# The worst case approach to UQ

**Houman Owhadi**

*"The gods to-day stand friendly, that we may,
Lovers of peace, lead on our days to age!
But, since the affairs of men rests still uncertain,
Let's reason with the worst that may befall."*

Julius Caesar, Act 5, Scene 1
William Shakespeare (1564 –1616)

CALTECH PSAAP Predictive Science Academic Alliance Program

California Institute of Technology 1891

NNSA National Nuclear Security Administration

ExMatEx

DARPA

Air Force Office of Scientific Research United States Air Force

**You want to certify that**

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

**Problem**

- You don't know $G$.

**and**

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

**You want to certify that**

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

**Problem**

- You don't know $G$.

**and**

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

**You only know**

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

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

**Compute** **Worst and best case**

optimal bounds $\mathbb{P}[G(X) \geq a]$ given available information.

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

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

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

$\mathcal{U}(\mathcal{A}) \leq \epsilon$: 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.

## Optimal Uncertainty Quantification

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. **Optimal Uncertainty Quantification**. *SIAM Review*, 55(2):271–345, 2013.

## Robust Optimization

A. Ben-Tal, L. El Ghaoui, and A.Nemirovski. ***Robust optimization***. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of **robust optimization**. *SIAM Rev.*, 53(3):464–501, 2011

A. Ben-Tal and A. Nemirovski. **Robust convex optimization**. *Math. Oper. Res.*, 23(4):769–805, 1998

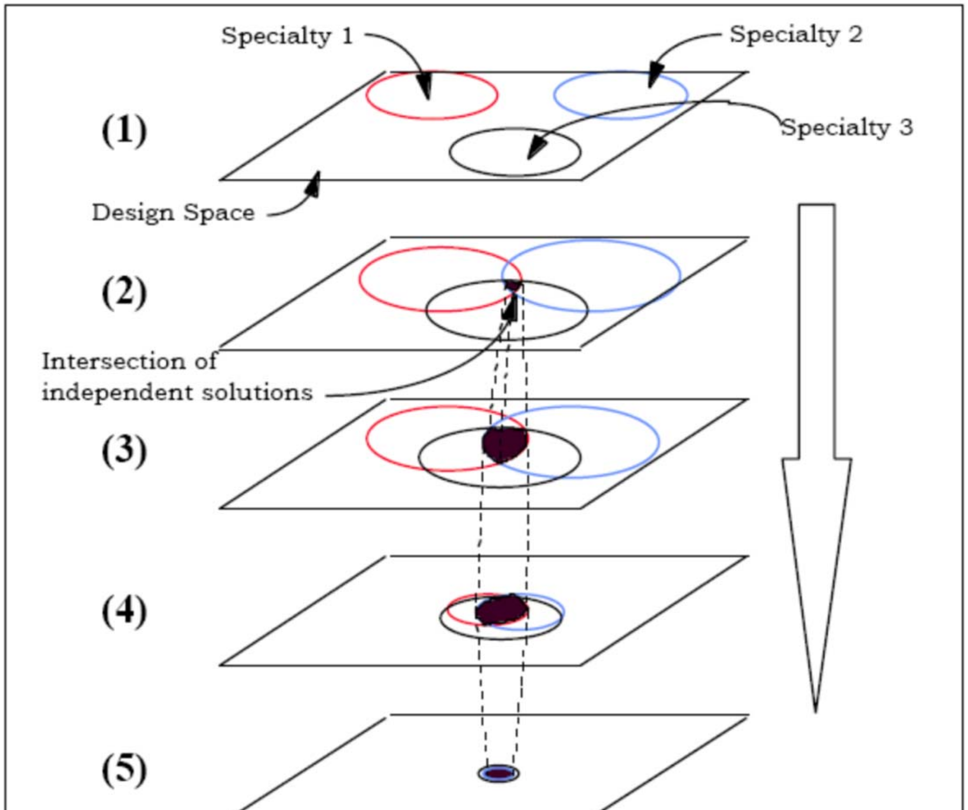
I. Elishakoff and M. Ohsaki. ***Optimization and Anti-Optimization of Structures Under Uncertainty***. World Scientific, London, 2010.

## Global Sensitivity Analysis

Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. (2008). *Global **Sensitivity Analysis***: The Primer. John Wiley & Sons.

# Set based design in the aerospace industry

Bernstein, J. I., 1998, Design Methods in the Aerospace Industry: Looking for Evidence of Set-Based Practices, Master of Science Thesis, Massachusetts Institute of Technology, 1998.



(Bernstein 1998)

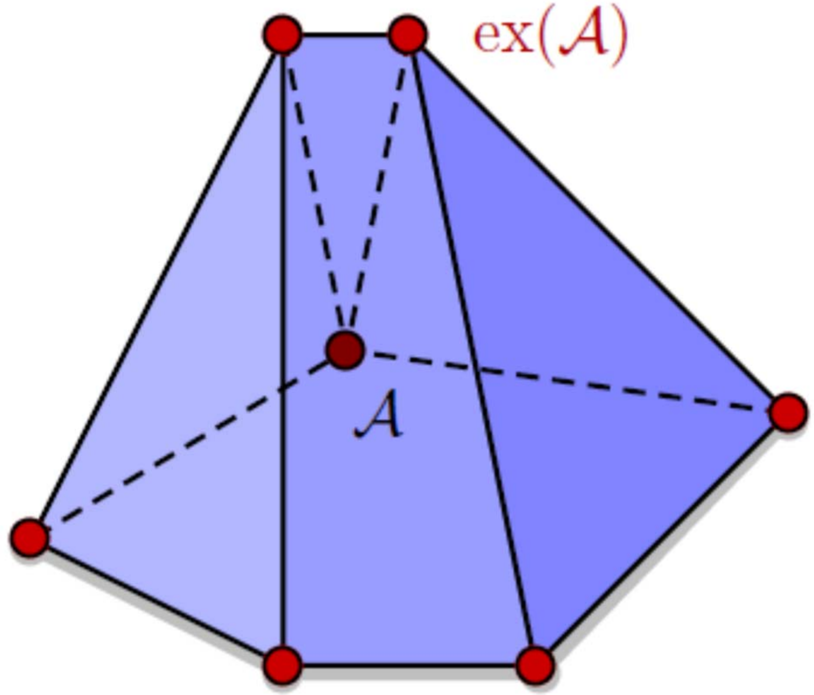
Bernstein JI (1998) *Design methods in the aerospace industry: looking for evidence of* ***set-based practices***. Master's thesis. Massachusetts Institute of Technology, Cambridge, MA.

# Set based design/analysis

David J. Singer, PhD., Captain Norbert Doerry, PhD., and Michael E. Buckley," *What is* ***Set-Based Design***? ," Presented at ASNE DAY 2009, National Harbor, MD., April 8-9, 2009. Also published in ASNE Naval Engineers Journal, 2009 Vol 121 No 4, pp. 31-43.

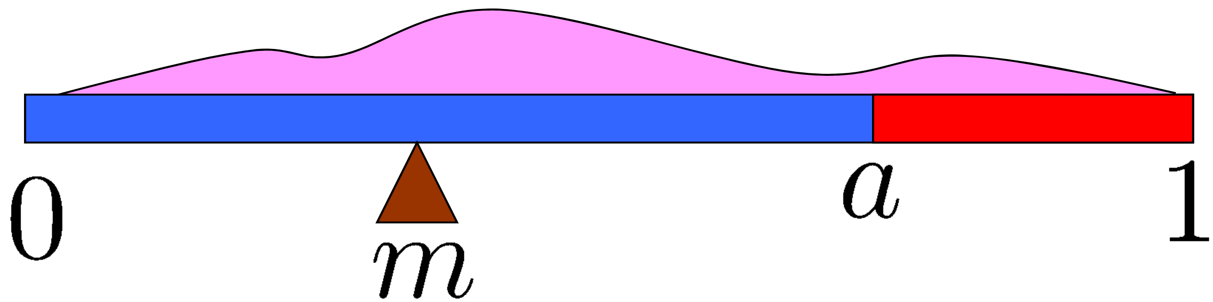
B. Rustem and Howe M. ***Algorithms for Worst-Case Design and Applications to Risk Management***. Princeton University Press, Princeton, 2002.

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




Optimization problems are a priori infinite dimensional, non-convex and highly constrained but as in linear programming, under general conditions, they can be reduced to finite dimensional families of extremal scenarios of $\mathcal{A}$ and the dimension of the reduced problem is proportional to the number of probabilistic inequalities describing $\mathcal{A}$

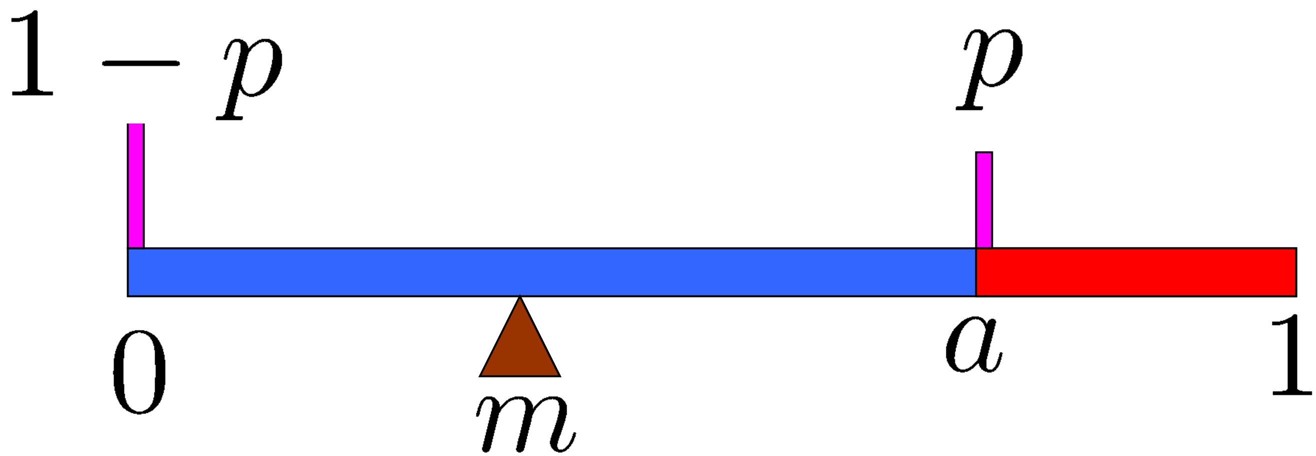
You are given one pound of playdoh, how much mass can you put above $a$ while keeping the seesaw balanced around $m$?




**P. L. Chebyshev**
**1821-1894**

**A. A. Markov**
**1856-1922**

**M. G. Krein**
**1907-1989**




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

**Answer** $\frac{m}{a}$

What is the least upper bound on $\mathbb{P}[X \geq a]$ if all that you know is that $\mathbb{P}$ is an unknown distribution on $[0, 1]$ having mean less than $m$


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

**Markov's inequality**

**Answer**

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

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

Can be considered as a generalization of classical Chebyshev inequalities

## History of classical inequalities

S. Karlin and W. J. Studden. ***Tchebycheff Systems: With Applications in Analysis and Statistics***. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

## Classical Markov-Krein theorem and classical works of Krein, Markov and Chebyshev

M. G. Krein. **The ideas of P. L. Cebysev and A. A. Markov in the theory of limiting values of integrals and their further development**. In E. B. Dynkin, editor, *Eleven papers on Analysis, Probability, and Topology*, American Mathematical Society Translations, Series 2, Volume 12, pages 1–122. American Mathematical Society, New York, 1959.

## Theory of majorization

A. W. Marshall and I. Olkin. ***Inequalities: Theory of Majorization and its Applications***, volume 143 of *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979

## Connections between Chebyshev inequalities and optimization theory


H. J. Godwin. **On generalizations of Tchebychef's inequality**. *J. Amer. Statist. Assoc.*, 50(271):923–945, 1955.

K. Isii. **On a method for generalizations of Tchebycheff's inequality**. *Ann. Inst. Statist. Math. Tokyo*, 10(2):65–88, 1959.

K. Isii. **On sharpness of Tchebycheff-type inequalities**. *Ann. Inst. Statist. Math.*, 14(1):185–197, 1962/1963.

K. Isii. **The extrema of probability determined by generalized moments. I. Bounded random variables**. *Ann. Inst. Statist. Math.*, 12(2):119–134; errata, 280, 1960.

I. Olkin and J. W. Pratt. A multivariate Tchebycheff inequality. *Ann. Math. Statist*, 29(1):226–234, 1958.

A. W. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *Ann. Math. Statist.*, 31(4):1001–1014, 1960.

A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its Applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979

## Connection between Chebyshev inequalities and optimization theory

E. B. Dynkin. Sufficient statistics and extreme points. *Ann. Prob.*, 6(5):705–730, 1978.

A. F. Karr. Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.*, 8(1):74–85, 1983.

H. Joe. **Majorization, randomness and dependence for multivariate distributions**. *Ann. Probab.*, 15(3):1217–1225, 1987.

J. E. Smith. **Generalized Chebychev inequalities: theory and applications in decision analysis**. *Oper. Res.*, 43(5):807–825, 1995.

D. Bertsimas and I. Popescu. **Optimal inequalities in probability theory: a convex optimization approach**. *SIAM J. Optim.*, 15(3):780–804 (electronic), 2005.

L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Rev.*, 49(1):52–64 (electronic), 2007

## Stochastic linear programming and Stochastic Optimization

G. B. Dantzig. **Linear programming under uncertainty**. *Management Sci.*, 1:197–206, 1955.

A. Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, pages 743–746, 1959

A. Madansky. Inequalities for stochastic linear programming problems. *Management science*, 6(2):197–204, 1960.

J. Zjackovja. On minimax solutions of stochastic linear programming problems. *Casopis Pest. Mat.*, 91:423–430, 1966.

C. C. Huang, W. T. Ziemba, and A. Ben-Tal. **Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming**. *Operations Research*, 25(2):315–325, 1977.

P. Kall. **Stochastric programming with recourse: upper bounds and moment problems: a review**. *Mathematical research*, 45:86–103, 1988

Y. Ermoliev, A. Gaivoronski, and C. Nedeva. **Stochastic optimization problems with incomplete information on distribution functions**. *SIAM Journal on Control and Optimization*, 23(5):697–716, 1985

J. R. Birge and R. J.-B. Wets. **Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse**. *Math. Prog. Stud.*, 27:54–102, 1986

## Chance constrained/distributionally robust optimization


A. A. Gaivoronski. **A numerical method for solving stochastic programming problems with moment constraints on a distribution function**. *Annals of Operations Research*, 31(1):347–369, 1991.

R. I. Bot N. Lorenz, and G. Wanka. **Duality for linear chance-constrained optimization problems**. *J. Korean Math. Soc.*, 47(1):17–28, 2010.

J. Goh and M. Sim. **Distributionally robust optimization and its tractable approximations**. *Oper. Res.*, 58(4, part 1):902–917, 2010

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Math. Program.*, 137(1-2, Ser.A):167–198, 2013.

L. Xu, B. Yu, and W. Liu. The distributionally robust optimization reformulation for stochastic complementarity problems. *Abstr. Appl. Anal.*, pages 7, 2014.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Oper. Res.*, 62(6):1358–1376, 2014.

G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. **A distributionally robust perspective on uncertainty quantification and chance constrained programming.** *Mathematical Programming*, 151(1):35–62, 2015


## Value at Risk


Artzner, P.; Delbaen, F.; Eber, J. M.; Heath, D. (1999). Coherent Measures of Risk. Mathematical Finance 9 (3): 203.

W. Chen, M. Sim, J. Sun, and C.-P. Teo. **From CVaR to uncertainty set: implications in joint chance-constrained optimization**. *Oper. Res.*, 58(2):470–485, 2010.

# Optimal Uncertainty Quantification

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. **Optimal Uncertainty Quantification**. *SIAM Review*, 55(2):271–345, 2013.

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi, and M. Ortiz. **Optimal uncertainty quantification for legacy data observations of Lipschitz functions.** *ESAIM Math. Model. Numer. Anal.*, 47(6):1657–1689, 2013.

S. Han, U. Topcu, M. Tao, H. Owhadi, and R. Murray. **Convex optimal uncertainty quantification: Algorithms and a case study in energy storage placement for power grids**. In *American Control Conference (ACC), 2013*, pages 1130–1137. IEEE, 2013

S. Han, M. Tao, U. Topcu, H. Owhadi, and R. M. Murray. **Convex optimal uncertainty quantification**. *SIAM Journal on Optimization*, 25(23):1368–1387, 2015.

J Chen, MD Flood, R Sowers. **Measuring the Unmeasurable: An Application of Uncertainty Quantification to Financial Portfolios**, OFR WP, 2015

L Ming, W Chenglin. **An improved algorithm for convex optimal uncertainty quantification with polytopic canonical form**. Control Conference (CCC), 2015

H. Owhadi, C. Scovel and T. Sullivan. Brittleness of Bayesian Inference under Finite Information in a Continuous World. Electronic Journal of Statistics, vol 9, pp 1-79, 2015. arXiv:1304.6772

H. Owhadi and Clint Scovel. Extreme points of a ball about a measure with finite support (2015). arXiv:1504.06745

## Our proof relies 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)

G. Winkler. On the integral representation in convex noncompact sets of tight measures. *Mathematische Zeitschrift*, 158(1):71–77, 1978

G. Winkler. Extreme points of moment sets. *Math. Oper. Res.*, 13(4):581–587, 1988.

H. von Weizsacker and G. Winkler. Integral representation in the set of solutions of a generalized moment problem. *Math. Ann.*, 246(1):23–32, 1979/80.

D. G. Kendall. Simplexes and vector lattices. *J. London Math. Soc.*, 37(1):365–371, 1962.

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. **Optimal Uncertainty Quantification**. *SIAM Review*, 55(2):271–345, 2013.

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

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

**Theorem**

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

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

# Further Reduction of optimization variables

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

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

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

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

# Another example: Optimal concentration inequality

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal Uncertainty Quantification. *SIAM Review*, 55(2):271–345, 2013.

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

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

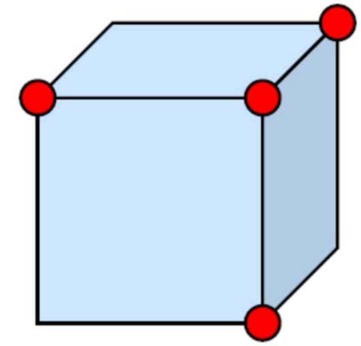
$$\mathcal{U}(\mathcal{A}_{MD}) := \sup_{(f,\mu)\in\mathcal{A}_{MD}} \mu[f(X) \ge a]$$

**McDiarmid inequality's**

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

# 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**

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

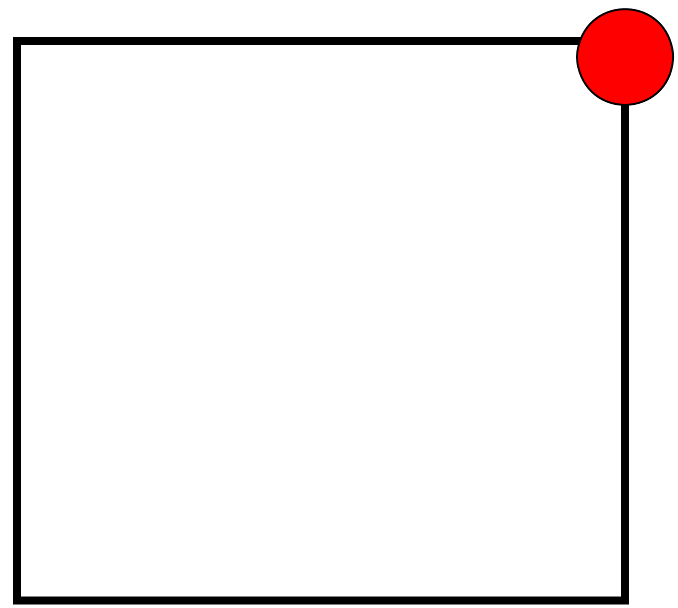
# Explicit Solution m=2

**Theorem** $m = 2$

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

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

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



# Explicit Solution m=2

**Theorem** $m = 2$

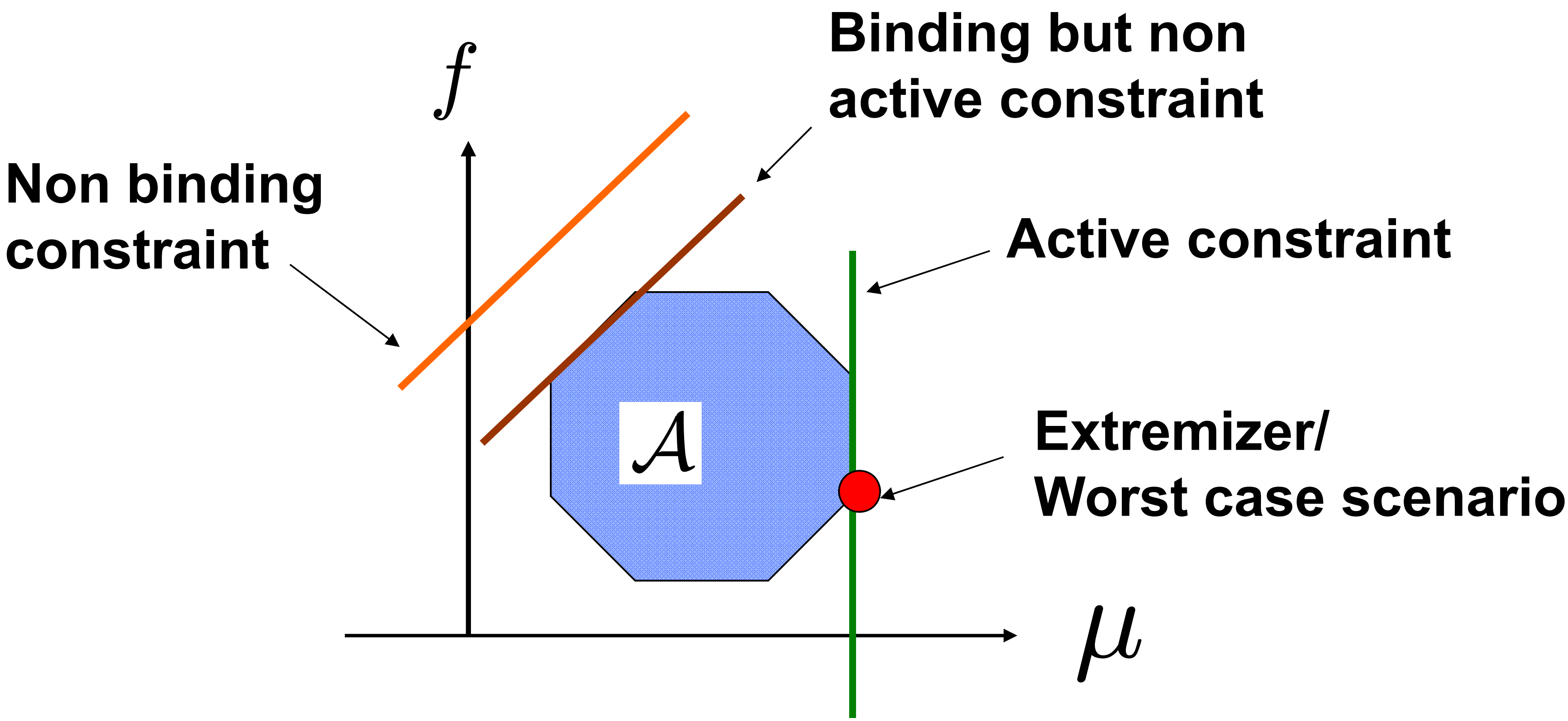
$$\mathcal{U}(\mathcal{A}_{MD}) = \begin{cases} 0 & \text{if} \quad D_1 + D_2 \leq a \\ \frac{(D_1+D_2-a)^2}{4D_1D_2} & \text{if} \quad |D_1 - D_2| \leq a \leq D_1 + D_2 \\ 1 - \frac{a}{\max(D_1,D_2)} & \text{if} \quad 0 \leq a \leq |D_1 - D_2| \end{cases}$$

**Corollary** If $D_1 \geq a + D_2$, then

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

**Each piece of information is a constraint on an optimization problem.**

**Optimization concepts (binding, active) transfer to UQ concepts**



## Optimal Hoeffding= Optimal McDiarmid for m=2

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

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

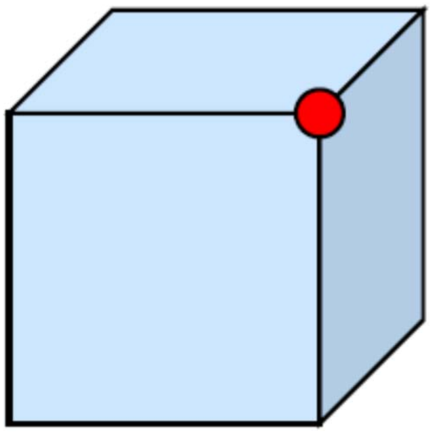
$$\mathcal{A}_{\mathrm{Hfd}} := \left\{ (f, \mu) \;\middle|\; \begin{array}{c} f = X_1 + \dots + X_m, \\ \mu \in \bigotimes_{i=1}^m \mathcal{M}([b_i - D_i, b_i]), \\ \mathbb{E}_\mu[f] \leq 0 \end{array} \right\}$$

# Explicit Solution m=3

**Theorem** $m = 3$

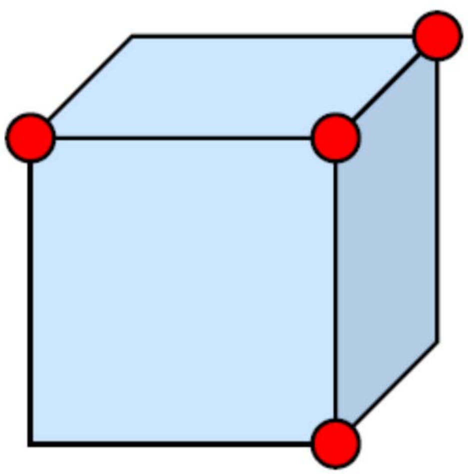
$$\mathcal{U}(\mathcal{A}_{MD}) = \max(\mathcal{F}_1, \mathcal{F}_2)$$

$\mathcal{F}_1$



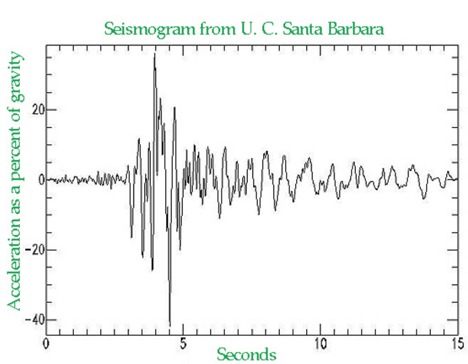
$$\mathcal{U}(\mathcal{A}_{\mathrm{MD}}) = \mathcal{U}(\mathcal{A}_{\mathrm{Hfd}})$$

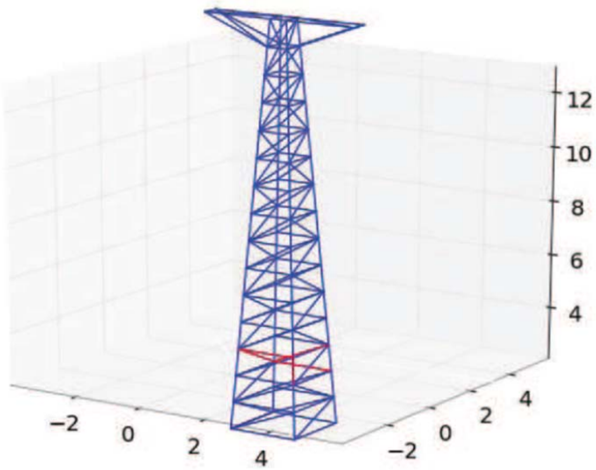
$\mathcal{F}_2$

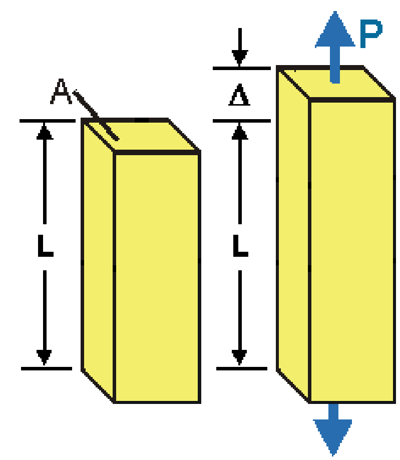


$$\mathcal{U}(\mathcal{A}_{\mathrm{MD}}) > \mathcal{U}(\mathcal{A}_{\mathrm{Hfd}})$$

