security Administration

From Optimal Uncertainty Quantification to the Scientific Computation of Optimal Statistical Estimators

Houman Owhadi

C. Scovel, T. Sullivan, M. McKerns and M.Ortiz.

Optimal Uncertainty Quantification. H. Owhadi, C. Scovel, T. Sullivan, M. McKerns and M. Ortiz. **SIAM Review** (Expository Research Papers)

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH UNITED STATES AIR FORCE

LLNL, October 2012

ExMatEx

The UQ challenge in the certification context

$$G : \chi \longrightarrow \mathbb{R}$$

$$X \longrightarrow G(X)$$

$$\mathbb{P} \in \mathcal{M}(\chi)$$

You want to certify that

$$\mathbb{P}[G(X) \geq a] \leq \epsilon$$

Problem

- You don't know G.

and

- You don't know $\mathbb{P}$

The UQ challenge in the certification context

$$G : \chi \longrightarrow \mathbb{R}$$

$$X \longrightarrow G(X)$$

$$\mathbb{P} \in \mathcal{M}(\chi)$$

You want to certify that

$$\mathbb{P}[G(X) \geq a] \leq \epsilon$$

You only know

$$(G, \mathbb{P}) \in \mathcal{A}$$

$$\mathcal{A} \subset \left\{ (f, \mu) \;\middle|\; \begin{array}{c} f : \mathcal{X} \to \mathbb{R}, \\ \mu \in \mathcal{P}(\mathcal{X}) \end{array} \right\}$$

Optimal bounds on $\mathbb{P}[G(X) \geq a]$

$$\mathcal{U}(\mathcal{A}) := \sup_{(f,\mu)\in\mathcal{A}} \mu[f(X) \geq a]$$

$$\mathcal{L}(\mathcal{A}) := \inf_{(f,\mu)\in\mathcal{A}} \mu[f(X) \geq a]$$

$$\mathcal{L}(\mathcal{A}) \leq \mathbb{P}[G(X) \geq a] \leq \mathcal{U}(\mathcal{A})$$

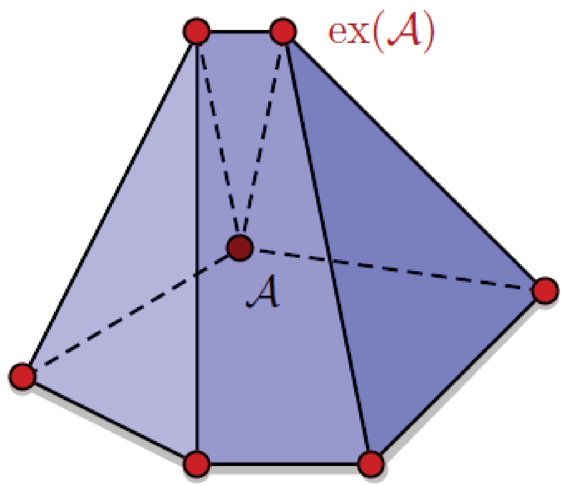
$\mathcal{U}(\mathcal{A}) \leq \epsilon$: Safe even in worst case.

$\epsilon < \mathcal{L}(\mathcal{A})$: Unsafe even in best case.

$\mathcal{L}(\mathcal{A}) \leq \epsilon < \mathcal{U}(\mathcal{A})$: Cannot decide.
Unsafe due to lack of information.

OUQ problems are a priori infinite dimensional, non-convex and highly constrained

But as in linear programming OUQ problems reduce to searches over finite dimensional families of extremal scenarios of $\mathcal{A}$



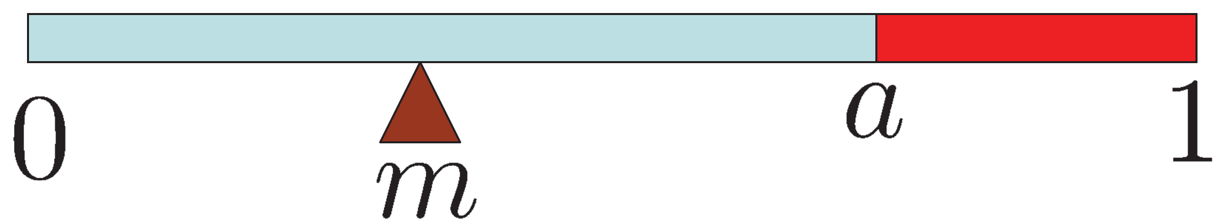
The dimension of the reduced problem is proportional to the number of probabilistic inequalities that describe $\mathcal{A}$

A simple example

What is the least upper bound on $\mathbb{P}[X \geq a]$

If all you know is $\mathbb{E}[X] \leq m$

and $\mathbb{P}[0 \leq X \leq 1] = 1$?

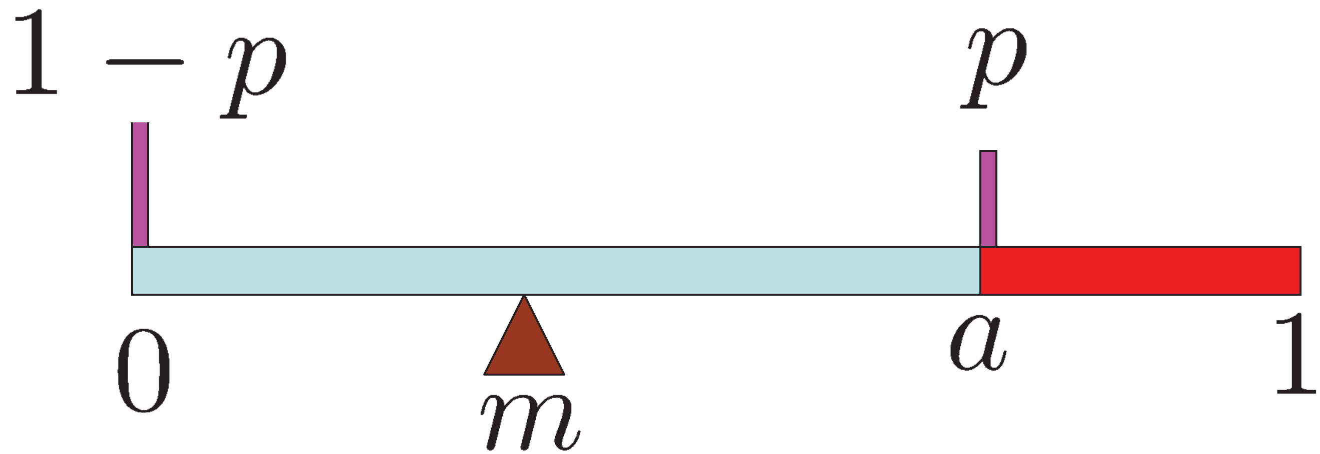


Answer

$$\sup_{\mu \in \mathcal{A}} \mu\left[X \geq a\right]$$

$$\mathcal{A} = \{\mu \in \mathcal{M}([0,1]) \,|\, \mathbb{E}_\mu[X] \leq m\}$$

You are given one pound of play-doh.
How much mass can you put above a while keeping the seesaw balanced around m?



Answer

$$\begin{cases} \max p \\ \text{subject to } a\, p \leq m \end{cases}$$

Markov's inequality

$$\sup_{\mu \in \mathcal{A}} \mu\left[X \geq a\right] = \frac{m}{a}$$

$$\mathcal{A} = \{\mu \in \mathcal{M}([0,1]) \,|\, \mathbb{E}_\mu[X] \leq m\}$$

Reduction theorems

$$\mathcal{A} = \left\{ (f,\mu) \;\middle|\; \begin{array}{c} f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \to \mathbb{R}, \\ \mu = \mu_1 \otimes \cdots \otimes \mu_m, \\ \mathcal{G}(f,\mu) \leq 0 \end{array} \right\}$$

$$\mathcal{G}(f,\mu) \leq 0 \Leftrightarrow \begin{cases} n' \text{ generalized moment constraints on } \mu, & \mathbb{E}_{\mu}[\varphi_j^f] \leq 0 \\ n_k \text{ generalized moment constraints on } \mu_k, & \mathbb{E}_{\mu_k}[\psi_{k,j}^f] \leq 0 \end{cases}$$

Theorem

$$\sup_{(f,\mu)\in\mathcal{A}} \mathbb{E}_{\mu}[q_f] = \sup_{(f,\mu)\in\mathcal{A}_\Delta} \mathbb{E}_{\mu}[q_f]$$

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

Reduction of optimization variables

$$\{f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X})\}$$

$$\left\{ f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X}) \,\middle|\, \mu = \sum_{i=1}^{k} \alpha_k \delta_{x_k} \right\}$$

$$\{f : \{1, 2, \ldots, n\} \to \mathbb{R},\ \mu \in \mathcal{P}(\{1, 2, \ldots, n\})\}$$

$$\{\{1, 2, \ldots, q\},\ \mu \in \mathcal{P}(\{1, 2, \ldots, n\})\}$$

Literature

$$\mathcal{U}(\mathcal{A}) := \sup_{(g,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_g]$$

Non-convex and infinite dimensional optimization problems

Can be considered as a **generalization of classical Chebyshev inequalities**

History of classical inequalities: Karlin, Studden (1966, Tchebycheff systems with applications in analysis and statistics)

Connection between Chebyshev inequalities and optimization theory

- Mulholland & Rogers (1958, Representation theorems for distribution functions)
- Godwin (1973, Manipulation of voting schemes: a general result)
- Isii (1959, On a method for generalization of Tchebycheff's inequality
  1960, The extrema of probability determined by generalized moments
  1962, On sharpness of Techebycheff-type inequalities)
- Olhin & Pratt (1958, A multivariate Tchebycheff inequality)
- Classical Markov-Krein theorem (Karlin, Studden, 1958)
- Dynkin (1978, Sufficient statistics & extreme points)
- Karr (1983, Extreme points of probability measures with applications)
- Artzner et al (1997, risk measures, value at risk, etc…)
- Betsimas & Popescu (2008, convex optimization approach to inequalities in prob. theo.

Literature

$$\mathcal{U}(\mathcal{A}) := \sup_{(g,\mu)\in\mathcal{A}} \mathbb{E}_{\mu}[q_g]$$

Our work: Further generalization to

- Independence constraints
- More general domains (Suslin spaces) (non metric, non compact)
- More general classes of functions (measurable) (non continuous, non-bounded)
- More general classes of probability measures
- More general constraints (inequalities, on measures and functions)

Theory of majorization

- Marshall & Olkin (1979, Inequalities: Theory of majorization and its applications)

Inequalities of

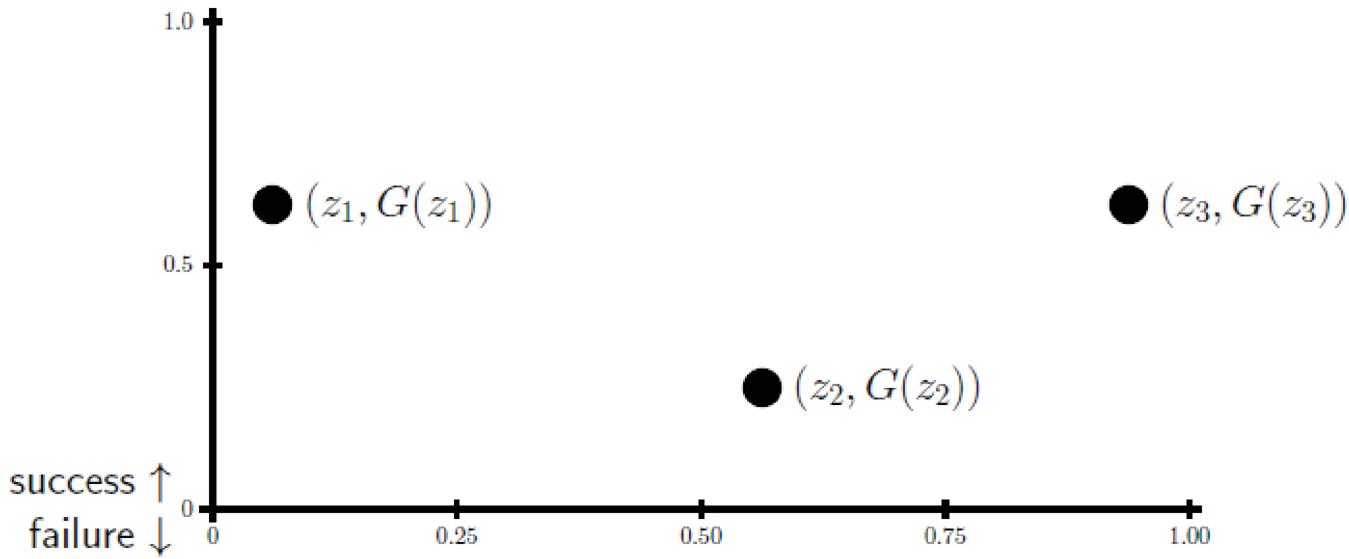
- Anderson (1955, the integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities)
- Hoeffding (1956, on the distribution of the number of successes in independent trials)
- Joe (1987, Majorization, randomness and dependence for multivariate distributions)
- Bentkus, Geuze, Van Zuijlen (2006, Optimal Hoeffding like inequalities under a symmetry assumption)
- Pinelis (2007, Exact inequalities for sums of asymmetric random variables with applications.
  2008, On inequalities for sums of bounded random variables)

Our proof rely on

- Winkler (1988, Extreme points of moment sets)
- Follows from an extension of Choquet theory (Phelps 2001, lectures on Choquet's theorem) by Von Weizsacker & Winkler (1979, Integral representation in the set of solutions of a generalized moment problem)
- Kendall (1962, Simplexes & Vector lattices)

Another simple example

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of G on $\mathcal{O} \subseteq [0, 1]$?



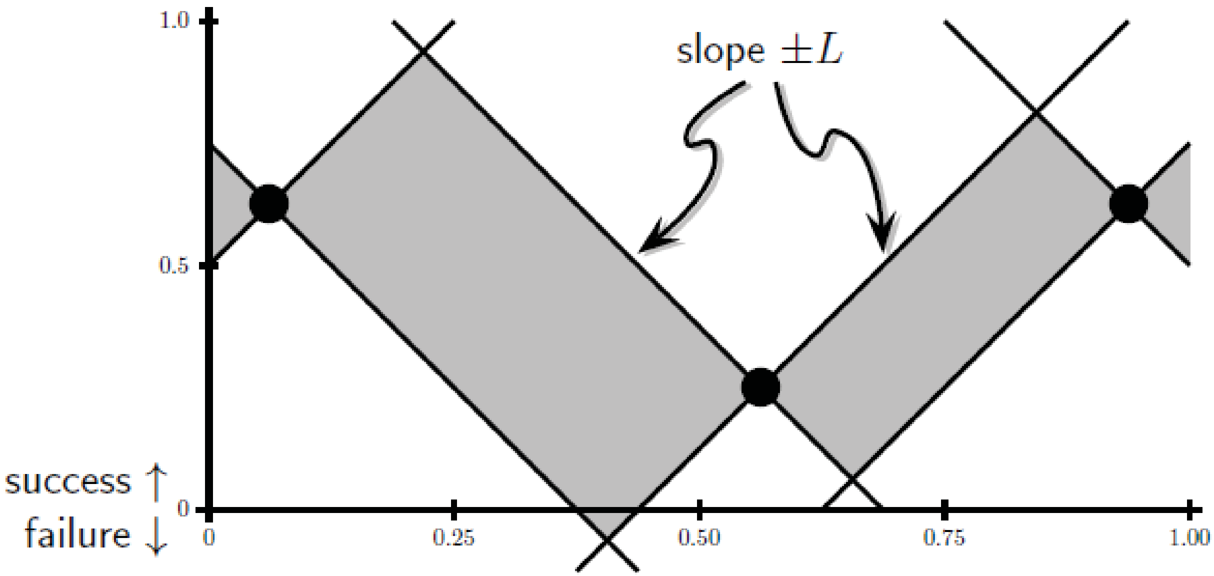
Sharpest Possible Answer...

With so little information, the only rigorous bounds that can be given are the trivial ones: $0 \leq \mathbb{P}[G(X) \leq 0] \leq 1$.

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi & M. Ortiz
“Optimal uncertainty quantification for legacy data observations of Lipschitz functions”

The effect of information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of G on $\mathcal{O} \subseteq [0, 1]$, and that $|G(x) - G(x')| \leq L|x - x'|$?

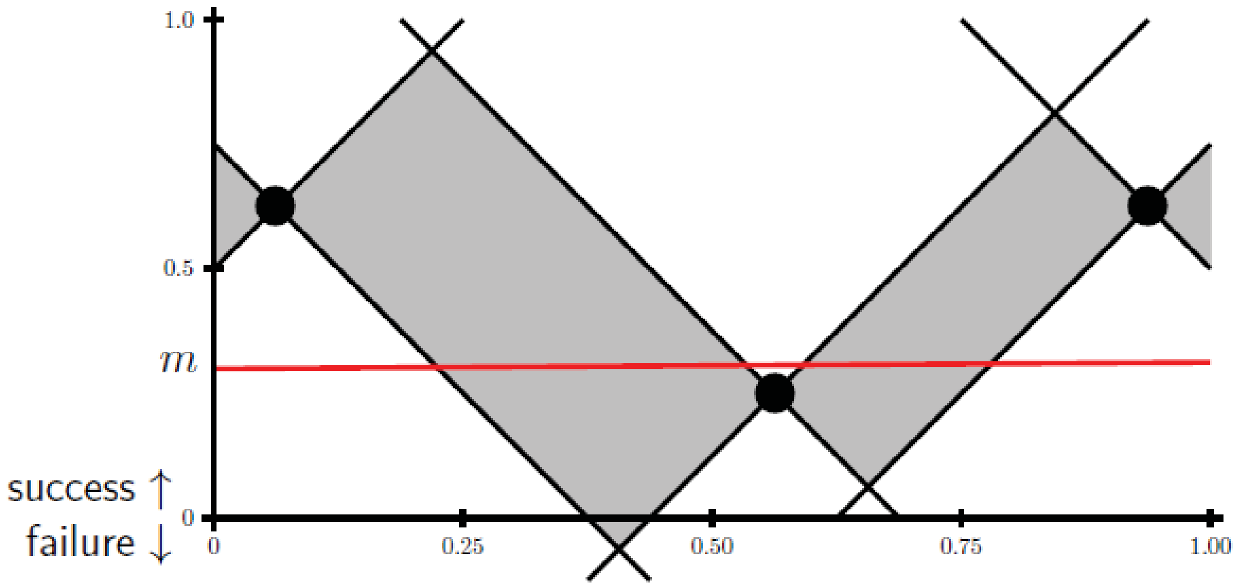


Sharpest Possible Answer...

...we might discover that $\mathbb{P}[G(X) \leq 0] = 0$ or $= 1$, but otherwise no improvement on the trivial bound $0 \leq \mathbb{P}[G(X) \leq 0] \leq 1$.

The effect of information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of G on $\mathcal{O} \subseteq [0,1]$, that $|G(x) - G(x')| \leq L|x - x'|$, and that $\mathbb{E}[G(X)] \geq m$?

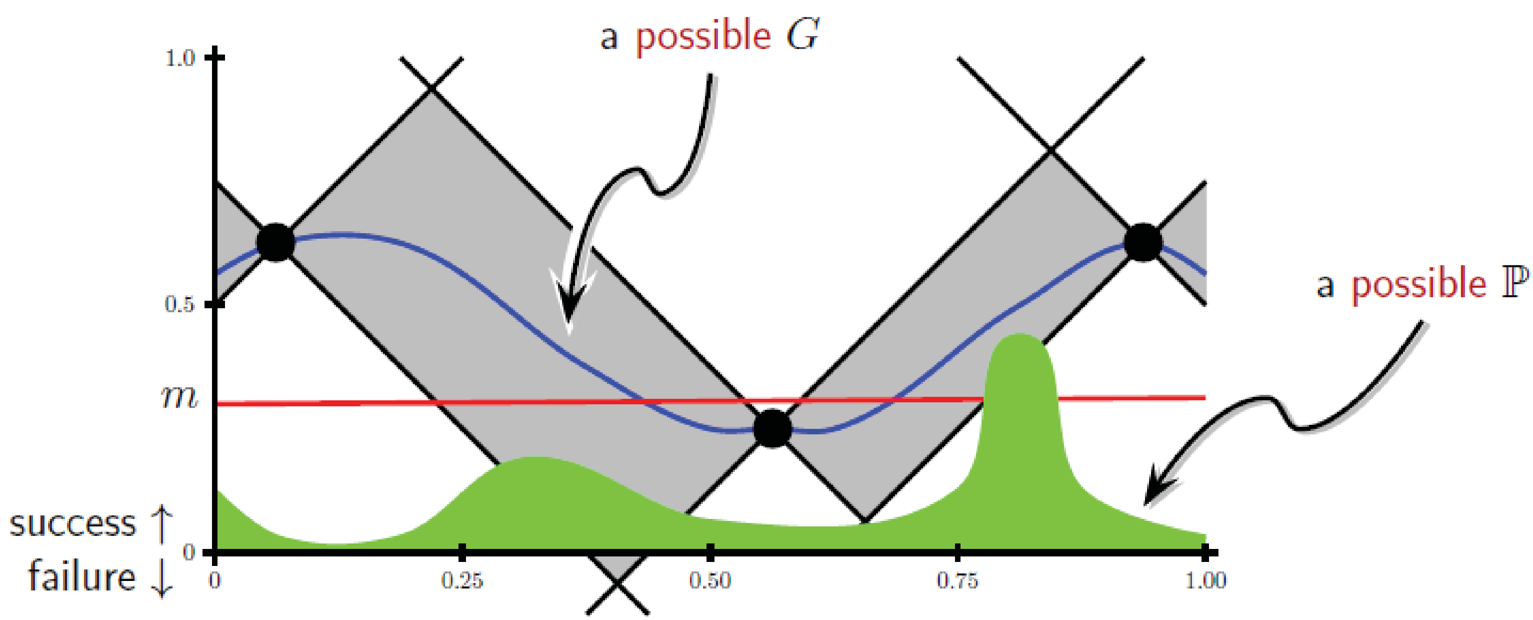


Sharpest Possible Answer. . .

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

The effect of information

What can be said about $\mathbb{P}[G(X) \leq 0]$ if all that is known are the values of G on $\mathcal{O} \subseteq [0, 1]$, that $|G(x) - G(x')| \leq L|x - x'|$, and that $\mathbb{E}[G(X)] \geq m$?



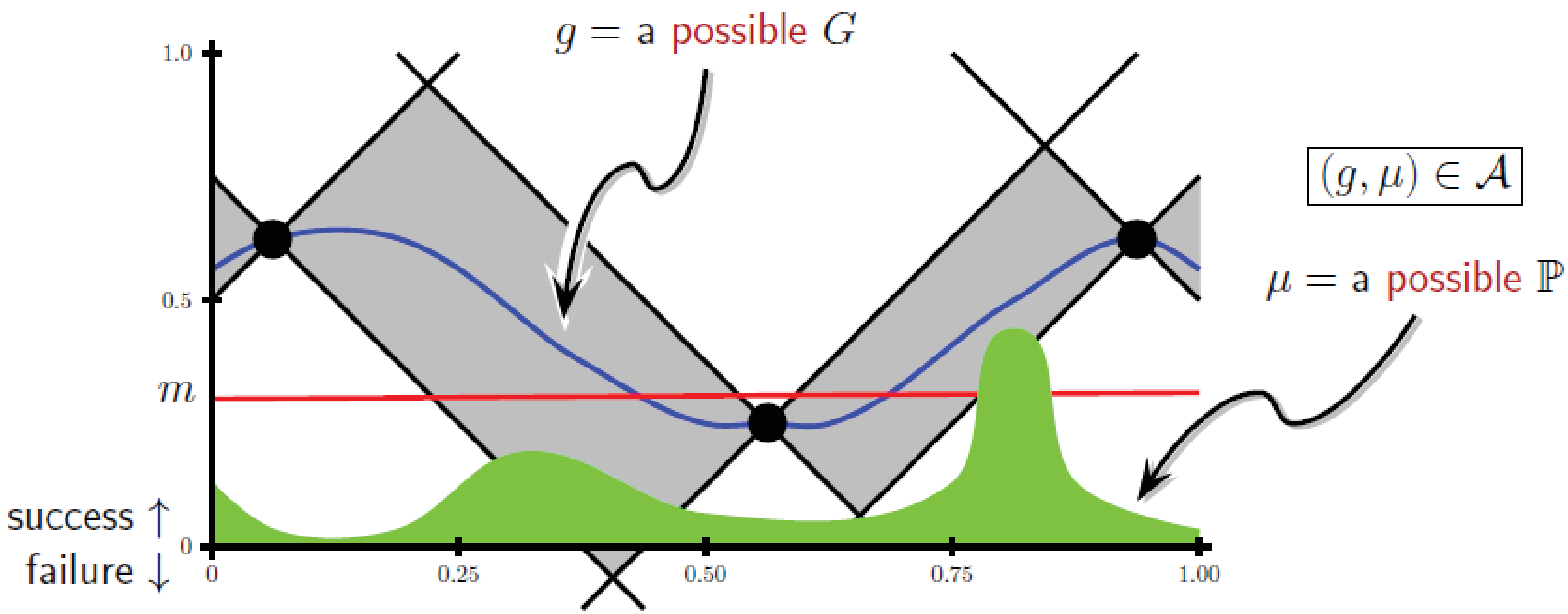
Sharpest Possible Answer...