# Seismic Safety Assessment of a Truss Structure






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

**Ground Acceleration**

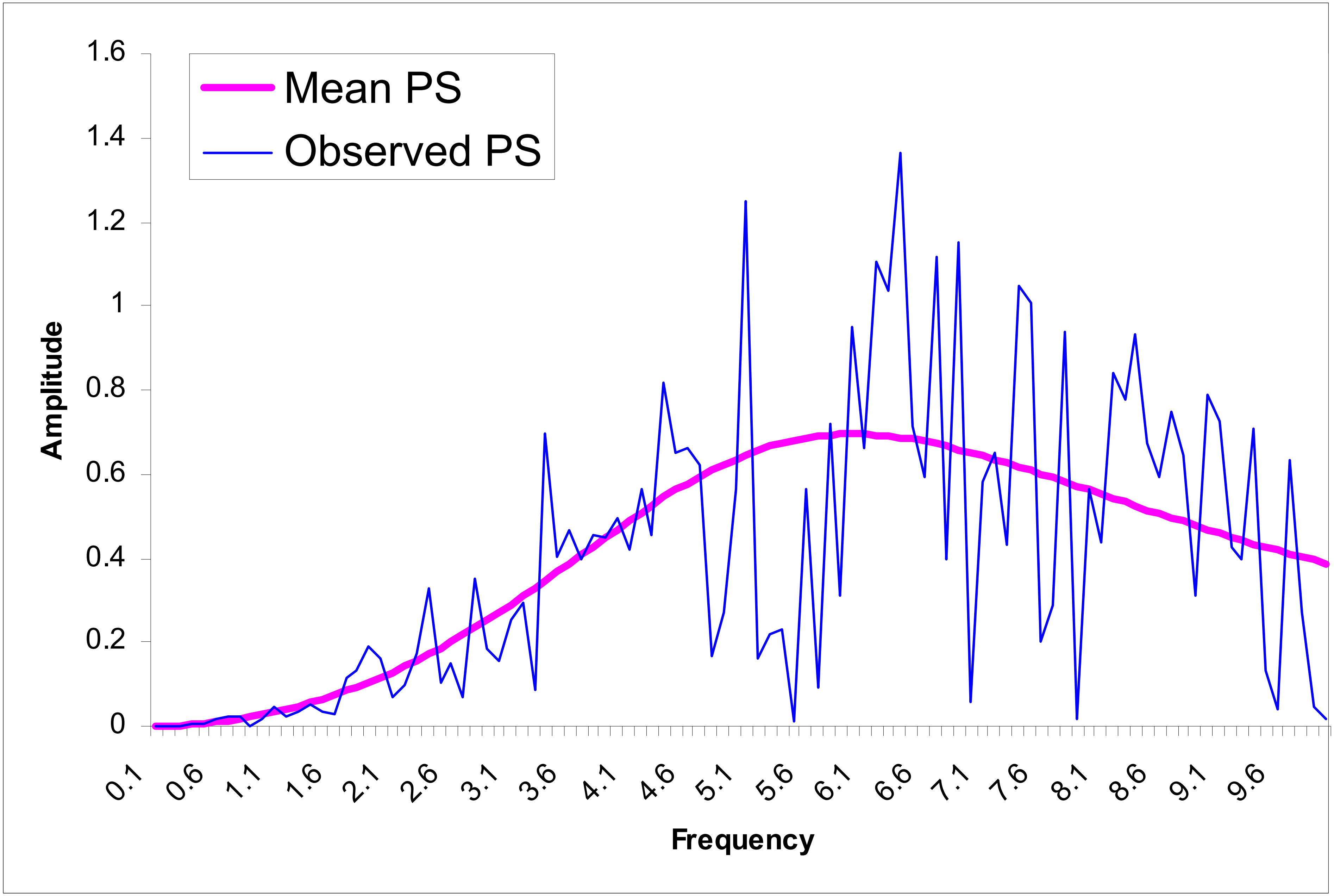
**F**

**min( Yield Strain - Axial Strain )**

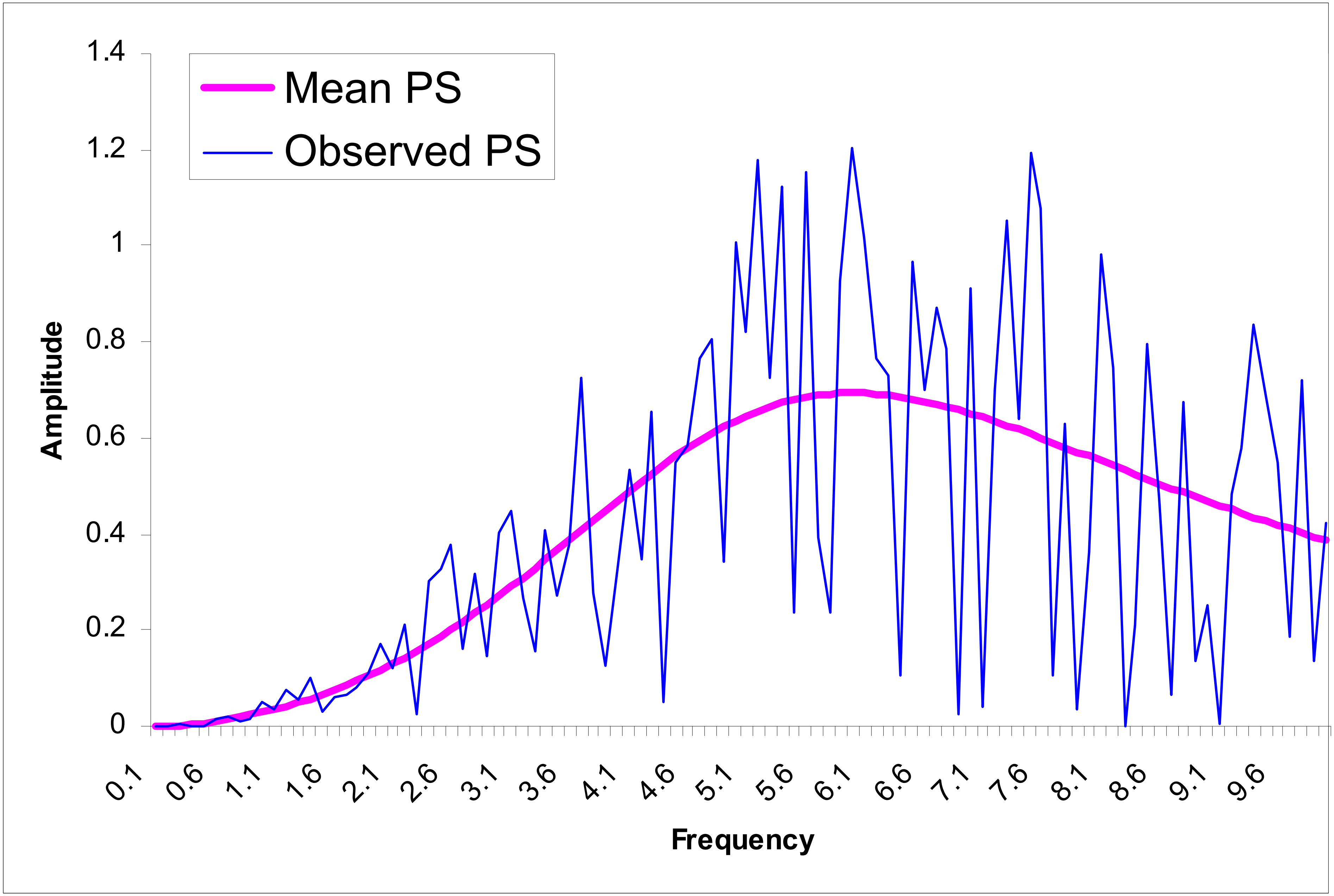
**We want to certify that**

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

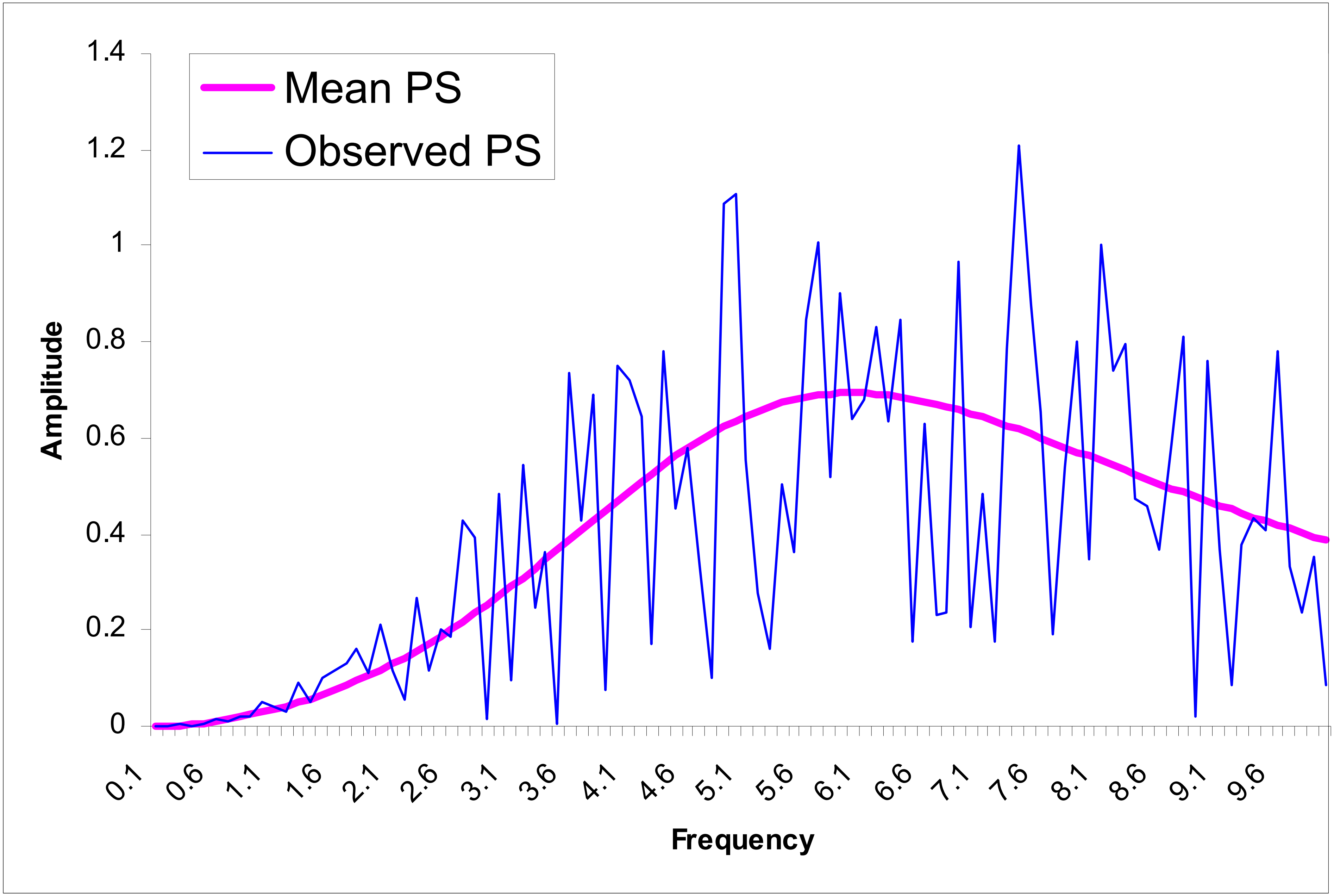
# Power Spectrum



# Power Spectrum

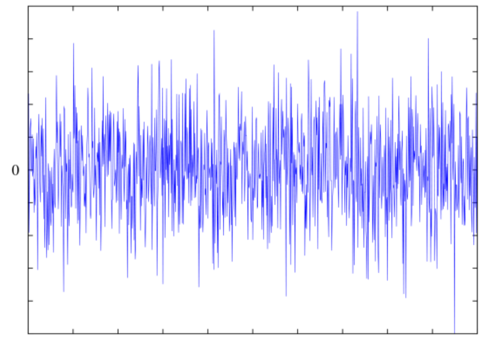


# Power Spectrum



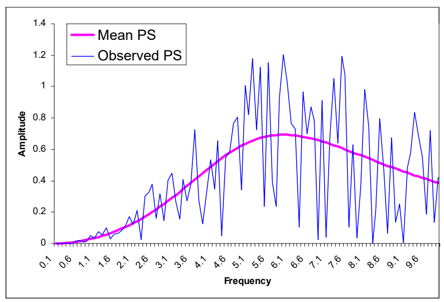
# Filtered White Noise Model

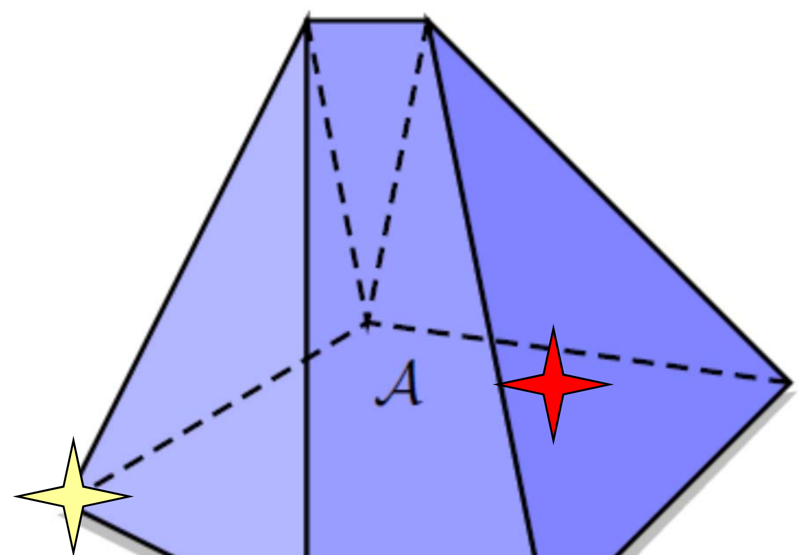
**White noise**



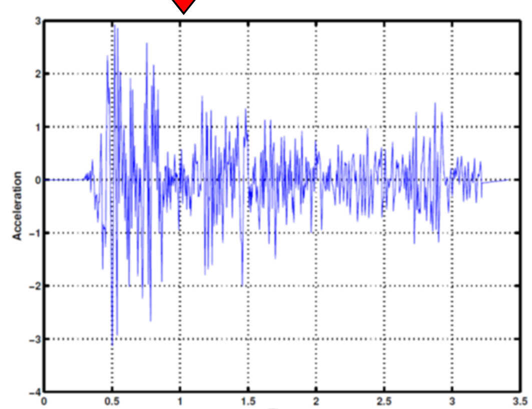
N. Lama, J. Wilsona, and G. Hutchinsona. Generation of synthetic earthquake accelograms using seismological modeling: a review. Journal of Earthquake Engineering, 4(3):321–354, 2000.

**Filter**

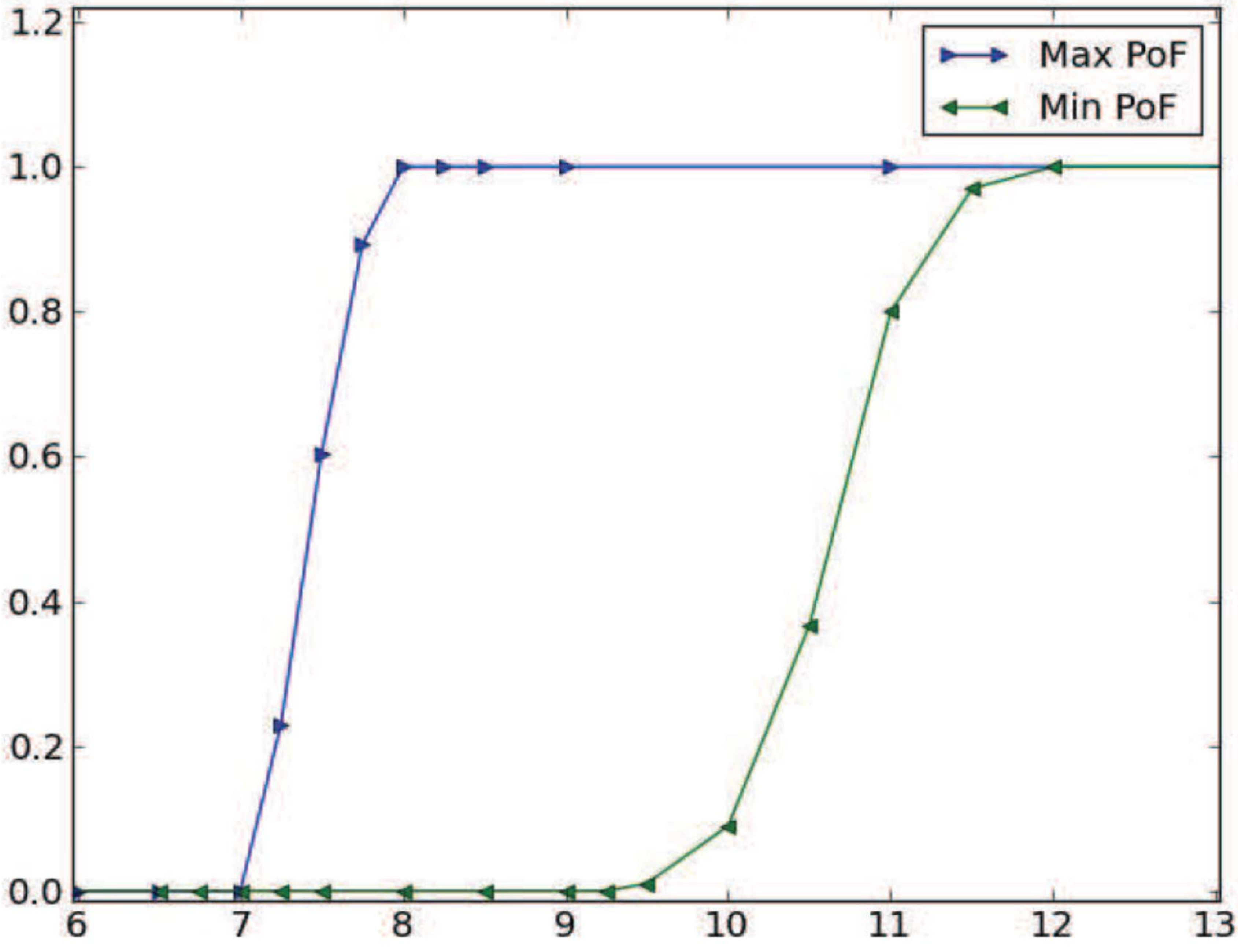


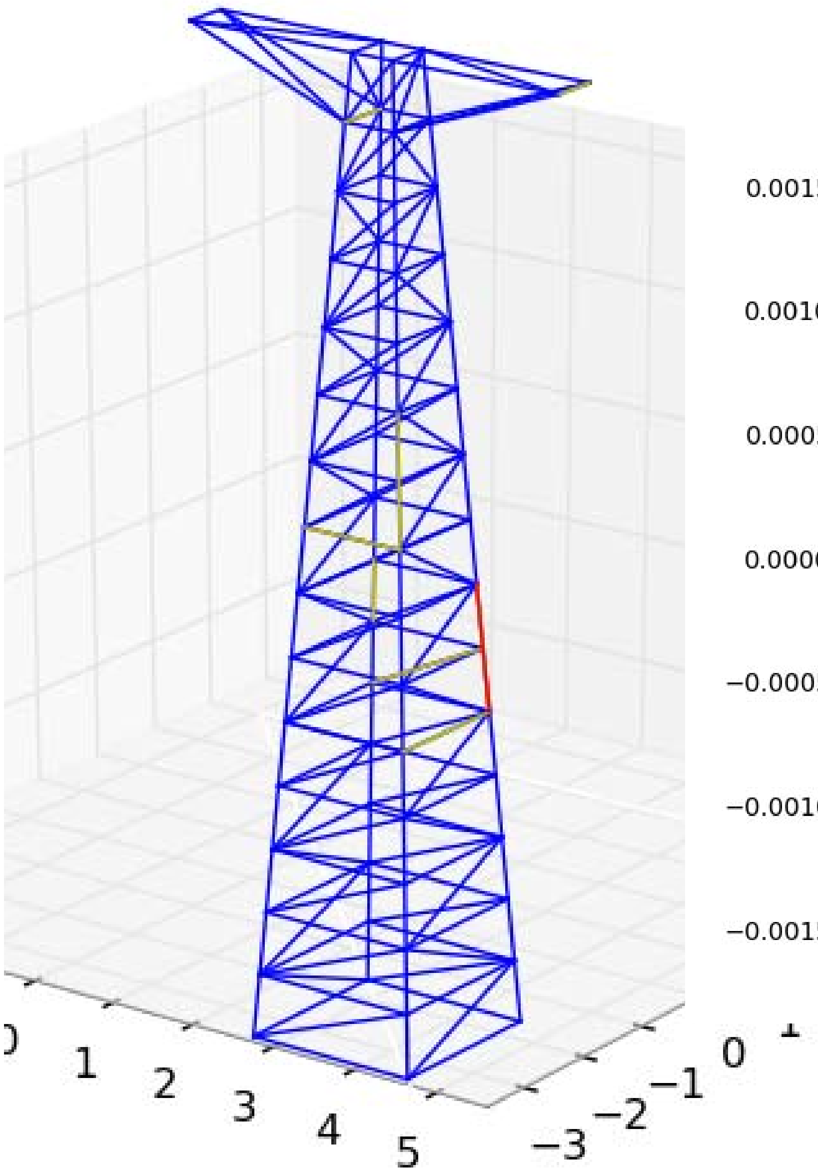

**Ground acceleration**

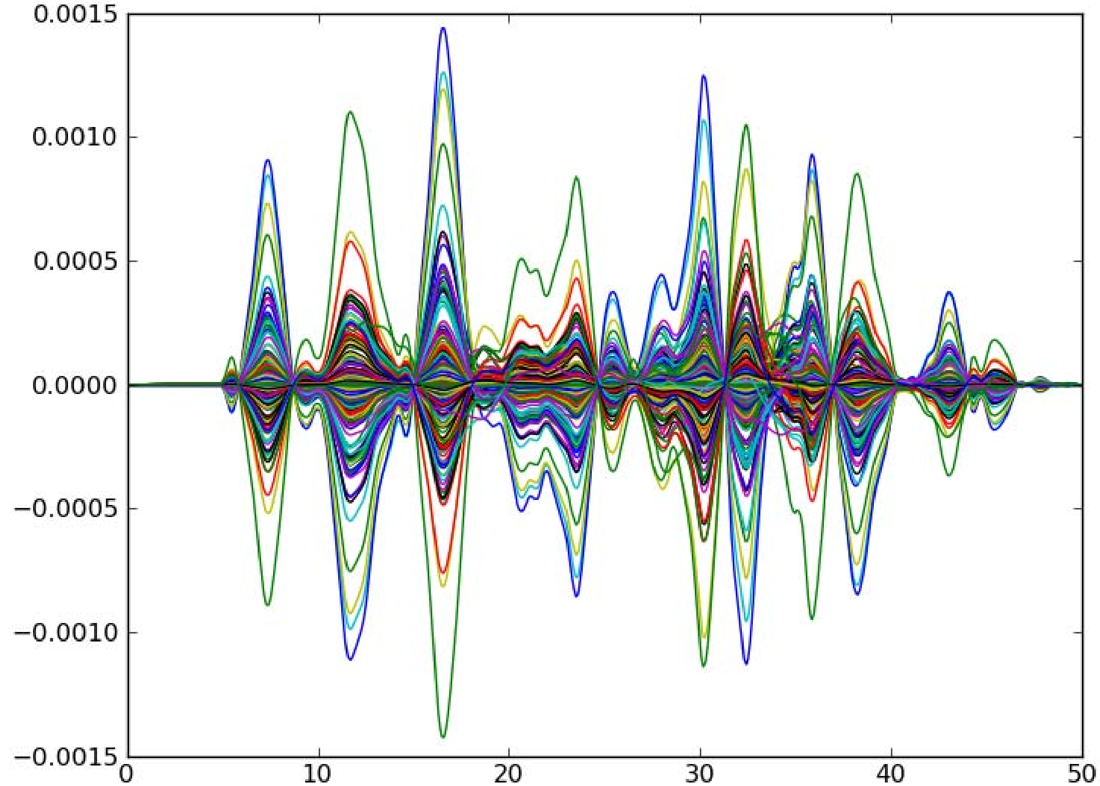


# Vulnerability Curves (vs earthquake magnitude)



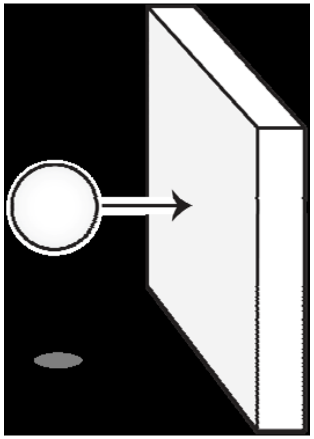
# Identification of the weakest elements

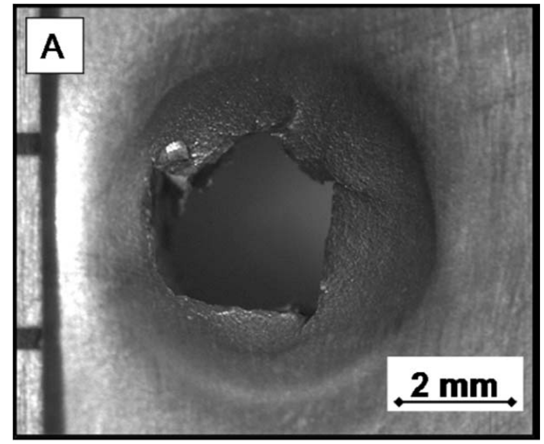



H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. **Optimal Uncertainty Quantification**. SIAM Review, 55(2):271–345, 2013.

# 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**

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

**Problem** We don't know $G$ nor $\mathbb{P}$.

# What do we know?

**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**

$$m_1 \leq \mathbb{E}[G] \leq m_2$$

**Bounds on the sensitivity of the response function w.r. to each variable**

$$\operatorname{Osc}_i G \leq D_i$$

$$\operatorname{Osc}_i G := \sup\left\{|G(x) - G(x')| \mid x_j = x'_j \text{ for } j \neq i\right\}$$

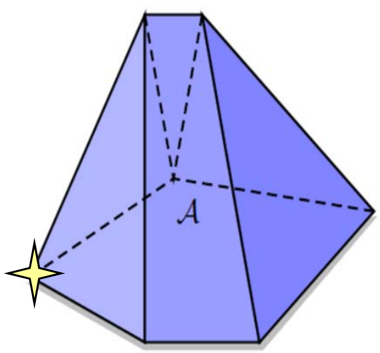
**We only know** $(G, \mathbb{P}) \in \mathcal{A}$

$$\mathcal{A} := \left\{ (f, \mu) \,\middle|\, \begin{array}{c} \mu = \mu_1 \otimes \mu_2 \otimes \mu_3, \\ m_1 \leq \mathbb{E}_\mu[f] \leq m_2 \\ \mathrm{Osc}_i\, f \leq D_i \end{array} \right\}$$

**Worst case bound**

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

**Reduction calculus**



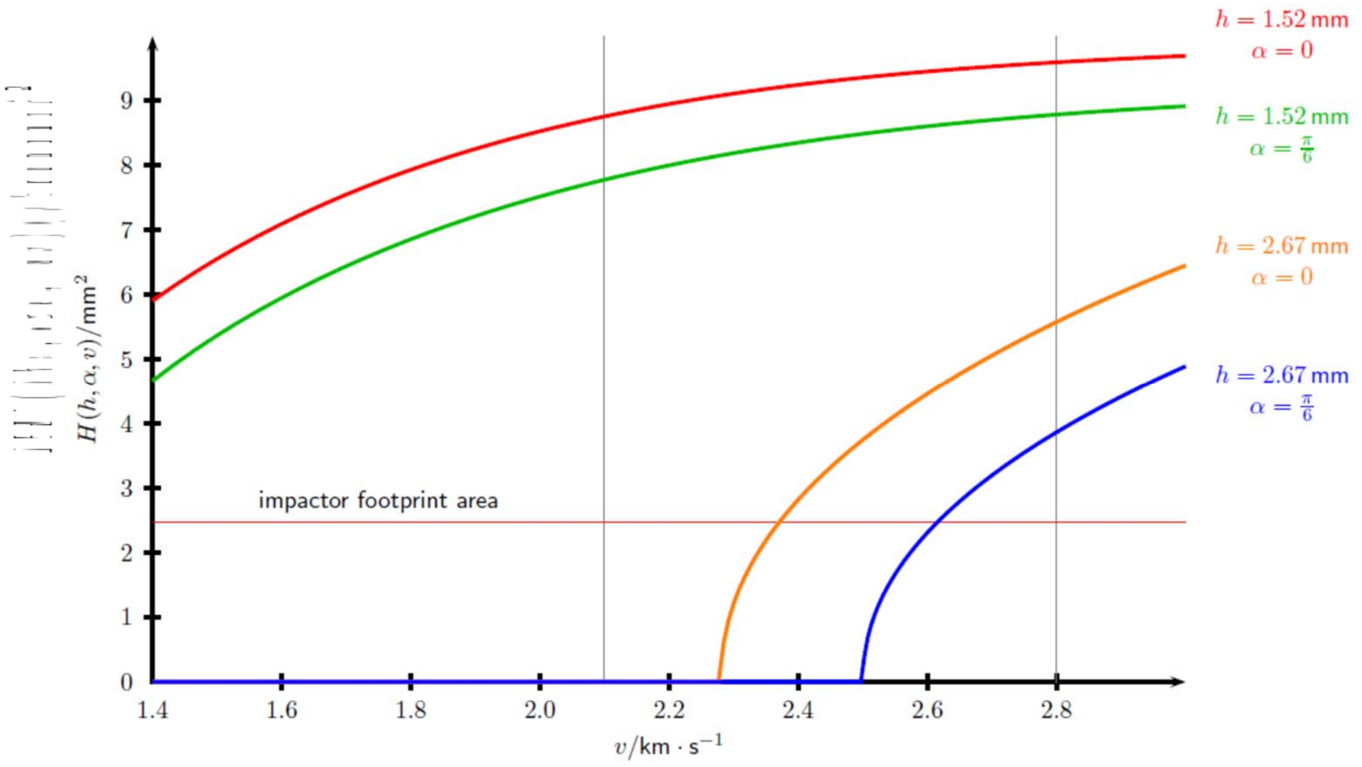

$$\mathcal{U}(\mathcal{A}) = 43.7\%$$

# What if we know the response function?

What if we know $G = H$?

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

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



## Optimal bound on the probability of non perforation

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

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

## Application of the reduction calculus

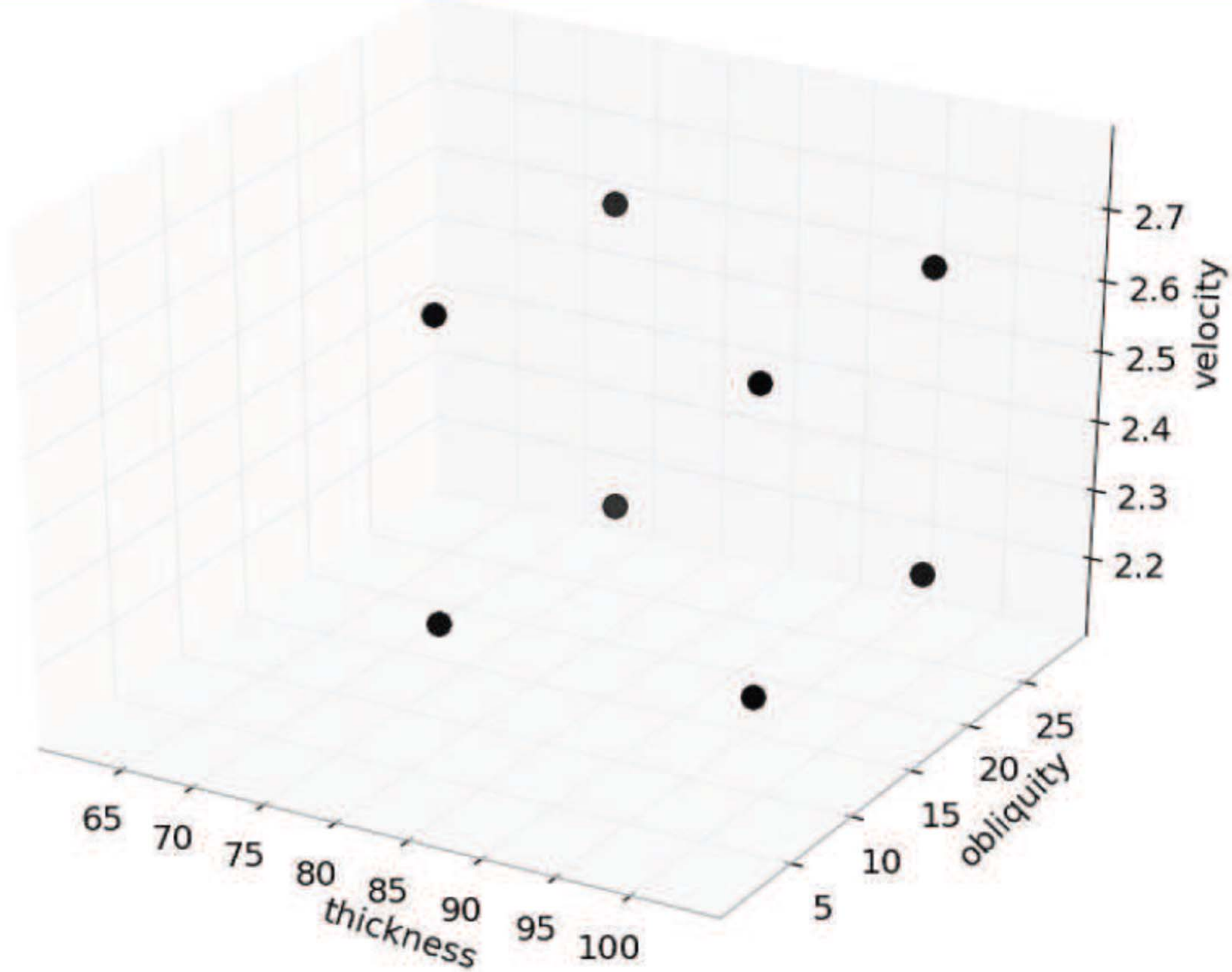
$$\mathcal{U}(\mathcal{A}) = \mathcal{U}(\mathcal{A}_\Delta)$$

$$\mathcal{A}_\Delta := \{(f,\mu) \in \mathcal{A} | \mu_k \text{ has support at only 2 points}\}$$

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

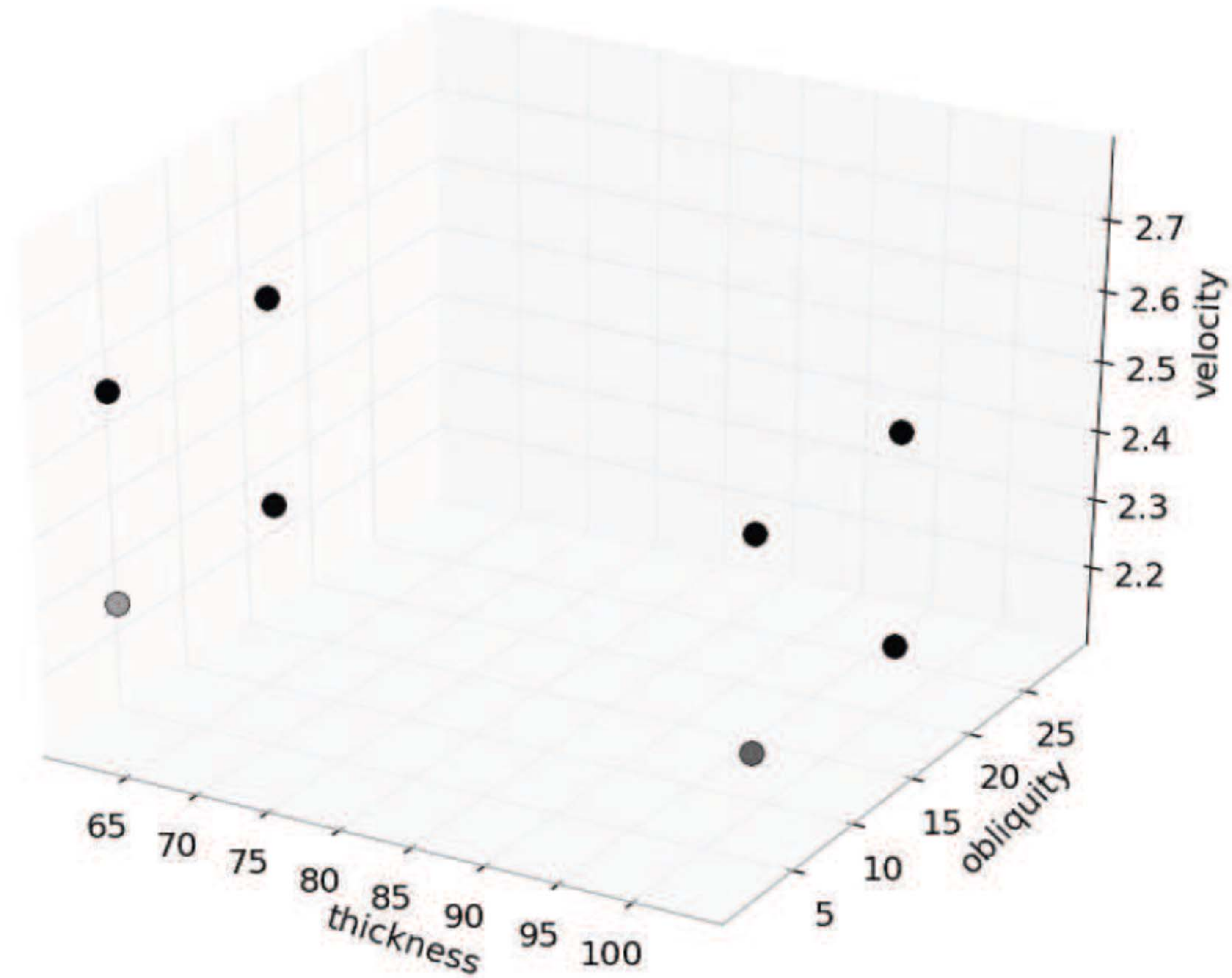
$$\mathcal{U}(\mathcal{A}) \overset{\text{num}}{=} 37.9\%$$