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

The reduced problem

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

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



Problem formulation

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

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

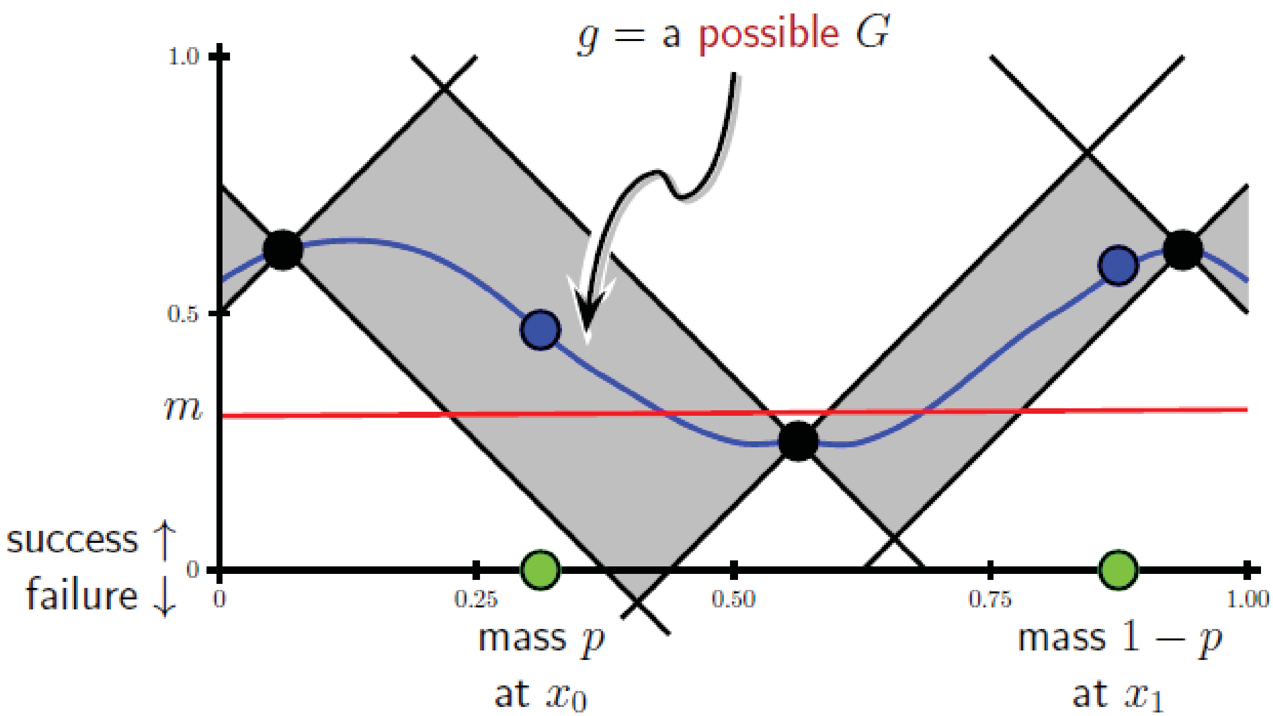
In other words, any (g, μ) for which g is L-Lipschitz, agrees with the legacy data, and has the right mean under μ could be $(G, \mathbb{P})$. The reduced admissible set, over which the quantity of interest has the same extreme values, is

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

The reduced problem

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

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

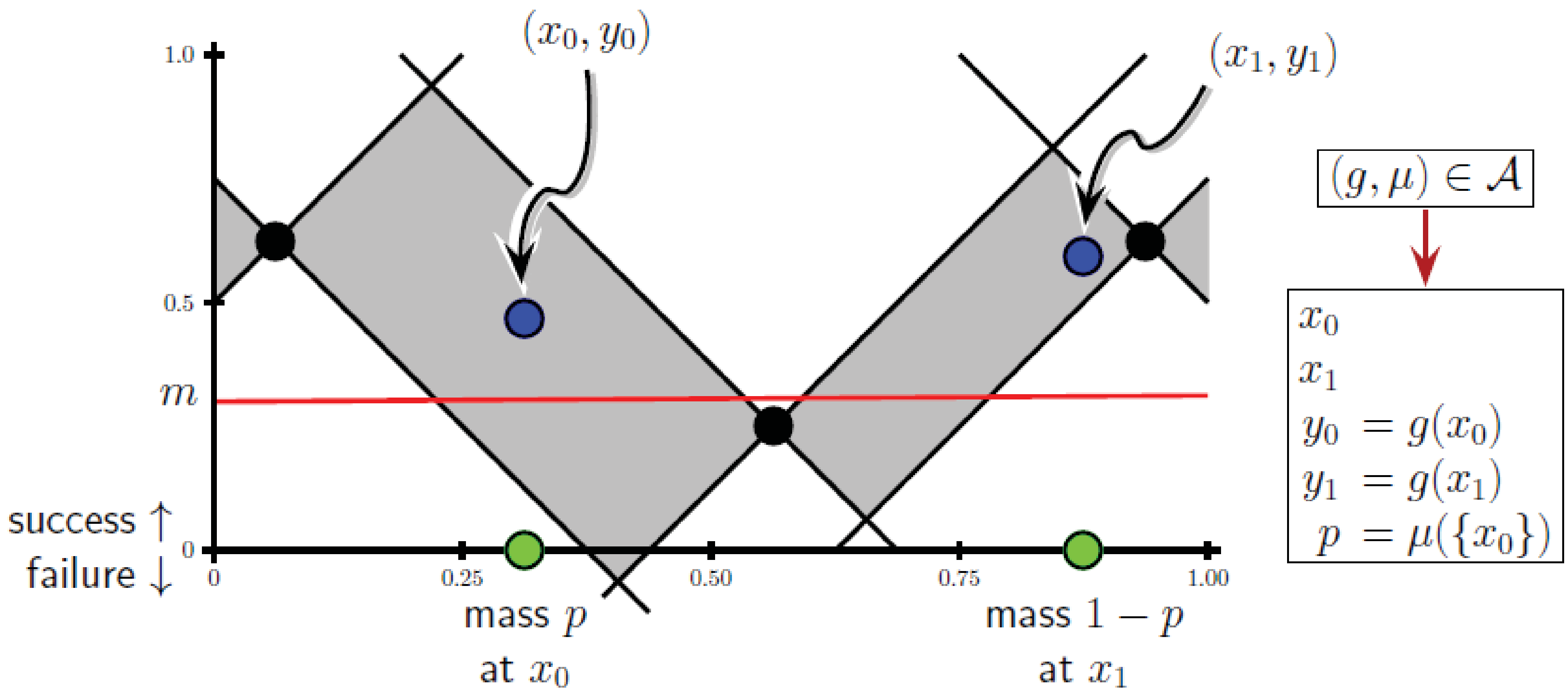


$(g, \mu) \in \mathcal{A}$

The reduced problem

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

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



Application: Optimal concentration inequality

$$\mathcal{A}_{MD} := \left\{ (f,\mu) \;\middle|\; \begin{array}{c} f\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \to \mathbb{R}, \\ \mu \in \mathcal{M}(\mathcal{X}_1) \otimes \cdots \otimes \mathcal{M}(\mathcal{X}_m), \\ \mathbb{E}_\mu[f] \le 0, \\ \mathrm{Osc}_i(f) \le D_i \end{array} \right\}$$

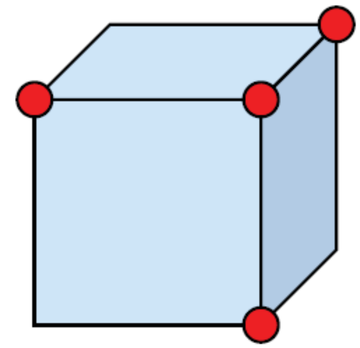
$$\mathrm{Osc}_i(f) := \sup_{(x_1,\ldots,x_m)\in\mathcal{X}} \sup_{x_i'\in\mathcal{X}_i} \big(f(\ldots,x_i,\ldots) - f(\ldots,x_i',\ldots)\big).$$

$$\mathcal{U}(\mathcal{A}_{MD}) := \sup_{(f,\mu)\in\mathcal{A}_{MD}} \mu[f(X) \ge a]$$

McDiarmid inequality $\mathcal{U}(\mathcal{A}_{MD}) \le \exp\left(-2\frac{a^2}{\sum_{i=1}^m D_i^2}\right)$

Reduction of optimization variables

$$\mathcal{A}_{\mathcal{C}} := \left\{ (C, \alpha) \;\middle|\; \begin{array}{c} C \subset \{0,1\}^m, \\ \alpha \in \otimes_{i=1}^m \mathcal{M}(\{0,1\}), \\ \mathbb{E}_\alpha[h^C] \leq 0 \end{array} \right\}$$



$$h^C : \{0,1\}^m \longrightarrow \mathbb{R}$$

$$t \longrightarrow a - \min_{s \in C} \sum_{i:\, s_i \neq t_i} D_i$$

$$\mathcal{U}(\mathcal{A}_{\mathcal{C}}) := \sup_{(C,\alpha) \in \mathcal{A}_{\mathcal{C}}} \alpha[h^C \geq a]$$

Theorem

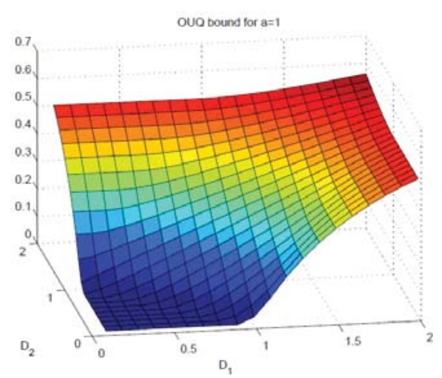
$$\mathcal{U}(\mathcal{A}_{MD}) = \mathcal{U}(\mathcal{A}_{\mathcal{C}})$$

Explicit Solution m=2

Theorem $m = 2$

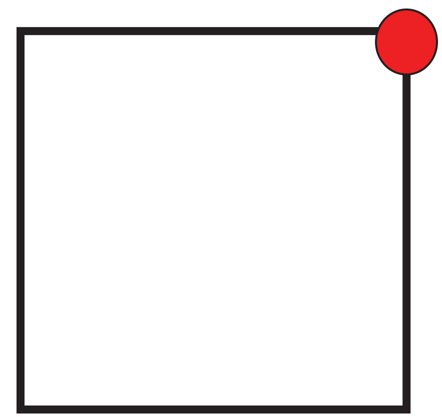
$$\mathcal{U}(\mathcal{A}_{MD}) = \begin{cases} 0 & \text{if} \quad D_1 + D_2 \le a \\ \frac{(D_1+D_2-a)^2}{4D_1D_2} & \text{if} \quad |D_1 - D_2| \le a \le D_1 + D_2 \\ 1 - \frac{a}{\max(D_1,D_2)} & \text{if} \quad 0 \le a \le |D_1 - D_2| \end{cases}$$

OUQ bound a=1



$$C = \{(1,1)\}$$

$$h^C(s) = a - (1 - s_1)D_1 - (1 - s_2)D_2$$



Corollary If $D_1 \ge a + D_2$, then

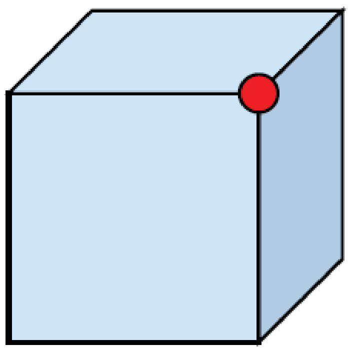
$$\mathcal{U}(\mathcal{A}_{MD})(a, D_1, D_2) = \mathcal{U}(\mathcal{A}_{MD})(a, D_1, 0)$$

Explicit Solution m=3

Theorem $m = 3 \qquad D_1 \geq D_2 \geq D_3$

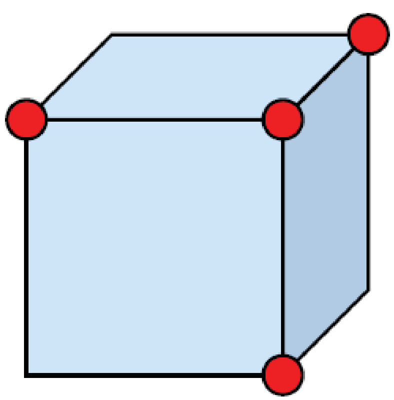
$$\mathcal{U}(\mathcal{A}_{MD}) = \max(\mathcal{F}_1, \mathcal{F}_2)$$

$\mathcal{F}_1$



$$\mathcal{F}_1 := \begin{cases} 0 & \text{if} \quad D_1 + D_2 + D_3 \leq a \\ \frac{(D_1+D_2+D_3-a)^3}{27 D_1 D_2 D_3} & \text{if} \quad D_1 + D_2 - 2D_3 \leq a \leq D_1 + D_2 + D_3 \\ \frac{(D_1+D_2-a)^2}{4 D_1 D_2} & \text{if} \quad D_1 - D_2 \leq a \leq D_1 + D_2 - 2D_3 \\ 1 - \frac{a}{\max(D_1, D_2)} & \text{if} \quad 0 \leq a \leq D_1 - D_2 \end{cases}$$

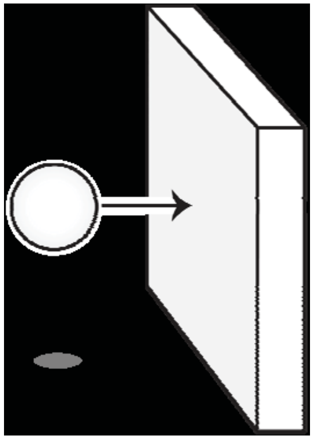
$\mathcal{F}_2$

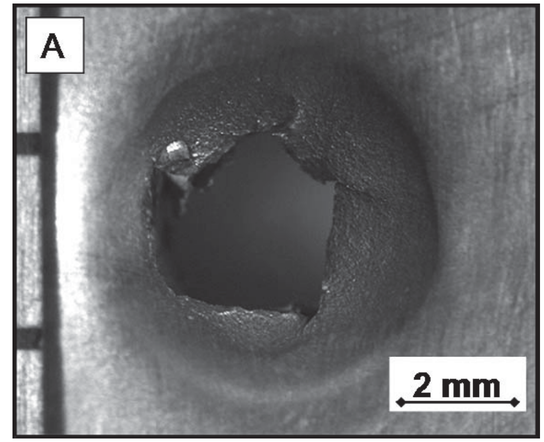


$$\mathcal{F}_2 := \max_{i \in \{1,2,3\}} \phi(\gamma_i)\psi(\gamma_i)$$

$$(1+\gamma)^3 - \frac{5D_2 - 2D_3}{2D_2 - D_3}(1+\gamma)^2 + \frac{4D_2 - a}{2D_2 - D_3} = 0,$$

Caltech Small Particle Hypervelocity Impact Range





$$(h, \alpha, v) \xrightarrow{\quad G \quad} G(h, \alpha, v)$$

Plate thickness

Plate Obliquity

Projectile velocity

Perforation area

We want to certify that

$$\mathbb{P}[G = 0] \leq \epsilon$$

Caltech Hypervelocity Impact Surrogate Model

Plate thickness $h \in \mathcal{X}_1 := [1.524, 2.667]\,\mathrm{mm},$

Plate Obliquity $\alpha \in \mathcal{X}_2 := [0, \frac{\pi}{6}],$

Projectile velocity $v \in \mathcal{X}_3 := [2.1, 2.8]\,\mathrm{km}\cdot\mathrm{s}^{-1}.$

Thickness, obliquity, velocity: independent random variables

Mean perforation area: in between 5.5 and 7.5 mm^2

Deterministic surrogate model for the perforation area (in mm^2)

$$H(h, \alpha, v) = K \left(\frac{h}{D_{\mathrm{p}}}\right)^{p} (\cos\alpha)^{u} \left(\tanh\left(\frac{v}{v_{\mathrm{bl}}} - 1\right)\right)_{+}^{m},$$

$H_0 = 0.5794\,\mathrm{km}\cdot\mathrm{s}^{-1}$, $s = 1.4004$, $n = 0.4482$, $K = 10.3936\,\mathrm{mm}^2$,

$p = 0.4757$, $u = 1.0275$, $m = 0.4682$.

$$v_{\mathrm{bl}} := H_0 \left(\frac{h}{(\cos\alpha)^{n}}\right)^{s}$$

Optimal bound on the probability of non perforation

$$\mathcal{A}_{\mathrm{McD}} := \left\{ (f,\mu) \;\middle|\; \begin{matrix} \mu = \mu_1 \otimes \mu_2 \otimes \mu_3, \\ 5.5\,mm^2 \leq \mathbb{E}_\mu[f] \leq 7.5\,mm^2, \\ \mathrm{Osc}_i(f) \leq \mathrm{Osc}_i(H) \text{ for } i = 1,2,3 \\ f \geq 0 \end{matrix} \right\}$$

$$\mathcal{U}(\mathcal{A}_{\mathrm{McD}}) := \sup_{(f,\mu)\in\mathcal{A}_{\mathrm{McD}}} \mu[f(X) = 0]$$

$$\mathbb{P}[H=0] \leq \mathcal{U}(\mathcal{A}_{\mathrm{McD}}) \leq \exp\left(-\frac{2m_1^2}{\sum_{i=1}^{3} \mathrm{Osc}_i(H)^2} \right) = 66.4\%.$$

$$\mathbb{P}[H=0] \leq \mathcal{U}(\mathcal{A}_{\mathrm{McD}}) = 43.7\%.$$

Optimal bound on the probability of non perforation

$$\mathcal{A} := \left\{ (f,\mu) \,\middle|\, \begin{array}{c} \mu = \mu_1 \otimes \mu_2 \otimes \mu_3, \\ 5.5\, mm^2 \leq \mathbb{E}_\mu[f] \leq 7.5\, mm^2, \\ f = H \end{array} \right\}$$

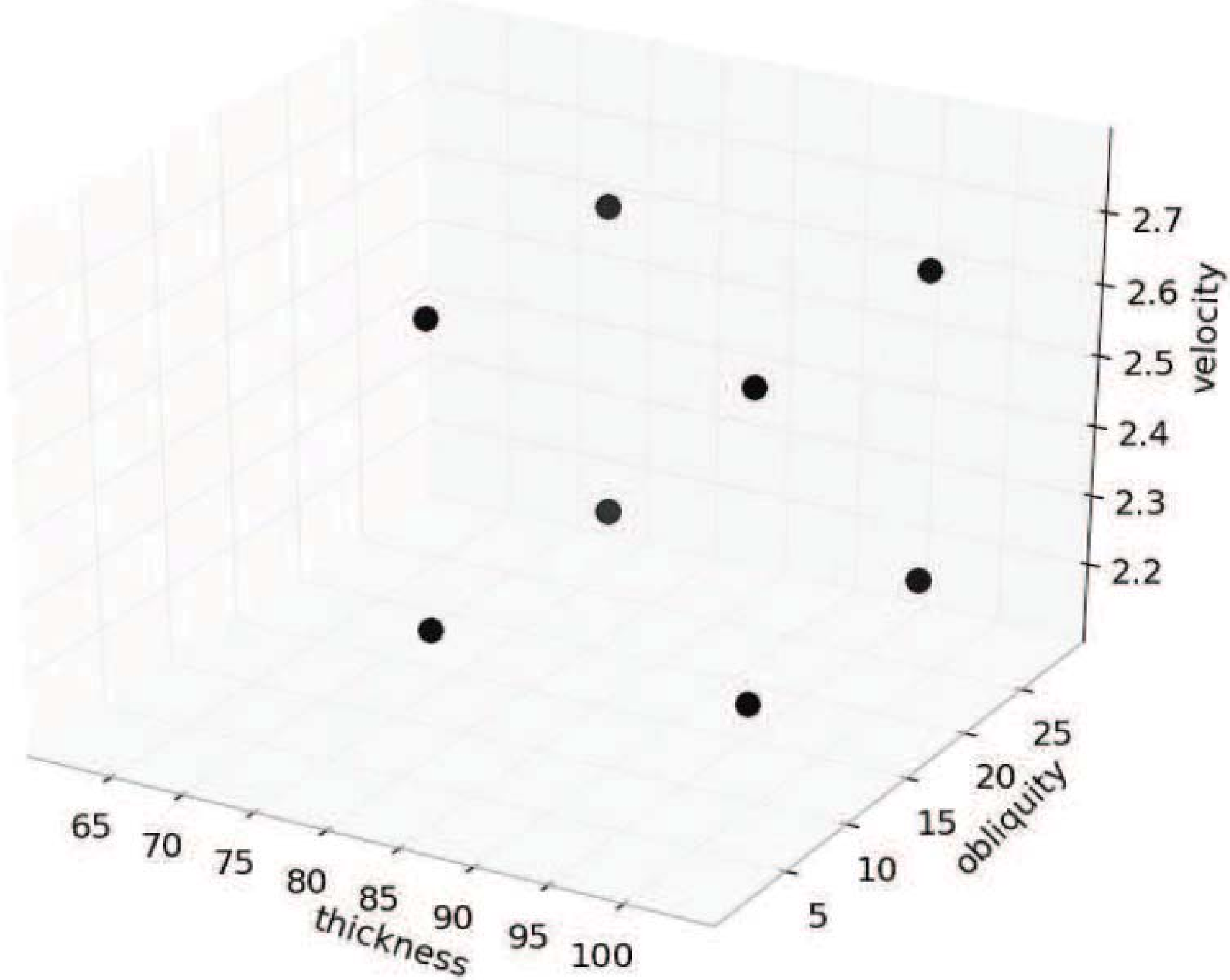
$$\mathcal{U}(\mathcal{A}) := \sup_{(f,\mu)\in\mathcal{A}} \mu[f(X) = 0]$$

Application of the reduction theorem

The measure of probability can be reduced to the tensorization of 2 Dirac masses on thickness, obliquity and velocity