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



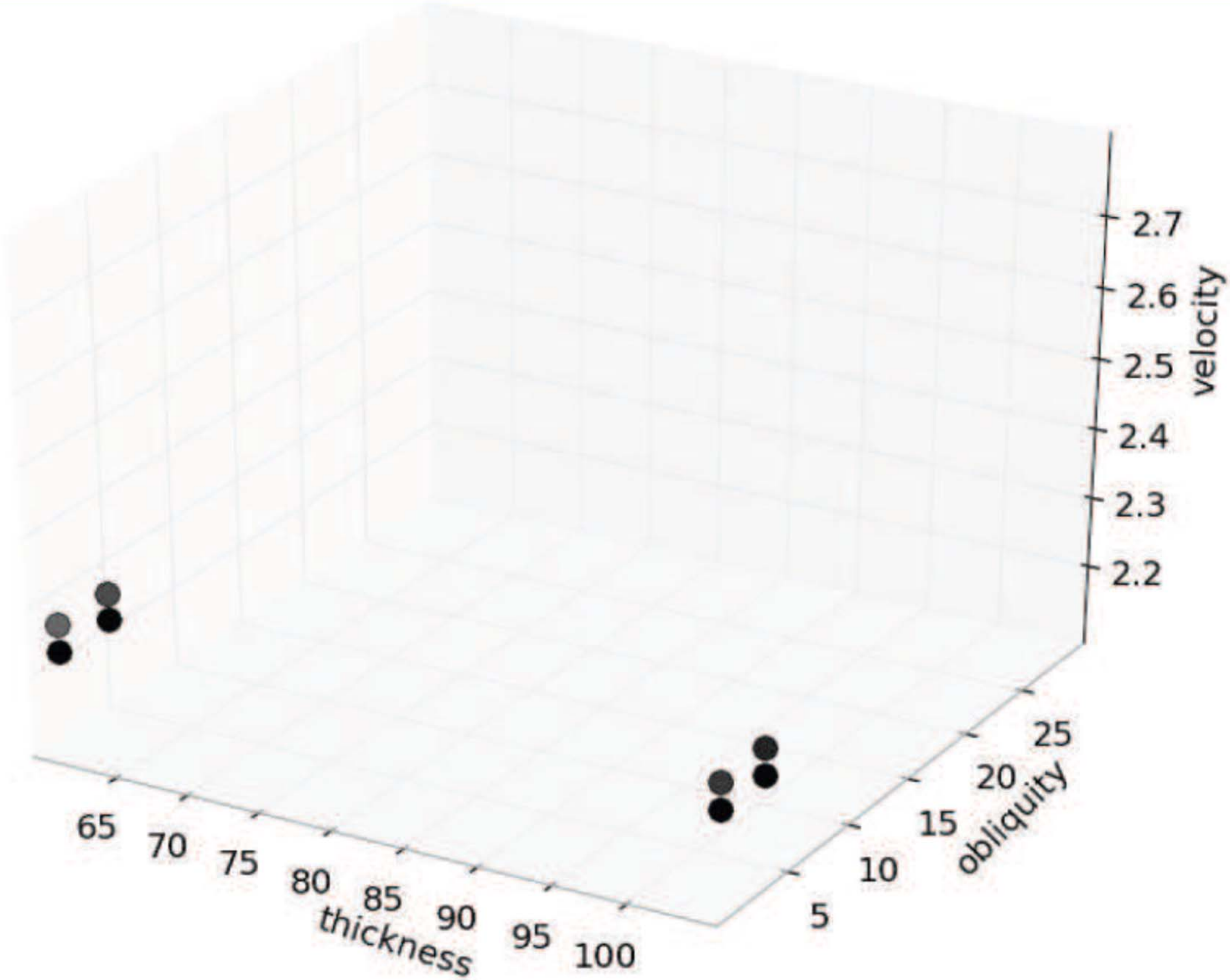

**Support Points at iteration 0**

# Numerical optimization



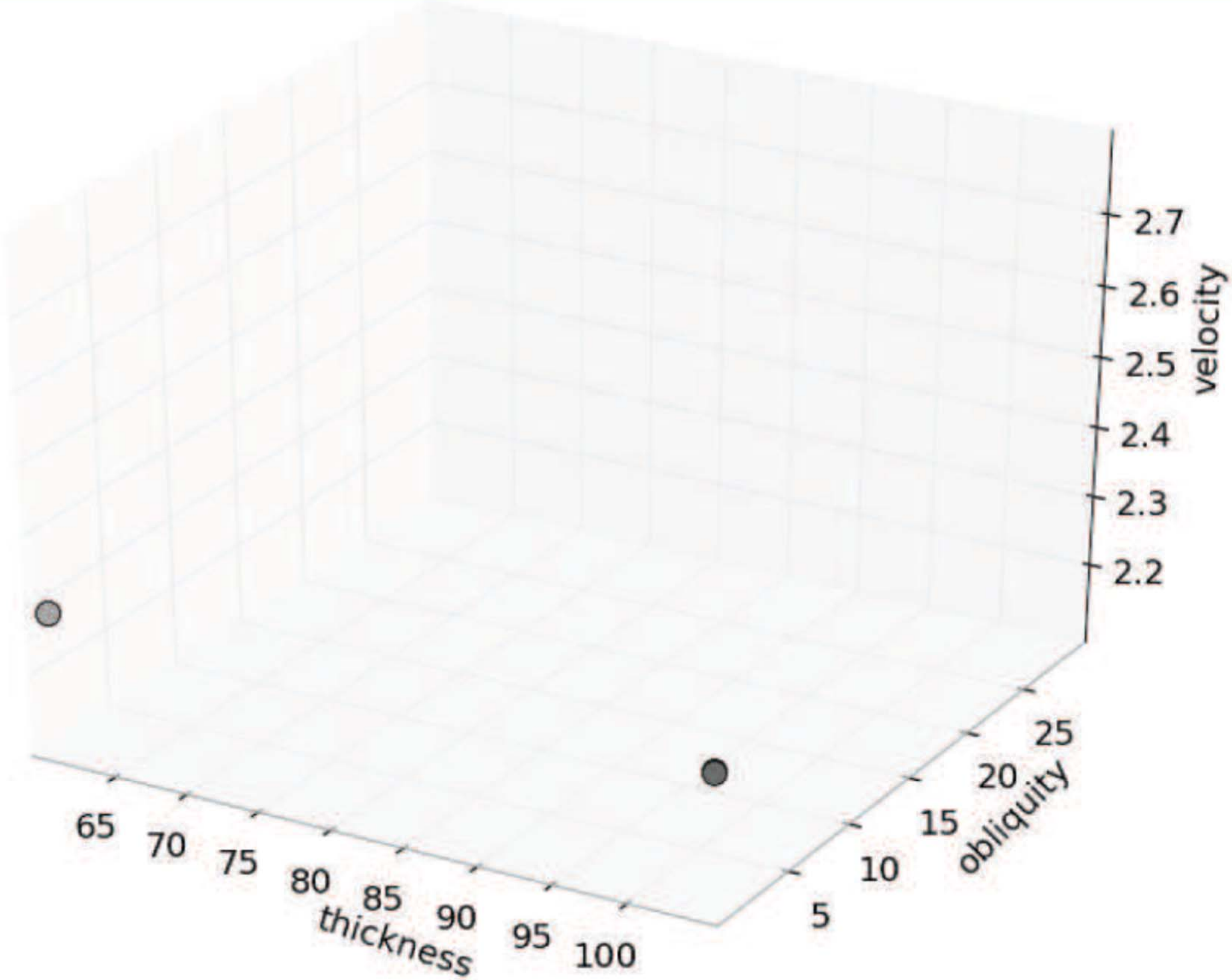

**Support Points at iteration 150**

# 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.

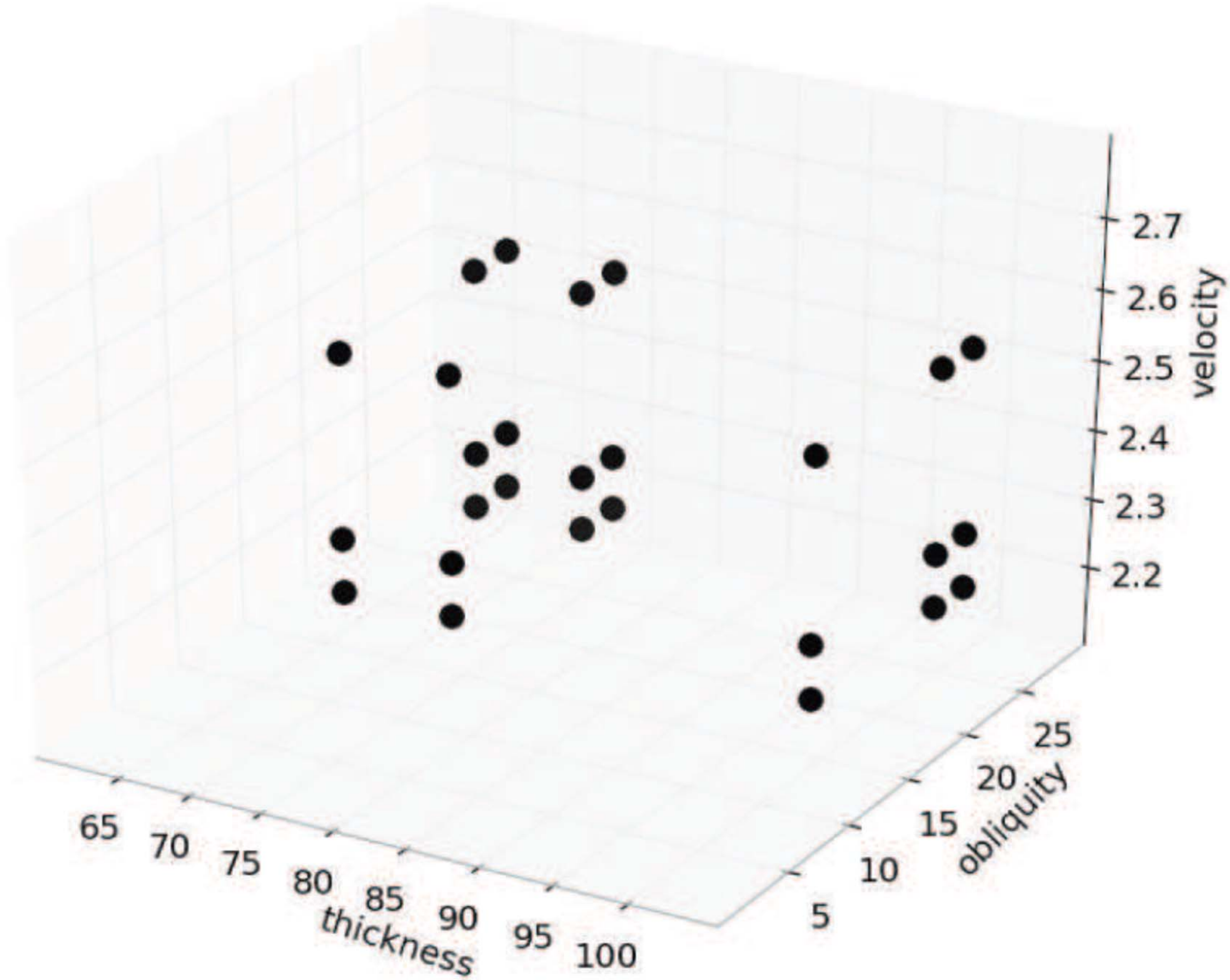
# Important observations

**Extremizers are singular**

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

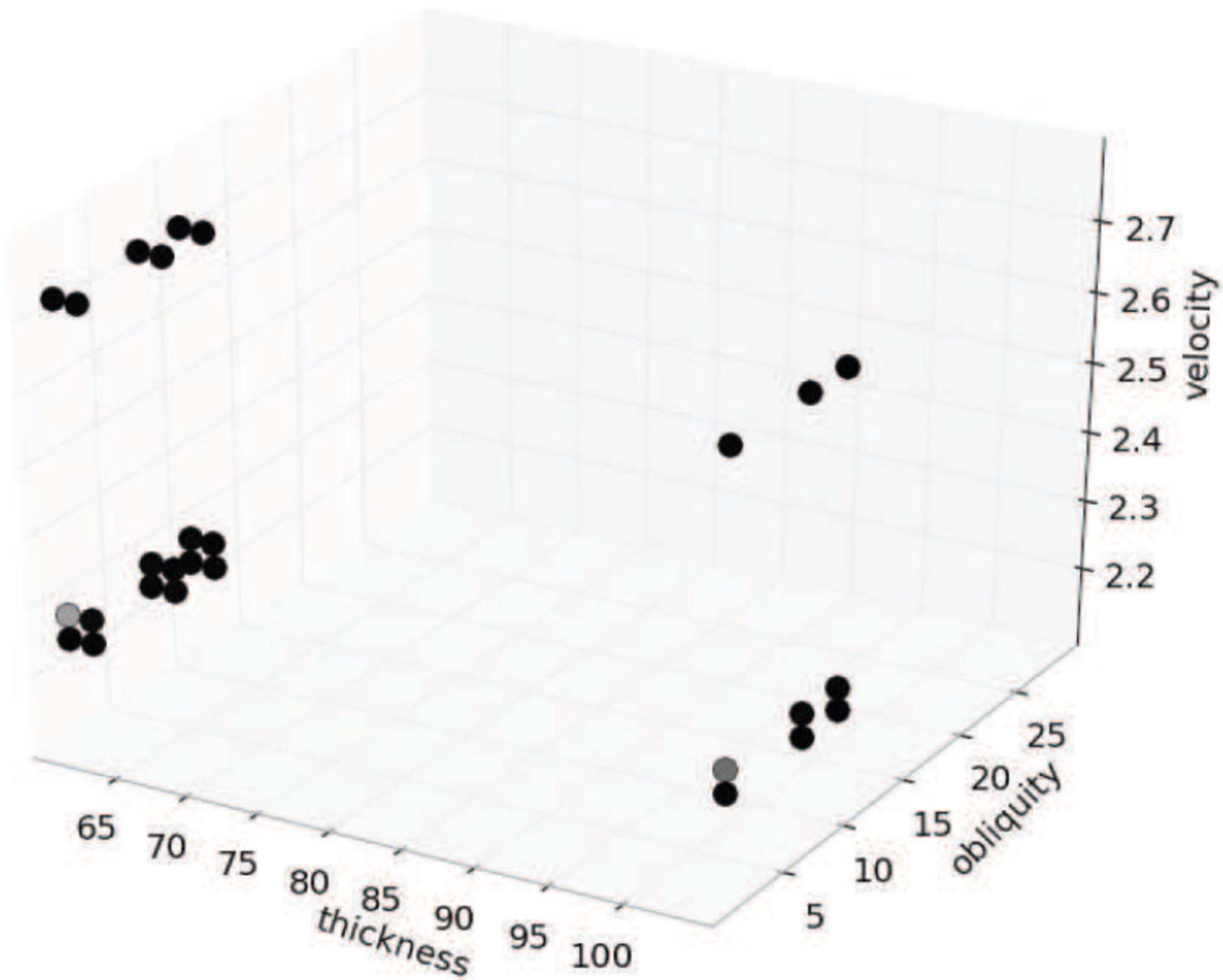
**Extremizers are attractors**

# Initialization with 3 support points per marginal



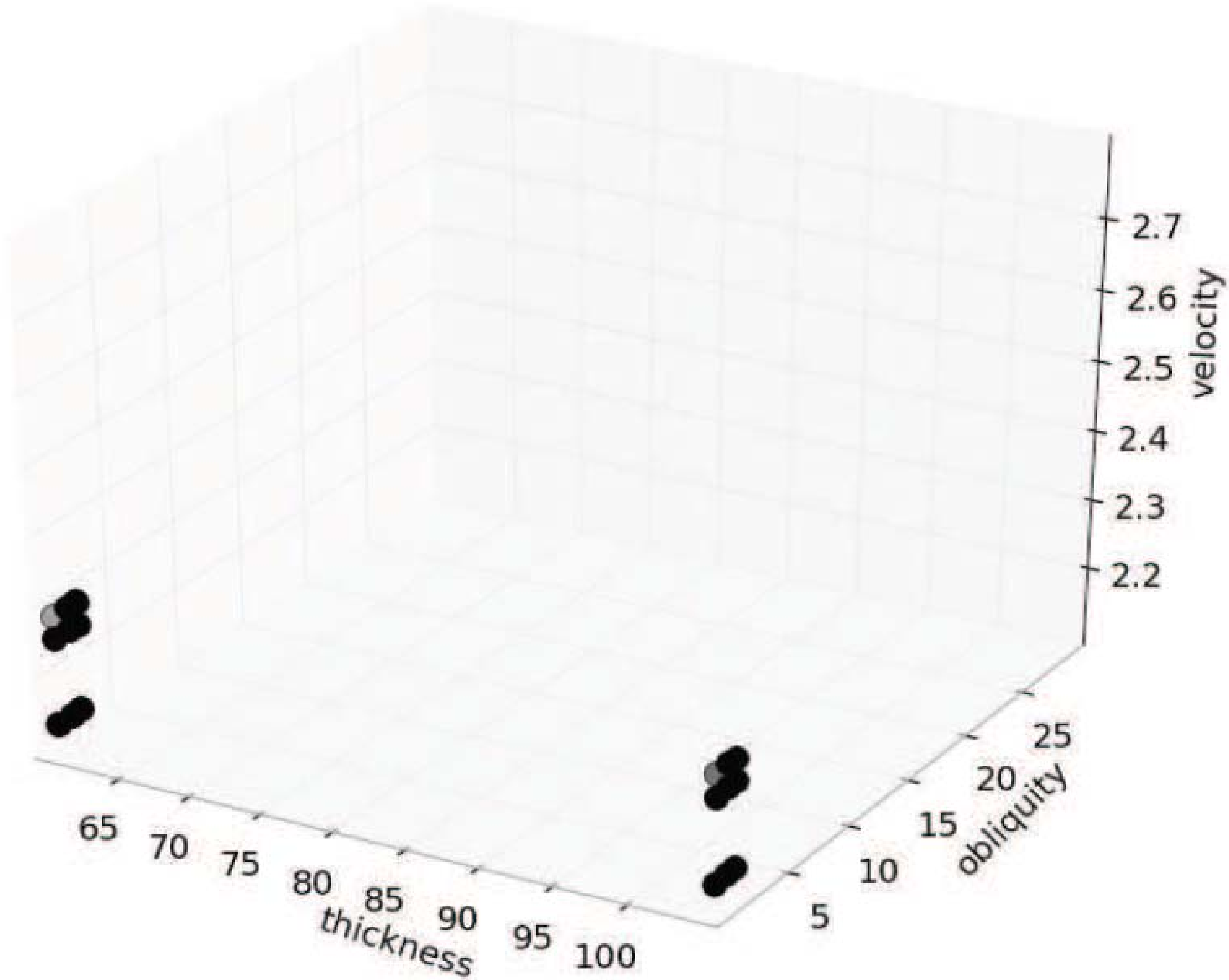

**Support Points at iteration 0**

## Initialization with 3 support points per marginal



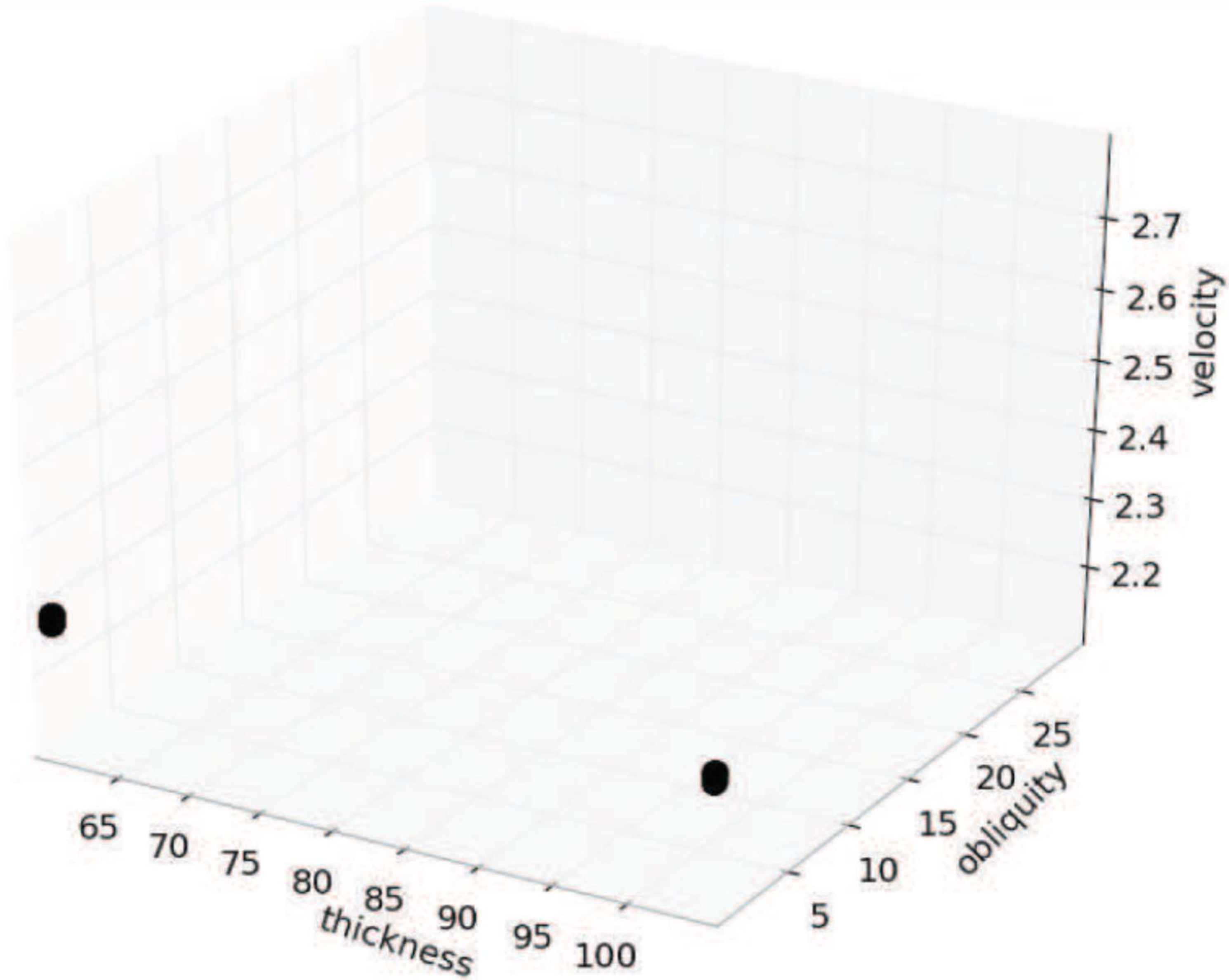

**Support Points at iteration 500**

## Initialization with 3 support points per marginal



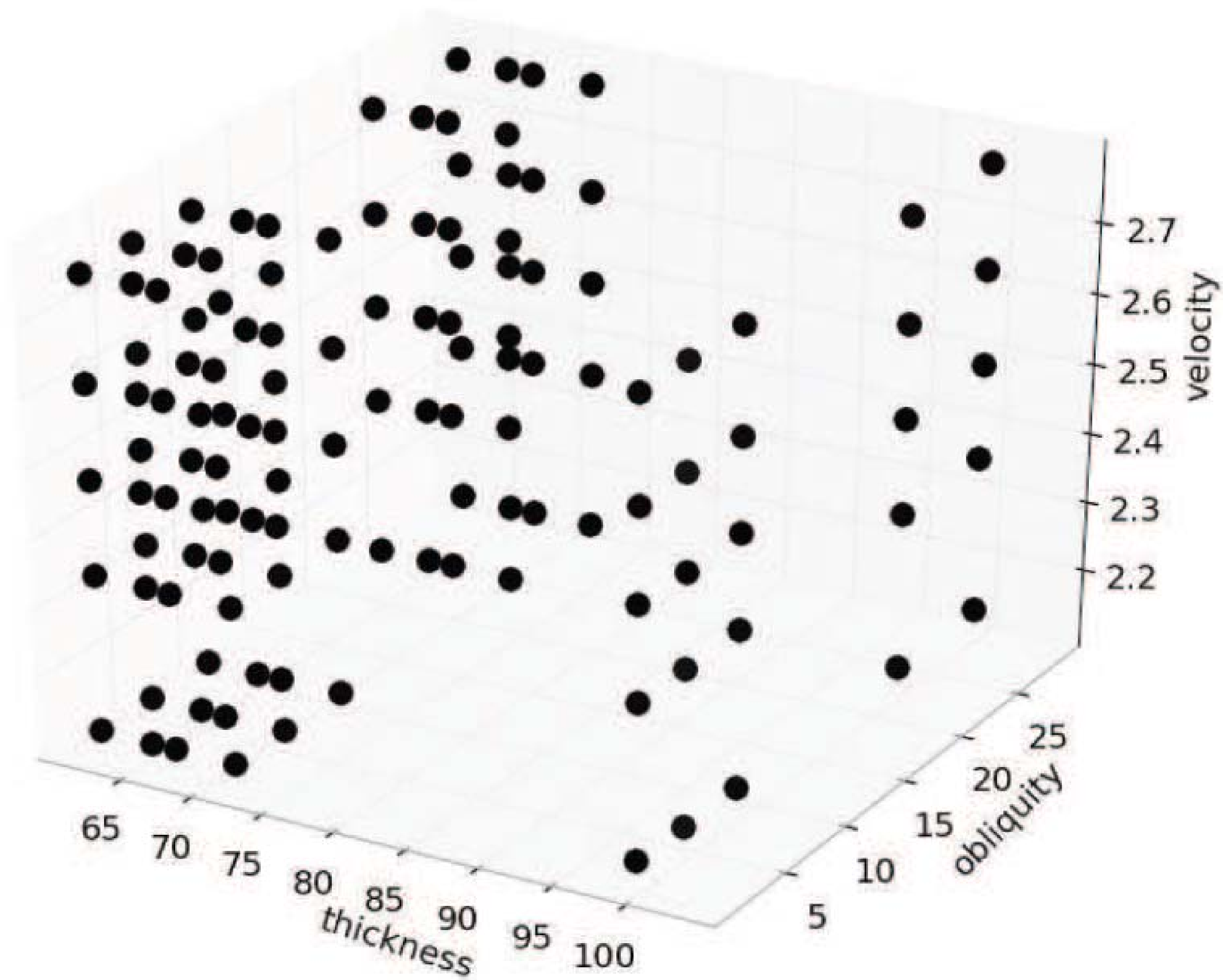

**Support Points at iteration 1000**

Initialization with 3 support points per marginal



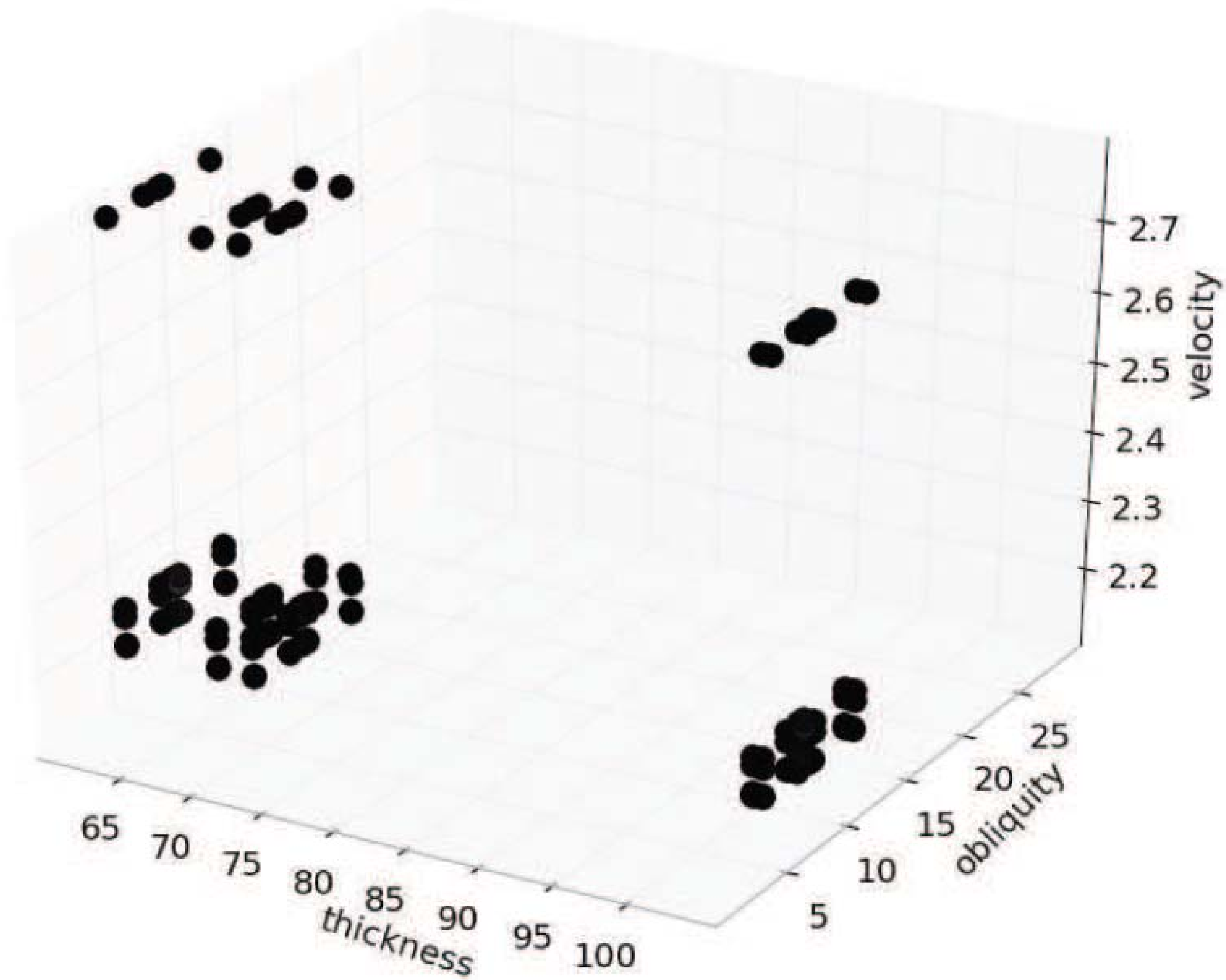

**Support Points at iteration 2155**

## Initialization with 5 support points per marginal



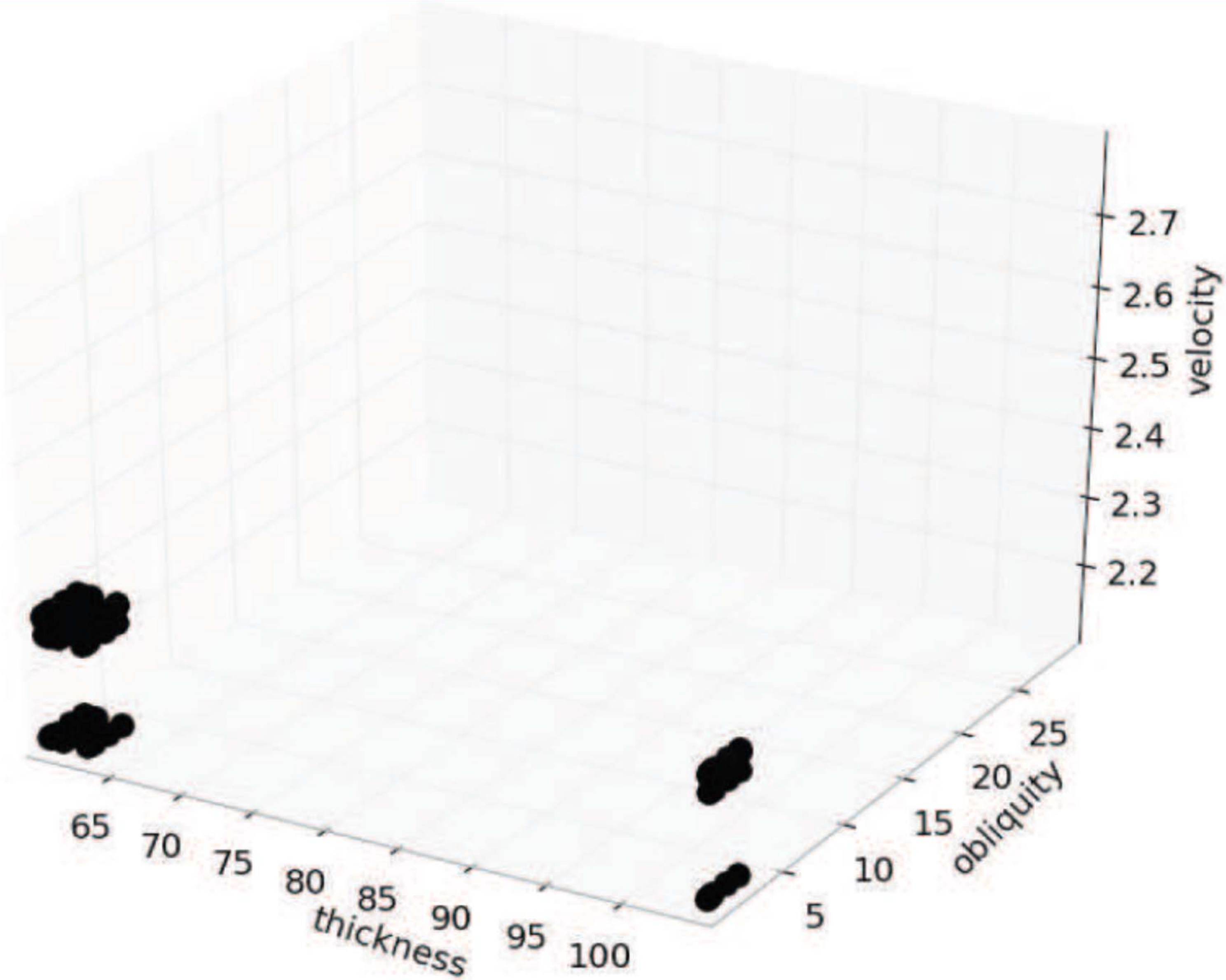

**Support Points at iteration 0**

# Initialization with 5 support points per marginal



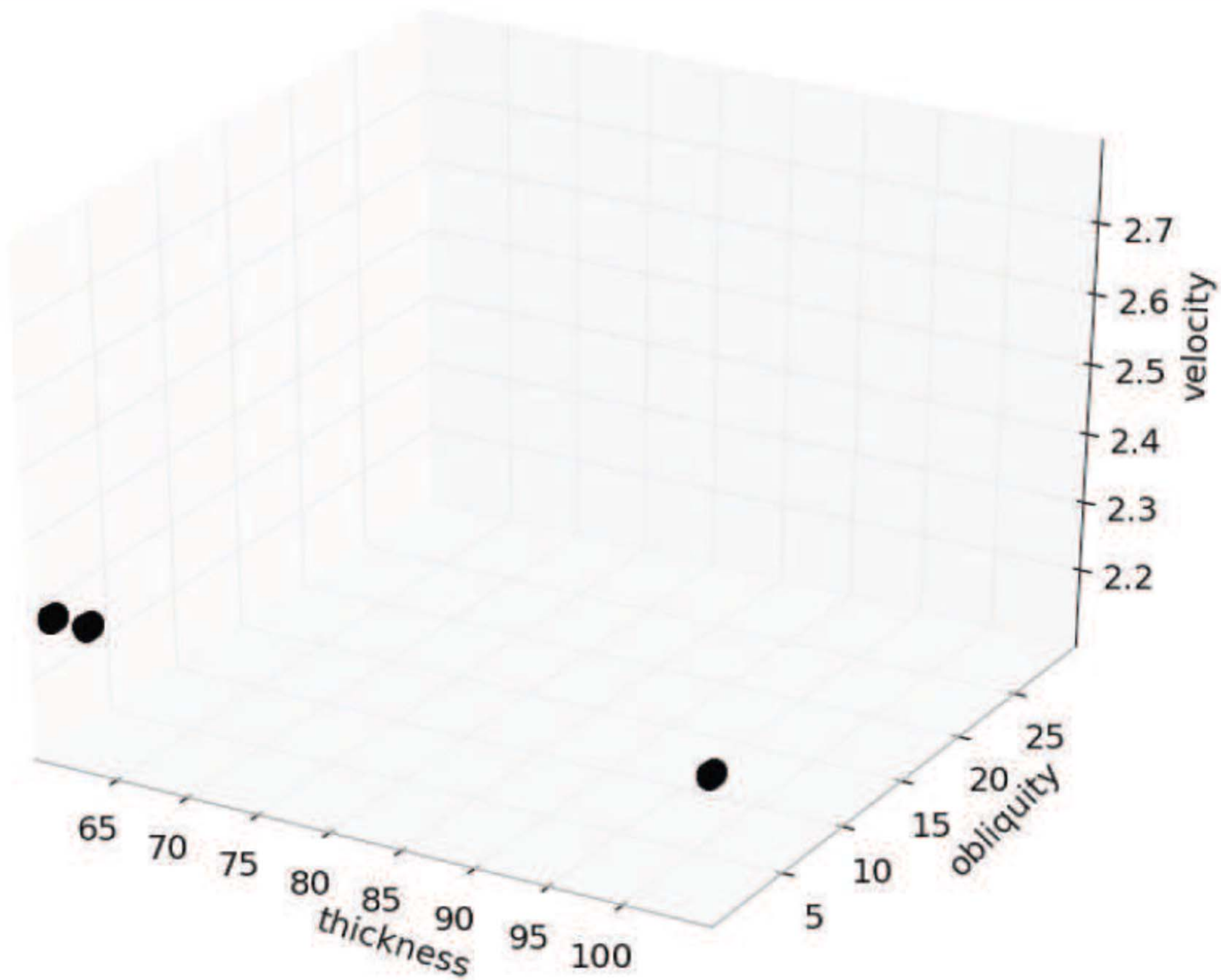

**Support Points at iteration 1000**

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 + Legacy data

**Objective**

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

**Constraints on input variables**

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

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

**Constraint on the mean perf. area**

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

**Modified Lipschitz continuity constraints 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

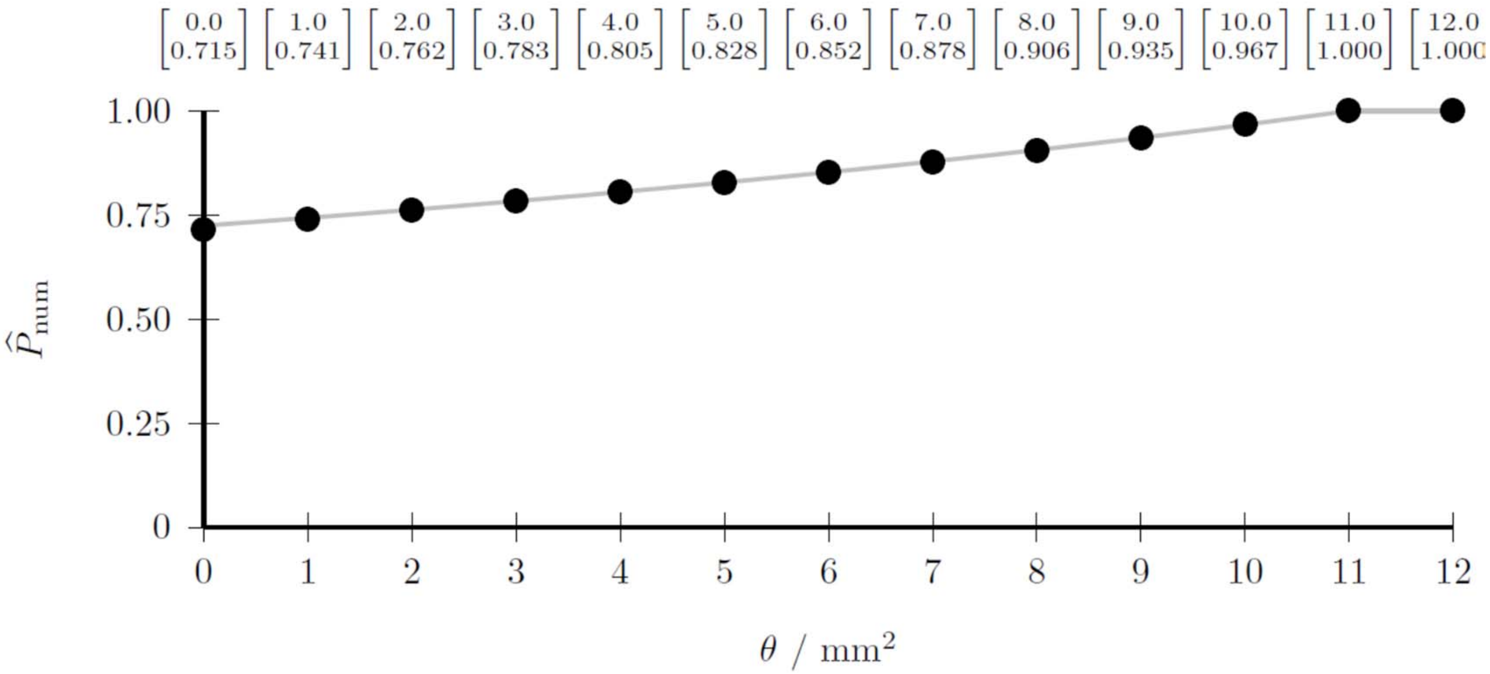
## 32 data points

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

These constrain the value of G at 32 points

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi, and M. Ortiz. Optimal uncertainty quantification for legacy data observations of Lipschitz functions. *ESAIM Math. Model. Numer. Anal.*, 47(6):1657–1689, 2013.

| ID | $h$ (inches) | $\alpha$ (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

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

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

**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.**

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi, and M. Ortiz. Optimal uncertainty quantification for legacy data observations of Lipschitz functions. *ESAIM Math. Model. Numer. Anal.*, 47(6):1657–1689, 2013.

| ID | $h$ (inches) | $\alpha$ (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 |

## What if we have model uncertainty?

### What do we want?

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

### What do we know?

Numerical model $(h,v) \rightarrow F(h,v)$

59 noisy data/experimental points $G(h_i, v_i)$

Expert judgement (+data points):

$\mathbb{E}[G(h,v)] \geq 12\,\mathrm{mm}^2$

$h$ and $v$ are independent random variables

$(h,v) \in [0.5, 3.0]\mathrm{mm} \times [4.5, 7.0]\mathrm{km/s}$

# PSAAP numerical model

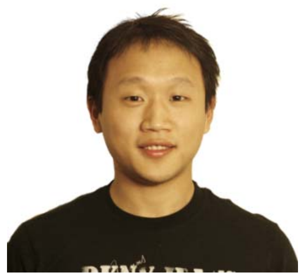
**Plate Obliquity=0**

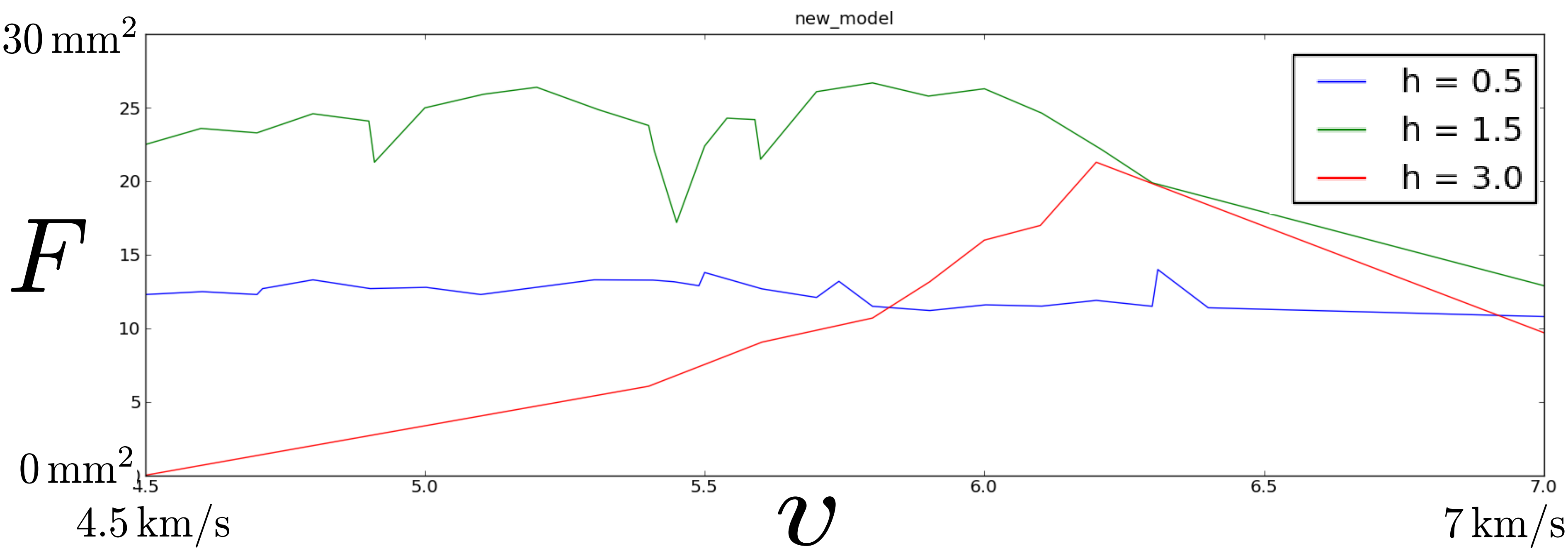
$$(h, v) \xrightarrow{\mathbf{F}} F(h, v)$$

**Plate thickness**

**Projectile velocity**

**Perforation area**





# 59 data/experimental points

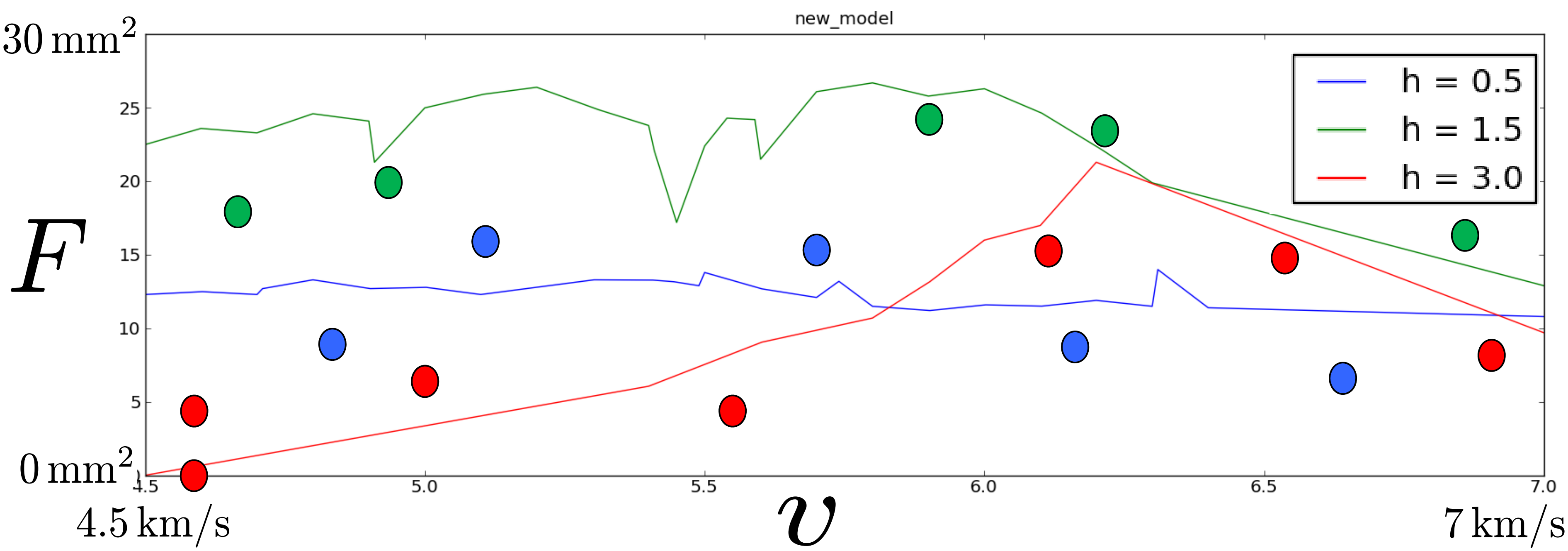
**Plate Obliquity=0**

$$(h, v) \xrightarrow{\mathbf{G}} G(h, v)$$

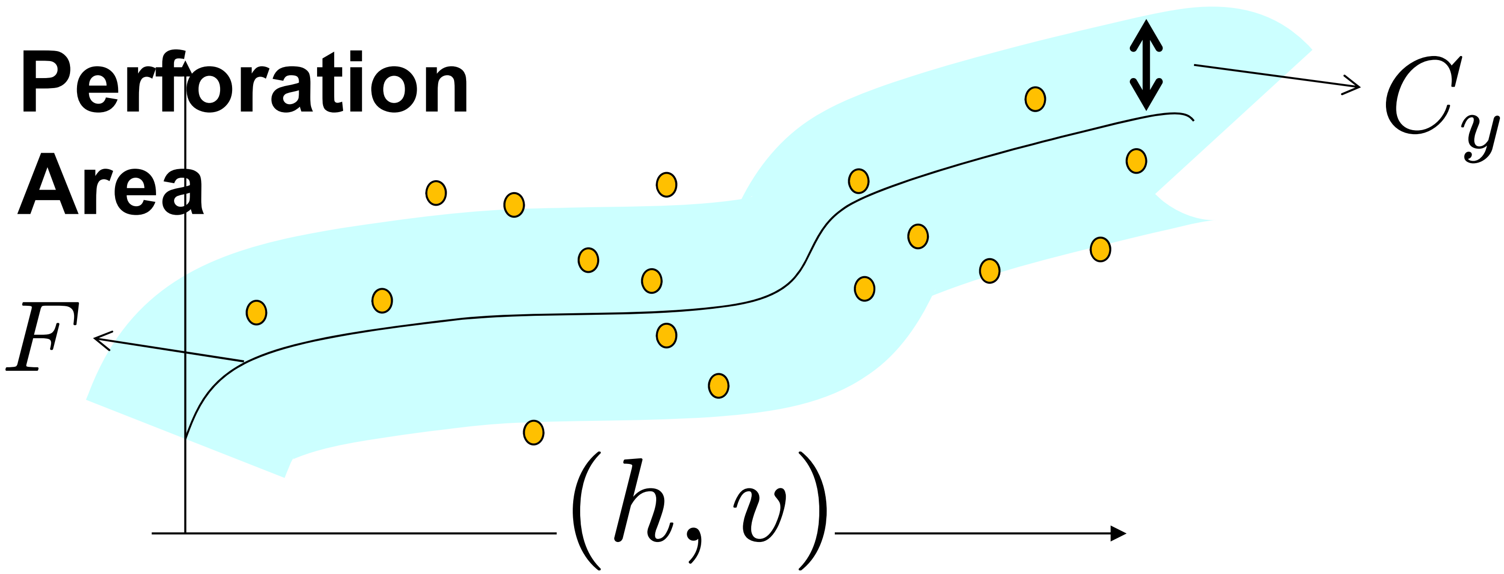
**Plate thickness**

**Projectile velocity**

**Perforation area**



# Confidence sausage around the model




With probability $p_1$ we have

$$\mathbb{P}[|G - F| \leq C_y] \geq p_2(p_1, C_y, data)$$

| $Cy$ | at prob. 0.78 | at prob. 0.85 | at prob. 0.94 |
|---|---|---|---|
| 3 | 0.3991 | 0.3822 | 0.3483 |
| 5 | 0.5325 | 0.5155 | 0.4816 |
| 7 | 0.6325 | 0.6155 | 0.5816 |
| 9 | 0.8325 | 0.8155 | 0.7816 |

# Admissible set

$$\mathcal{A} := \left\{ (g,\mu) \;\middle|\; \begin{array}{c} g\colon \mathcal{X} \to \mathcal{Y}, \\ \mu = \mu_h \otimes \mu_v, \\ \mathcal{X} := [0.5, 3.0]\text{mm} \times [4.5, 7.0]\text{km/s}, \\ \mathbb{E}_\mu[g(h,v,\alpha)] \geq 12.0\text{mm}^2, \\ \|g - F\|_\infty \leq C_y \end{array} \right\}$$

# Confidence sausage

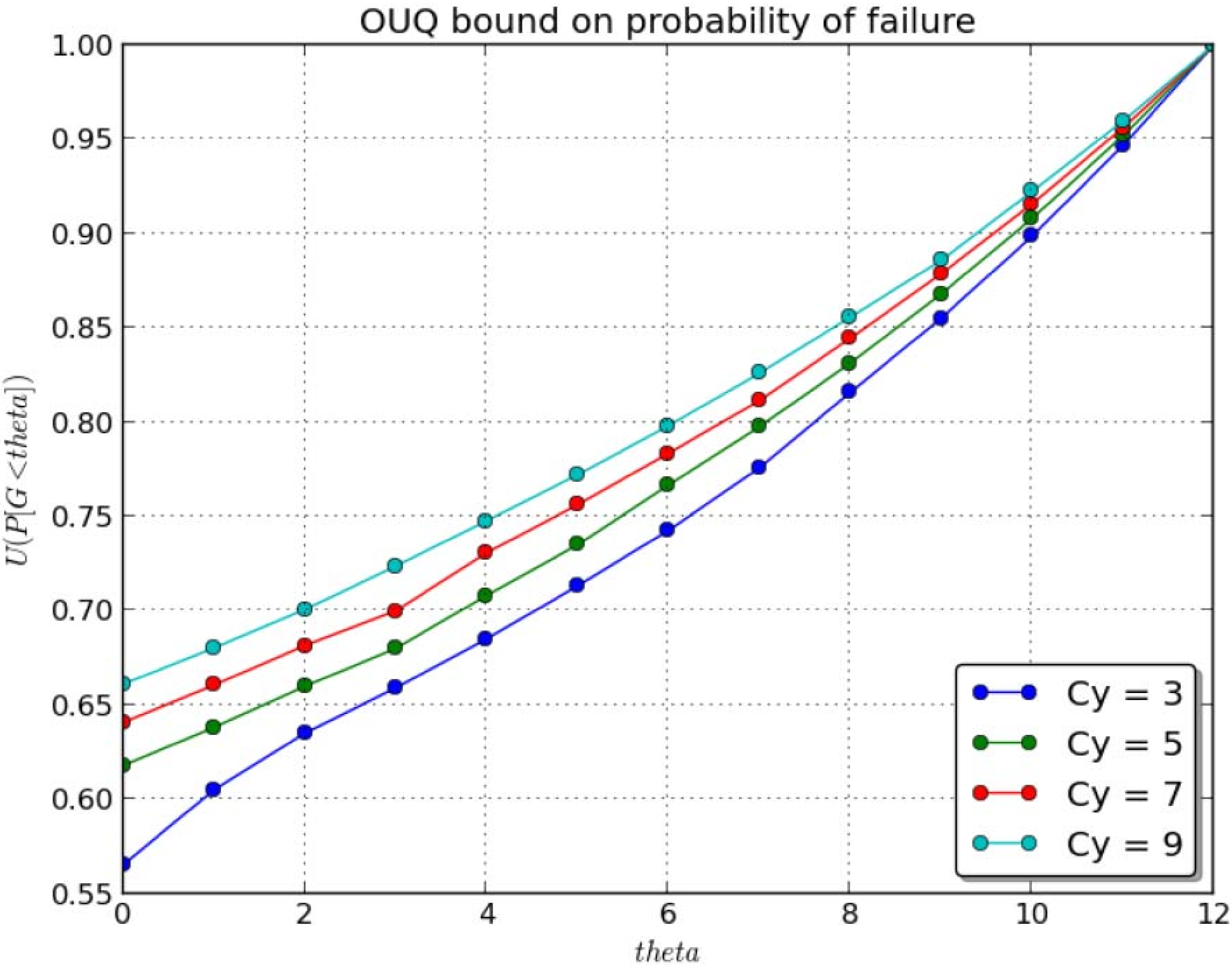
With probability $p_1$ we have

$$\mathbb{P}[|G - F| \leq C_y] \geq p_2(p_1, C_y, data)$$

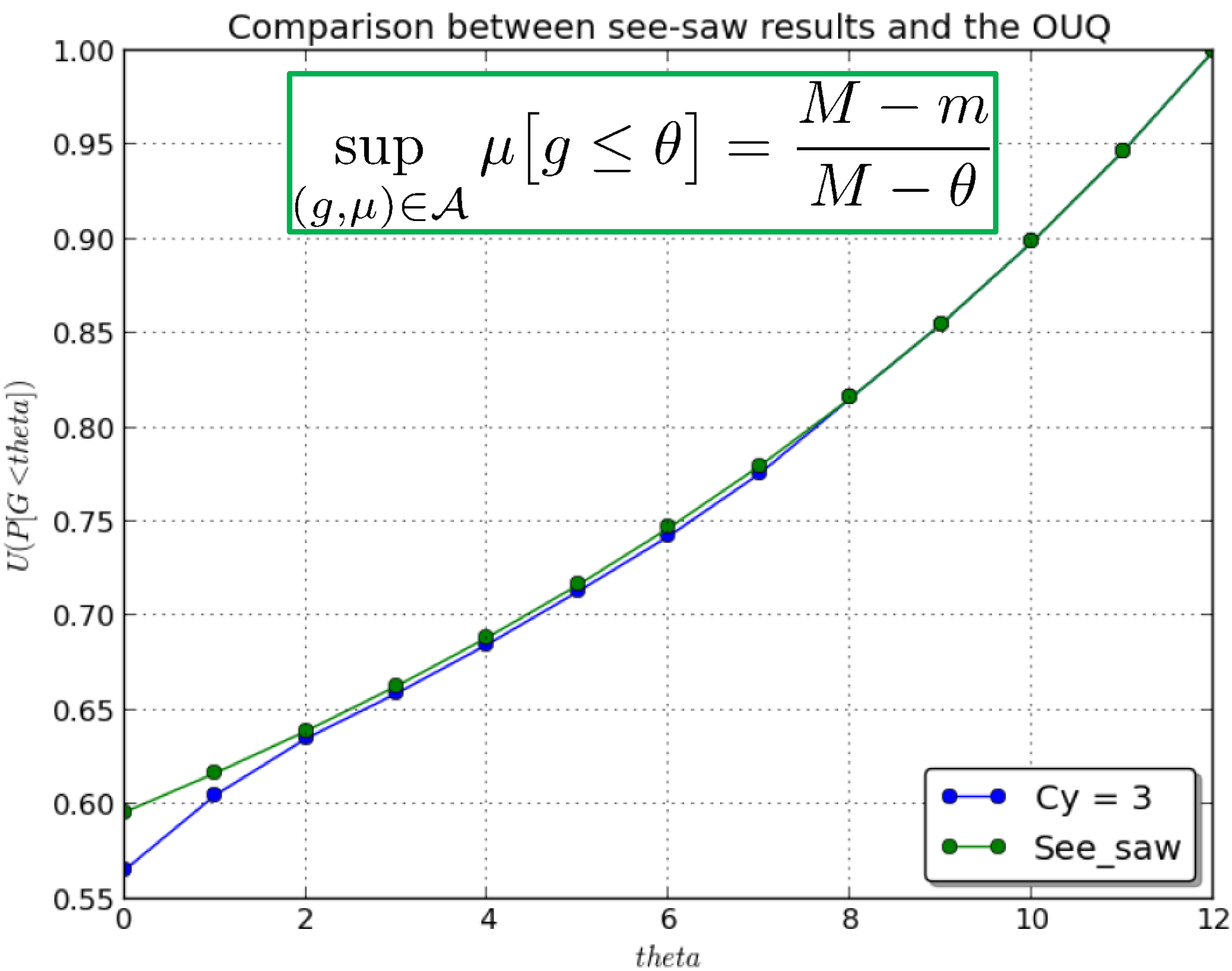
For $C_y = 5$ and $p_1 = 0.85$ we have $p_2 = 0.51$