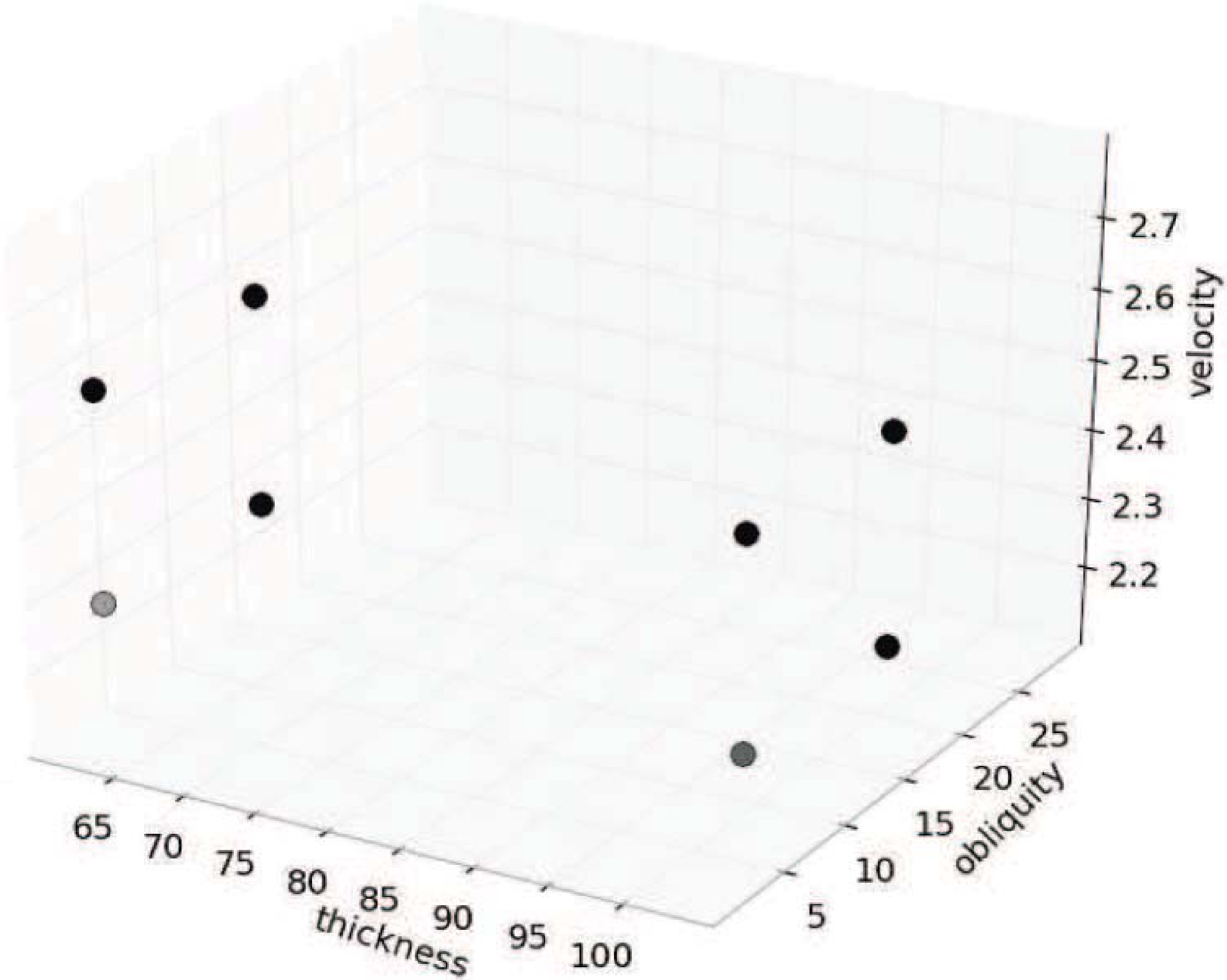
$$\mathcal{U}(\mathcal{A}) \overset{\text{num}}{=} 37.9\%$$

The optimization variables can be reduced to the tensorization of 2 Dirac masses on thickness, obliquity and velocity



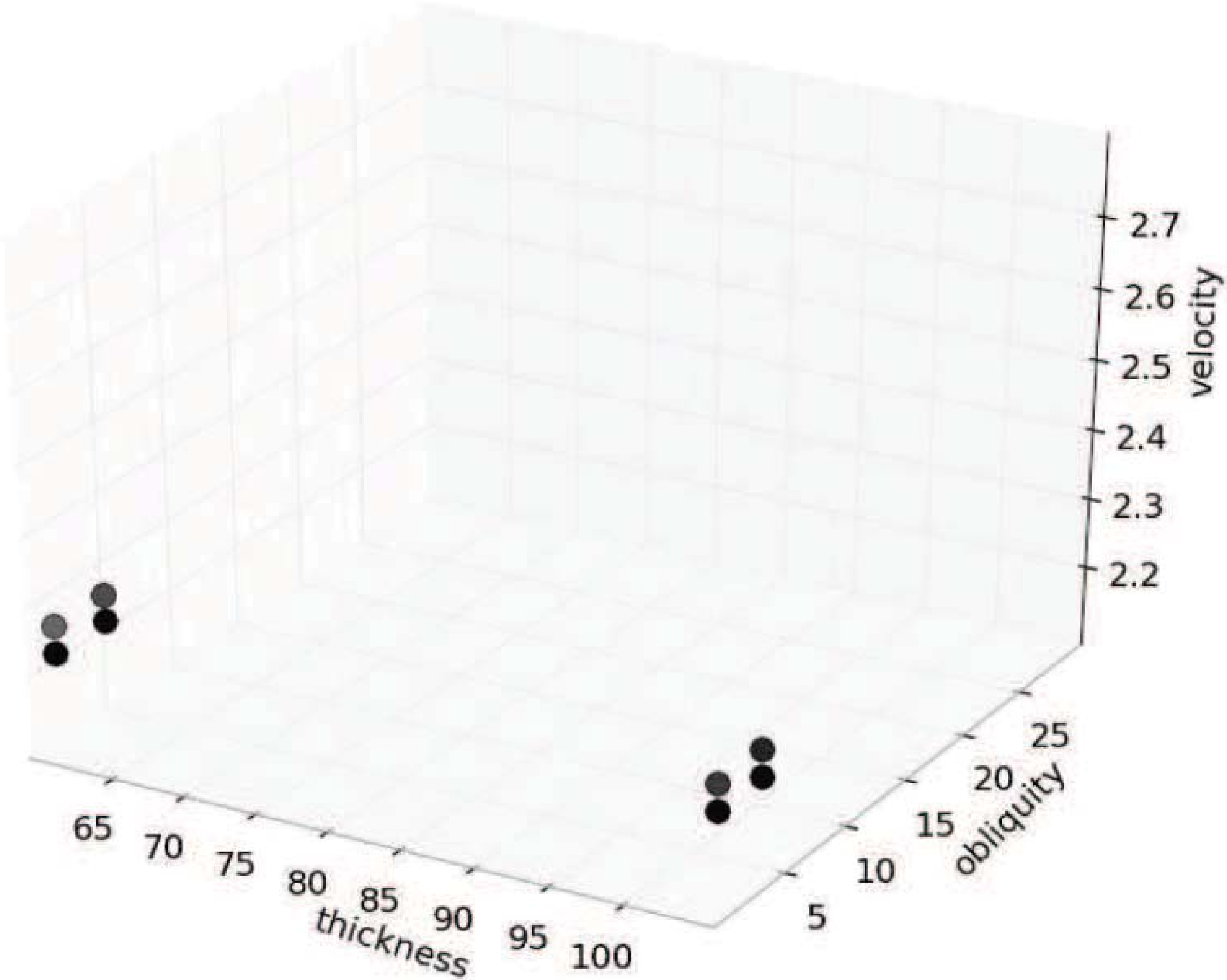
Support Points at iteration 0

Numerical optimization



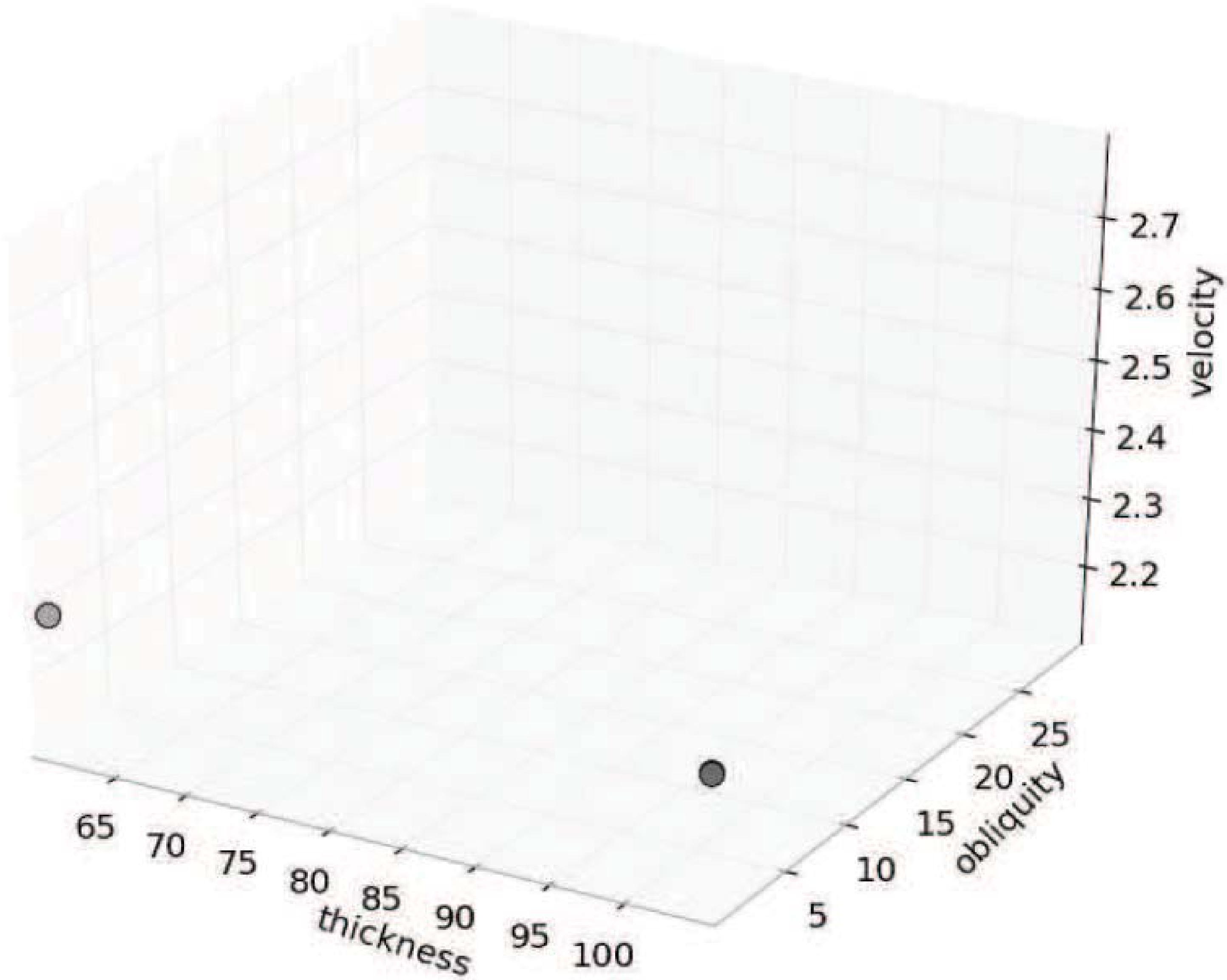
Support Points at iteration 150

Numerical optimization



Support Points at iteration 200

Velocity and obliquity marginals each collapse to a single Dirac mass. The plate thickness marginal collapses to have support on the extremes of its range.



Iteration 1000

Probability non-perforation maximized by distribution supported on minimal, not maximal, impact obliquity. Dirac on velocity at a non extreme value.