# What we compute

$$\sup_{(g,\mu)\in\mathcal{A}} \mu[g \leq \theta]$$



OUQ bound on probability of failure
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8
10
12
theta
U(P[G <theta])
Cy = 3
Cy = 5
Cy = 7
Cy = 9



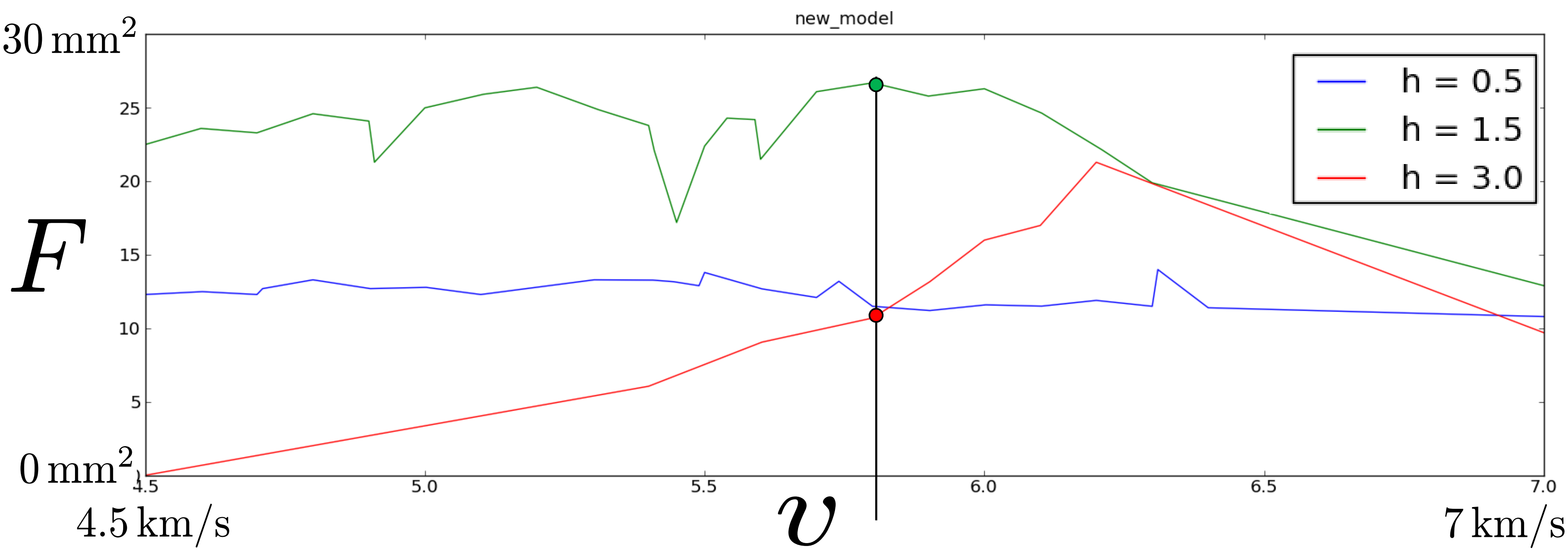
Comparison between see-saw results and the OUQ
$$\sup_{(g,\mu)\in\mathcal{A}} \mu[g \leq \theta] = \frac{M-m}{M-\theta}$$
$U(P[G < theta])$
theta
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
2
4
6
8
10
12
Cy = 3
See_saw

$$\sup_{(g,\mu)\in\mathcal{A}} \mu[g \leq \theta] = \frac{M-m}{M-\theta}$$

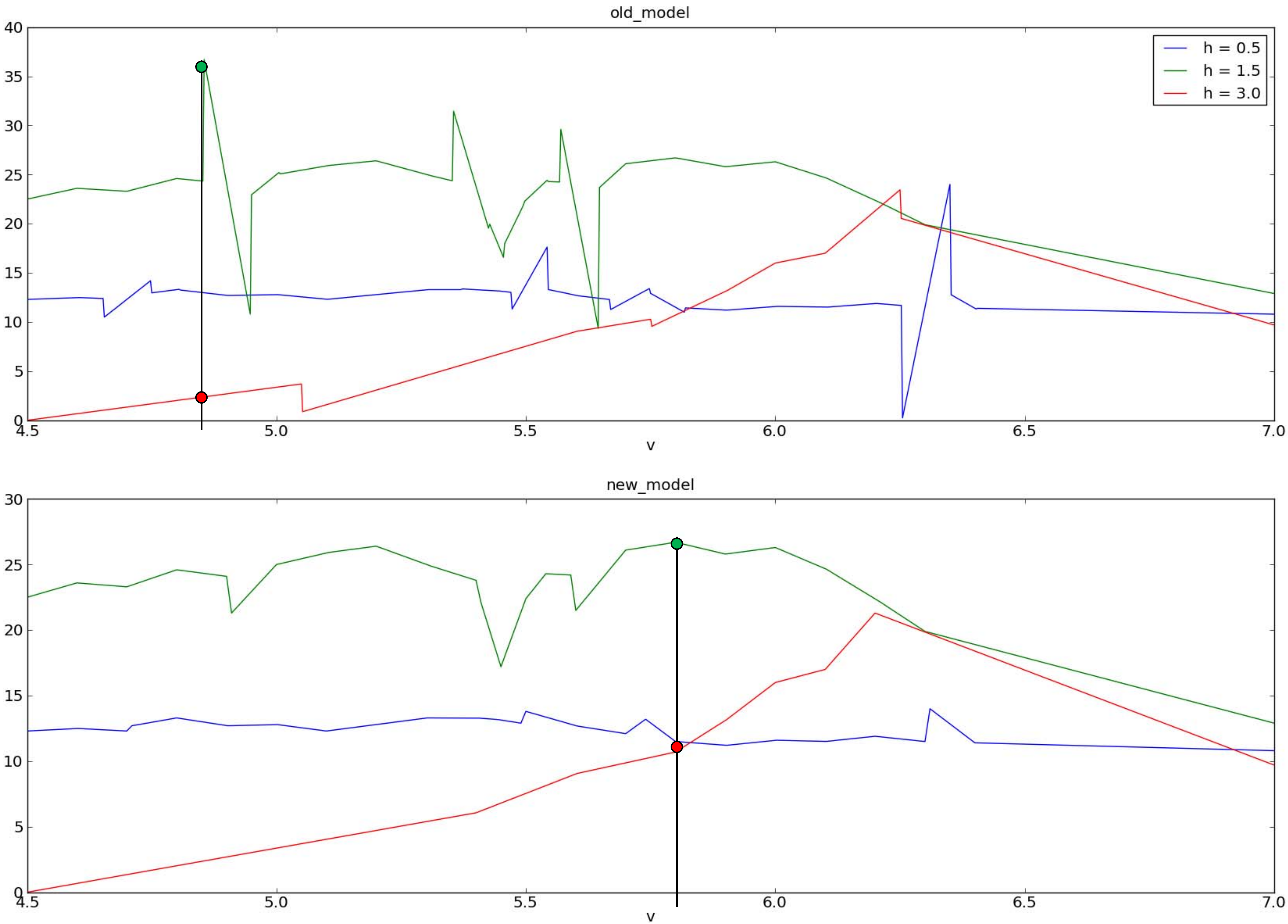
For $C_y = 3$ and $\theta \geq 8$ the model impacts the least upper bound only through its maximum value $M$

At the extremum

$\mu_v$ collapses to a single mass of Dirac



# The extremizers led to the identification of a bug in an old model



## Caltech PSAAP Center UQ analysis


L. J. Lucas, H. Owhadi, and M. Ortiz. Rigorous verification, validation, uncertainty quantification and certification through concentration-of-measure inequalities. *Comput. Methods Appl. Mech. Engrg.*, 197(51-52):4591–4609, 2008.

M. M. McKerns, L. Strand, T. J. Sullivan, A. Fang, and M. A. G. Aivazis. Building a framework for predictive science. In *Proceedings of the 10th Python in Science Conference (SciPy 2011)*, 2011.

P.-H. T. Kamga, B. Li, M. McKerns, L. H. Nguyen, M. Ortiz, H. Owhadi, and T. J. Sullivan. Optimal uncertainty quantification with model uncertainty and legacy data. *Journal of the Mechanics and Physics of Solids*, 72:1–19, 2014

A. A. Kidane, A. Lashgari, B. Li, M. McKerns, M. Ortiz, H. Owhadi, G. Ravichandran, M. Stalzer, and T. J. Sullivan. Rigorous model-based uncertainty quantification with application to terminal ballistics. Part I: Systems with controllable inputs and small scatter. *Journal of the Mechanics and Physics of Solids*, 60(5):983–1001, 2012.

T. J. Sullivan, M. McKerns, D. Meyer, F. Theil, H. Owhadi, and M. Ortiz. **Optimal uncertainty quantification for legacy data observations of Lipschitz functions.** *ESAIM Math. Model. Numer. Anal.*, 47(6):1657–1689, 2013.

H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. **Optimal Uncertainty Quantification**. *SIAM Review*, 55(2):271–345, 2013.

**Reduced numerical optimization problems solved using**

- **mystic: http://trac.mystic.cacr.caltech.edu/project/mystic**
  - **a highly-configurable optimization framework**

- **pathos: http://trac.mystic.cacr.caltech.edu/project/pathos**
  - **a distributed parallel graph execution framework providing a high-level programmatic interface to heterogeneous computing**

**Mike McKerns**

# Important observations

**In presence of incomplete information on the distribution of input variables the dependence of the least upper bound on the accuracy of the model is very weak**

**We need to extract as much information as possible from the sample/experimental data on the underlying distributions**

**How do  we reason with the worst in presence of data sampled from an unknown distribution?**

**Quantity of Interest** $\Phi(\mu^\dagger) = \mu^\dagger[X \geq a]$

$\mu^\dagger$: Unknown or partially known measure of probability on $\mathbb{R}$

**You know** $\mu^\dagger \in \mathcal{A}$

**You observe** $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$

$n$ i.i.d samples from $\mu^\dagger$

**Problem:**

Find the best estimate of $\Phi(\mu^\dagger)$

$$\theta(d)$$

**Player I**

Chooses

$\mu \in \mathcal{A}$

*Max*

**Player II**

Sees $d \sim \mu^n$

Chooses $\theta$

*Min*

$$\mathcal{E}(\mu, \theta)$$

**Mean squared error**

$$\mathcal{E}(\mu, \theta) = \mathbb{E}_{d \sim \mu^n}\Big[\big[\theta(d) - \Phi(\mu)\big]^2\Big]$$

**Confidence error**

$$\mathcal{E}(\mu, \theta) = \mathbb{P}_{d \sim \mu^n}\Big[\big|\theta(d) - \Phi(\mu)\big| \geq r\Big]$$

# Game theory and statistical decision theory

**John Von Neumann**

**Abraham Wald**

J. Von Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100(1):295–320, 1928

J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, New Jersey, 1944.

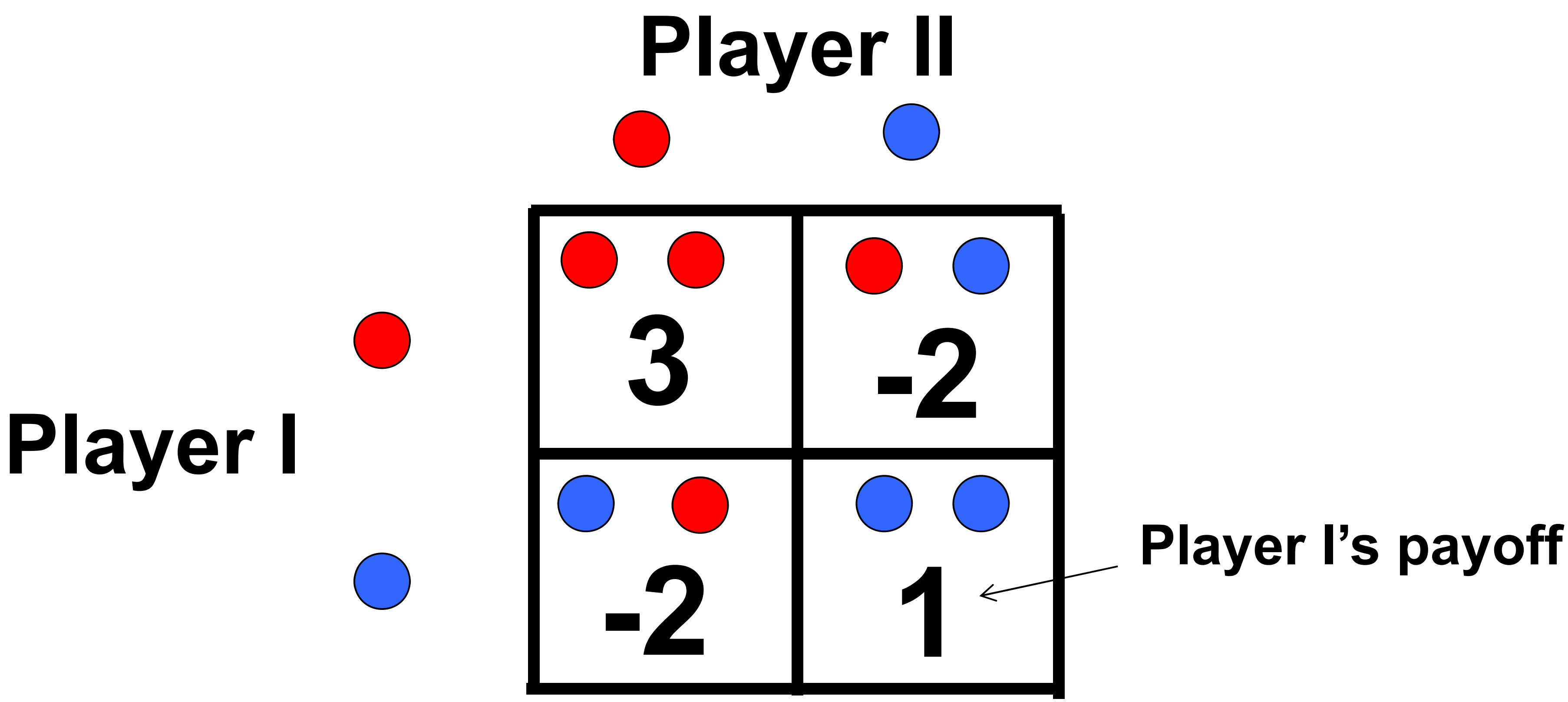
A. Wald. Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Statist.*, 10(4):299–326, 1939.

A. Wald. Statistical decision functions which minimize the maximum risk. *Ann. of Math. (2)*, 46:265–280, 1945.

A. Wald. An essentially complete class of admissible decision functions. *Ann. Math. Statistics*, 18:549–555, 1947.

A. Wald. Statistical decision functions. *Ann. Math. Statistics*, 20:165–205, 1949.

# Deterministic zero sum game

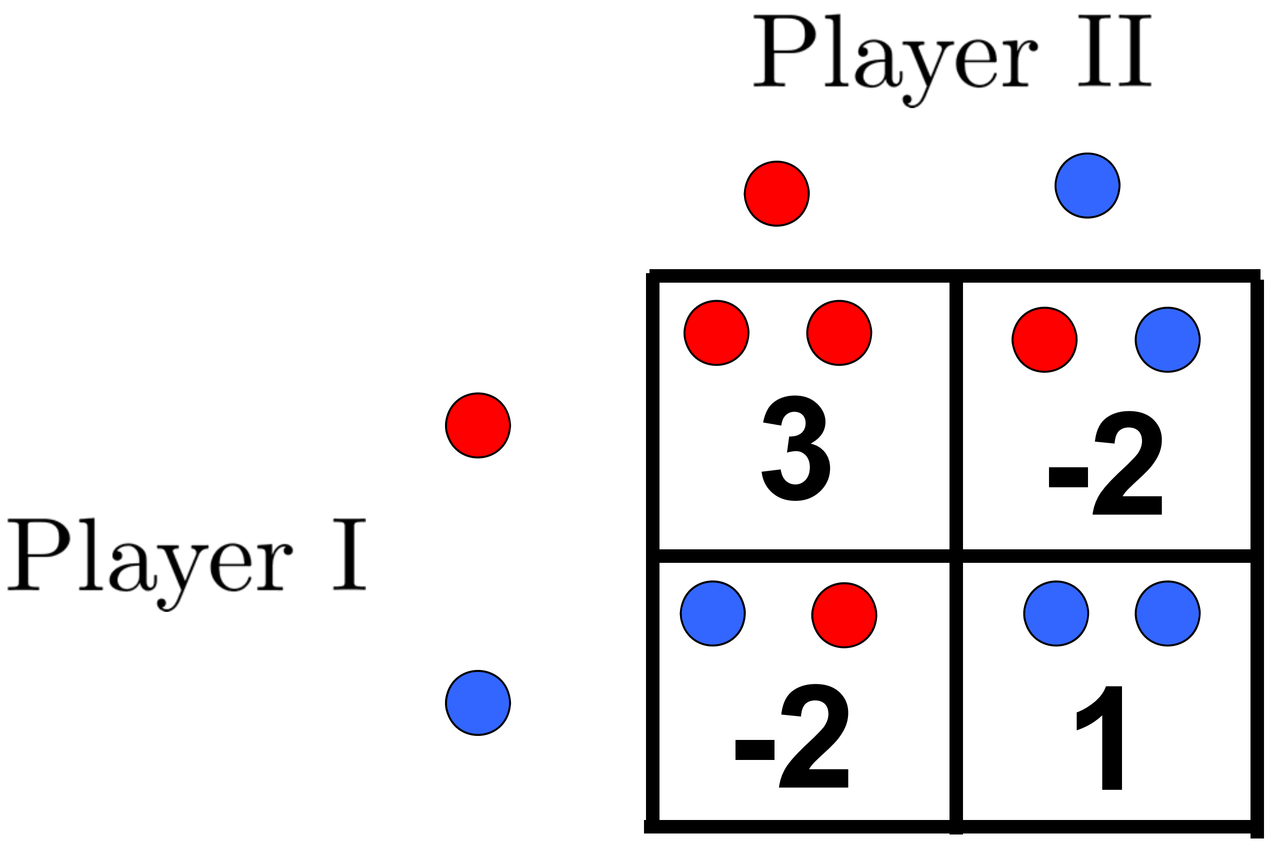



**Player I & II both have a blue and a red marble**

**At the same time, they show each other a marble**

## How should I & II play the game?

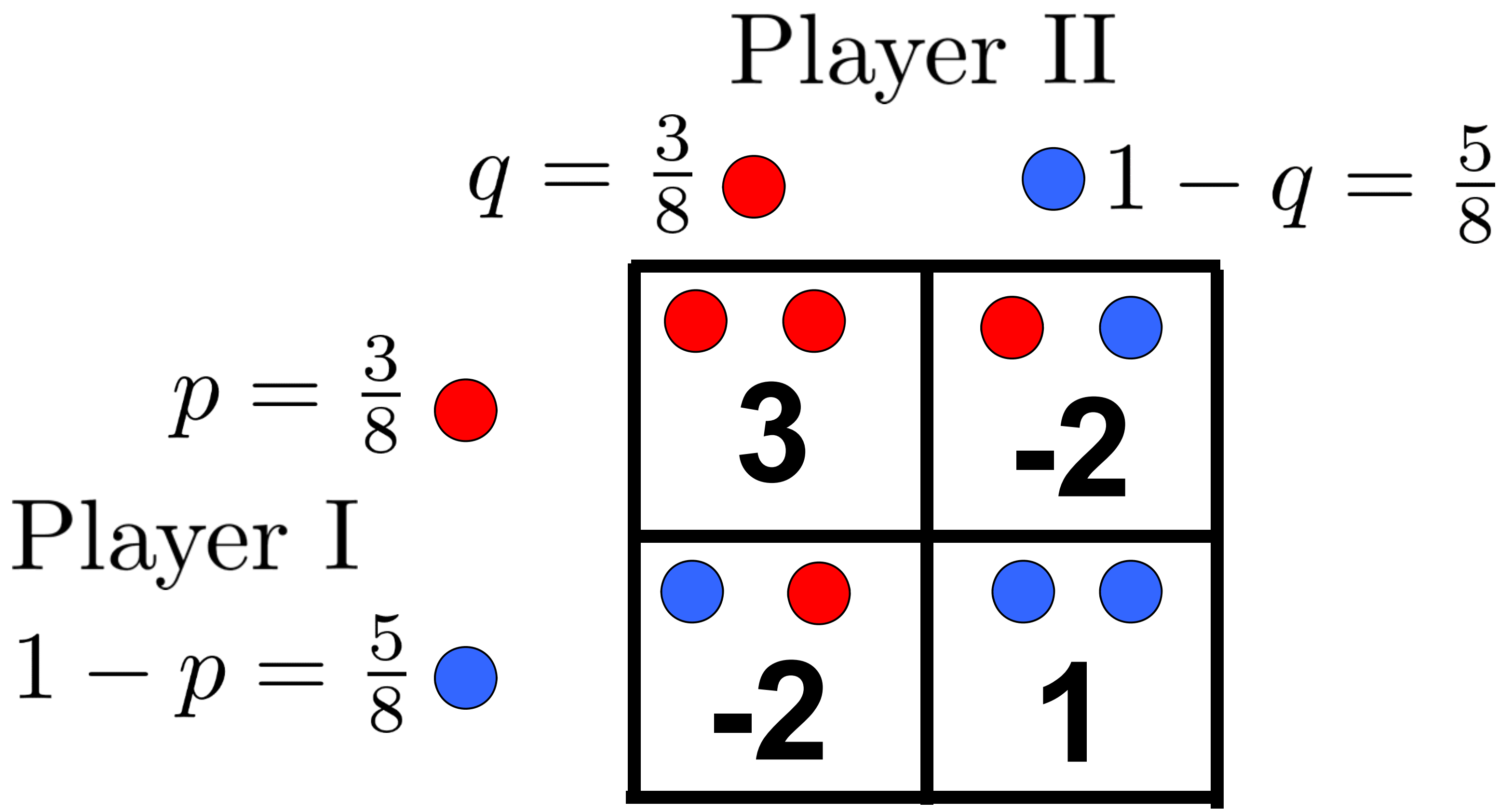
# Pure strategy solution




**II should play blue and loose 1 in the worst case**

**I should play red and loose 2 in the worst case**

# Mixed strategy (repeated game) solution




**II should play red with probability 3/8 and win 1/8 on average**

$$\text{Player } I\text{'s expected payoff} = 3pq + (1-p)(1-q) - 2p(1-q) - 2q(1-p)$$

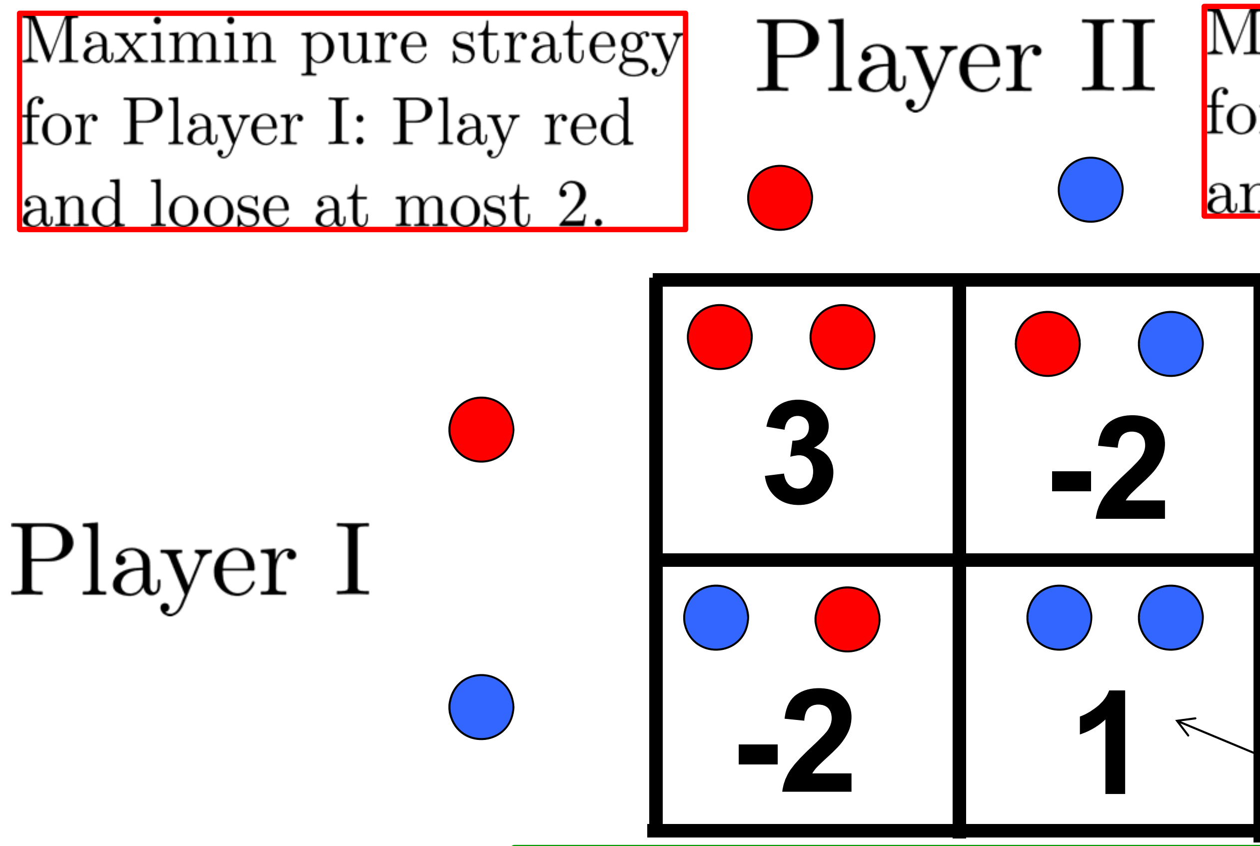
$$= 1 - 3q + p(8q - 3) = -\tfrac{1}{8} \text{ for } q = \tfrac{3}{8}$$

**I should play red with probability 3/8 and loose 1/8 on average**

$$\max\min \neq \min\max$$

Maximin pure strategy for Player I: Play red and loose at most 2.

Minimax pure strategy for Player II: Play blue and loose at most 1.





J. Von Neumann

$$\max\min = \min\max$$

Maximin mixed strategy for Player I: Play red with probability $\frac{3}{8}$ and loose exactly $\frac{1}{8}$ on average.

Minimax mixed strategy for Player II: Play red with probability $\frac{3}{8}$ and win exactly $\frac{1}{8}$ on average.