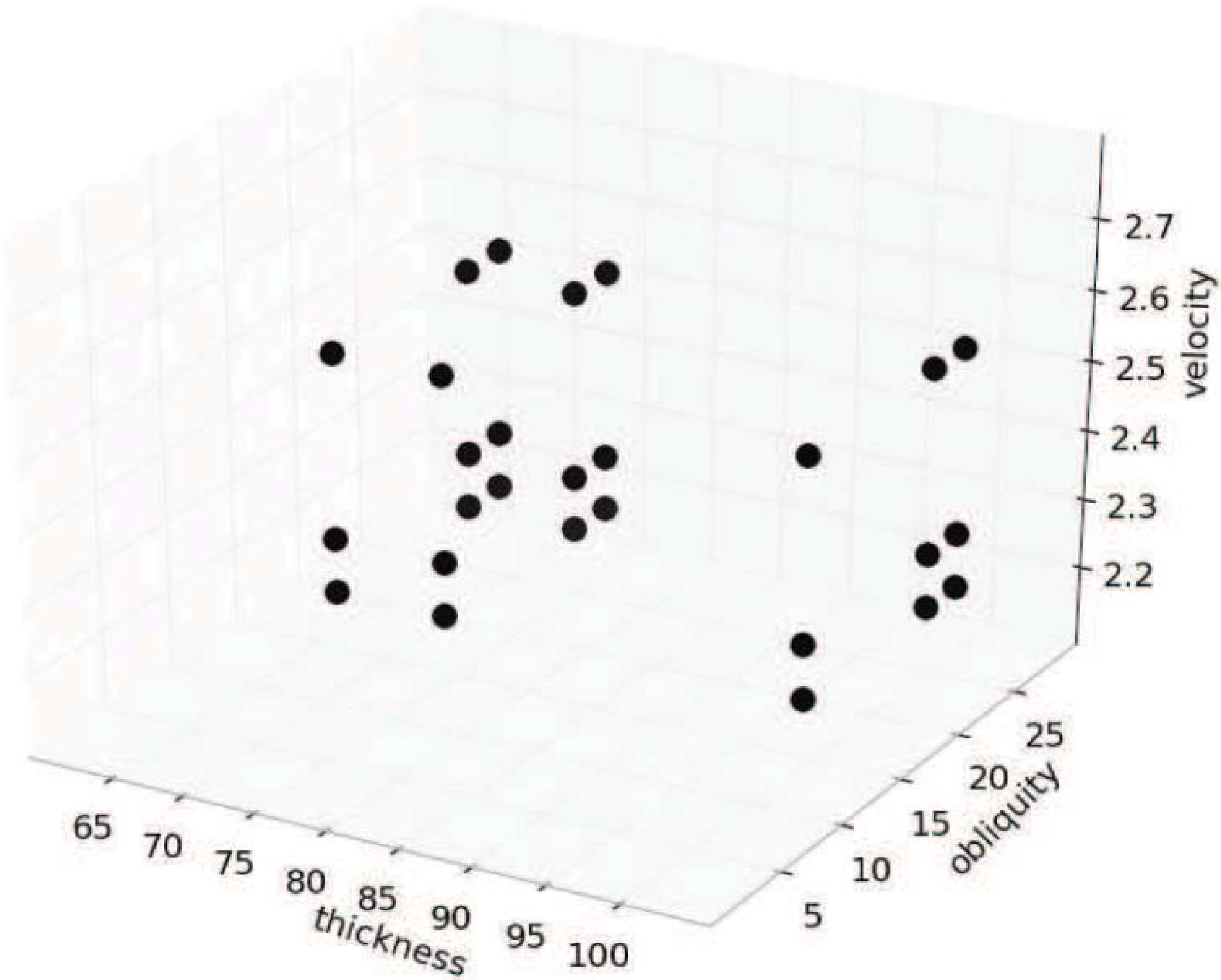
Important observations

Extremizers are singular

They identify key players
i.e. vulnerabilities of the physical system

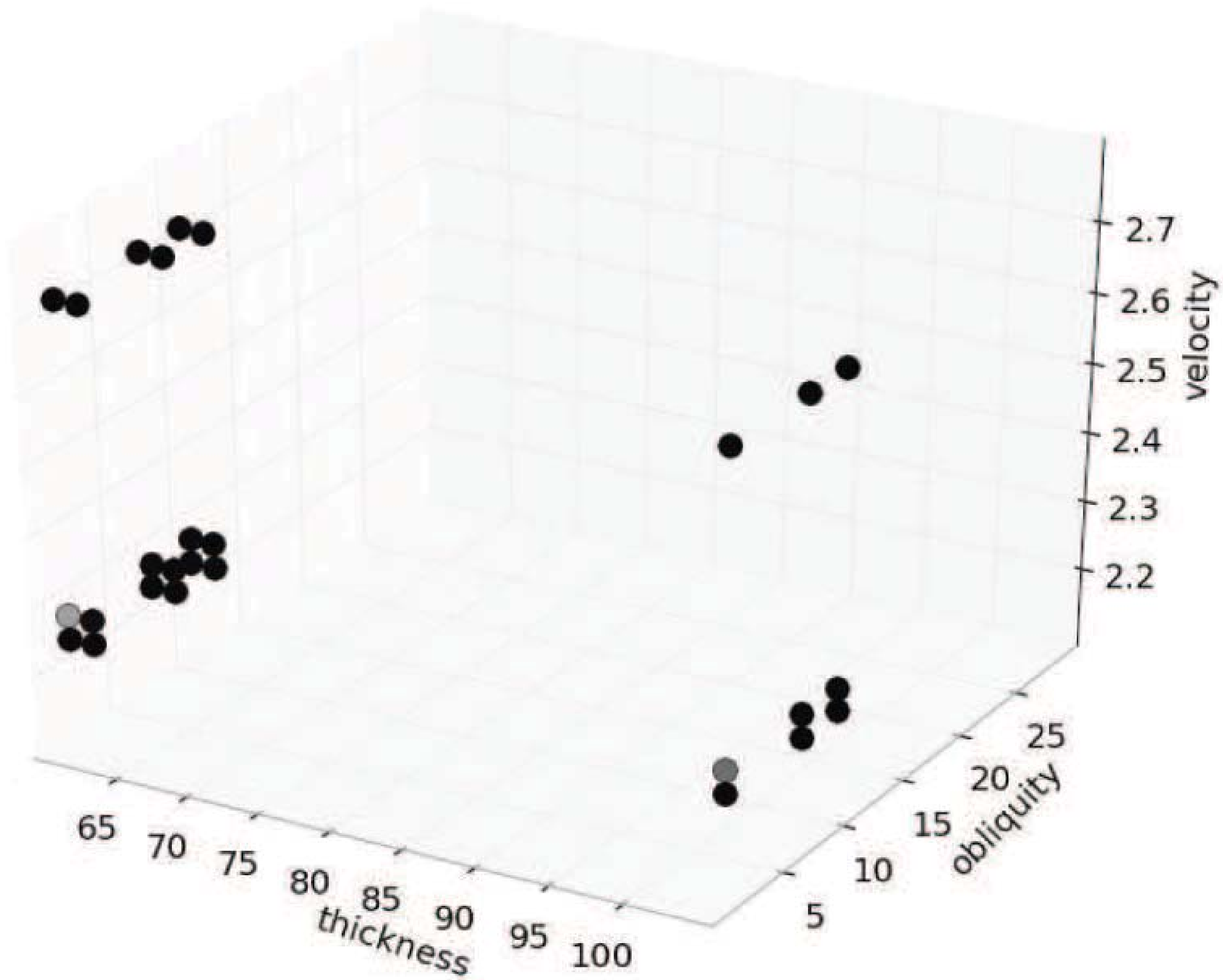
Extremizers are attractors

Initialization with 3 support points per marginal



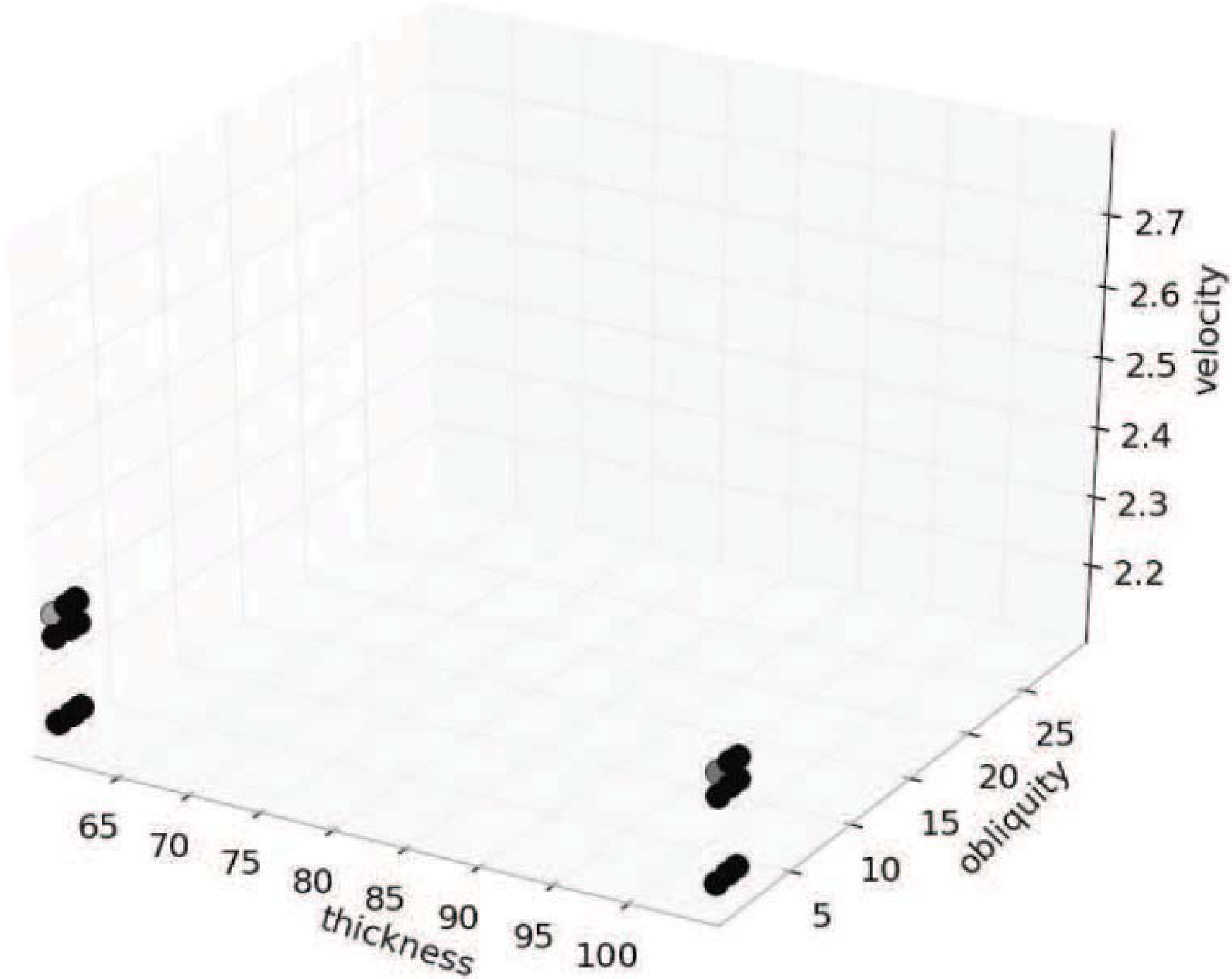
Support Points at iteration 0

Initialization with 3 support points per marginal



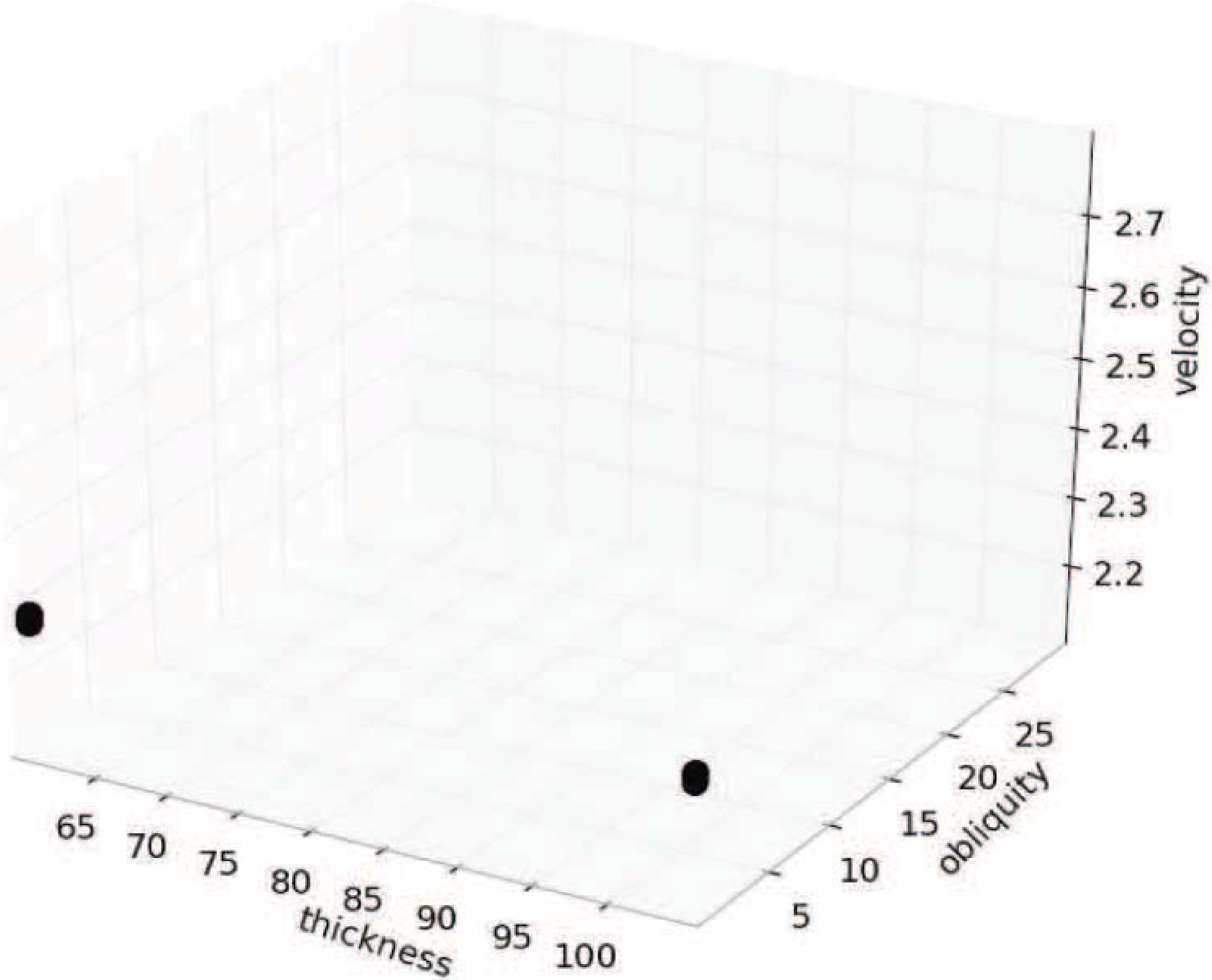
Support Points at iteration 500

Initialization with 3 support points per marginal



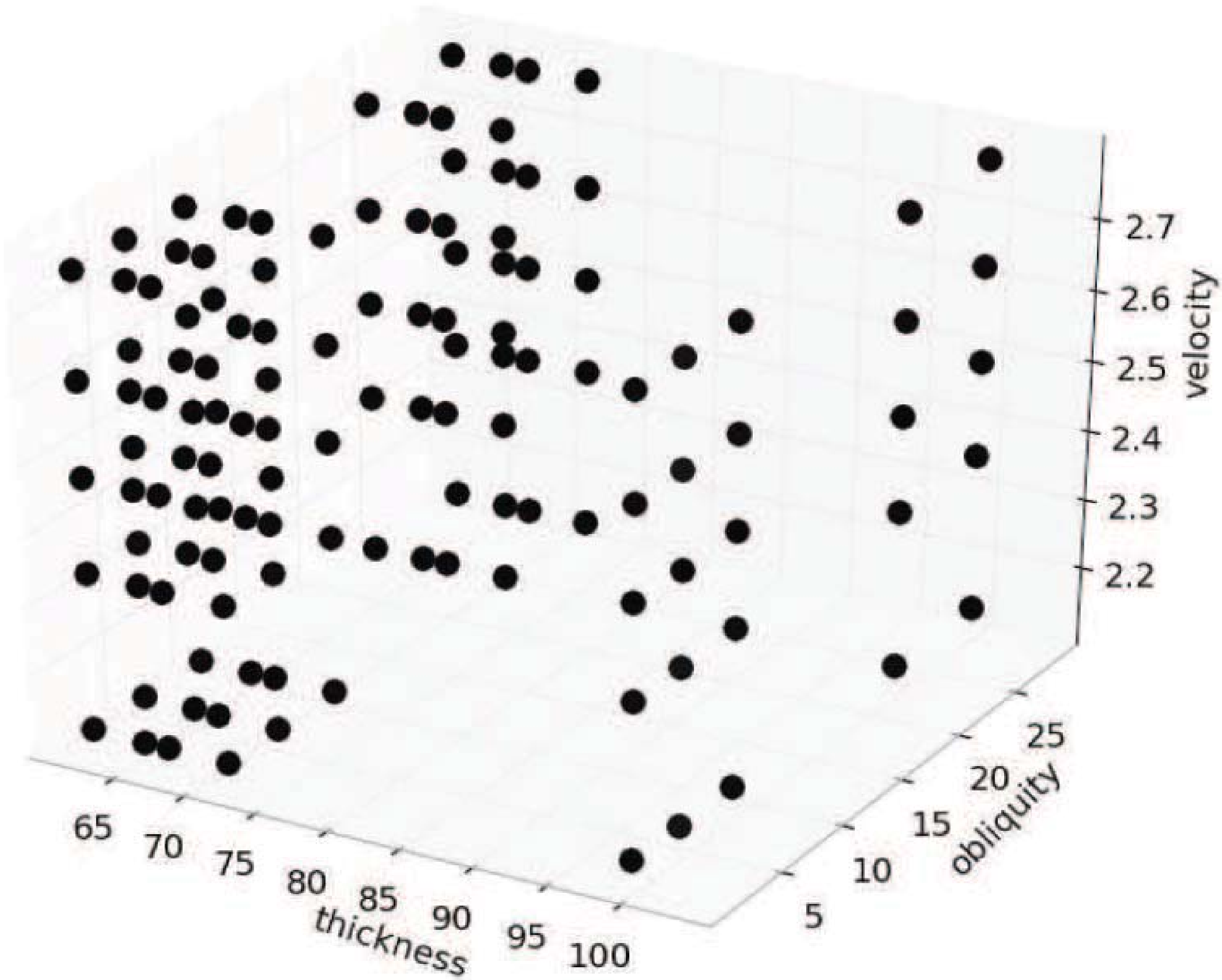
Support Points at iteration 1000

Initialization with 3 support points per marginal



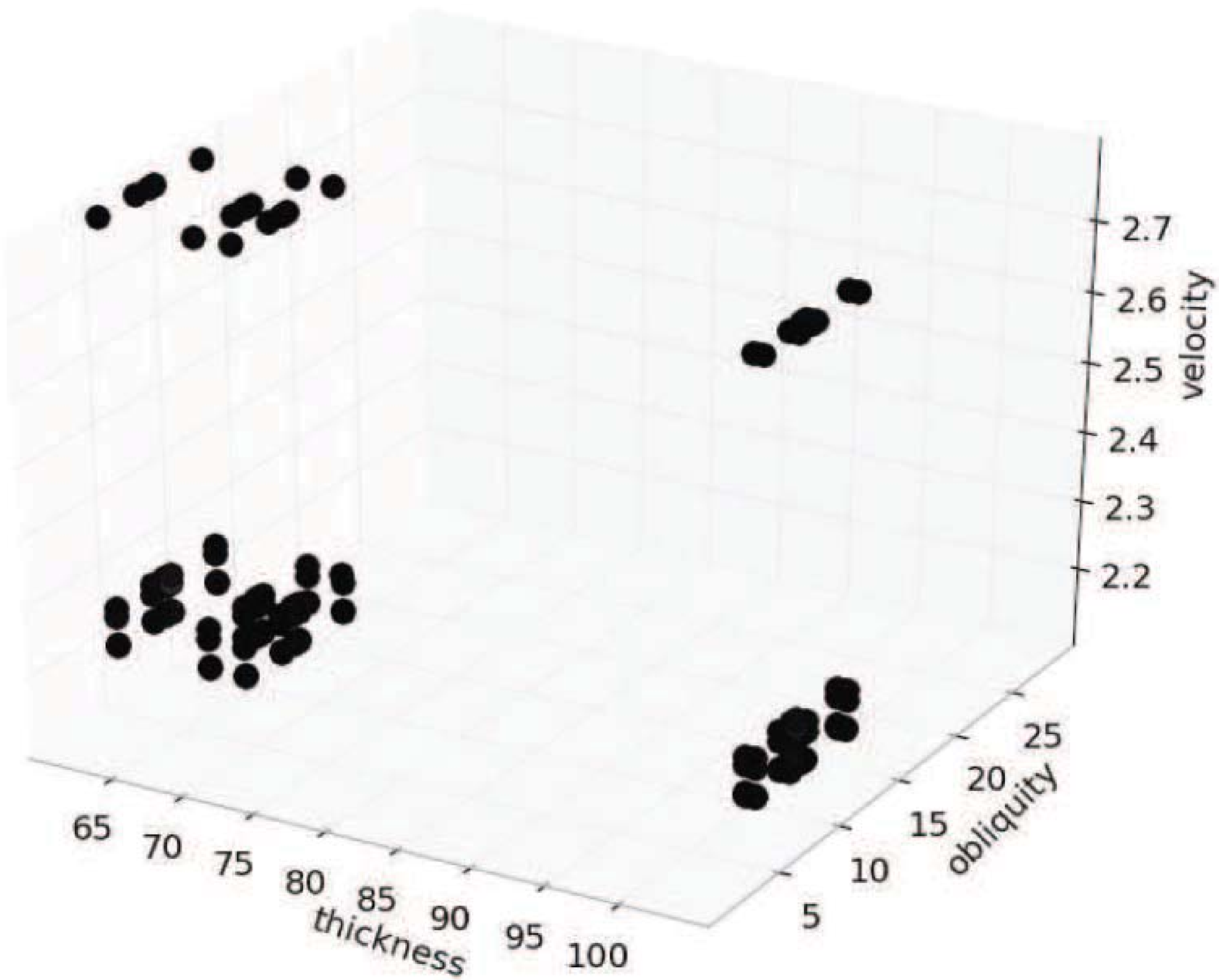
Support Points at iteration 2155

Initialization with 5 support points per marginal



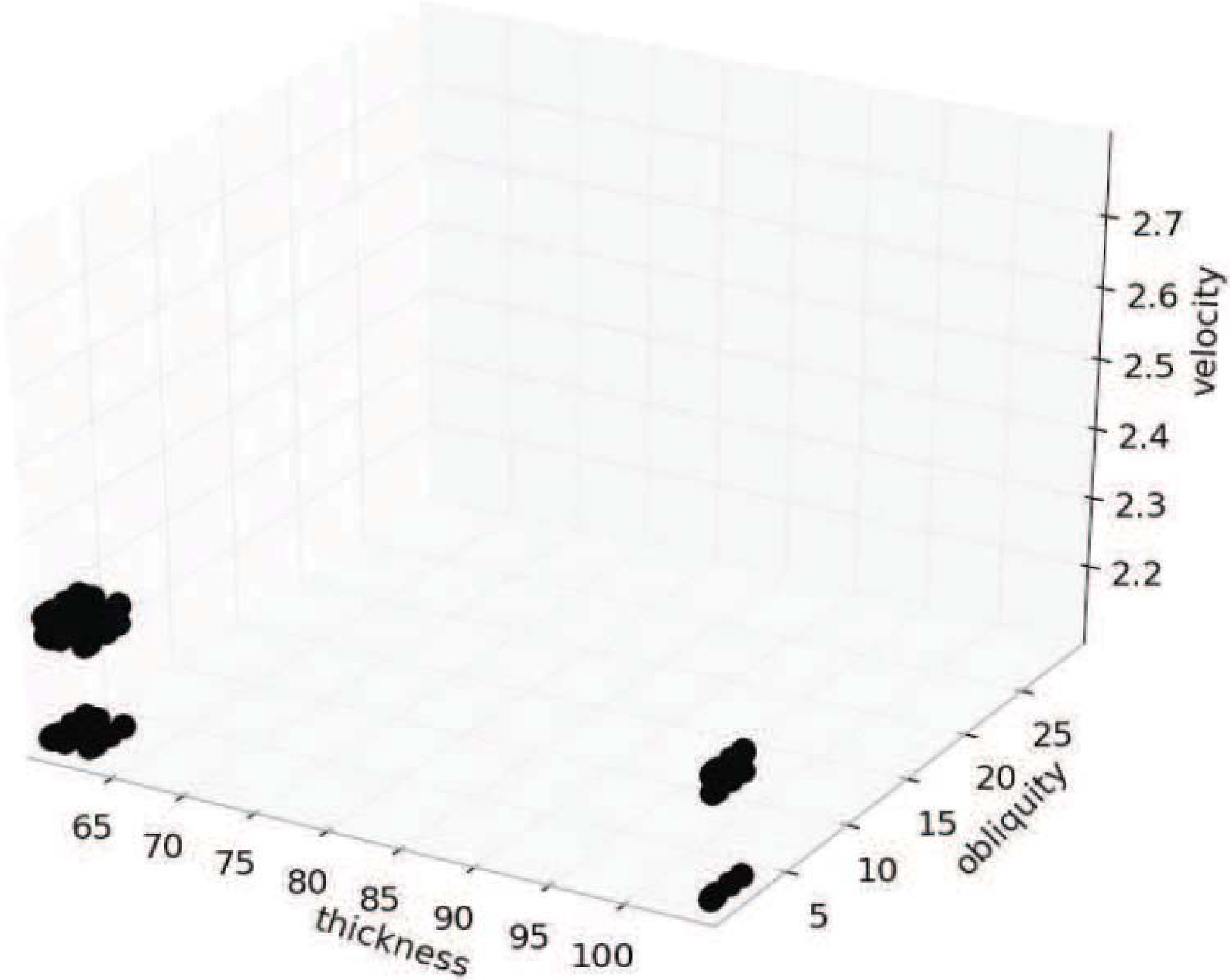
Support Points at iteration 0

Initialization with 5 support points per marginal



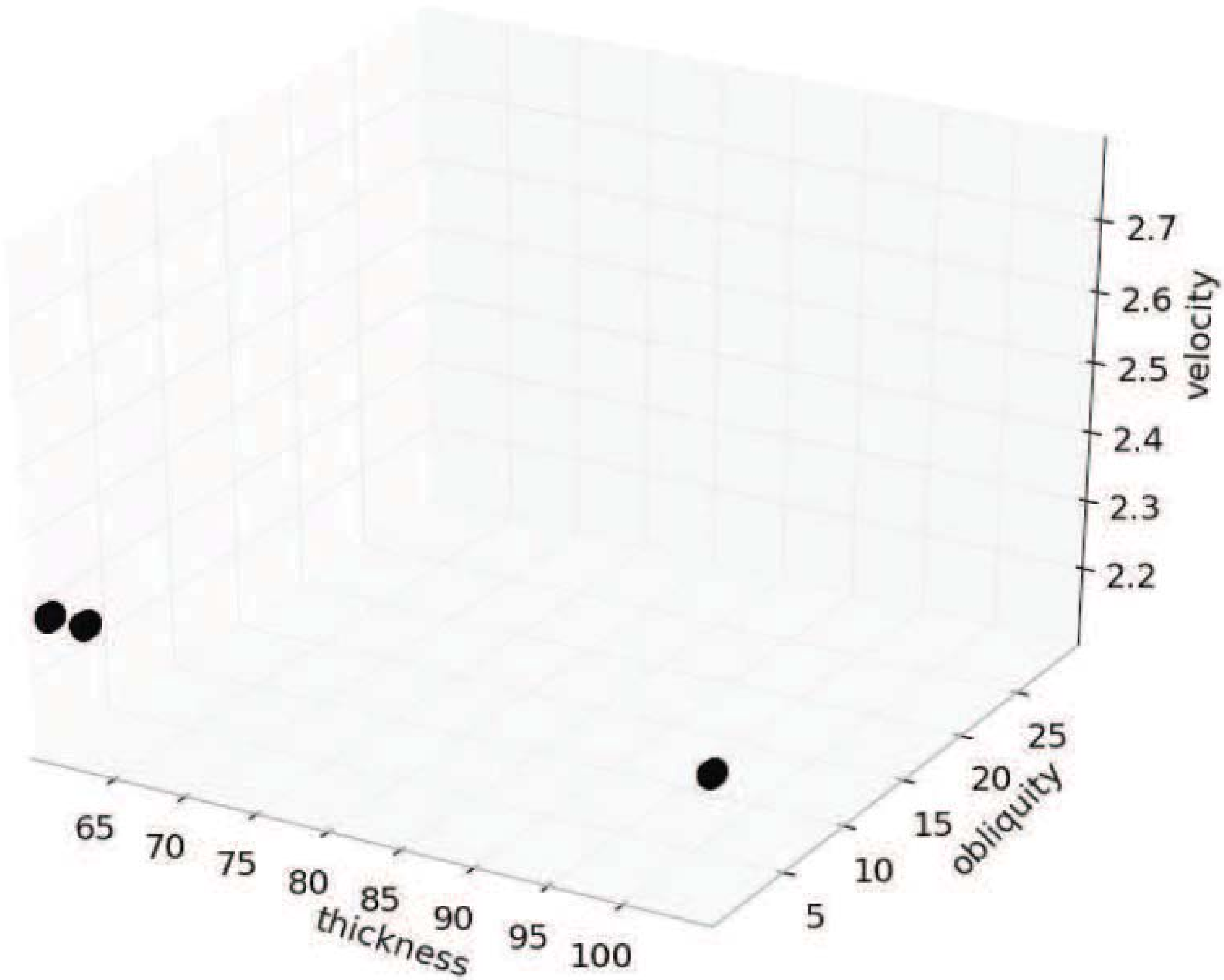
Support Points at iteration 1000

Initialization with 5 support points per marginal



Support Points at iteration 3000

Initialization with 5 support points per marginal



Support Points at iteration 7100

Unknown response function G

Objective

$$\text{We want least upper bound on } \mathbb{P}[G(h,\alpha,v) \leq \theta]$$

Constrain on input variables

$$h, \alpha, v\text{: independent}$$

$$(h,\alpha,v) \in [0.062, 0.125]\,\text{in} \times [0,30]\,\text{deg} \times [2300, 3200]\,\text{m}/$$

Constrain on the mean perf. area

$$\mathbb{E}[G(h,\alpha,v)] \geq 11.0\,\text{mm}^2$$

Modified Lipschitz continuity constrain on response function

$$|G(h,\alpha,v) - G(h',\alpha',v')| \leq d_L((h,\alpha,v),(h',\alpha',v')) + T,$$

$$d_L((h,\alpha,v),(h',\alpha',v')) := L_h|h-h'| + L_\alpha|\alpha-\alpha'| + L_v|v-v'|$$

$$L := (L_h, L_\alpha, L_v), \qquad T := 1.0\,\text{mm}^2,$$

$$L_h := 175.0\,\text{mm}^2/\text{in}, \quad L_\alpha := 0.075\,\text{mm}^2/\text{deg}, \quad L_v := 0.1\,\text{mm}^2/(\text{m/s}).$$

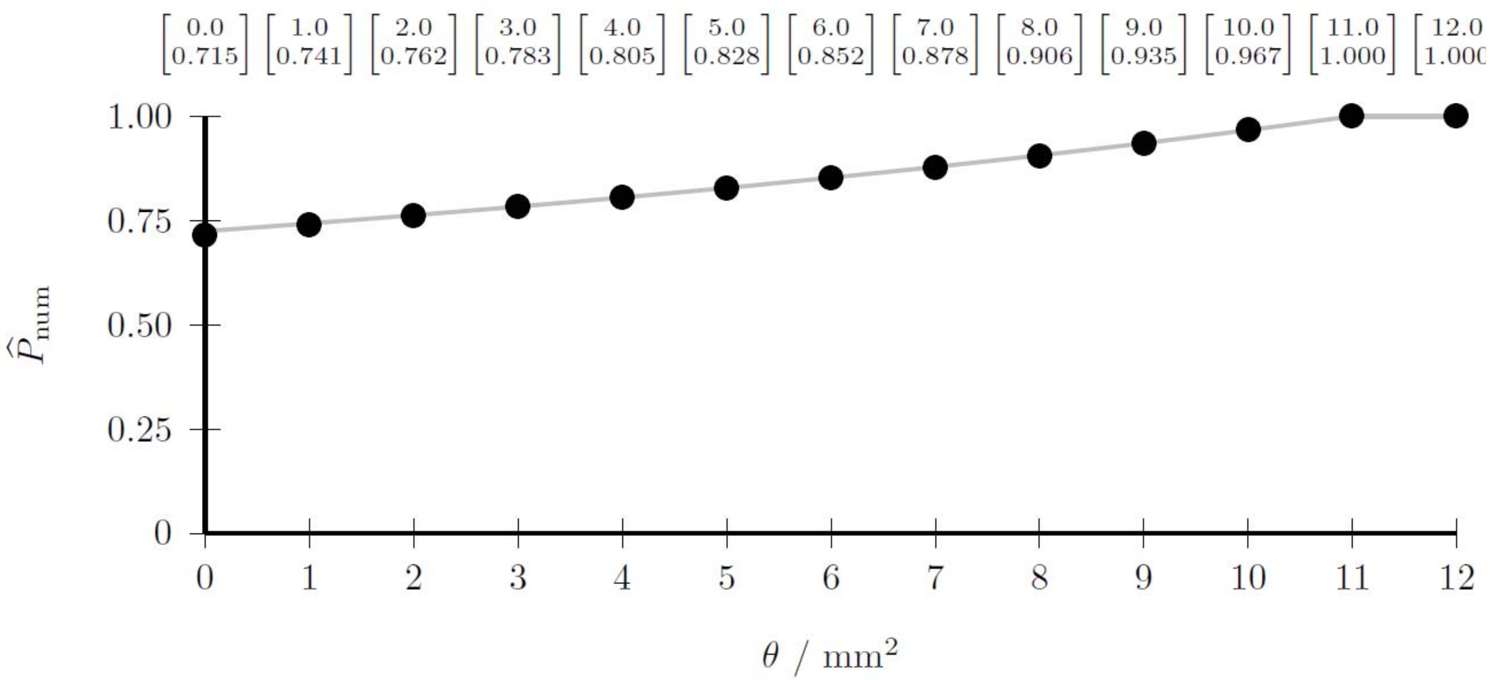
Legacy Data

32 data points

(steel-on-aluminium shots A48–A81) from summer 2010 at Caltech's SPHIR facility:

These constrain the value of G at 32 points

ID	h (inches)	α (degrees)	v (m/s)	$G(h, \alpha, v)$ (mm^2)
A48	0.062	0.0	2288.0	7.73
A49	0.125	30.0	2840.0	13.38
A50	0.125	0.0	2556.0	11.83
A51	0.062	30.0	2329.0	6.31
A52	0.062	0.0	2363.0	7.78
A53	0.125	0.0	2326.0	9.26
A54	0.125	30.0	3235.0	15.98
A55	0.062	0.0	2686.0	9.86
A56	0.062	30.0	2728.0	11.35
A57	0.062	30.0	2627.0	12.09
A58	0.125	30.0	2531.0	11.24
A60	0.125	0.0	2363.0	9.93
A61	0.062	0.0	2707.0	9.96
A62	0.062	30.0	2756.0	11.07
A63	0.062	0.0	2614.0	9.02
A64	0.125	0.0	2439.0	10.52
A65	0.062	0.0	2485.0	8.56
A66	0.125	0.0	2607.0	12.46
A67	0.125	30.0	3036.0	15.36
A68	0.125	30.0	2325.0	8.15
A69	0.062	30.0	2702.0	10.81
A70	0.062	30.0	2473.0	9.52
A71	0.121	30.0	2520.0	9.47
A72	0.121	0.0	2439.0	10.19
A73	0.121	30.0	2366.0	9.42
A74	0.121	30.0	2402.0	8.68
A75	0.062	30.0	2413.0	9.19
A77	0.062	30.0	2756.0	11.32
A78	0.121	30.0	2432.0	10.00
A79	0.062	30.0	2393.0	9.29
A80	0.121	30.0	2479.0	9.53
A81	0.060	30.0	2356.0	8.27



Least upper bound on $\mathbb{P}[G(h,\alpha,v) \leq \theta]$

The numerical results demonstrate agreement with the Markov bound

$$\mathbb{P}[G(h,\alpha,v) \leq \theta] \leq \frac{M-m}{M-\theta},$$

$$M := \sup_{(h,\alpha,v)\in\mathcal{X}} \inf_{z\in\mathcal{O}} \left(G(z) + d_L(z,(h,\alpha,v)) + T\right) \approx 39.895\,\mathrm{mm}^2$$

Only 2 data points out of 32 carry information about the optimal bound

Legacy Data

32 data points

(steel-on-aluminium shots A48–A81) from summer 2010 at Caltech's SPHIR facility:

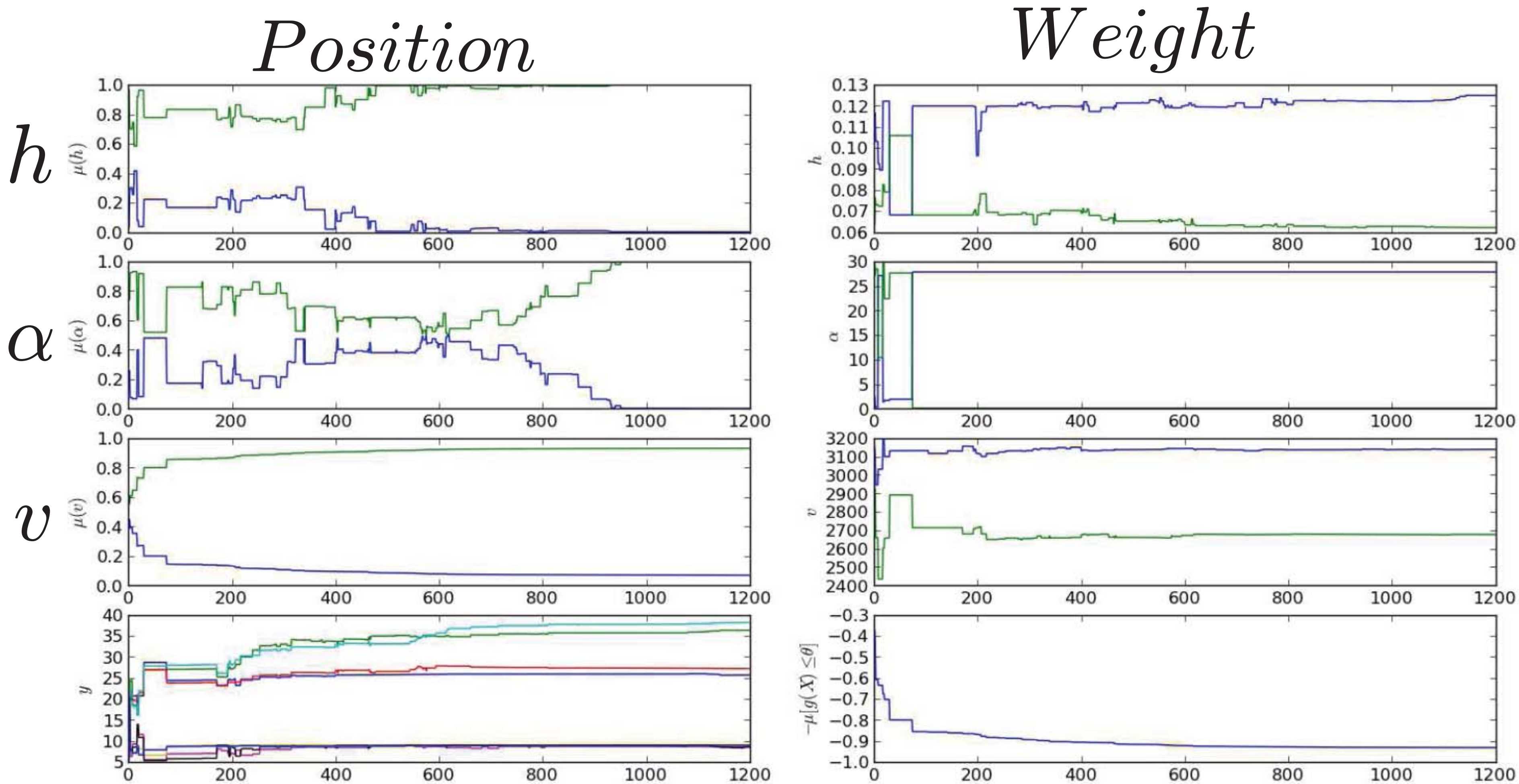
Only A54 and A67 carry information

The other 30 data points carry
no information about least upper bound
and could have be ignored.

ID	h (inches)	α (degrees)	v (m/s)	$G(h,\alpha,v)$ (mm^2)
A48	0.062	0.0	2288.0	7.73
A49	0.125	30.0	2840.0	13.38
A50	0.125	0.0	2556.0	11.83
A51	0.062	30.0	2329.0	6.31
A52	0.062	0.0	2363.0	7.78
A53	0.125	0.0	2326.0	9.26
A54	0.125	30.0	3235.0	15.98
A55	0.062	0.0	2686.0	9.86
A56	0.062	30.0	2728.0	11.35
A57	0.062	30.0	2627.0	12.09
A58	0.125	30.0	2531.0	11.24
A60	0.125	0.0	2363.0	9.93
A61	0.062	0.0	2707.0	9.96
A62	0.062	30.0	2756.0	11.07
A63	0.062	0.0	2614.0	9.02
A64	0.125	0.0	2439.0	10.52
A65	0.062	0.0	2485.0	8.56
A66	0.125	0.0	2607.0	12.46
A67	0.125	30.0	3036.0	15.36
A68	0.125	30.0	2325.0	8.15
A69	0.062	30.0	2702.0	10.81
A70	0.062	30.0	2473.0	9.52
A71	0.121	30.0	2520.0	9.47
A72	0.121	0.0	2439.0	10.19
A73	0.121	30.0	2366.0	9.42
A74	0.121	30.0	2402.0	8.68
A75	0.062	30.0	2413.0	9.19
A77	0.062	30.0	2756.0	11.32
A78	0.121	30.0	2432.0	10.00
A79	0.062	30.0	2393.0	9.29
A80	0.121	30.0	2479.0	9.53
A81	0.060	30.0	2356.0	8.27

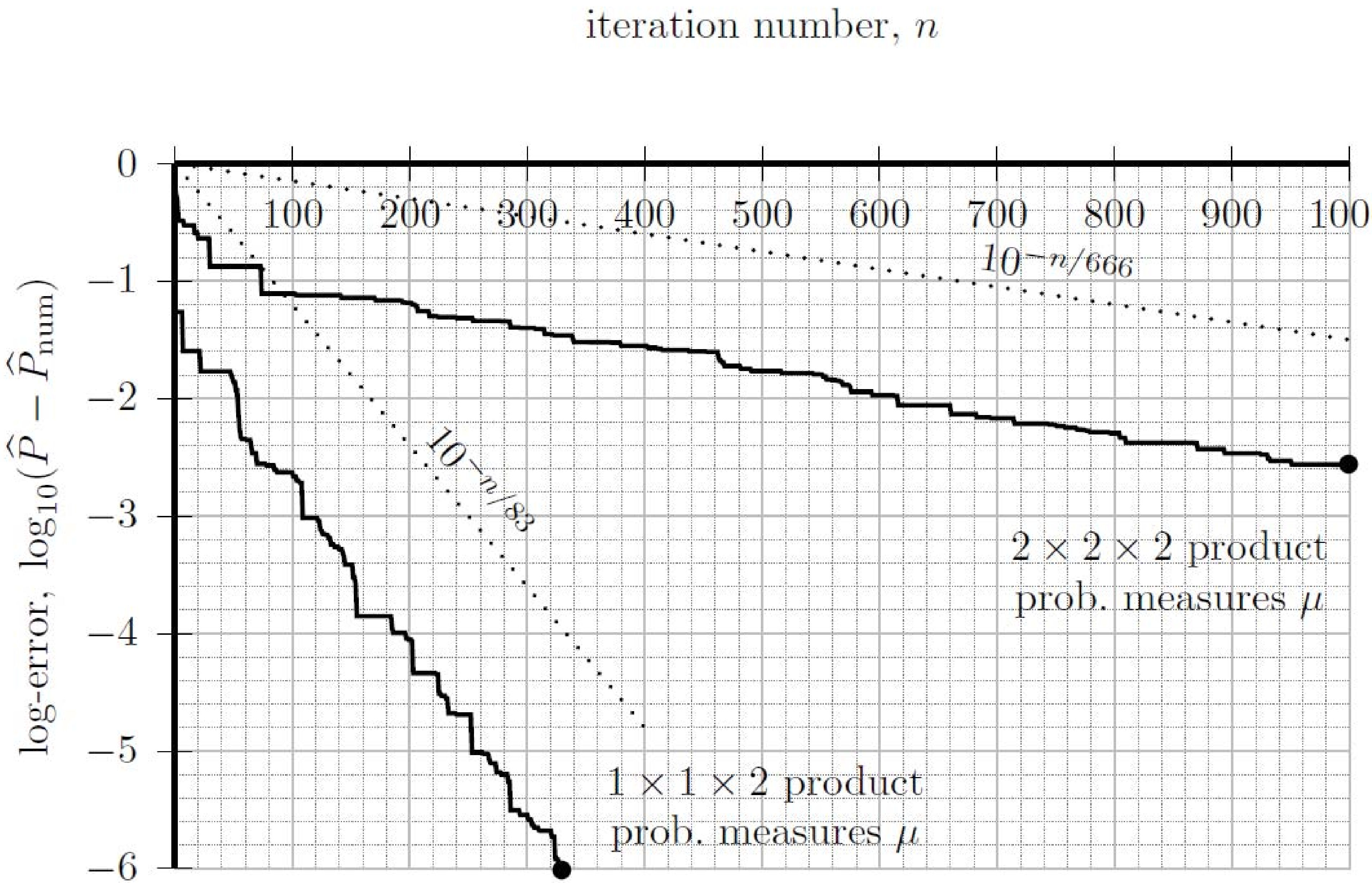
Dimensional collapse

$\theta = 9\,\mathrm{mm}^2$



μ collapses from a $2 \times 2 \times 2$ measure to a $1 \times 1 \times 2$ measure

At the optimum only the v marginal has support on 2 points

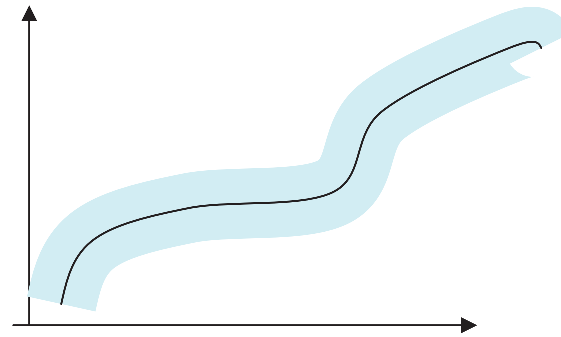


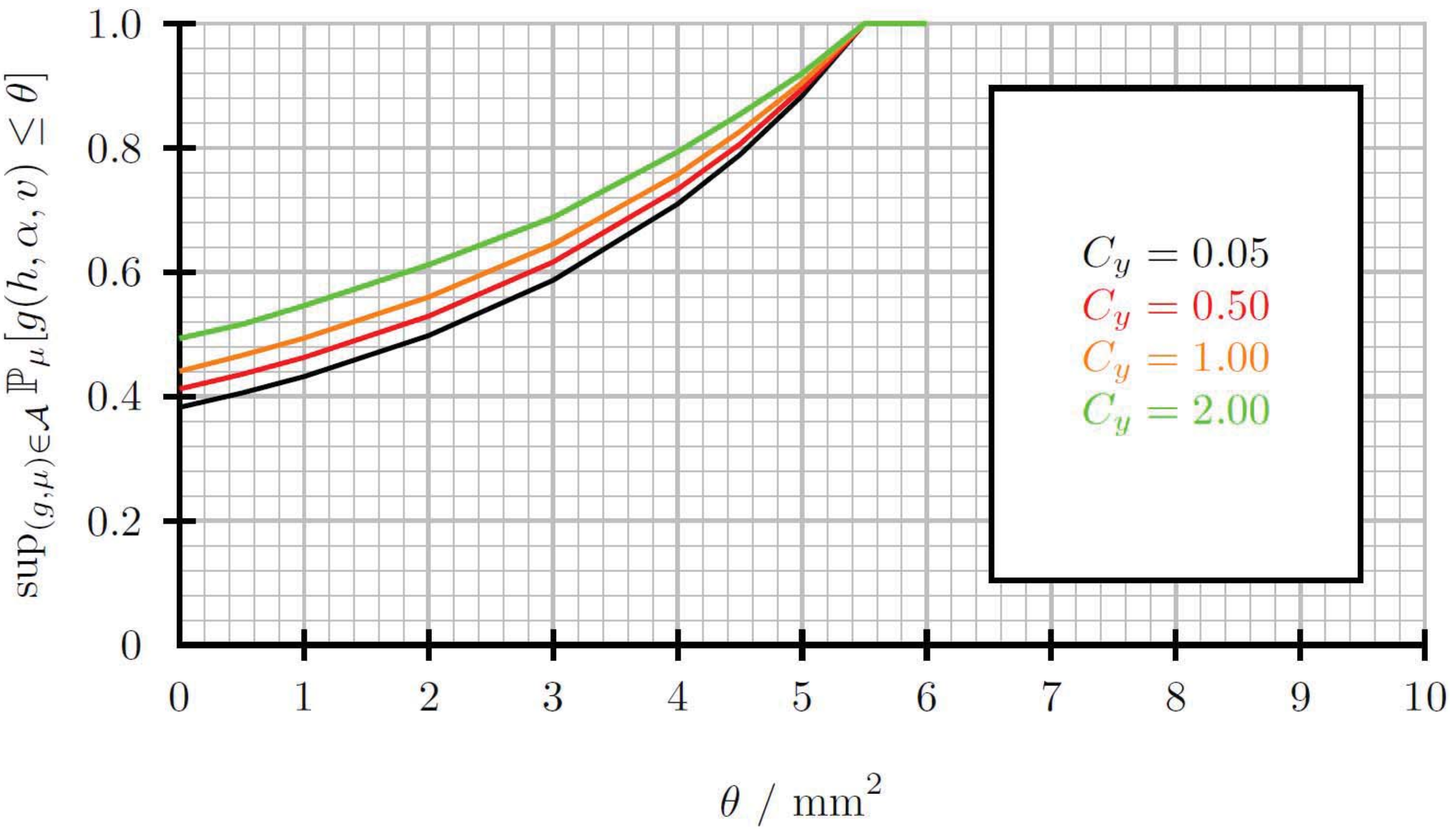
iteration number, n
0
100
200
300
400
500
600
700
800
900
1000
−1
−2
−3
−4
−5
−6
log-error, $\log_{10}(\widehat{P} - \widehat{P}_{\mathrm{num}})$
$10^{-n/666}$
$10^{-n/83}$
$2 \times 2 \times 2$ product
prob. measures μ
$1 \times 1 \times 2$ product
prob. measures μ

OUQ with sausage around a model

$$\mathcal{X} := [60, 105]\,\mathrm{mil} \times [0, 30]\,\mathrm{deg} \times [2.1, 2.8]\,\mathrm{km/s}$$

$$\mathcal{A} := \left\{ (g, \mu) \;\middle|\; \begin{array}{c} g\colon \mathcal{X} \to \mathbb{R}, \\ \mu = \mu_h \otimes \mu_\alpha \otimes \mu_v \in \mathcal{P}(\mathcal{X}), \\ \mathbb{E}_\mu[g(h, \alpha, v)] \geq 5.5\,\mathrm{mm}^2, \\ \|g - F_{\mathrm{StStSurr}}\|_\infty \leq C_y \end{array} \right\}$$





Optimal bounds for other admissible sets

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

Should we compare those bounds to the true P.O.F.?

One should be careful with such comparisons in presence of asymmetric information

The real question is how to construct a selective information set A.

Selection of the most decisive experiment

$$\mathcal{A} = \mathcal{A}_{\text{safe}} \cup \mathcal{A}_{\text{unsafe}}$$

$$\mathcal{A}_{\text{safe}} = \{(\mu, f) \in \mathcal{A} \colon \mu[f(X) \geq a] \leq \epsilon\}$$

$$\mathcal{A}_{\text{unsafe}} = \{(\mu, f) \in \mathcal{A} \colon \mu[f(X) \geq a] > \epsilon\}$$

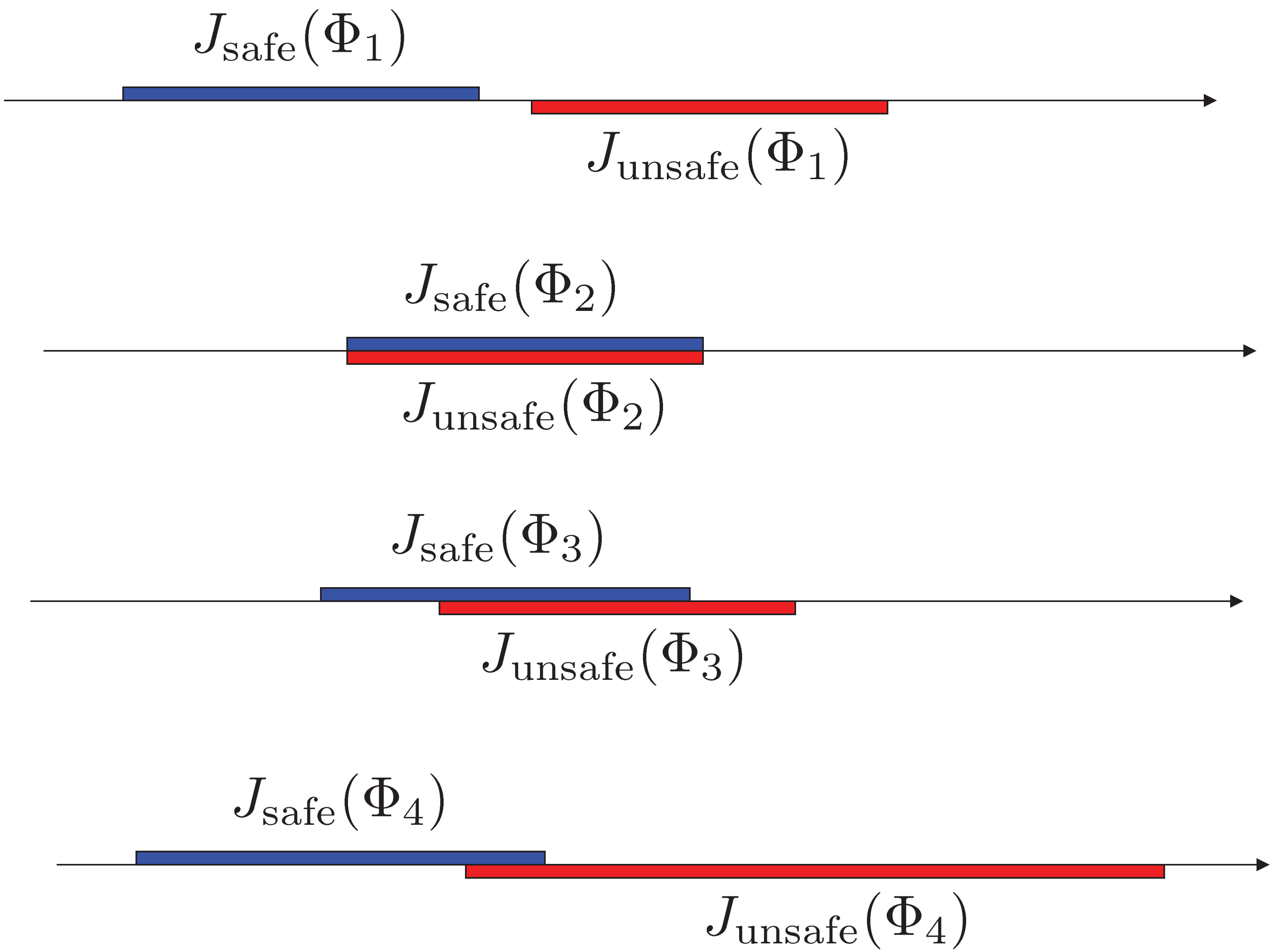
Experiments $\Phi(G, \mathbb{P})$

Ex: $\Phi_1(G, \mathbb{P}) = \mathbb{P}[X \in A]$ $\Phi_2(G, \mathbb{P}) = \mathbb{E}_{\mathbb{P}}[G]$

$$J_{\text{safe}}(\Phi) := \left[\inf_{f, \mu \in \mathcal{A}_{\text{safe}}} \Phi(f, \mu), \sup_{f, \mu \in \mathcal{A}_{\text{safe}}} \Phi(f, \mu) \right]$$

$$J_{\text{unsafe}}(\Phi) := \left[\inf_{f, \mu \in \mathcal{A}_{\text{unsafe}}} \Phi(f, \mu), \sup_{f, \mu \in \mathcal{A}_{\text{unsafe}}} \Phi(f, \mu) \right]$$

Selection of the most decisive experiment

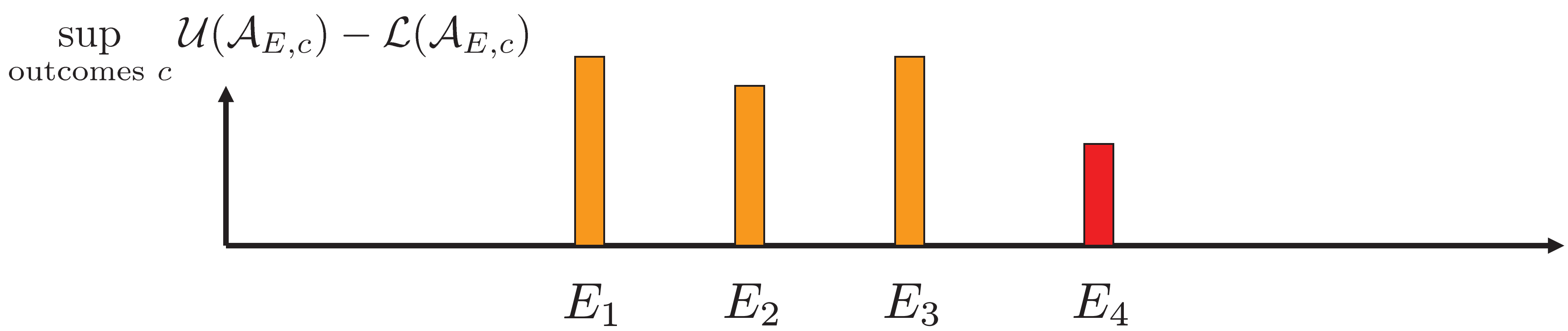


Selection of the most predictive experiment

$$\mathcal{L}(\mathcal{A}) \leq \mathbb{P}[G(X) \geq a] \leq \mathcal{U}(\mathcal{A})$$

- If your objective is to have an "accurate" prediction of $\mathbb{P}[G(X) \leq \theta]$ in the sense that $\mathcal{U}(\mathcal{A}) - \mathcal{L}(\mathcal{A})$ is small, then proceed as follows:
- Let $\mathcal{A}_{E,c}$ denote those scenarios in $\mathcal{A}$ that are compatible with obtaining outcome c from experiment E.
- The experiment that is most predictive even in the worst case is defined by a minimax criterion: we seek

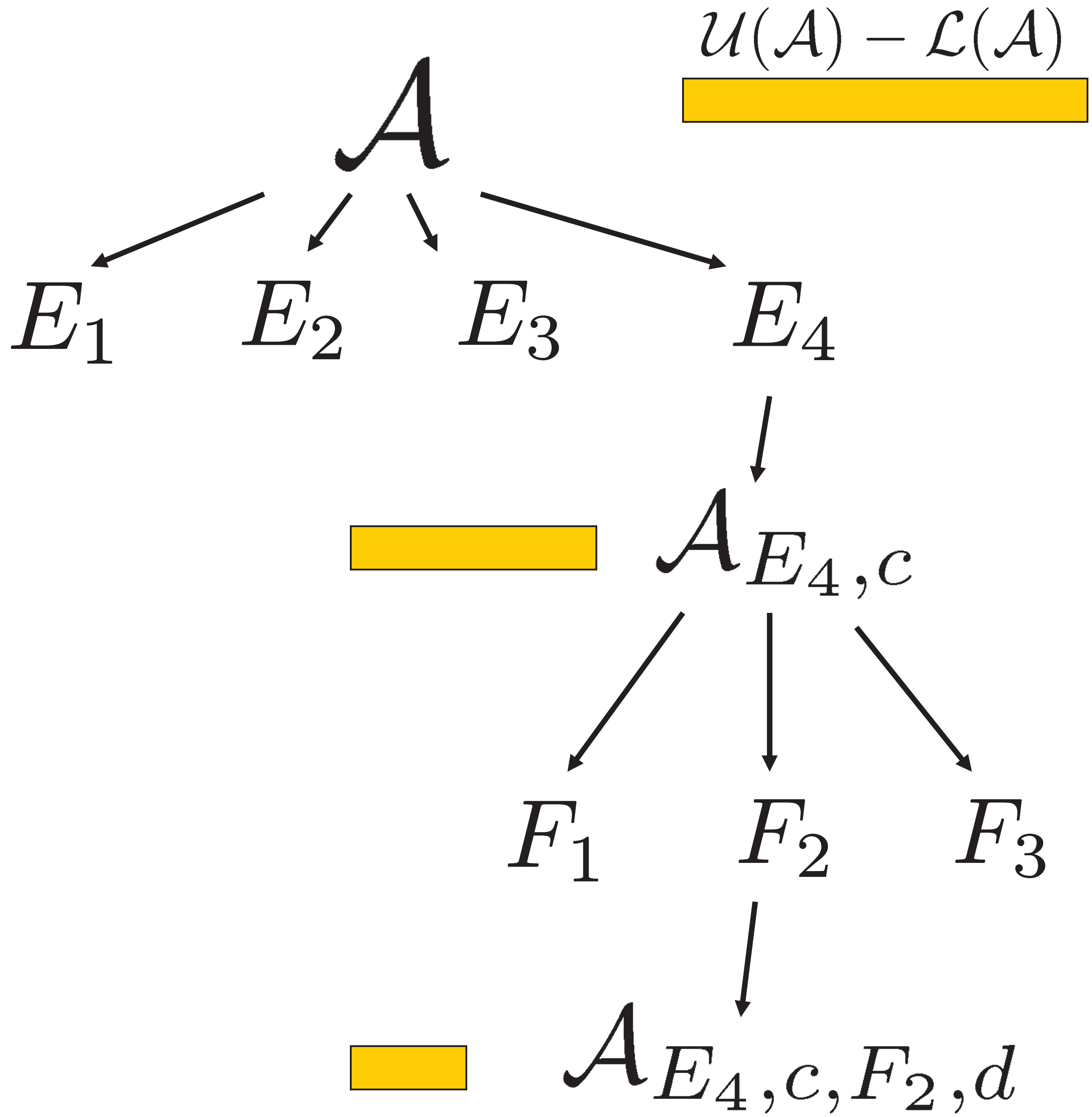
$$E^* \in \operatorname*{arg\,min}_{\text{experiments } E} \left(\sup_{\text{outcomes } c} \left(\mathcal{U}(\mathcal{A}_{E,c}) - \mathcal{L}(\mathcal{A}_{E,c}) \right) \right).$$



- This idea of experimental selection can be extended to plan several experiments in advance, *i.e.* to plan campaigns of experiments.