Player I chooses $\mu \in \mathcal{A}$

Player II chooses $\theta$

Max

Min

$$\mathcal{E}(\mu, \theta)$$

**Pure strategy solution for Player II**

**Optimal bound on the statistical error**

$$\max_{\mu \in \mathcal{A}} \mathcal{E}(\mu, \theta)$$

**Optimal statistical estimators**

$$\min_{\theta} \max_{\mu \in \mathcal{A}} \mathcal{E}(\mu, \theta)$$

Not a saddle point: $\min_{\theta} \max_{\mu \in \mathcal{A}} \mathcal{E}(\mu, \theta) \neq \max_{\mu \in \mathcal{A}} \min_{\theta} \mathcal{E}(\mu, \theta) = 0$

Player I chooses $\mu \in \mathcal{A}$

Player II chooses $\theta$

Max

Min

$$\mathcal{E}(\mu, \theta)$$

**Mixed strategy (repeated game) solution for Player I**

$$\mu \sim \pi_I \in \mathcal{M}(\mathcal{A})$$

**Mixed strategy (repeated game) solution for Player II**

Choose $\theta$ at random and minimize

$$\mathbb{E}_{\mu \sim \pi_I,\, \hat{\theta}}\left[\mathcal{E}(\mu, \hat{\theta})\right]$$

Saddle point: $\min_{\hat{\theta}} \max_{\pi_I \in \mathcal{M}(\mathcal{A})} \mathbb{E}_{\mu \sim \pi_I, \hat{\theta}}\left[\mathcal{E}(\mu, \theta)\right] = \max_{\pi_I \in \mathcal{M}(\mathcal{A})} \min_{\hat{\theta}} \mathbb{E}_{\mu \sim \pi_I, \hat{\theta}}\left[\mathcal{E}(\mu, \theta)\right]$

Bayesian estimator (with prior $\pi \in \mathcal{M}(\mathcal{A})$)

$$\theta_\pi(d) = \mathbb{E}_{\mu \sim \pi,\, d' \sim \mu^n}\left[\Phi(\mu) \middle| d' = d\right]$$

**Theorem** If the loss is quadratic, i.e.

$$\mathcal{E}(\mu, \theta) = \mathbb{E}_{d \sim \mu^n}\left[\left[\theta(d) - \Phi(\mu)\right]^2\right]$$

then for all prior $\pi \in \mathcal{M}(\mathcal{A})$

$$\underbrace{\min_{\theta} \max_{\mu} \mathcal{E}(\mu, \theta)}_{\text{Minimal loss in non cooperative game}} \geq \underbrace{\mathbb{E}_{\mu \sim \pi}\left[\mathcal{E}(\mu, \theta_\pi)\right]}_{\text{Variance of Bayesian estimator}}$$

**Can we have equality?**

**Theorem** If the loss is quadratic, i.e.

$$\mathcal{E}(\mu,\theta) = \mathbb{E}_{d\sim\mu^n}\Big[\big[\theta(d) - \Phi(\mu)\big]^2\Big]$$

then the optimal $\hat{\theta}$ is non-random and lives in the (classical) Bayesian class of estimators

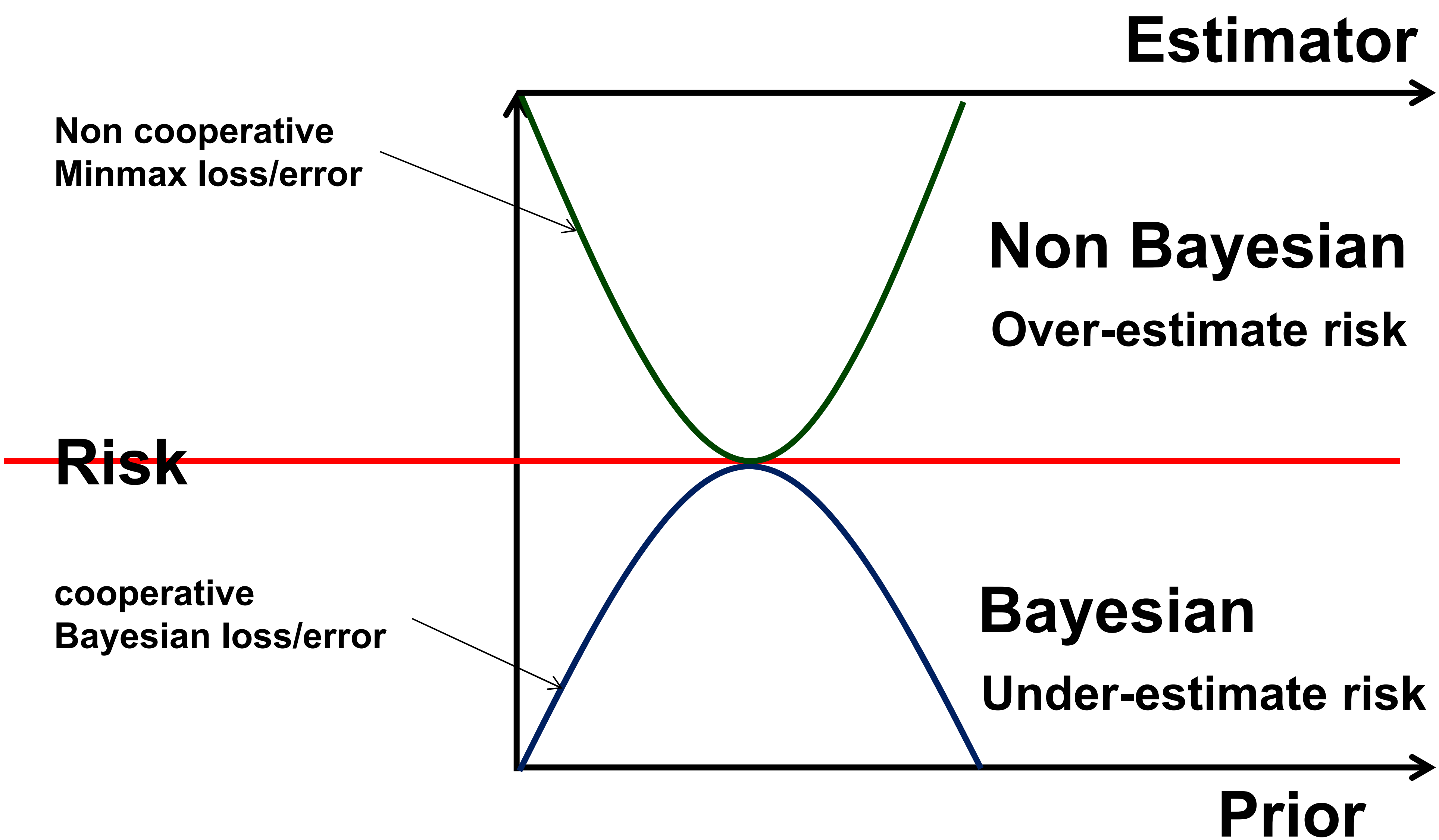
$$\min_{\theta}\max_{\mu}\mathcal{E}(\mu,\theta) = \max_{\pi}\mathbb{E}_{\mu\sim\pi}\Big[\mathcal{E}(\mu,\theta_\pi)\Big]$$

**The best mixed strategy for I and II = worst prior for II**

**The best estimator is not random if the loss function is strictly convex**

A. Dvoretzky, A. Wald, and J. Wolfowitz. Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *Ann. Math. Statist.*, 22(1):1–21, 1951.

# Complete class theorem



## Further generalization of Statistical decision theory


L. J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46:55–67, 1951.

L. Le Cam. An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.*, 26:69–81, 1955

L. D. Brown. Minimaxity, more or less. In *Statistical Decision Theory and Related Topics V*, pages 1–18. Springer, 1994.

L. D. Brown. An essay on statistical decision theory. *Journal of the American Statistical Association*, 95(452):1277–1281, 2000.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989

A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optim. Methods Softw.*, 17(3):523–542, 2002.

M. Sniedovich. The art and science of modeling decision-making under severe uncertainty. *Decis. Mak. Manuf. Serv.*, 1(1-2):111–136, 2007

M. Sniedovich. A classical decision theoretic perspective on worst-case analysis. *Appl. Math.*, 56(5):499–509, 2011.

## Impact in econometrics and social sciences

O. Morgenstern. Abraham Wald, 1902-1950. *Econometrica: Journal of the Econometric Society*, pages 361–367, 1951.

G. Tintner. Abraham Wald's contributions to econometrics. *Ann. Math. Statistics*, 23:21–28, 1952.

R. Leonard. *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*. Cambridge University Press, 2010.

## If we want to make decision theory practical for UQ we need to introduce computational complexity constraints

H. Owhadi and C. Scovel. Towards Machine Wald. *Handbook for Uncertainty Quantication*, 2016. arXiv:1508.02449.

**How do we do that?**

**Is there a natural relation between game theory, computational complexity and numerical approximations?**

# A simple approximation problem

Approximate solution $x$ of

$$Ax = b$$

$A$: Known $n \times n$ symmetric positive definite matrix

$b$: Unknown element of $\mathbb{R}^n$

Based on the information that

$$\Phi x = y$$

$\Phi$: Known $m \times n$ rank $m$ matrix $(m < n)$

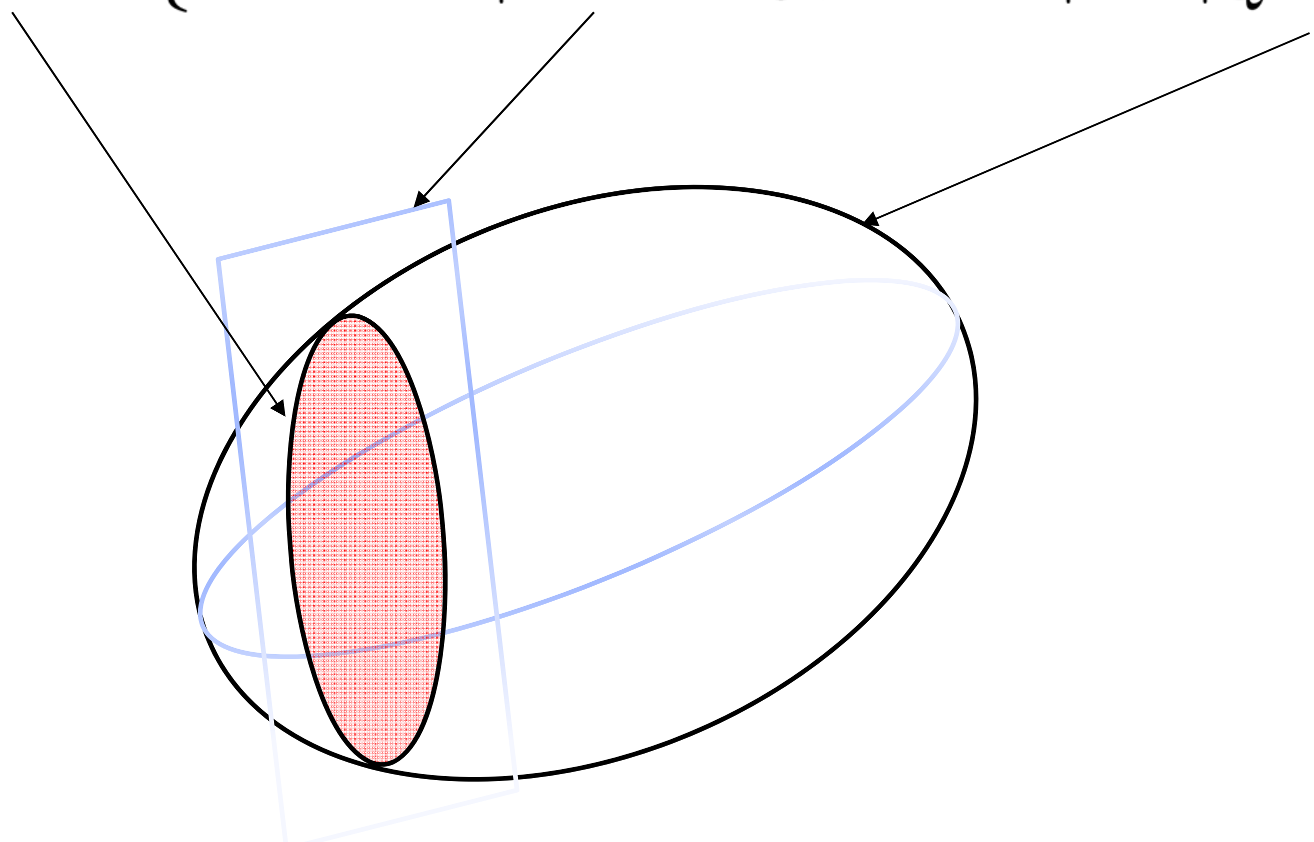
$$b^T Q^{-1} b \leq 1$$

$y$: Known element of $\mathbb{R}^m$

$Q$: Known $n \times n$ symmetric positive definite matrix

# Set of candidates (ambiguity set) for $x$

$$\mathcal{A} = \left\{ z \in \mathbb{R}^n \mid \Phi z = y, \text{ and } |Az|_{Q^{-1}} \leq 1 \right\}$$



$$|b|_{Q^{-1}} := \sqrt{b^T Q^{-1} b}$$

Classical numerical analysis minimax solution

$z^*$ minimizing

$$\min_{z^* \in \mathcal{A}} \max_{z \in \mathcal{A}} \|z - z^*\|$$

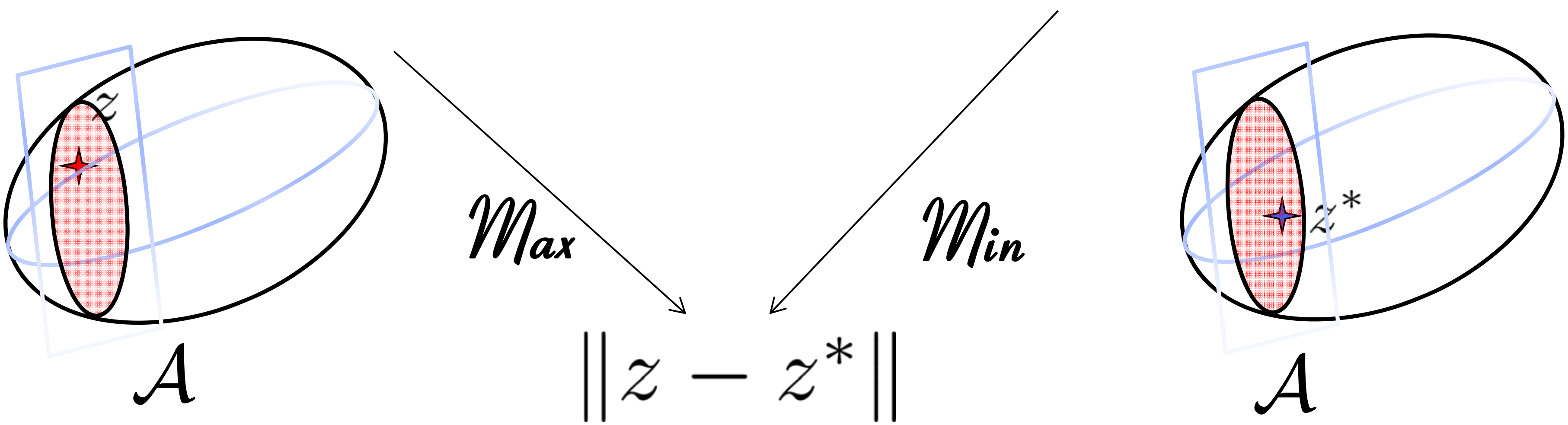
Looks like a game

Player I:

Chooses $z \in \mathcal{A}$

Player II:

Chooses $z^* \in \mathcal{A}$



Classical numerical analysis minimax solution

$z^*$ minimizing

$$\min_{z^* \in \mathcal{A}} \max_{z \in \mathcal{A}} \|z - z^*\|$$

---

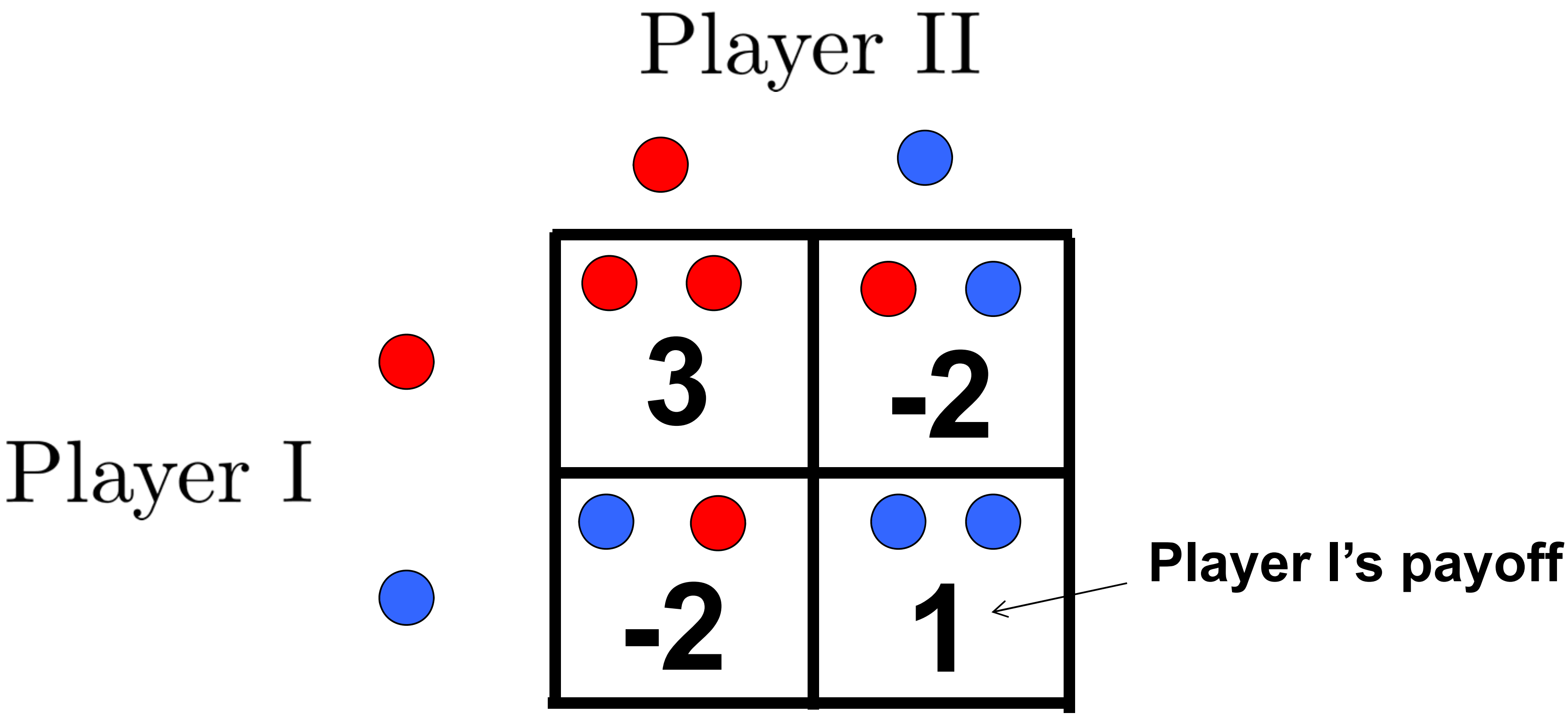
No saddle point in the numerical analysis formulation!

$$\max_{z^* \in \mathcal{A}} \min_{z \in \mathcal{A}} \|z - z^*\| = 0$$

$$\min_{z^* \in \mathcal{A}} \max_{z \in \mathcal{A}} \|z - z^*\| \neq \max_{z^* \in \mathcal{A}} \min_{z \in \mathcal{A}} \|z - z^*\|$$

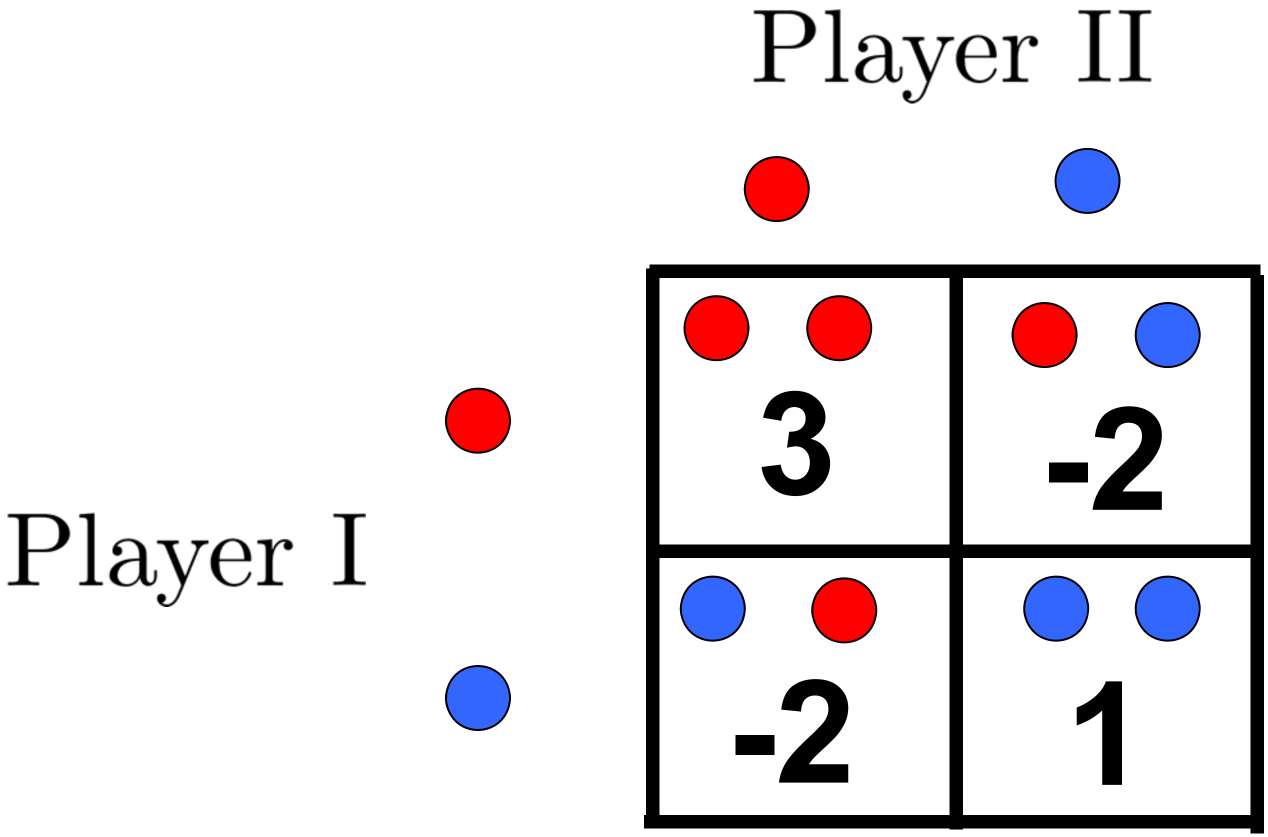
Why should we care?

# Deterministic zero sum game




**How should I and II play the game?**

## Pure strategy (classical numerical analysis) solution




**II should play blue and loose 1 in the worst case**

**I should play red and loose 2 in the worst case**

## Mixed strategy (repeated game) solution

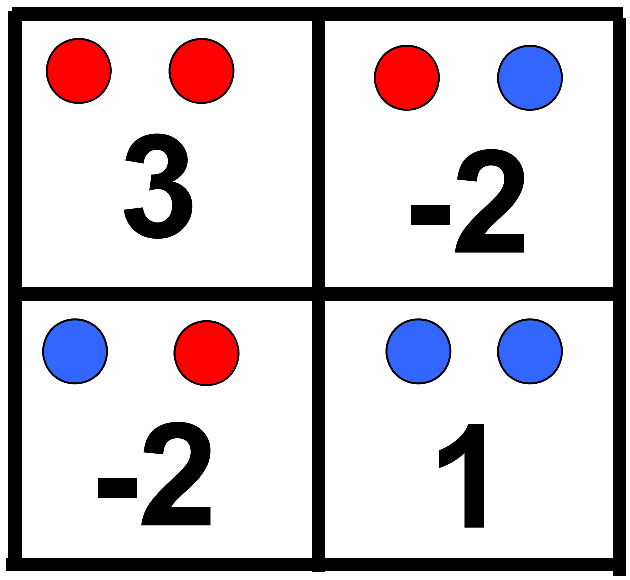
Player II

$q = \frac{3}{8}$ $\quad$ $1 - q = \frac{5}{8}$

Player I

$p = \frac{3}{8}$

$1 - p = \frac{5}{8}$




**II should play red with probability 3/8 and win 1/8 on average**

$$\text{Player } I\text{'s expected payoff} = 3pq + (1-p)(1-q) - 2p(1-q) - 2q(1-p)$$

$$= 1 - 3q + p(8q - 3) = -\frac{1}{8} \text{ for } q = \frac{3}{8}$$

**I should play red with probability 3/8 and loose 1/8 on average**

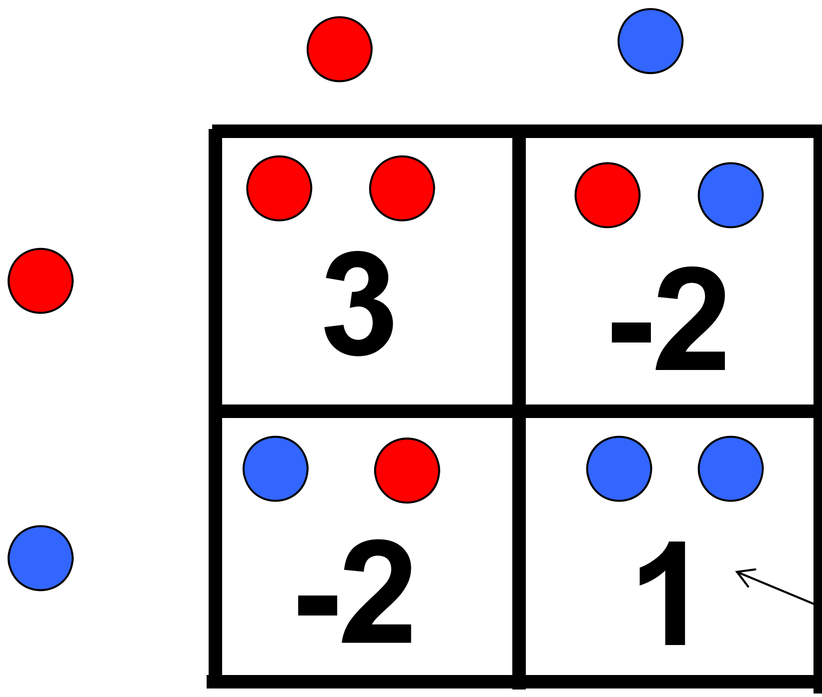
Maximin pure strategy for Player I: Play red and loose at most 2.

$\min\max \neq \max\min$

Minimax pure strategy for Player II: Play blue and loose at most 1.

Player II

Player I




Player I's payoff

J. Von Neumann

Maximin mixed strategy for Player I: Play red with probability $\frac{3}{8}$ and loose exactly $\frac{1}{8}$ on average.

Minimax mixed strategy for Player II: Play red with probability $\frac{3}{8}$ and win exactly $\frac{1}{8}$ on average.

$\min\max = \max\min$

# Game theoretic formulation

$$Ax = b$$

Player I
chooses
$b \in \mathbb{R}^n$
$b^T Q^{-1} b \leq 1$

Player II
sees $y = \Phi x$
chooses $x^*$

*Max* → $\|x - x^*\|$ ← *Min*

**Abraham Wald**

**Continuous game but as in decision theory under compactness it can be approximated by a finite game**

## Best strategy: lift minimax to measures

$$Ax = b$$

Player I chooses $b \in \mathbb{R}^n$, $b^T Q^{-1} b \leq 1$

Player II sees $y = \Phi x$, chooses $x^*$

Max → $\|x - x^*\|$ ← Min

**The best strategy for I is to play at random**

**Player II's best strategy live in the Bayesian class of estimators**

## Player II's mixed strategy

$$Ax = b \quad \Leftrightarrow \quad AX = \xi$$

$$\xi \sim \mathcal{N}(0, Q)$$

## Player II's bet

$$x^* = \mathbb{E}\big[X \big| \Phi X = \Phi x\big]$$

Player II's recovery error on $x_i$

$$\big| x_i - \mathbb{E}[X_i | \Phi X = \Phi x] \big| \quad \text{unknown}$$

Player II's stochastic error assuming that Player I is selecting $x$ at random with the same prior distribution

$$\big| X_i - \mathbb{E}[X_i | \Phi X] \big|$$

random variable with known distribution

# Player II's mixed strategy

$$Ax = b \quad \Longleftrightarrow \quad AX = \xi$$

$$\xi \sim \mathcal{N}(0, Q)$$

# Theorem

$$\left|x_i - \mathbb{E}[X_i|\Phi X = y]\right| \leq \sqrt{\mathbb{E}\left[\left|X_i - \mathbb{E}[X_i|\Phi X]\right|^2\right]\, b^T Q^{-1} b}$$

unknown deterministic error

known Standard Deviation of stochastic error

known compactness bound on $b$

Owhadi 2015, Multi-grid with rough coefficients and Multiresolution PDE decomposition from Hierarchical Information Games, arXiv:1503.03467, SIAM Review (to appear)

## Main Question

**Can we turn the process of discovery of a scalable numerical method into a UQ problem and, to some degree, solve it as such in an automated fashion?**

**Can we use a computer, not only to implement a numerical method but also to find the method itself?**

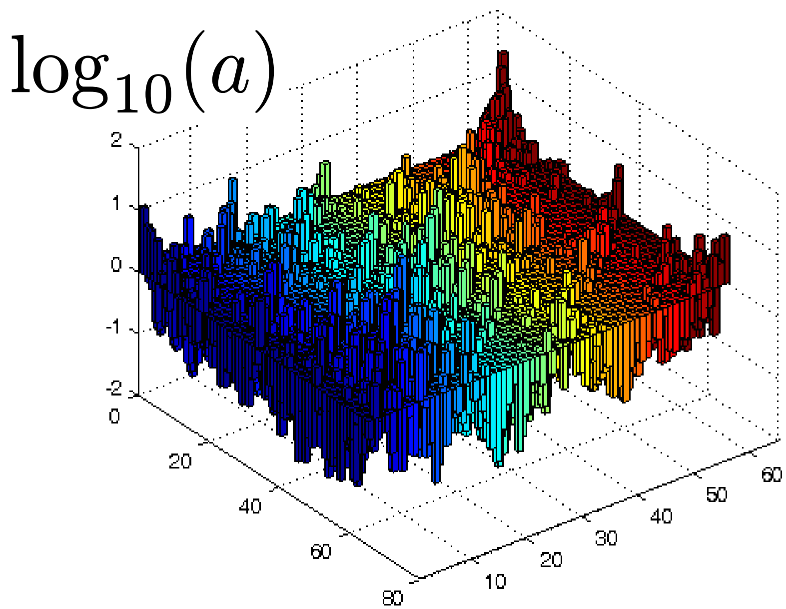
**Example: Find a method for solving (1) as fast as possible to a given accuracy**

(1)
$$\begin{cases} -\operatorname{div}(a\nabla u) = g, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$

$\Omega \subset \mathbb{R}^d$ $\qquad$ $\partial\Omega$ is piec. Lip.

$a$ unif. ell.

$a_{i,j} \in L^\infty(\Omega)$



## Multigrid Methods

Multigrid: [Fedorenko, 1961, Brandt, 1973, Hackbusch, 1978]

## Multiresolution/Wavelet based methods

[Brewster and Beylkin, 1995, Beylkin and Coult, 1998, Averbuch et al., 1998]

- **Linear complexity with smooth coefficients**

**Problem** **Severely affected by lack of smoothness**

## Robust/Algebraic multigrid

[Mandel et al., 1999,Wan-Chan-Smith, 1999,
Xu and Zikatanov, 2004, Xu and Zhu, 2008], [Ruge-Stüben, 1987]
[Panayot - 2010]

### Stabilized Hierarchical bases, Multilevel preconditioners

[Vassilevski - Wang, 1997, 1998]

[Panayot - Vassilevski, 1997]

[Chow - Vassilevski, 2003]

[Aksoylu- Holst, 2010]

- Some degree of robustness but problem remains open with rough coefficients

**Why?** **Interpolation operators are unknown**

Don't know how to bridge scales with rough coefficients!

## Low Rank Matrix Decomposition methods

Fast Multipole Method: [Greengard and Rokhlin, 1987]

Hierarchical Matrix Method: [Hackbusch et al., 2002]

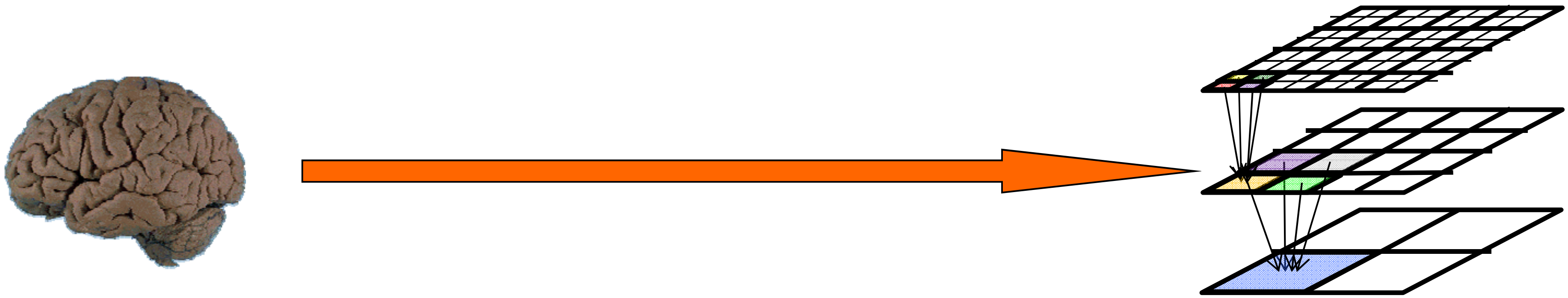
[Bebendorf, 2008]:

$$N \ln^{2d+8} N \text{ complexity}$$

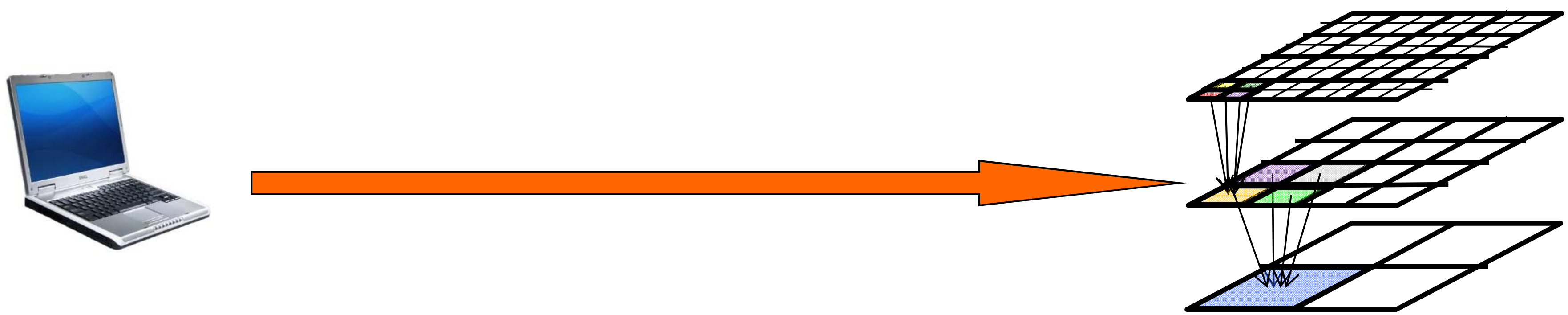
To achieve grid-size accuracy in $L^2$-norm

# Common theme between these methods

**Their process of discovery is based on intuition, brilliant insight, and guesswork**



**Can we turn this process of discovery into an algorithm?**



**Answer:** YES

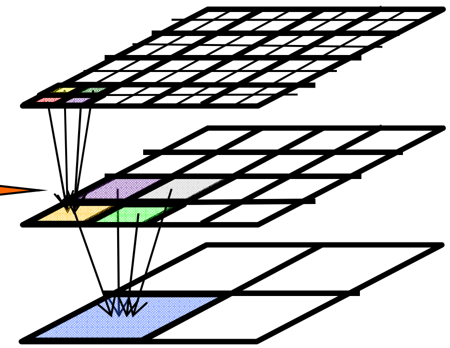
Compute fast

Compute with partial information

Play adversarial Information game

Identify game

Find optimal strategy



Owhadi 2015, Multi-grid with rough coefficients and Multiresolution PDE decomposition from Hierarchical Information Games, arXiv:1503.03467, SIAM Review (to appear)

**Resulting method:**

This is a theorem

$N \ln^{3d} N$ complexity

To achieve grid-size accuracy in $H^1$-norm

Subsequent solves: $N \ln^{d+1} N$ complexity

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases}$$

**Resulting method:**

$$H_0^1(\Omega) = \mathfrak{W}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \cdots \oplus_a \mathfrak{W}^{(k)} \oplus_a \cdots$$

$$<\psi, \chi>_a := \int_\Omega (\nabla\psi)^T a \nabla\chi = 0 \text{ for } (\psi, \chi) \in \mathfrak{W}^{(i)} \times \mathfrak{W}^{(j)},\ i \neq j$$

**Theorem** For $v \in \mathfrak{W}^{(k)}$

$$\frac{C_1}{2^k} \leq \frac{\|v\|_a}{\|\operatorname{div}(a\nabla v)\|_{L^2(\Omega)}} \leq \frac{C_2}{2^k}$$

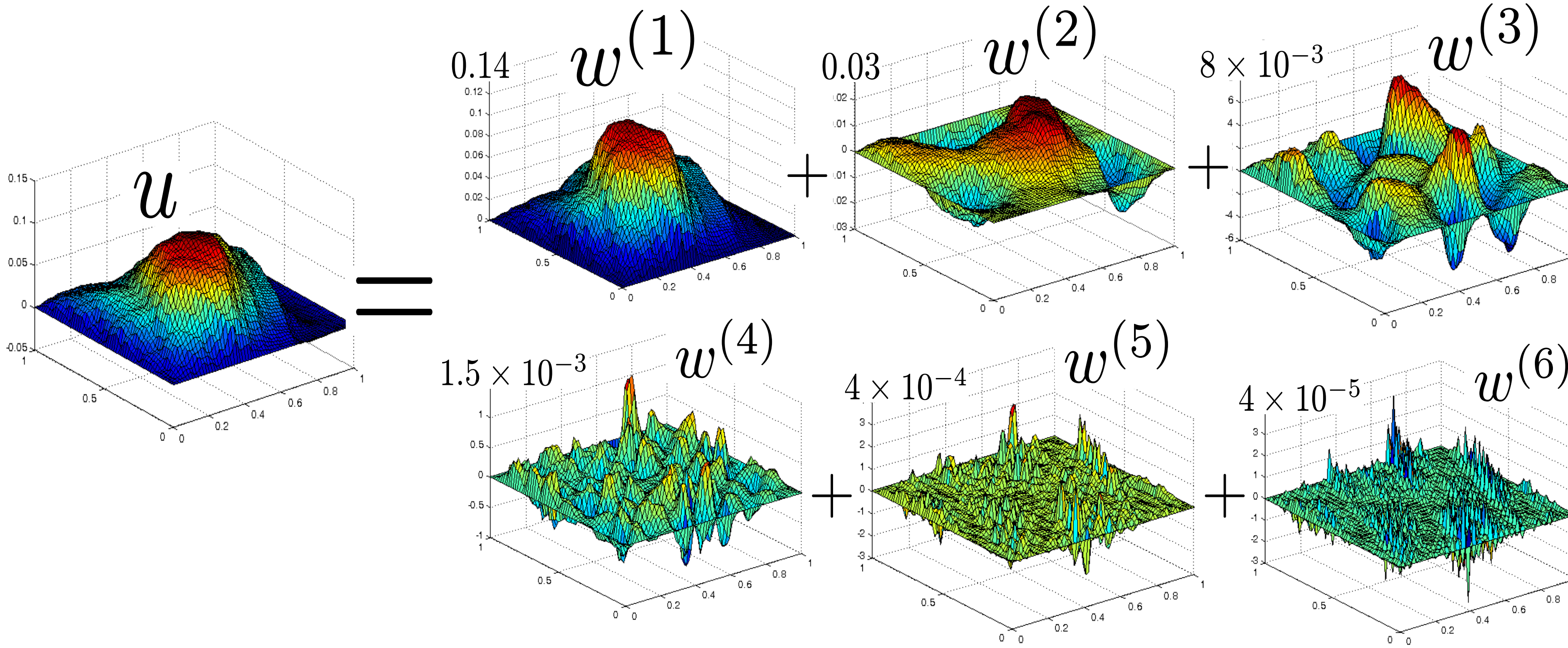
$$\|v\|_a^2 :=< v, v >_a = \int_\Omega (\nabla v)^T a \nabla v$$

**Looks like an eigenspace decomposition**

$$u = w^{(1)} + w^{(2)} + \cdots + w^{(k)} + \cdots$$

$$w^{(k)} = \text{F.E. sol. of PDE in } \mathfrak{W}^{(k)}$$

Can be computed independently




Multiresolution decomposition of solution space

$$u = w^{(1)} + w^{(2)} + \cdots + w^{(k)} + \cdots$$

$w^{(k)}$ = F.E. sol. of PDE in $\mathfrak{W}^{(k)}$

Can be computed independently

$B^{(k)}$: Stiffness matrix of PDE in $\mathfrak{W}^{(k)}$

**Theorem**

$$\frac{\lambda_{\max}(B^{(k)})}{\lambda_{\min}(B^{(k)})} \leq C$$

⬇

Just relax in $\mathfrak{W}^{(k)}$ to find $w^{(k)}$

**Quacks like an eigenspace decomposition**

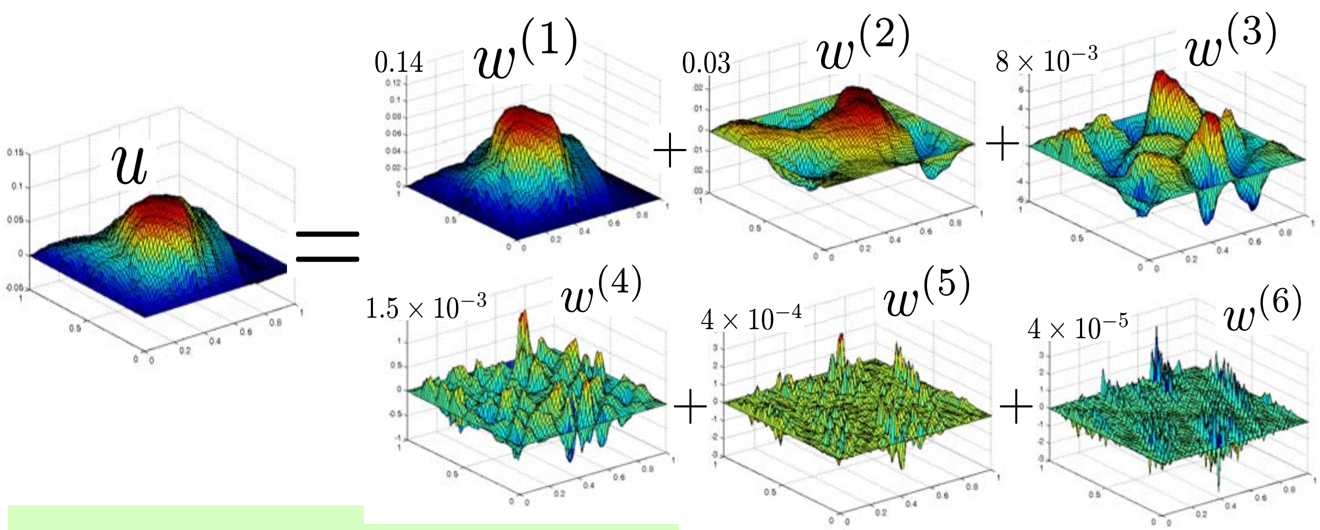
# Application to time dependent problems

[Owhadi-Zhang 2016, From gamblets to near FFT-complexity solvers for wave and parabolic PDEs with rough coefficients]

$$\mu(x)\partial_t^2 u - \operatorname{div}(a\nabla u) = g(x,t)$$

$$\mu(x)\partial_t u - \operatorname{div}(a\nabla u) = g(x,t)$$

Hyperbolic and parabolic PDEs with rough coefficients can be solved in $\mathcal{O}(N \ln^{3d} N)$ (near FFT) complexity




**Swims like an eigenspace decomposition**

$\mathfrak{V}$: F.E. space of $H_0^1(\Omega)$ of dim. $N$

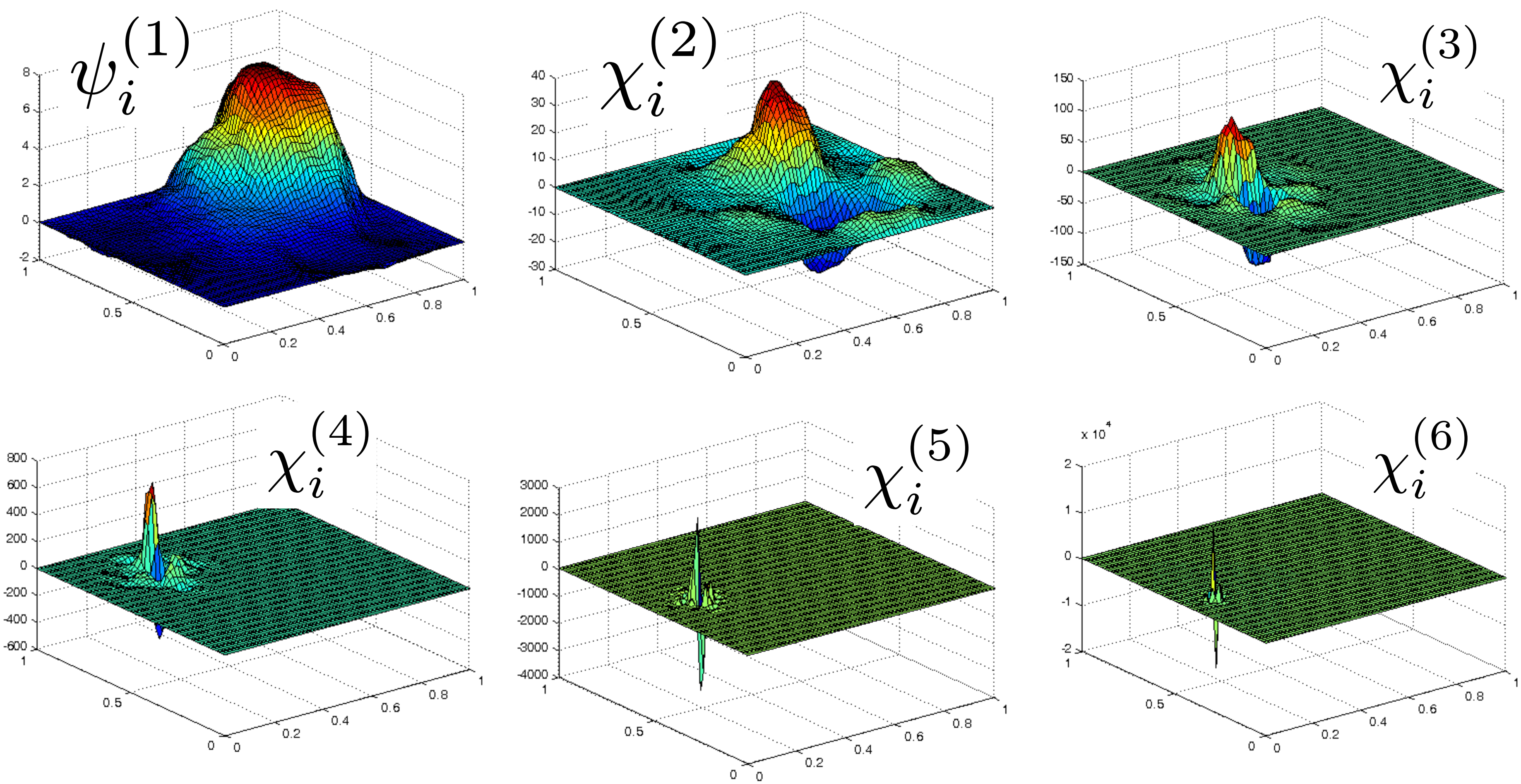
**Theorem** The decomposition

$$\mathfrak{V} = \mathfrak{W}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \cdots \oplus_a \mathfrak{W}^{(k)}$$

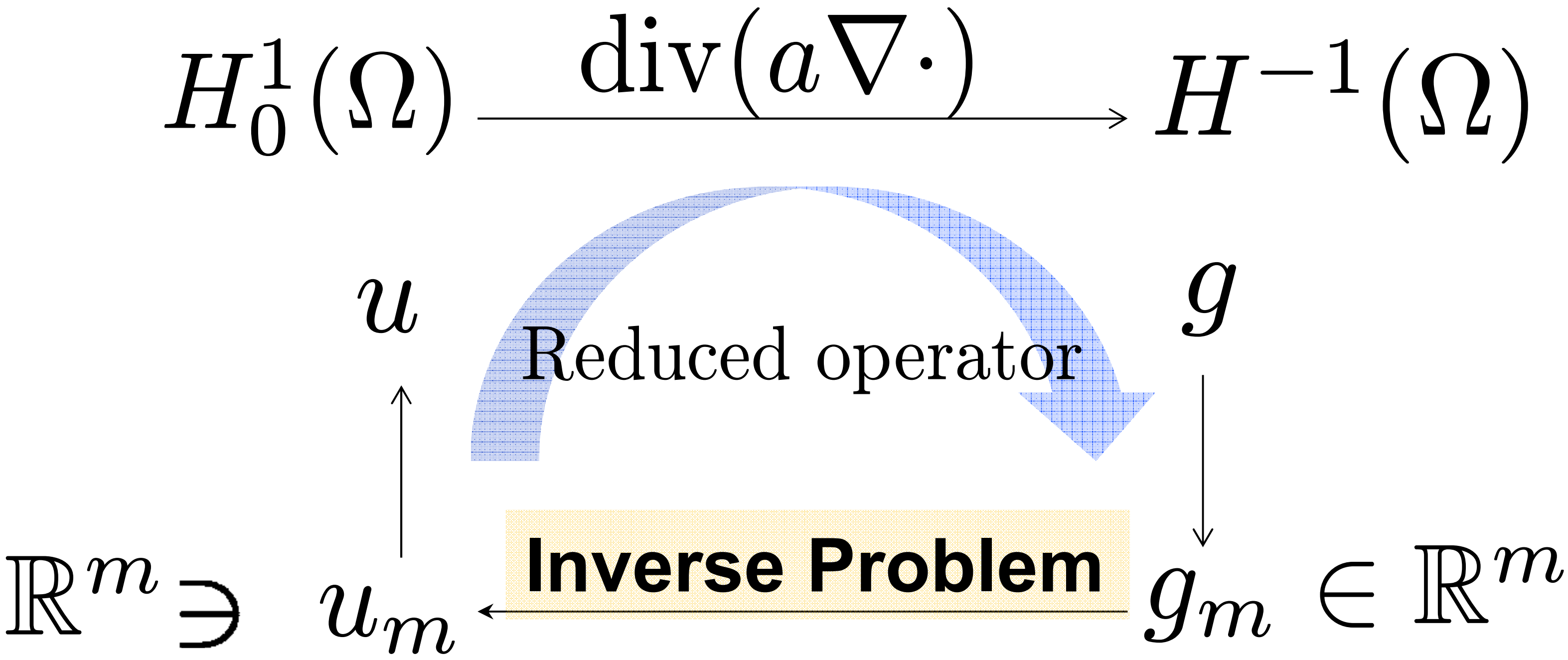
Can be performed and stored in

$$\mathcal{O}(N \ln^{3d} N) \text{ operations}$$

**Doesn't have the complexity of an eigenspace decomposition**




**Basis functions look like and behave like wavelets: Localized and can be used to compress the operator and locally analyze the solution space**




Numerical implementation requires computation with partial information.

$$\phi_1, \ldots, \phi_m \in L^2(\Omega)$$

$$u_m = \left(\int_\Omega \phi_1 u, \ldots, \int_\Omega \phi_m u\right)$$

$$u_m \in \mathbb{R}^m \xrightarrow{\text{Missing information}} u \in H_0^1(\Omega)$$

**Discovery process**

**Identify underlying information game**

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases}$$

Measurement functions: $\phi_1, \ldots, \phi_m \in L^2(\Omega)$

## Player I

Chooses

$g \in L^2(\Omega)$

$\|g\|_{L^2(\Omega)} \leq 1$

## Player II

Sees $\int_\Omega u\phi_1, \ldots, \int_\Omega u\phi_m$

Chooses $u^* \in L^2(\Omega)$

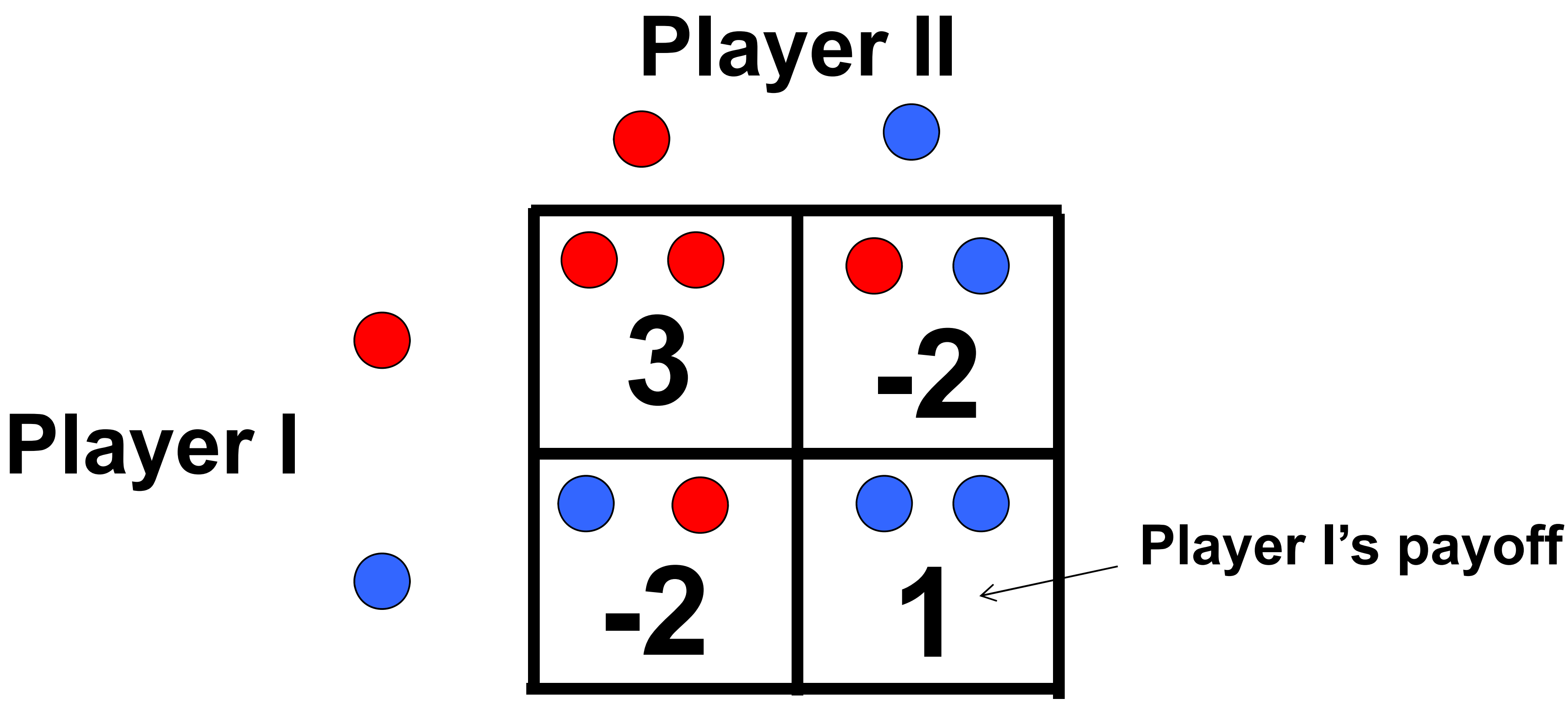
Max

Min

$$\|u - u^*\|_a$$

$\|f\|_a^2 := \int_\Omega (\nabla f)^T a \nabla f$

# Deterministic zero sum game




**Player I & II both have a blue and a red marble**

**At the same time, they show each other a marble**

## How should I & II play the (repeated) game?

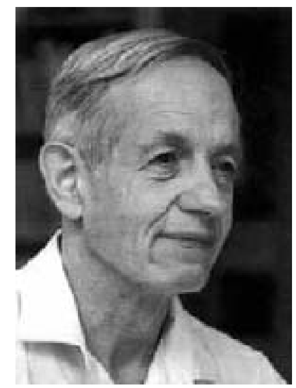
## Game theory

**Optimal strategies are mixed strategies**

**Optimal way to play is at random**

| | | **Player II** $q$ (red) | $1-q$ (blue) |
|---|---|---|---|
| **Player I** | $p$ (red) | 3 | -2 |
| | $1-p$ (blue) | -2 | 1 |

**John Von Neumann**



**John Nash**

Player $I$'s expected payoff

$$= 3pq + (1-p)(1-q) - 2p(1-q) - 2q(1-p)$$

$$= 1 - 3q + p(8q-3) \; = -\frac{1}{8} \quad \text{for } q = \frac{3}{8}$$

**Player A**

Chooses
$g \in L^2(\Omega)$
$\|g\|_{L^2(\Omega)} \leq 1$

**Player B**

Sees $\int_\Omega u\phi_1, \ldots, \int_\Omega u\phi_m$
Chooses $u^* \in L^2(\Omega)$

$$\left\|u - u^*\right\|_a$$

**Continuous game but as in decision theory under compactness it can be approximated by a finite game**

**Abraham Wald**

**The best strategy for A is to play at random**

**Player B's best strategy live in the Bayesian class of estimators**

# Player II's class of mixed strategies

**Pretend that player I is choosing g at random**

$$g \in L^2(\Omega) \quad \Longleftrightarrow \quad \xi\text{: Random field}$$

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases} \quad \Longleftrightarrow \quad \begin{cases} -\operatorname{div}(a\nabla v) = \xi \text{ in } \Omega, \\ v = 0 \text{ on } \partial\Omega, \end{cases}$$

# Player II's bet

$$u^*(x) := \mathbb{E}\Big[v(x) \Big| \int_\Omega v(y)\phi_i(y)\,dy = \int_\Omega u(y)\phi_i(y)\,dy, \forall i\Big]$$

# Player II's optimal strategy?

Player II's best bet? $\Longleftrightarrow$ min max problem over distribution of $\xi$

**Computational efficiency** → $\xi \sim \mathcal{N}(0,\Gamma)$

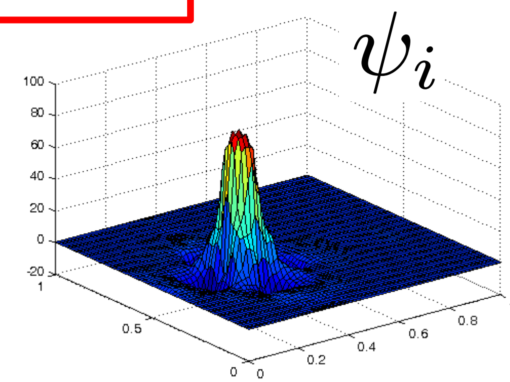
↓

**Elementary gambles form deterministic basis functions for player B's bet**

↓

**Theorem**

$$u^*(x) = \sum_{i=1}^{m} \psi_i(x) \int_\Omega u(y)\phi_i(y)\,dy$$




**Gamblets**

$\psi_i$: Elementary gambles/bets

Player II's bet if $\int_\Omega u\phi_j = \delta_{i,j}$, $j = 1,\ldots,m$

$$\psi_i(x) := \mathbb{E}_{\xi\sim\mathcal{N}(0,\Gamma)}\Big[v(x)\Big|\int_\Omega v(y)\phi_j(y)\,dy = \delta_{i,j},\ j\in\{1,\ldots,m\}\Big]$$

**What are these gamblets?**

# Depend on

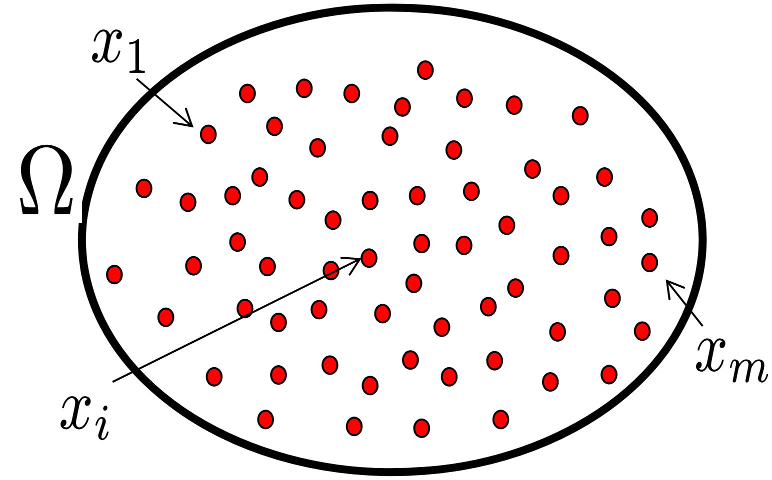
- $\Gamma$: Covariance function of $\xi$ (Player B's decision)
- $(\phi_i)_{i=1}^m$: Measurements functions (rules of the game)

**Example**

[Owhadi, SIAM MMS, 2015]
Bayesian Numerical Homogenization

$$\Gamma(x,y) = \delta(x-y)$$

$$\phi_i(x) = \delta(x-x_i)$$




$a = I_d$ ⟺ $\psi_i$: Polyharmonic splines

[Harder-Desmarais, 1972] [Duchon 1976, 1977,1978]

$a_{i,j} \in L^\infty(\Omega)$ ⟺ $\psi_i$: Rough Polyharmonic splines

[Owhadi-Zhang-Berlyand 2013]

**What is Player II's best strategy?**

**What is Player II's best choice for**

$$\Gamma(x,y)=\mathbb{E}\big[\xi(x)\xi(y)\big] \quad \text{?}$$

$$\boxed{\Gamma=\mathcal{L}}$$

$\Longleftrightarrow$

$$\int_\Omega \xi(x)f(x)\,dx \sim \mathcal{N}(0,\|f\|_a^2)$$

$$\|f\|_a^2 := \int_\Omega (\nabla f)^T a \nabla f$$

$\Updownarrow$

$$\boxed{\mathcal{L}=-\operatorname{div}(a\nabla\cdot)}$$

**Why?** $\Longleftrightarrow$ **See algebraic generalization**

## The recovery is optimal (Galerkin projection)

**Theorem** If $\Gamma = \mathcal{L}$ then

$u^*(x)$ is the F.E. solution of (1) in $\text{span}\{\mathcal{L}^{-1}\phi_i | i = 1, \ldots, m\}$

$$\|u - u^*\|_a = \inf_{\psi \in \text{span}\{\mathcal{L}^{-1}\phi_i : i \in \{1,\ldots,m\}\}} \|u - \psi\|_a$$

$$\mathcal{L} = -\operatorname{div}(a\nabla \cdot)$$

**(1)**
$$\begin{cases} -\operatorname{div}(a\nabla u) = g, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$

## Optimal variational properties

**Theorem**

$\sum_{i=1}^m w_i \psi_i$ minimizes $\|\psi\|_a$
over all $\psi$ such that $\int_\Omega \phi_j \psi = w_j$ for $j = 1, \ldots, m$