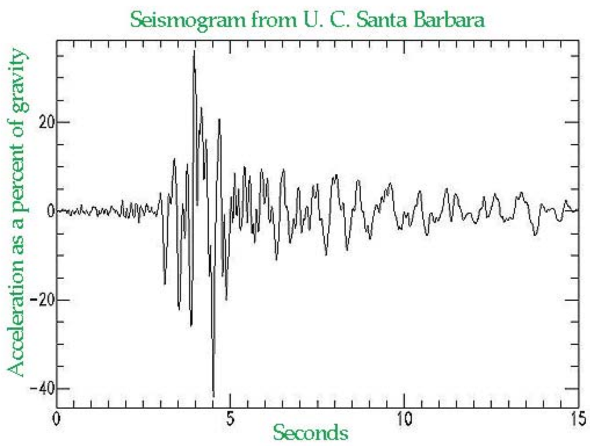


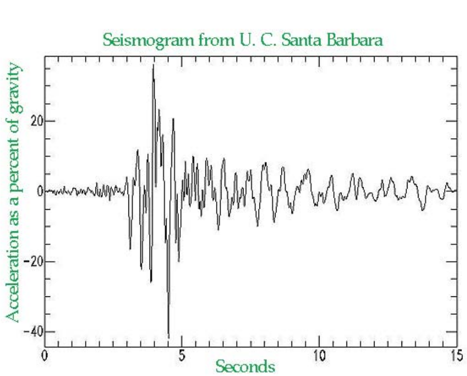



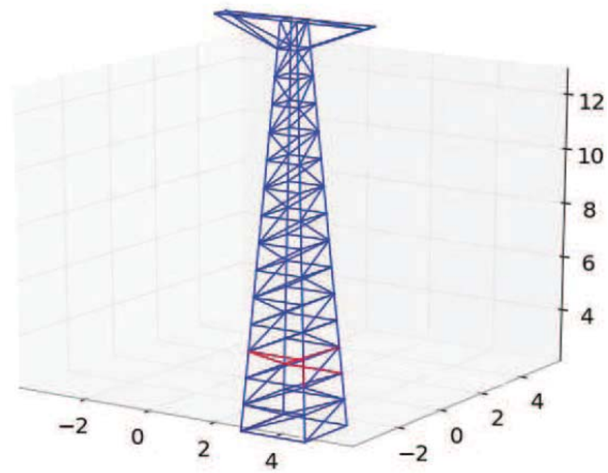
Plan several experiments in advance, i.e. campaigns of experiments

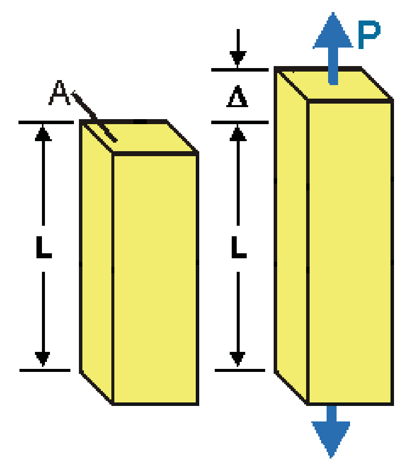
- This is a kind of infinite-dimensional *Cluedo*, played on spaces of admissible scenarios, against our lack of perfect information about reality, and made tractable by the reduction theorems.

Seismic Safety Assessment of Truss Structures









$$a(t) \xrightarrow{\quad F \quad} F(a)$$

Ground Acceleration

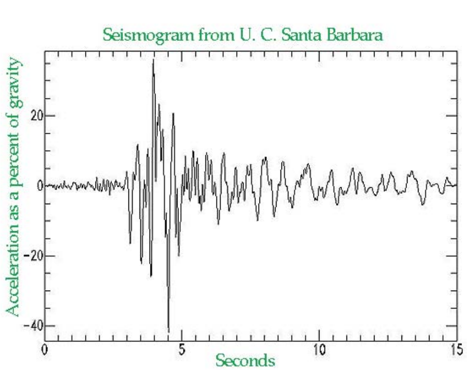
F

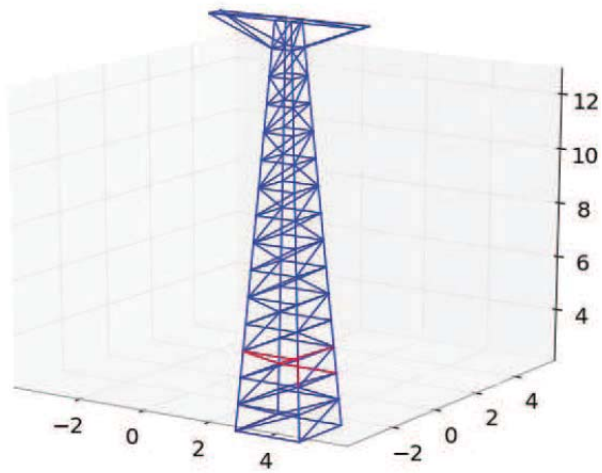
min(Yield Strain - Axial Strain)

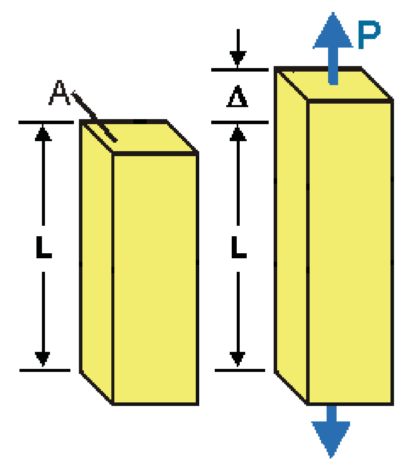
$$F(a) = \min_i (S_i - \|Y_i\|_\infty)$$

S_i: Yield strain of member i

$Y_i(t)$: Axial strain of member i







$$a(t)$$

Ground Acceleration

F

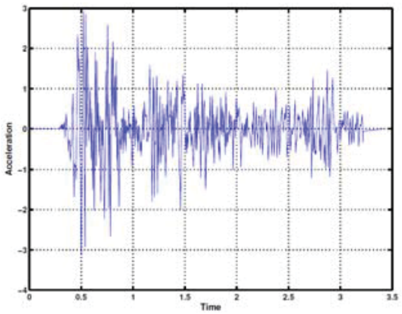
$$F(a)$$

min(Yield Strain - Axial Strain)

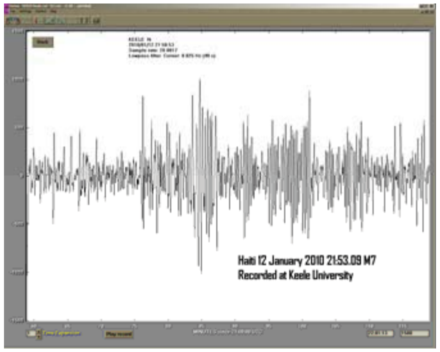
We want to certify that

$$\mathbb{P}\left[F(a) \leq 0\right] \leq \epsilon$$

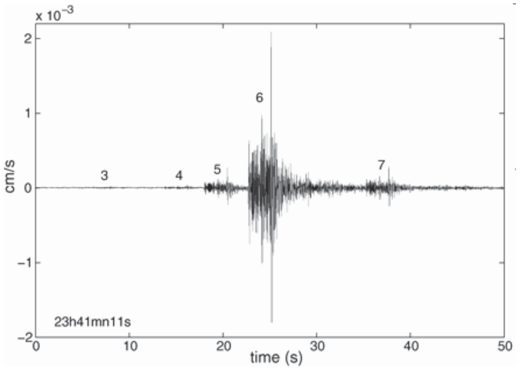
Historical Data Method



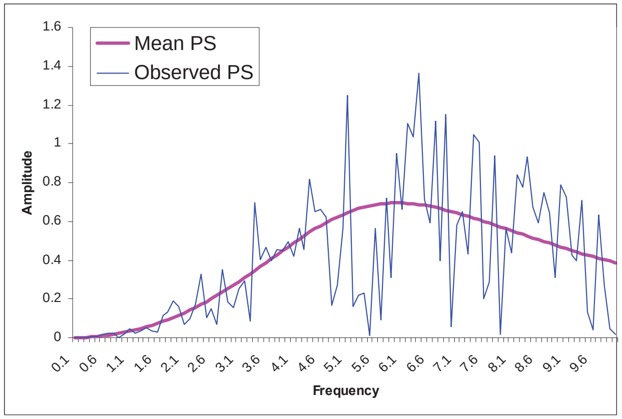
1940 Elcentro

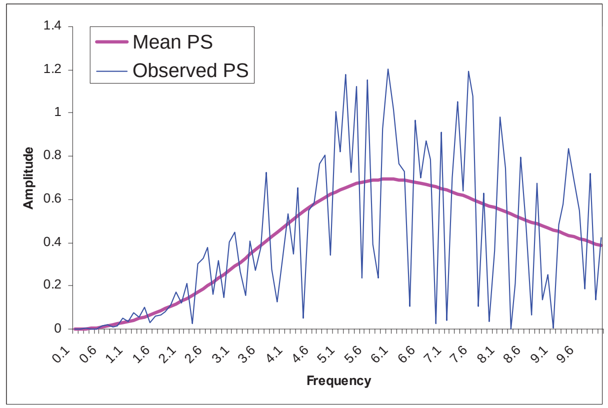


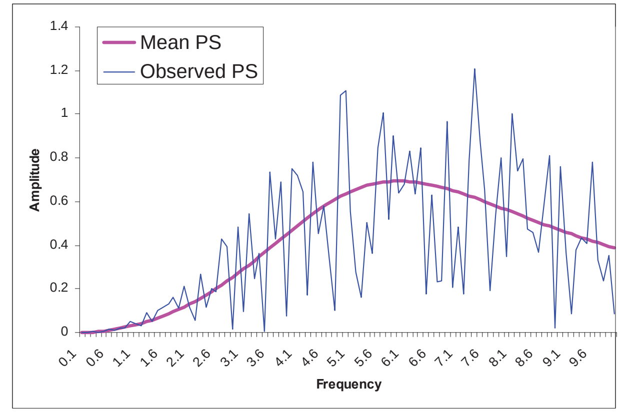
2010 Haiti



1999 Izmit



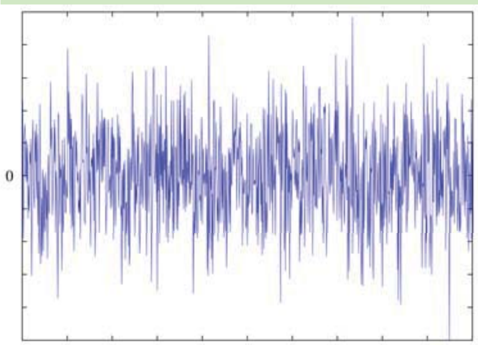




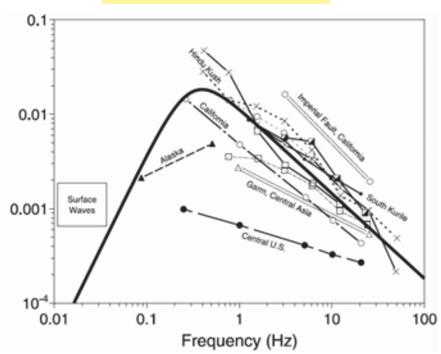
Matsuda-Asano shape function (mean power spectrum)

$$s(\omega) := \frac{\omega_g^2 \omega^2}{(\omega_g^2 - \omega^2)^2 + 4\xi_g^2 \omega_g^2 \omega^2}$$

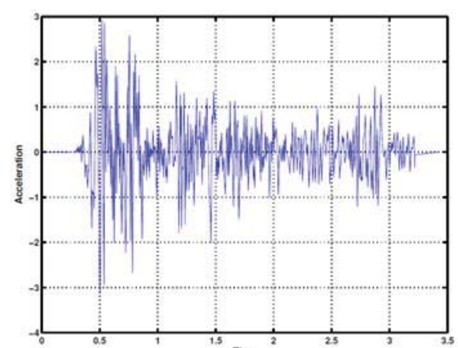
White noise



Filter



Ground acceleration

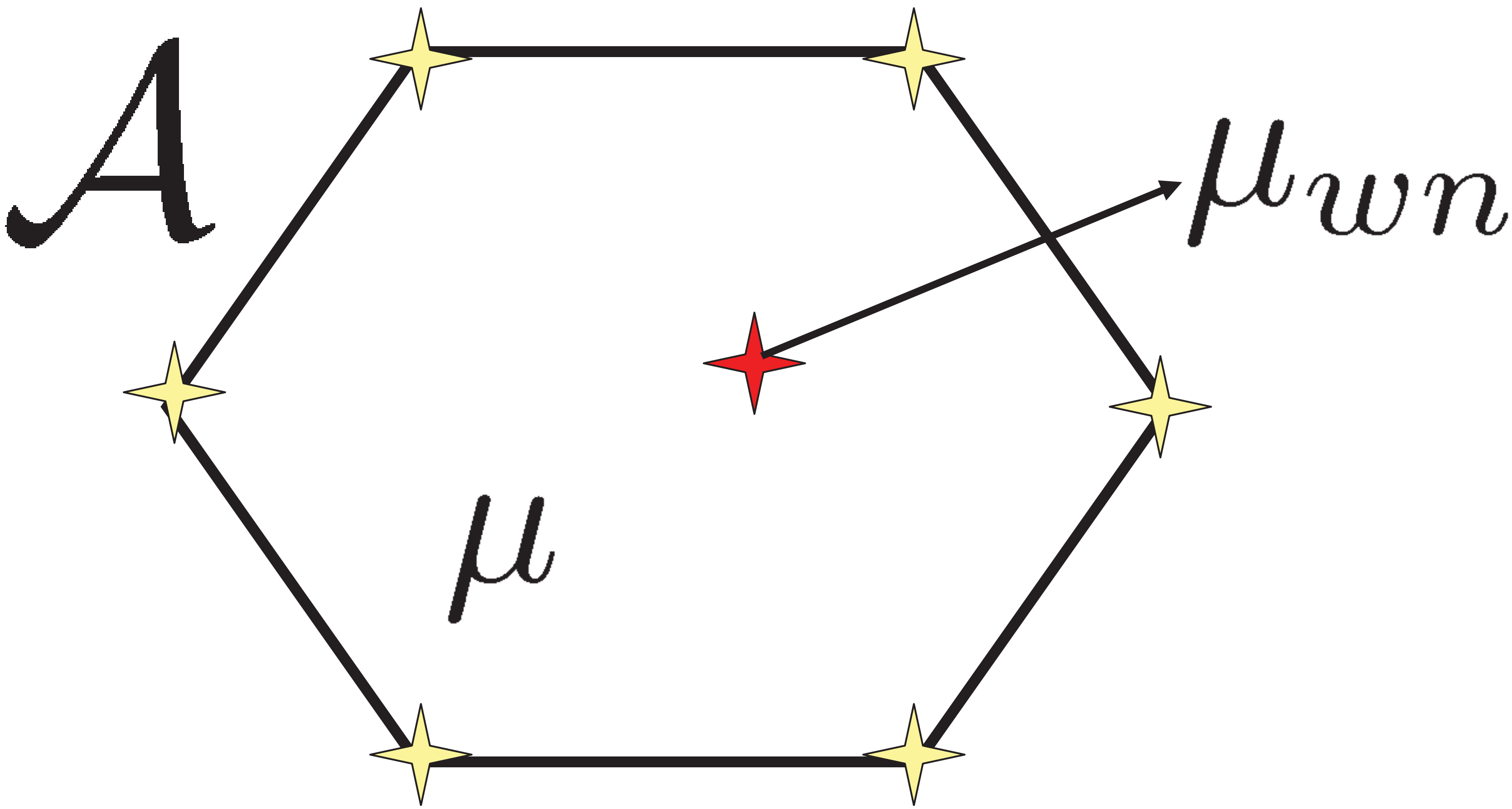


OUQ vs Filtered White Noise

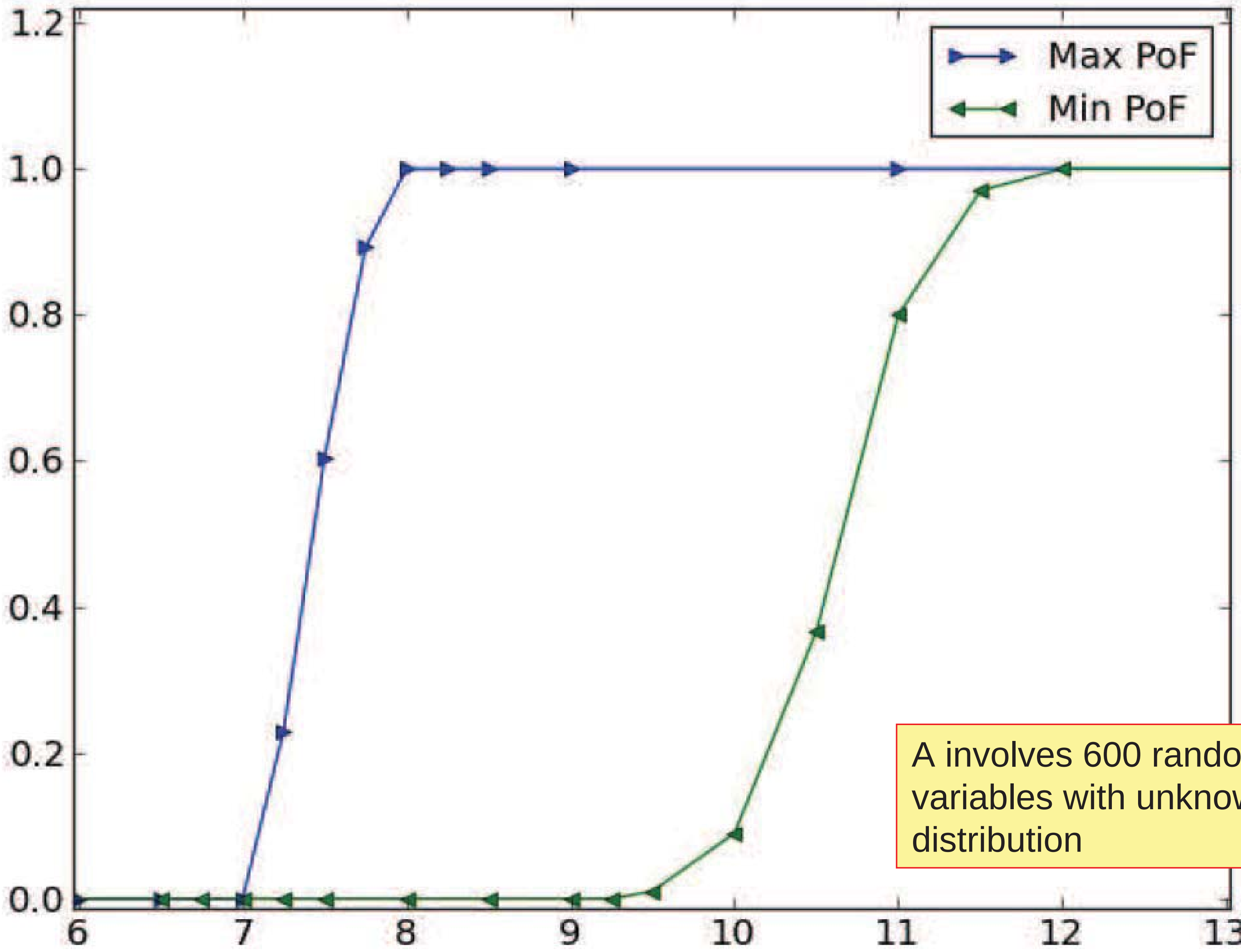
$\mathcal{A}$: Set of measures μ on a

Maximum grounded acceleration bounded

Mean power spectrum given



Vulnerability Curves (vs earthquake magnitude)



A involves 600 random variables with unknown distribution

Modeling in the frequency domain

$$a := \sum_{k=1}^{W} \Big((A_{6k-5}, A_{6k-4}, A_{6k-3}) \cos(2\pi\omega_k t) + (A_{6k-2}, A_{6k-1}, A_{6k}) \sin(2\pi\omega_k t) \Big)$$

$$\omega_k := \frac{k}{T} \quad T = 20\,\mathrm{s} \quad W := 100$$

$$\frac{1}{T} \int_0^T |a|^2 \, dt \leq \frac{a_{\max}^2}{2}$$

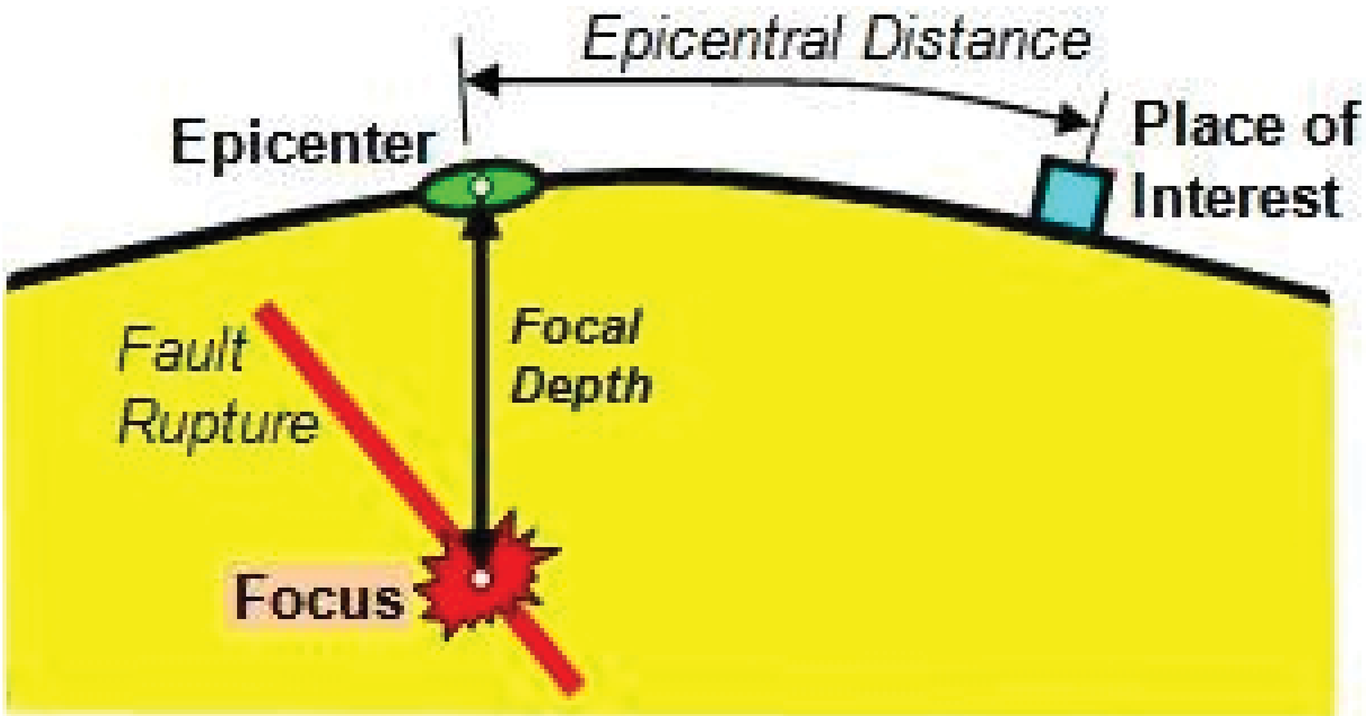
$$\Updownarrow$$

$$A := (A_1, \ldots, A_{6W})$$

$$\mathbb{P}\big[A \in B(0, a_{\max})\big] = 1$$

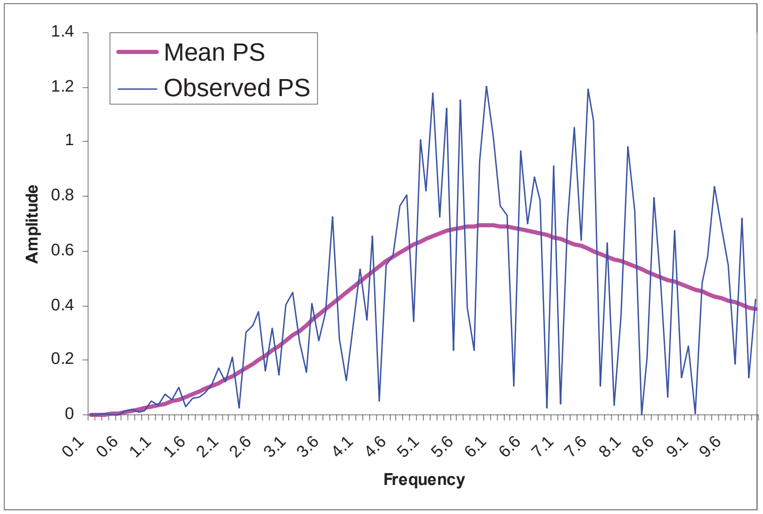
Esteva's semi-empirical expression

$$a_{\max} := \frac{a_0 e^{\lambda M_L}}{(R_0 + R)^2}$$



R: source to site distance

$$\mathbb{E}[A_i^2] = b_i$$



$$b_{6k-j} = \frac{a_{\max}^2}{12} \frac{s(\omega_k)}{\sum_{n=1}^{W} s(\omega_n)}] \qquad j \in \{0, \ldots, 5\}$$

Matsuda-Asano shape function

$$s(\omega) := \frac{\omega_g^2 \omega^2}{(\omega_g^2 - \omega^2)^2 + 4\xi_g^2 \omega_g^2 \omega^2}$$

ω_g: natural frequency of the site.

ξ_g: natural damping factor of the site.

Modeling in the frequency domain

Number of truss structure (electric tower) members : **198**

Number of random Fourier coefficients (with unknown pdf): **600**

Dimension of the Reduced Problem **1200**

Reduced problem solved with a Differential Evolution Algorithm modified to use large-scale parallel computing resources

Differential Evolution Algorithm population size **40**

High performance computer cluster: **88 cores**

shc (PSAAP) with **11 core-4 nodes** (44 total)
foxtrot (DANSE) with **4 core-12 nodes, 11/12** (44 total)

Convergence time: **15 hours**

Number of iterations: **2000**

Number of function evaluations: **35,000 to 50,000**

Punch lines
and
Important points to remember

OUQ is the business of finding optimal bounds on quantities of interest given the information at hand.

You want to estimate $\Phi(G, \mathbb{P})$

Example $\Phi(f, \mu) := \mu[f \geq a]$

You only know $(G, \mathbb{P}) \in \mathcal{A}$

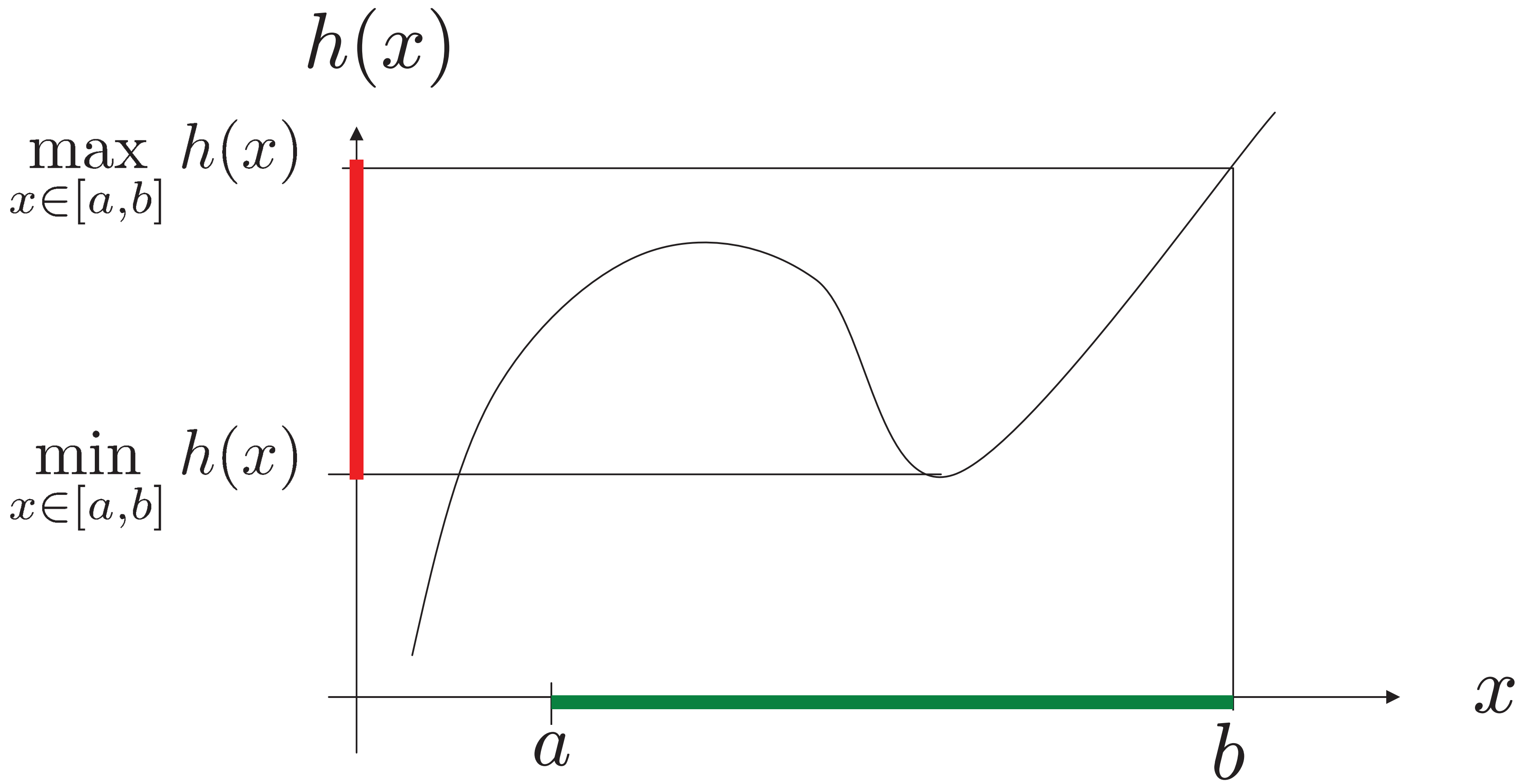
You compute $\Phi(G, \mathbb{P})$

$$\mathcal{U}(\mathcal{A}) := \sup_{(f,\mu)\in\mathcal{A}} \Phi(f, \mu)$$

$$\mathcal{L}(\mathcal{A}) := \inf_{(f,\mu)\in\mathcal{A}} \Phi(f, \mu)$$

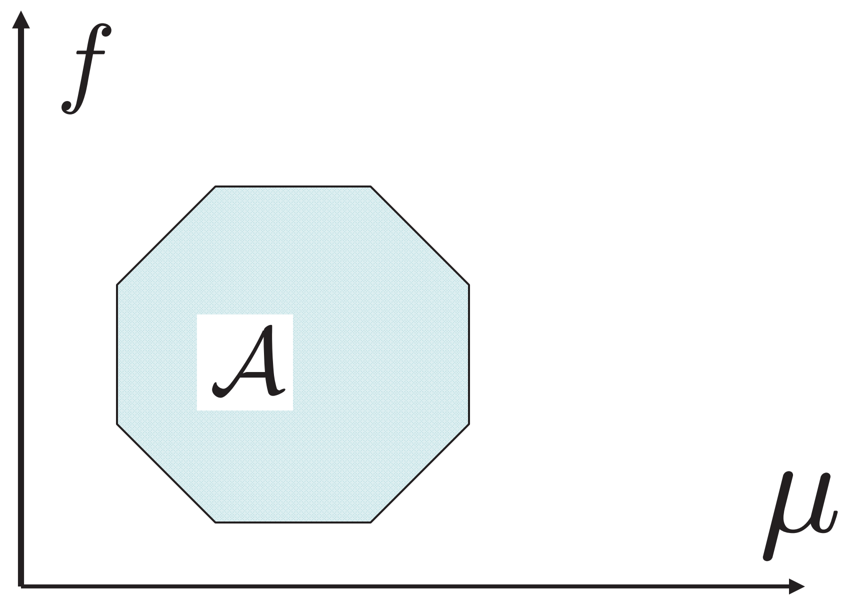
The first thing to do in UQ problem is to identify the quantity of interest(Φ) whose value we are tying to predict and quantify the information at hand (define the information set $\mathcal{A}$).

OUQ can be seen as form of interval/sensitivity analysis but instead of maximizing and minimizing $h(x)$ over $x \in [a, b]$ we maximize and minimize $\Phi(f, \mu)$ over $(f, \mu) \in \mathcal{A}$.



OUQ can be seen as form of interval/sensitivity analysis but instead of maximizing and minimizing $h(x)$ over $x \in [a,b]$ we maximize and minimize $\Phi(f,\mu)$ over $(f,\mu) \in \mathcal{A}$.

$$(G, \mathbb{P}) \in \mathcal{A}$$



$$\inf_{(\mu,f)\in\mathcal{A}} \Phi(f,\mu) \leq \Phi(G,\mathbb{P}) \leq \sup_{(\mu,f)\in\mathcal{A}} \Phi(f,\mu)$$

OUQ problems are not directly computationally tractable (optimization variables are infinite dimensional) but using the reduction theorems found in OUQ we can turn them into (computationally tractable) finite dimensional optimization problems.

$$\{f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X})\}$$

$$\left\{ f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X}) \,\middle|\, \mu = \sum_{i=1}^{k} \alpha_k \delta_{x_k} \right\}$$

$$\{f : \{1, 2, \ldots, n\} \to \mathbb{R},\ \mu \in \mathcal{P}(\{1, 2, \ldots, n\})\}$$

$$\{\{1, 2, \ldots, q\},\ \mu \in \mathcal{P}(\{1, 2, \ldots, n\})\}$$

Even after reduction these problems can be very large, highly nonlinear and non-convex so we need Mystic to solve them and Pathos to run Mystic on large computer clusters (without the need to adapt Mystic to the cluster).

- mystic:
  - a highly-configurable optimization framework
- pathos:
  - a distributed parallel graph execution framework providing a high-level programmatic interface to heterogeneous computing
- OUQ + mystic + pathos:
  - calculations of uncertainties cast as highly-constrained massively parallel global optimization problems

OUQ can drive experimental planning

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

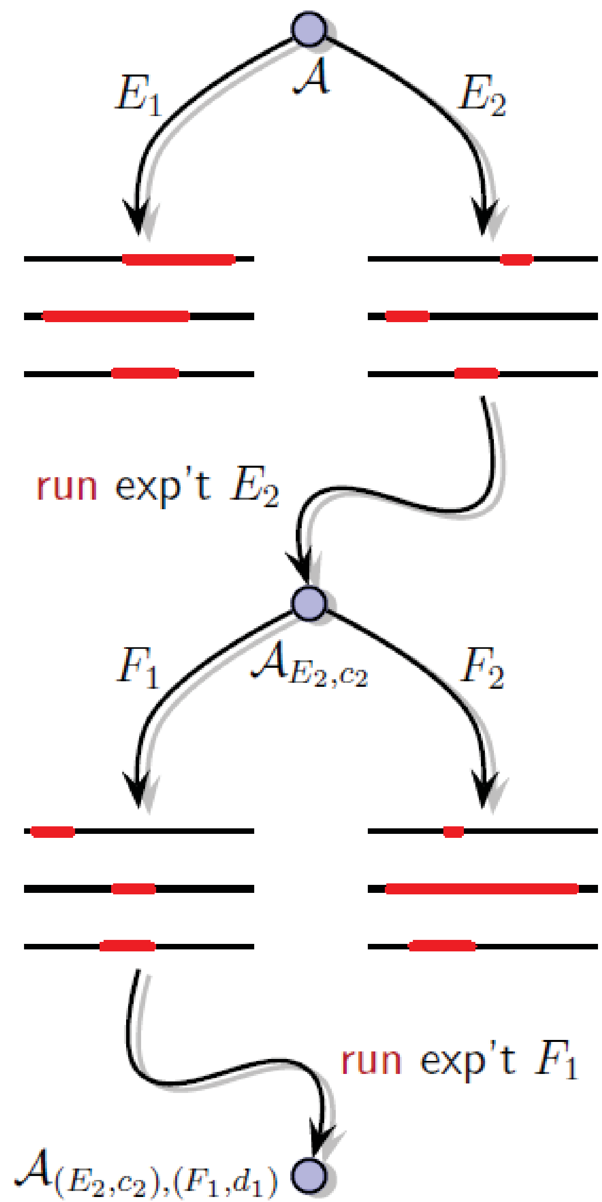
$$\mathcal{R}(q|\mathcal{A}) := \sup_{(f,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_f] - \inf_{(f,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q_f]$$

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

Let $\mathcal{A}_{E,c}$ denotes those scenarios in c that are consistent with getting outcome c for some experiment c.

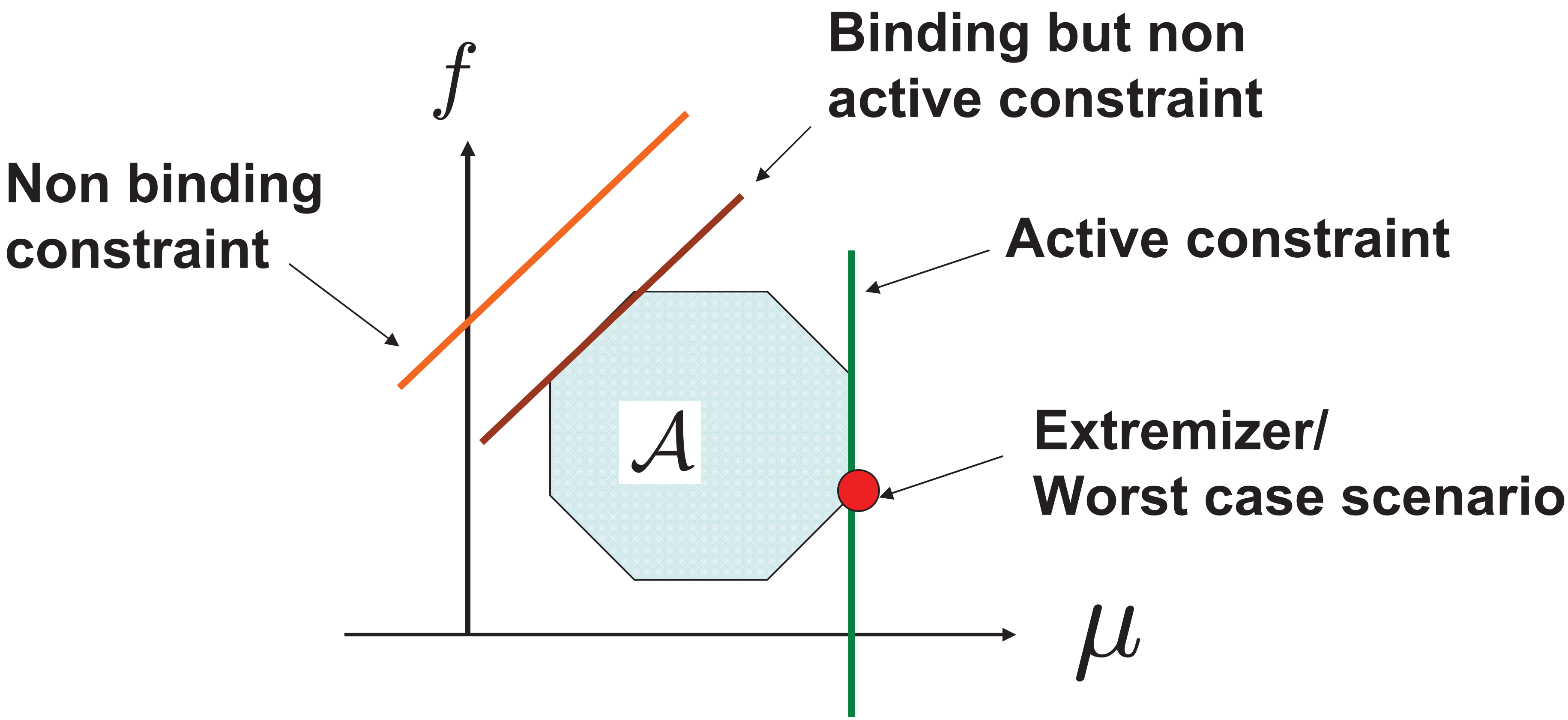
The optimal next experiment E^* satisfies a minmax criterion, i.e. E^* is the most predictive even in the least predictive outcome:

E^* mimimizes $\sup_{(f,\mu)\in\mathcal{A}} \mathcal{R}(q|\mathcal{A})$



In OUQ each piece of information is a constraint on an optimization problem.

Optimization concepts (binding, active) transfer to UQ concepts



With OUQ information may not propagate through hierarchies

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

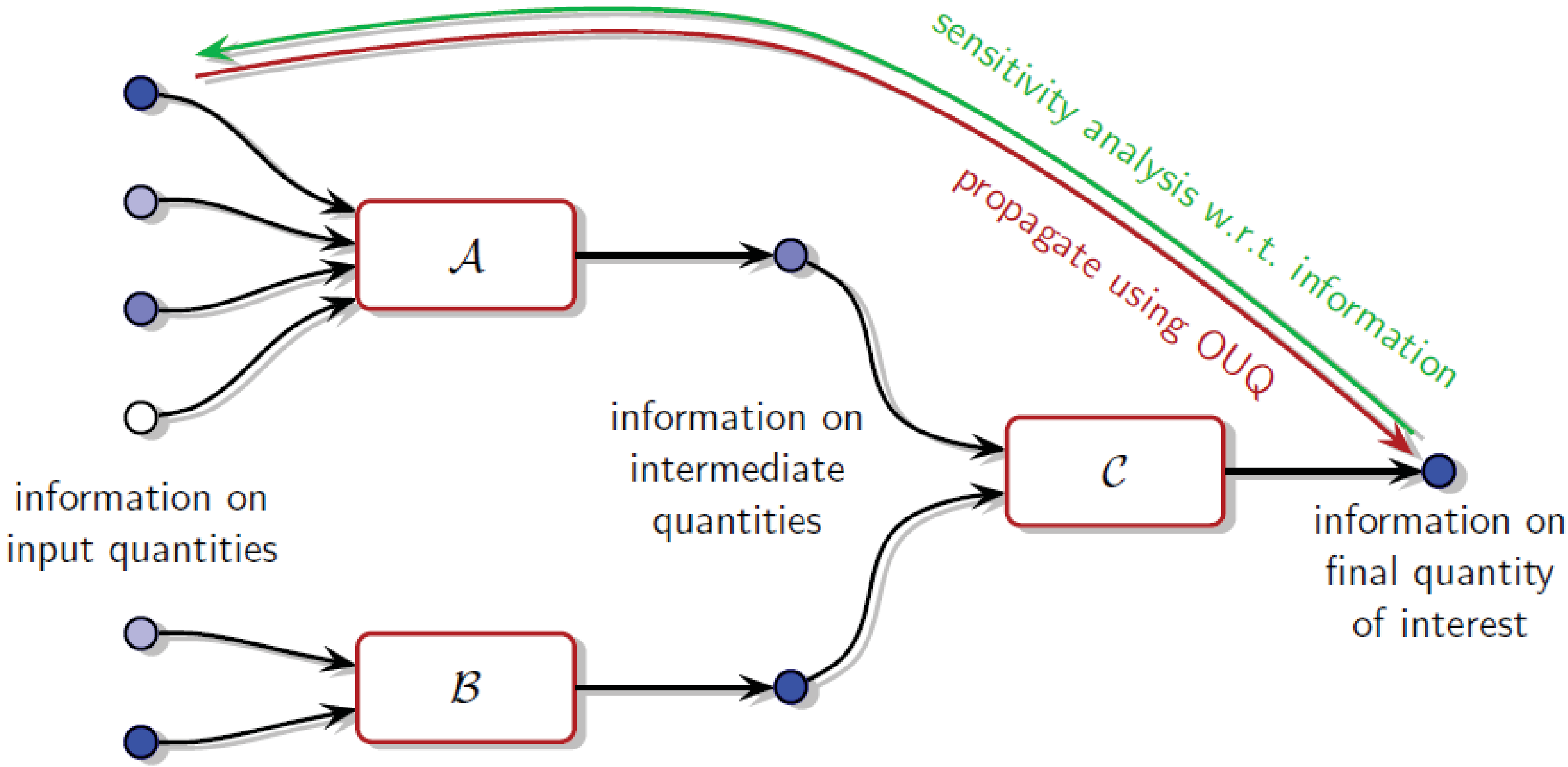
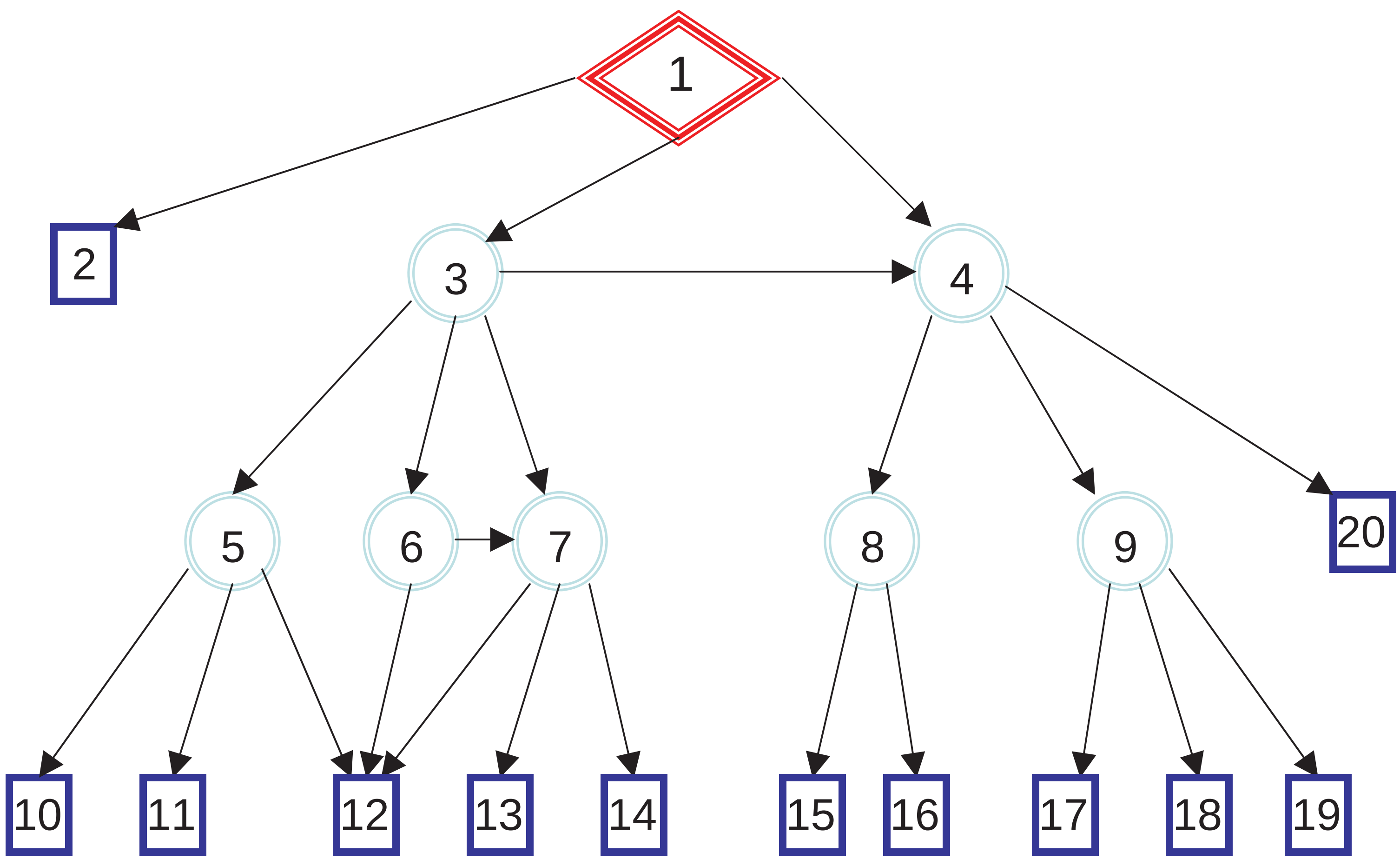
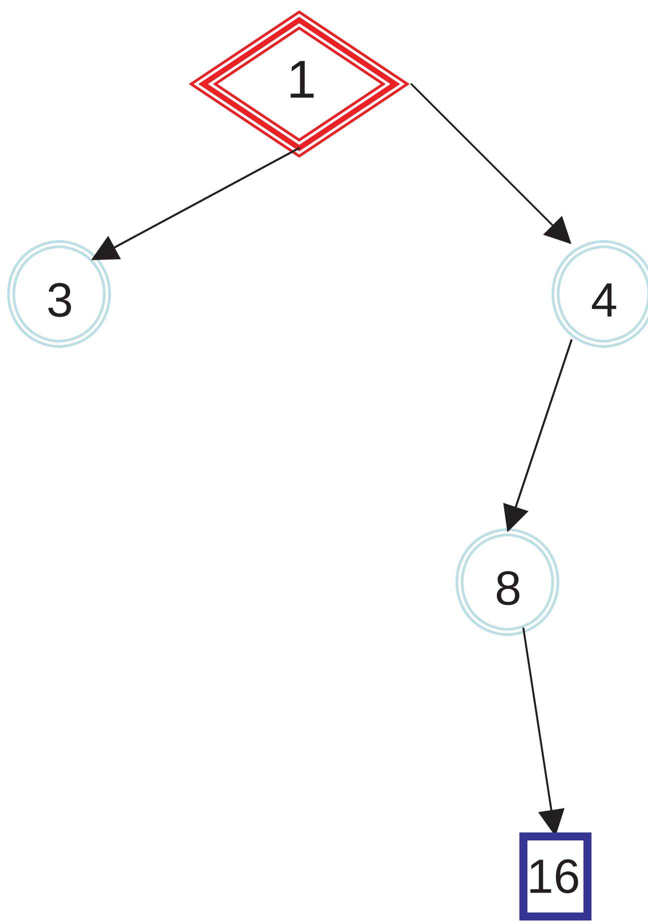


Figure: Because OUQ is a *sharp* information propagation scheme, the results of sensitivity analysis ("inverse OUQ") give non-trivial insights into the roles of the various pieces of input information. Some inputs may even be irrelevant!

OUQ leads to sparse information trees/graphs



OUQ leads to sparse information trees/graphs



OUQ is well adapted to exascale computing

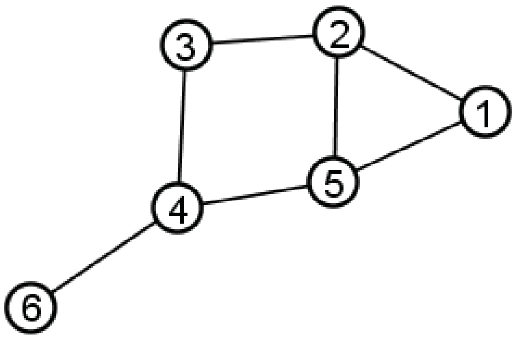
OUQ optimization problems can naturally be divided into smaller ones, which can then be solved concurrently

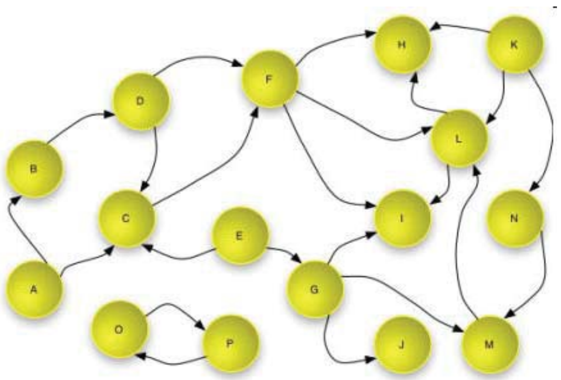
These problems have a natural implementation on massively parallel computing clusters

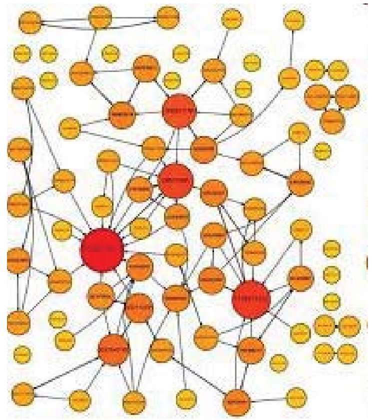
Each new piece of information acts as a new constraint for OUQ optimization problems

New information can be added and/or modified on the fly

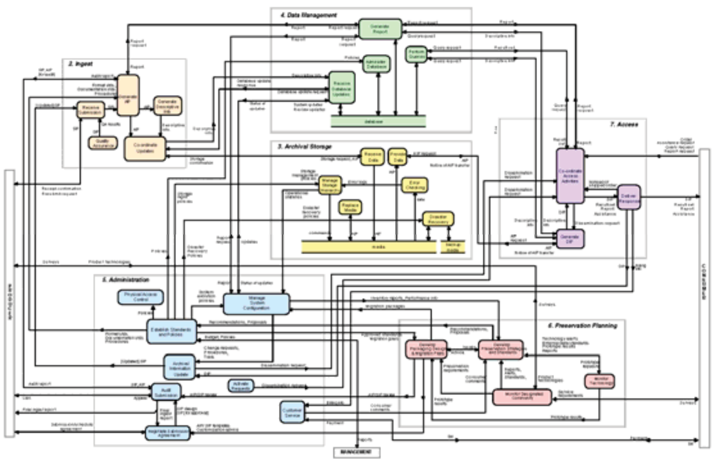
Information can be coded and processed at different levels of complexity

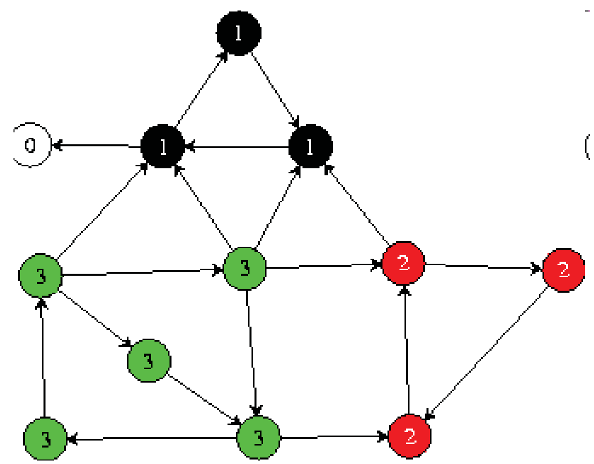






OUQ is well adapted to exascale computing





OUQ bounds are sharp and identify (ir)relevant information

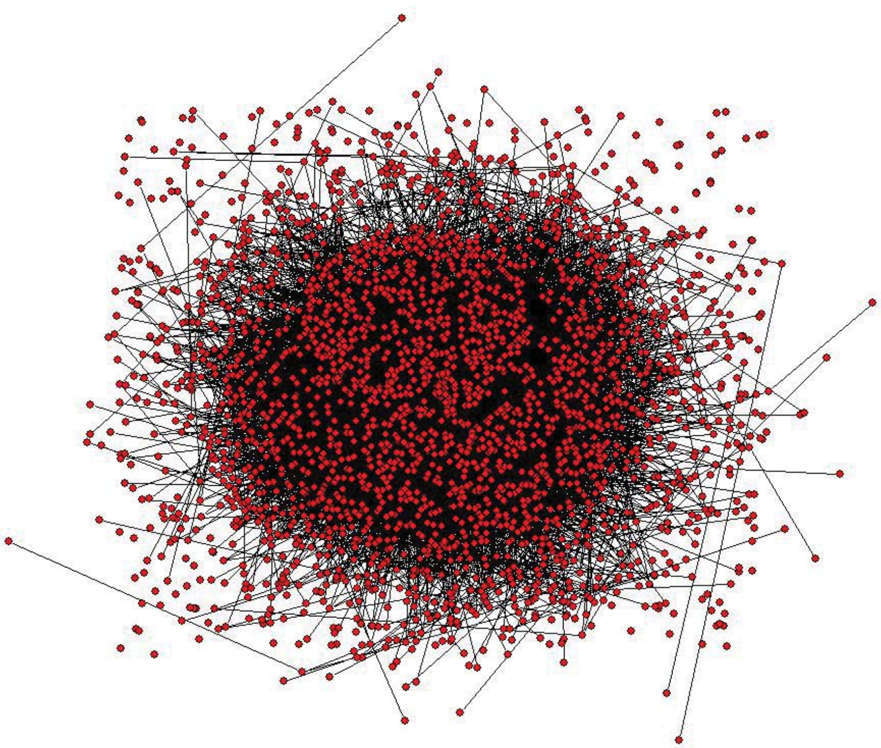
It is not necessary to code all that is known (“too much information kills information”)

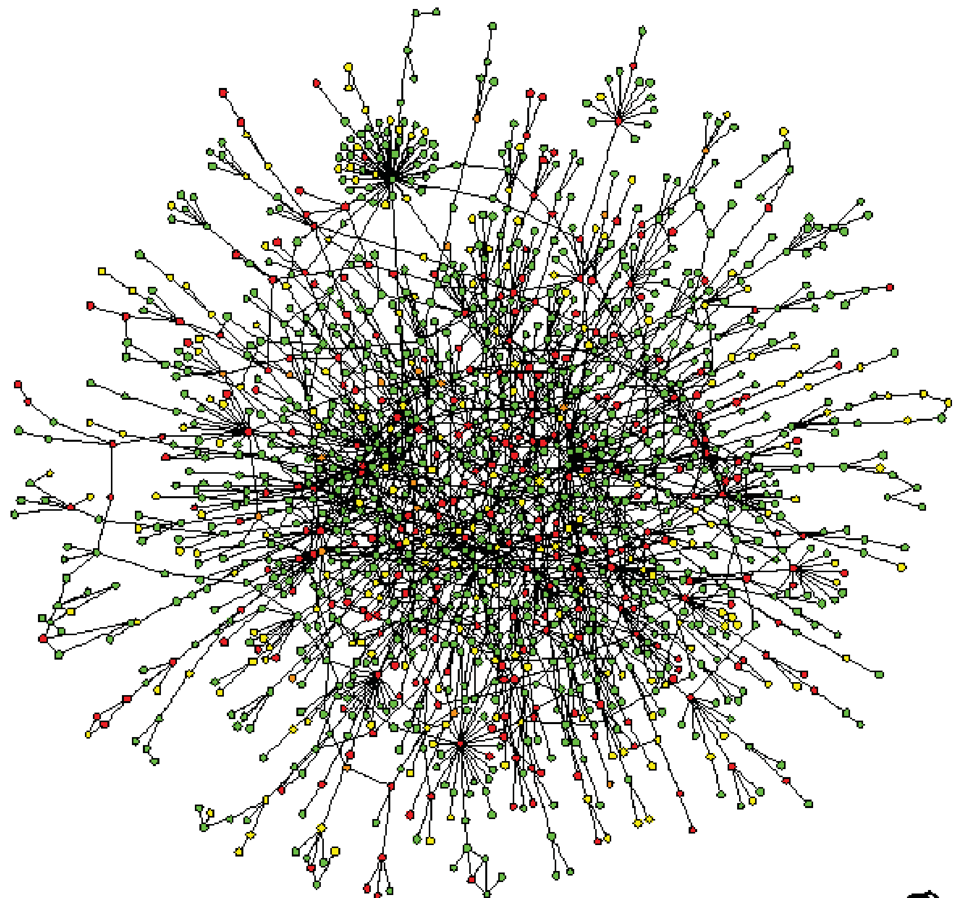
We can use exascale and OUQ to design a scheme where information is coded and processes at different levels of complexity and the most relevant/important elements are coded/processed first.

Exascale computing can lead to a new paradigm for scientific investigation (optimal strategies of experimental design, hierarchical information processing, new language)

UQ can be applied to Exascale computing in several places

Exascale computing will allow us to quantify uncertainties and compute optimal intervals of confidence and make optimal decisions for very complex systems (for such systems the reduced optimization problems would still be very large).

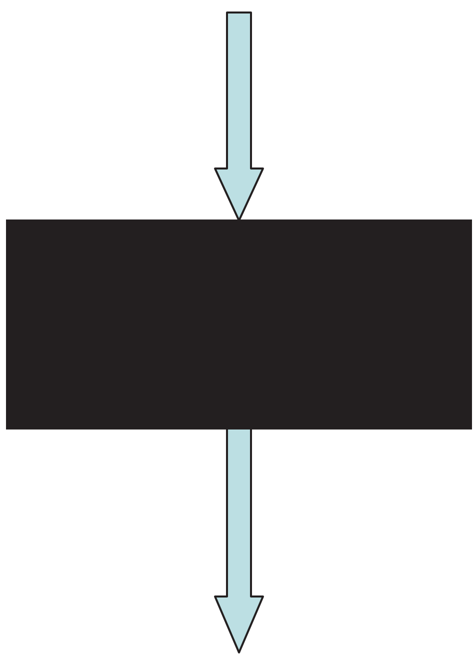


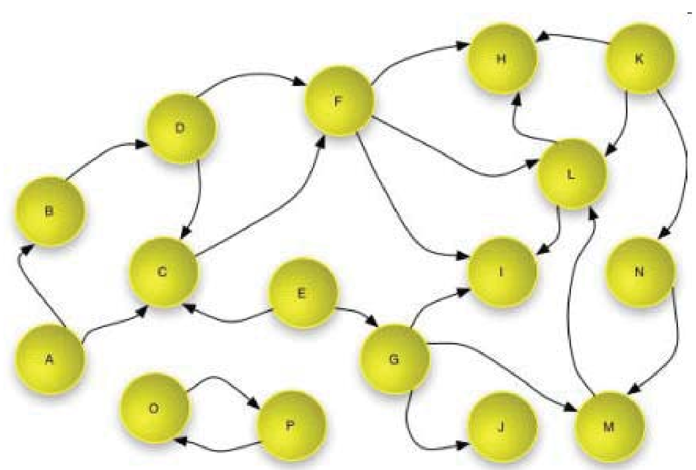


Pajek

UQ can be applied to Exascale computing in several places

Mike McKerns is currently developing an OUQ app that will allow for a OUQ analysis of other proxy apps (treated as black box input output systems), these OUQ app should allow for the identification of key variables, major vulnerabilities and sources of uncertainties in these other apps and it is designed to be user friendly.





UQ can be applied to Exascale computing in several places

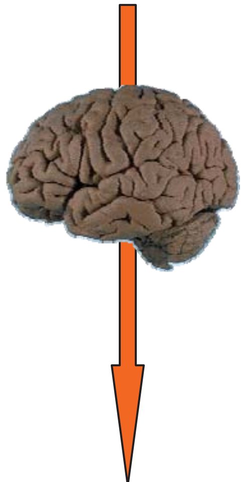
If we combine Exascale with the generalization of OUQ to sample data, then we will be able to compute digital libraries for optimal statistical tests and play information wars/games

UQ can also be applied to Exascale in many other specific places, but for these other applications it is very important to identify the QOI and the information at hand (this requires a close collaboration with someone from LLNL or LANL).

Scientific Computation of Optimal Statistical Estimators

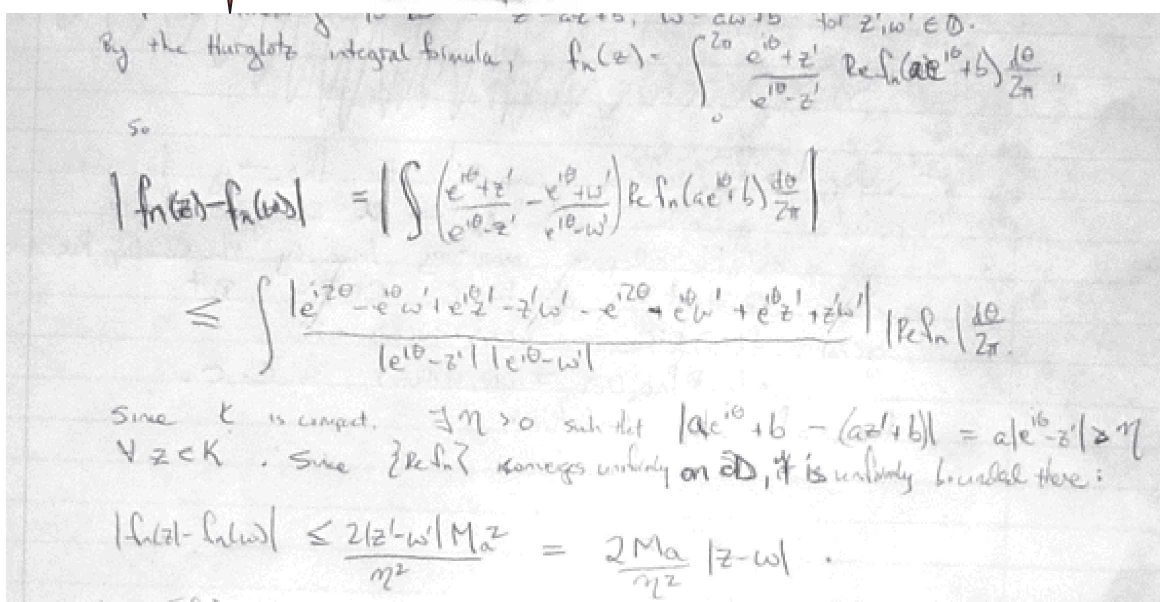
Solving PDEs: Two centuries ago

$$\Delta u = f$$



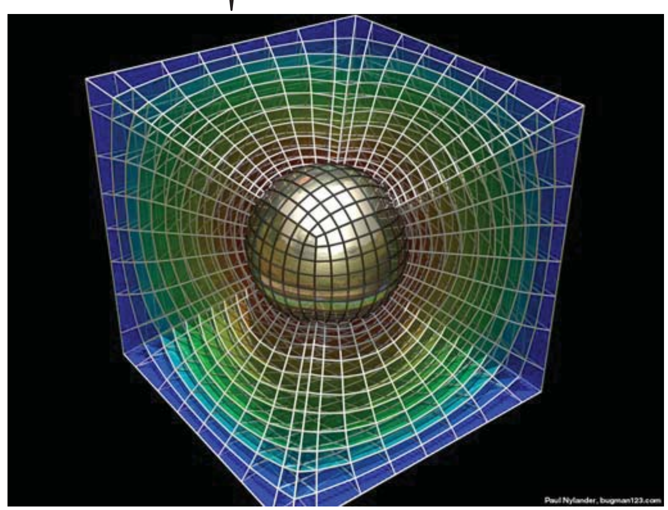
A. L. Cauchy
(1789-1857)

S. D. Poisson
(1781-1840)



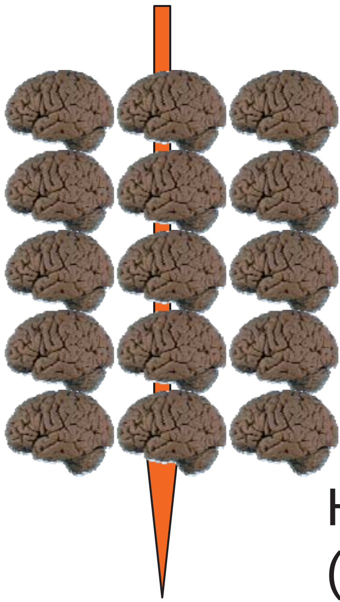
Solving PDEs: Now.

$$\Delta u = f$$

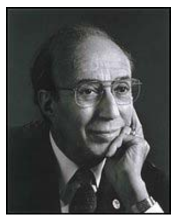


Paradigm shift

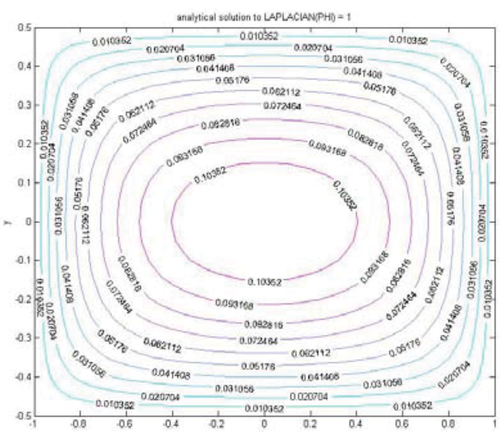
$$\Delta u = f$$



J. V. Neumann
(1903-1957)



H. Goldstine
(1913-2004)



Where are we at in finding statistical estimators?

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

where

χ^2 is Chi-squared,
$\sum$ stands for summation,
o is the observed values,
e is the expected values.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2							
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08

Where are we at in finding statistical estimators?

Estimate

$$\Phi(G, \mathbb{P})$$

$$\theta$$

Available information

$$(\mathbb{P}, G) \in \mathcal{A}$$

+ (sample) data

+ (sample) data

$$\theta(data)$$

Scientific computing of optimal statistical estimators