## Variational characterization

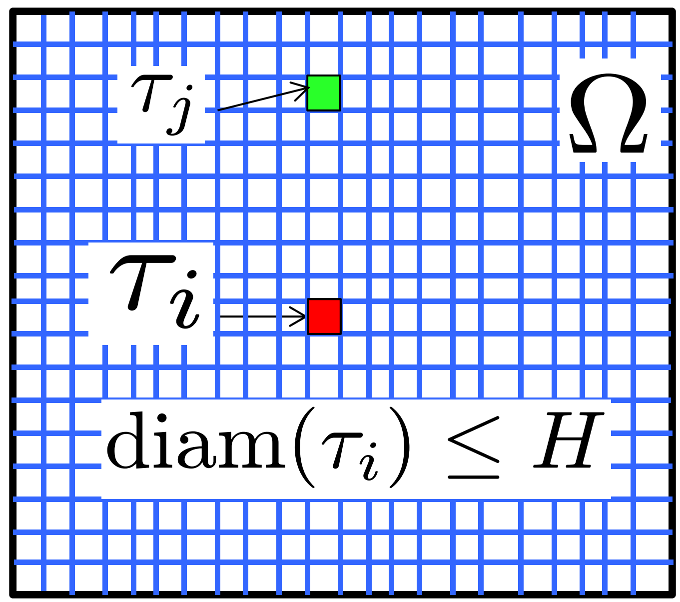
**Theorem** $\psi_i$: Unique minimizer of

$$\begin{cases} \text{Minimize} & \|\psi\|_a \\ \text{Subject to} & \psi \in H_0^1(\Omega) \text{ and } \int_\Omega \phi_j \psi = \delta_{i,j}, \quad j = 1, \ldots, m \end{cases}$$

## Selection of measurement functions

**Example** Indicator functions of a Partition of $\Omega$ of resolution $H$

$$\phi_i = 1_{\tau_i}$$

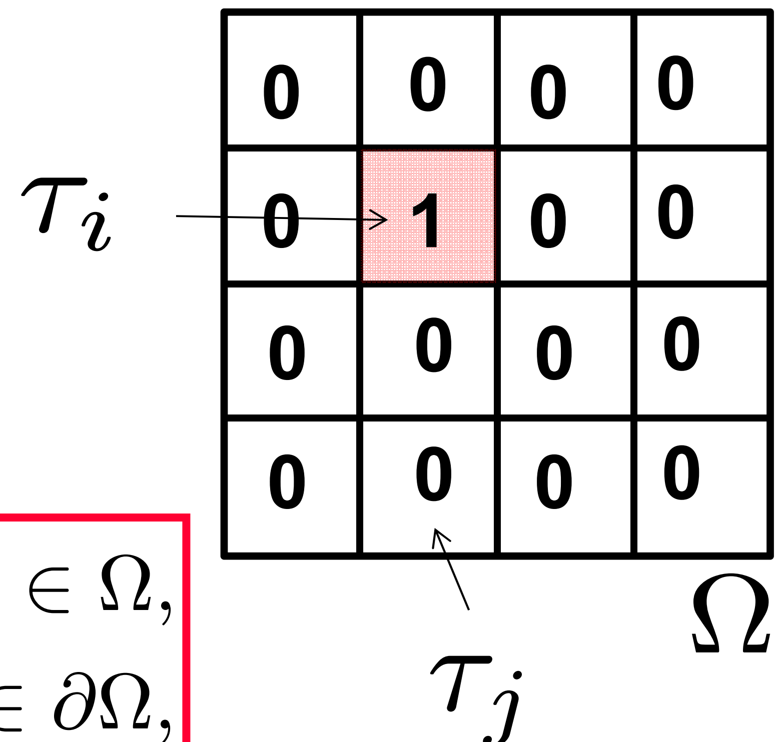



**Theorem**

$$\|u - u^*\|_a \leq \frac{H}{\lambda_{\min}(a)} \|g\|_{L^2(\Omega)}$$

## Elementary gamble

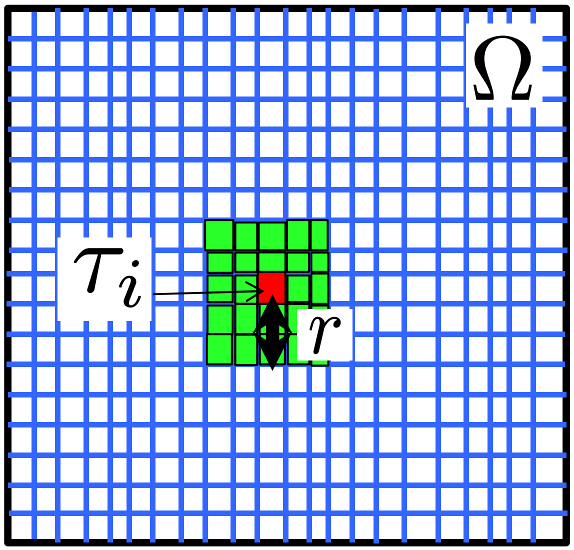
$\psi_i$ Your best bet on the value of $u$ given the information that $\int_{\tau_i} u = 1$ and $\int_{\tau_j} u = 0$ for $j \neq i$




**(1)**

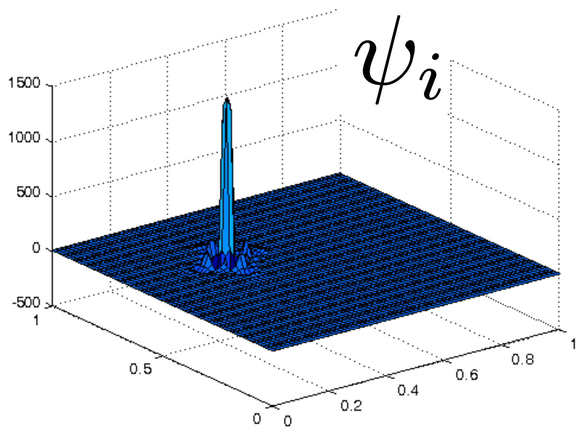
$$\begin{cases} -\operatorname{div}(a\nabla u) = g, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$

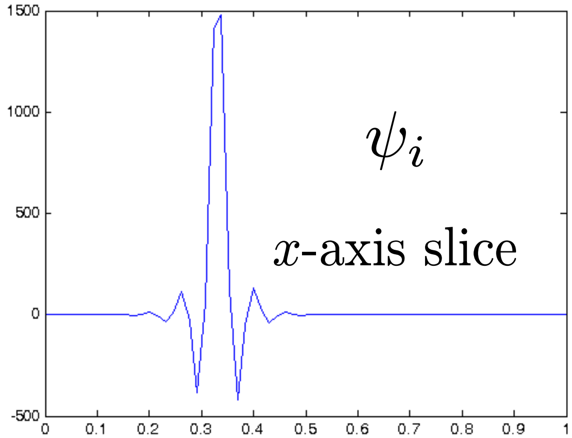
# Exponential decay of gamblets

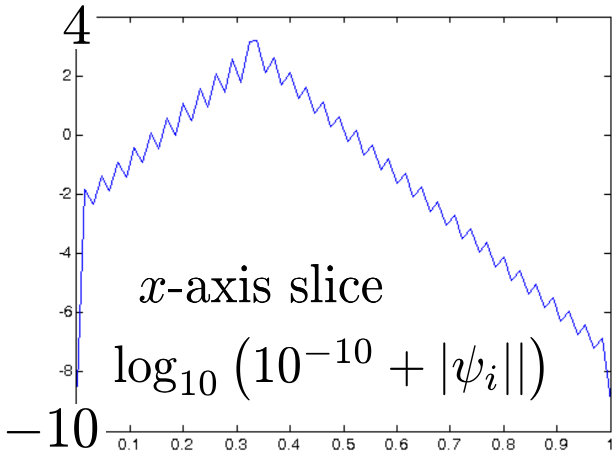



## Theorem

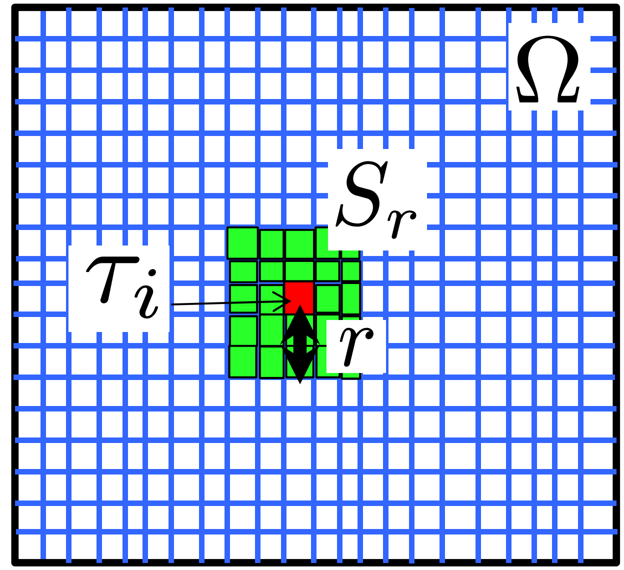
$$\int_{\Omega\cap(B(\tau_i,r))^c}(\nabla\psi_i)^T a\nabla\psi_i \leq e^{-\frac{r}{lH}}\|\psi_i\|_a^2$$



## Localization of the computation of gamblets




$\psi_i^{\text{loc},r}$: Minimizer of

$$\begin{cases} \text{Minimize} & \|\psi\|_a \\ \text{Subject to} & \psi \in H^1_0(S_r) \text{ and } \int_{S_r} \phi_j \psi = \delta_{i,j} \\ & \quad \text{for } \tau_j \in S_r \end{cases}$$

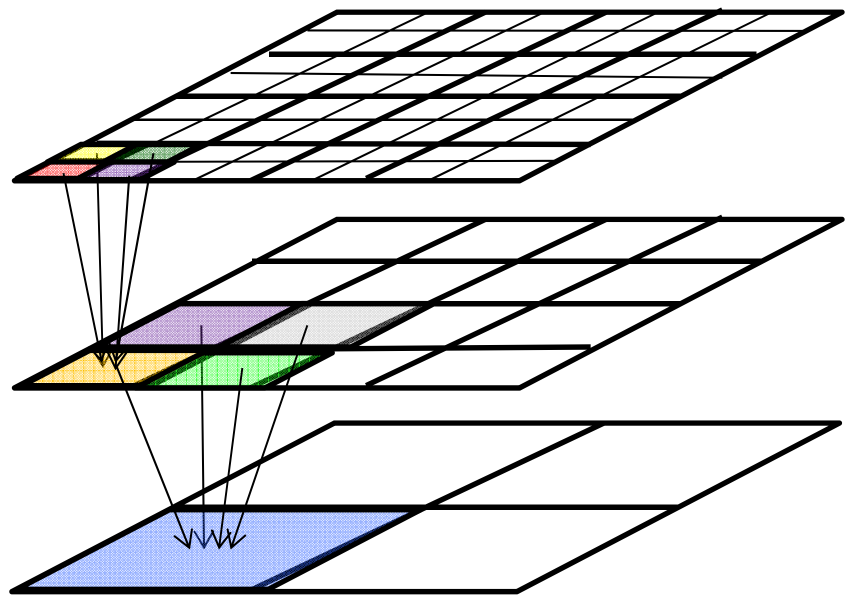
No loss of accuracy if localization $\sim H \ln \frac{1}{H}$

$$u^{*,\text{loc}}(x) = \sum_{i=1}^{m} \psi_i^{\text{loc,r}}(x) \int_\Omega u(y)\phi_i(y)\, dy$$

**Theorem** If $r \geq CH \ln \frac{1}{H}$

$$\|u - u^{*,\text{loc}}\|_a \leq \frac{1}{\sqrt{\lambda_{\min}(a)}} H \|g\|_{L^2(\Omega)}$$

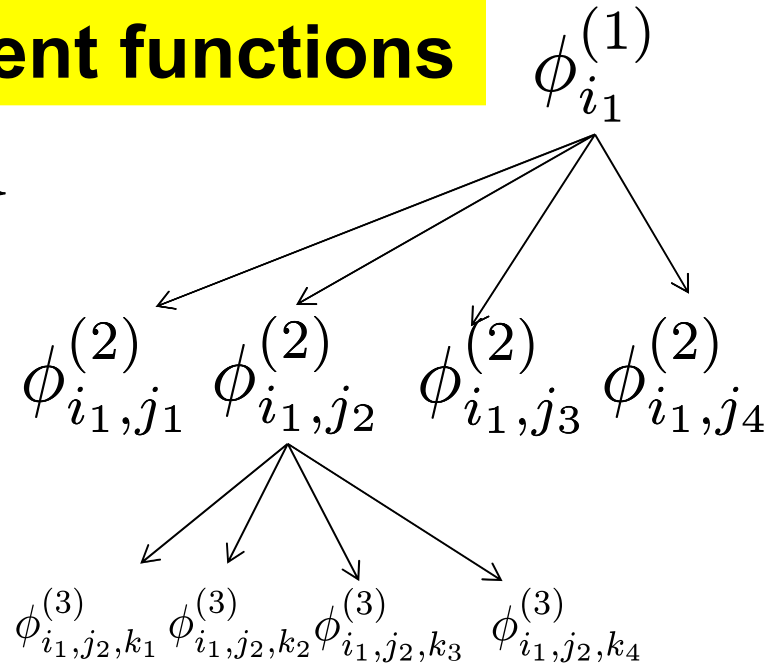
# Formulation of the hierarchical game



# Hierarchy of nested Measurement functions

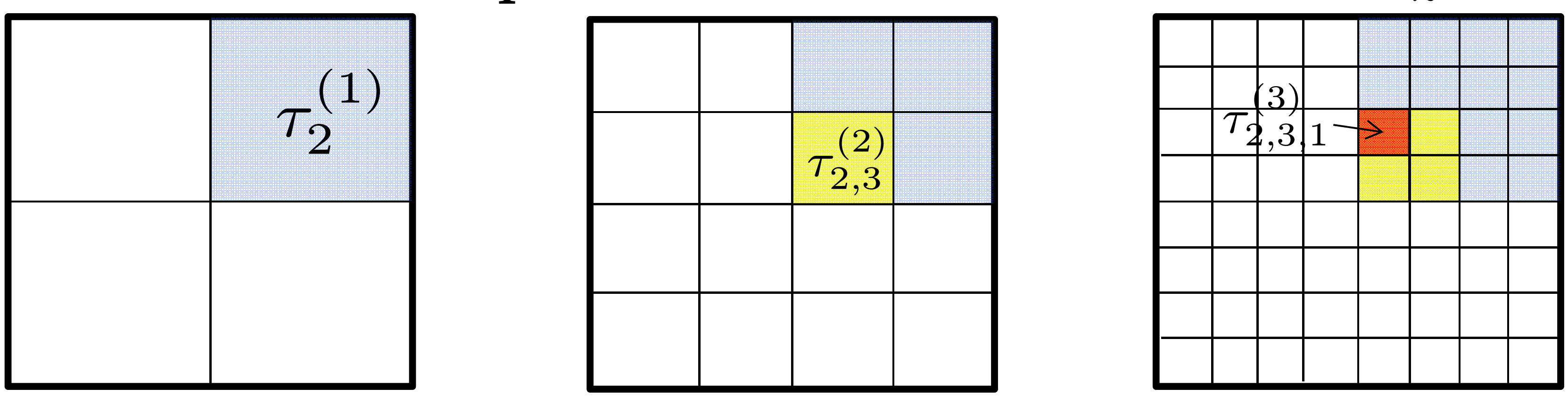
$$\phi^{(k)}_{i_1,\ldots,i_k} \text{ with } k \in \{1,\ldots,q\}$$

$$\phi^{(k)}_{i} = \sum_j c_{i,j}\phi^{(k+1)}_{i,j}$$



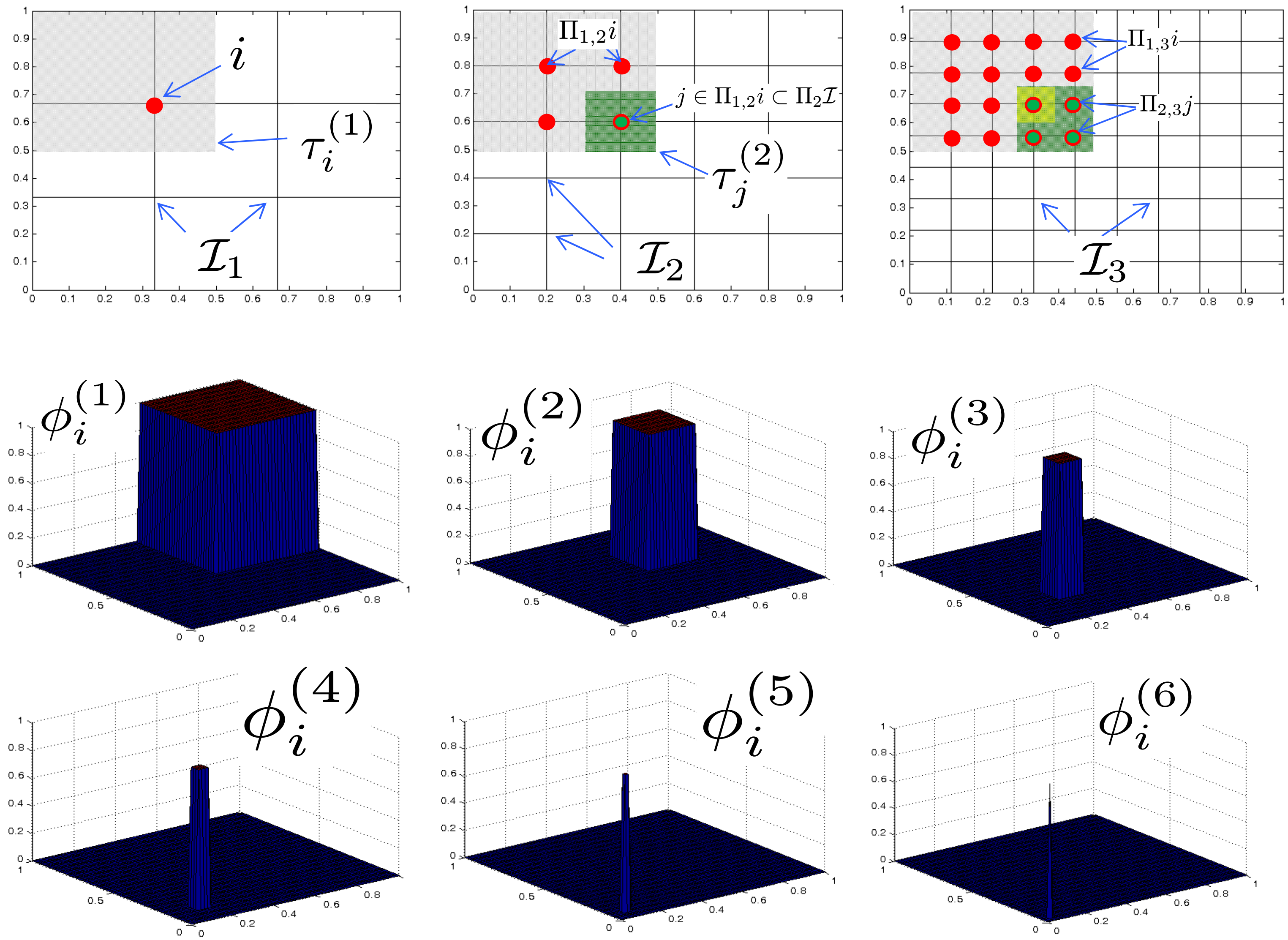

## Example

$\phi^{(k)}_{i}$ : Indicator functions of a hierarchical nested partition of $\Omega$ of resolution $H_k = 2^{-k}$




$$\phi^{(1)}_2 = 1_{\tau^{(1)}_2} \qquad \phi^{(2)}_{2,3} = 1_{\tau^{(2)}_{2,3}} \qquad \phi^{(3)}_{2,3,1} = 1_{\tau^{(3)}_{2,3,1}}$$

# In the discrete setting simply aggregate elements (as in algebraic multigrid)



# Formulation of the hierarchy of games

**Player I**

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases}$$

**Player II**

Chooses

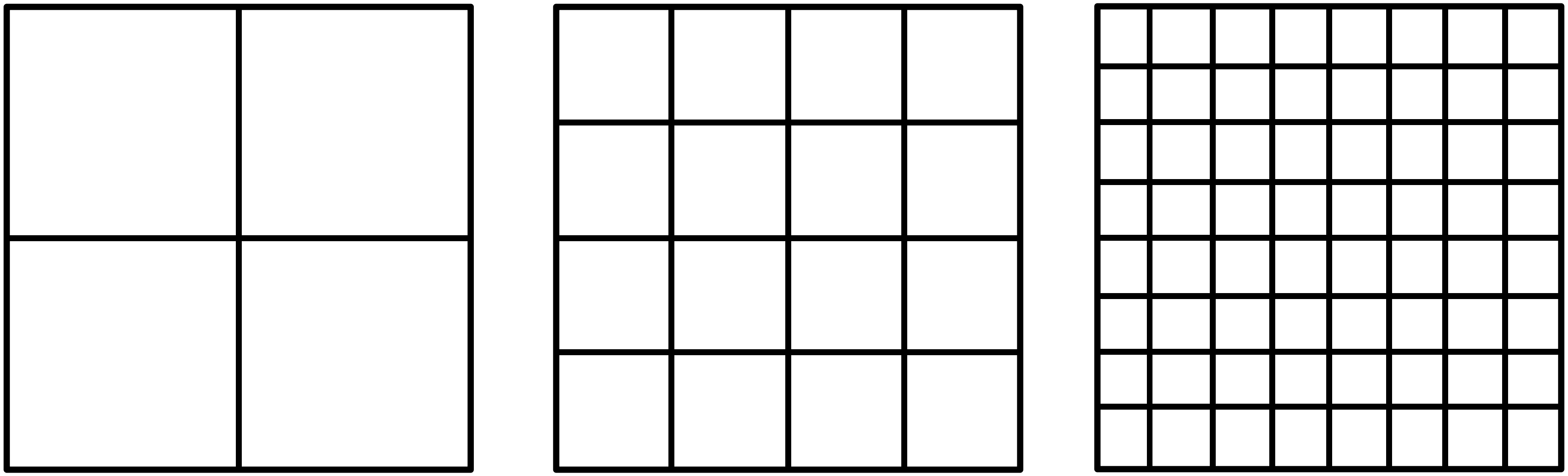
$g \in L^2(\Omega)$

$\|g\|_{L^2(\Omega)} \le 1$

Sees $\{\int_\Omega u\phi_i^{(k)},\ i \in \mathcal{I}_k\}$

Must predict

$u$ and $\{\int_\Omega u\phi_j^{(k+1)},\ j \in \mathcal{I}_{k+1}\}$



## Player II's best strategy

$$\xi \sim \mathcal{N}(0, \mathcal{L})$$

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases} \Longleftrightarrow \begin{cases} -\operatorname{div}(a\nabla v) = \xi \text{ in } \Omega, \\ v = 0 \text{ on } \partial\Omega, \end{cases}$$

## Player II's bets

$$u^{(k)}(x) := \mathbb{E}\Big[v(x)\Big| \int_\Omega v(y)\phi_i^{(k)}(y)\,dy = \int_\Omega u(y)\phi_i^{(k)}(y)\,dy,\ i \in \mathcal{I}_k\Big]$$

**The sequence of approximations forms a martingale under the mixed strategy emerging from the game**

$$\mathcal{F}_k = \sigma\Big(\int_\Omega v\phi_i^{(k)},\ i \in \mathcal{I}_k\Big) \qquad v^{(k)}(x) := \mathbb{E}\Big[v(x)\Big|\mathcal{F}_k\Big]$$

**Theorem**

$$\mathcal{F}_k \subset \mathcal{F}_{k+1}$$

$$v^{(k)}(x) := \mathbb{E}\Big[v^{(k+1)}(x)\Big|\mathcal{F}_k\Big]$$

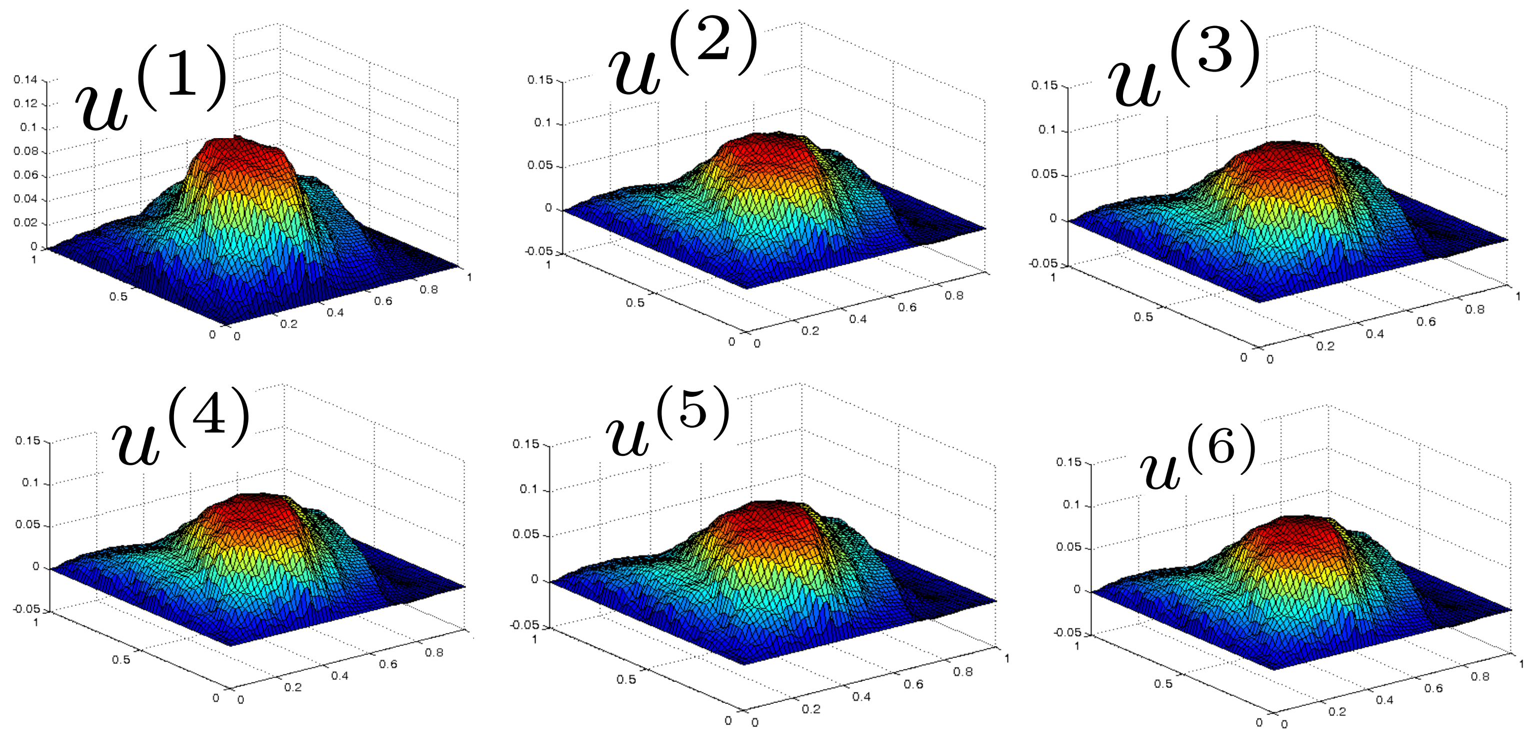
# Player II's best strategy

$$\xi \sim \mathcal{N}(0, \mathcal{L})$$

$$\begin{cases} -\operatorname{div}(a\nabla u) = g \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{cases} \Longleftrightarrow \begin{cases} -\operatorname{div}(a\nabla v) = \xi \text{ in } \Omega, \\ v = 0 \text{ on } \partial\Omega, \end{cases}$$

# Player II's bets

$$u^{(k)}(x) := \mathbb{E}\Big[v(x)\Big| \int_\Omega v(y)\phi_i^{(k)}(y)\,dy = \int_\Omega u(y)\phi_i^{(k)}(y)\,dy,\ i \in \mathcal{I}_k\Big]$$



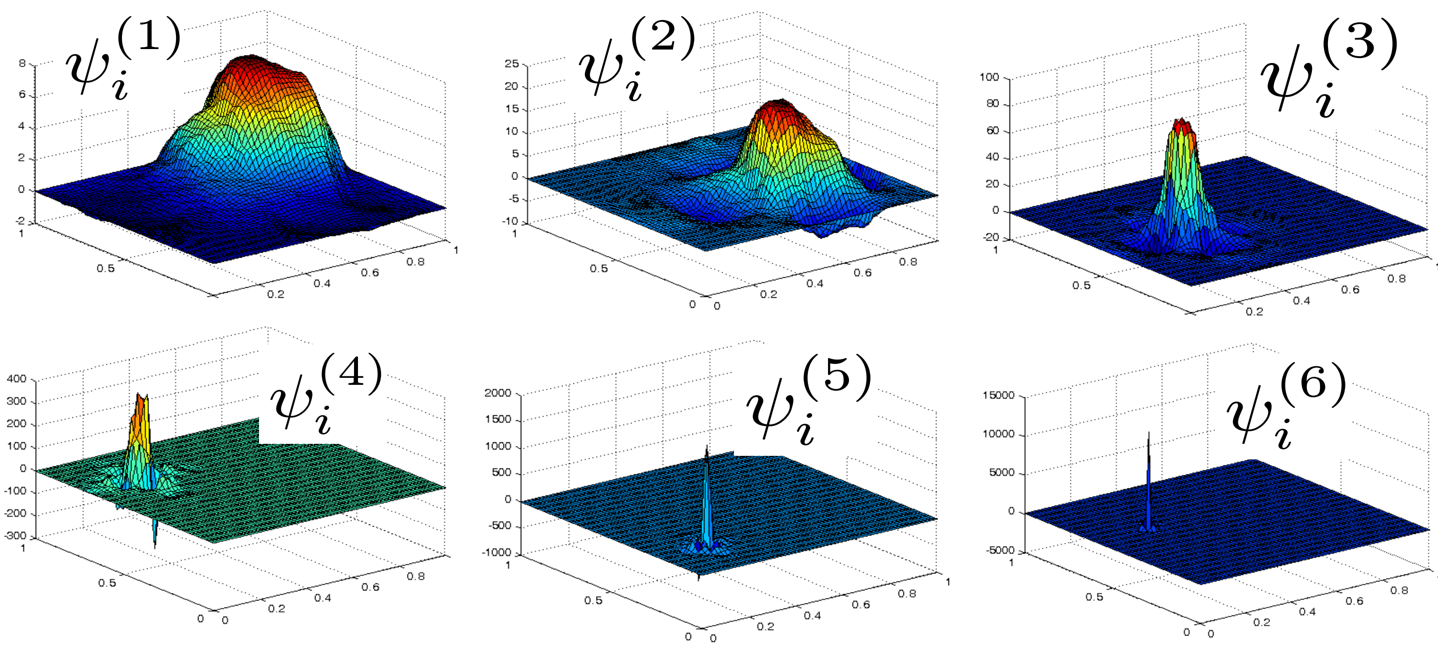
**Gamblets** **Elementary gambles form a hierarchy of deterministic basis functions for player II's hierarchy of bets**

**Theorem**

$$u^{(k)}(x) = \sum_i \psi_i^{(k)}(x) \int_\Omega u(y)\phi_i^{(k)}(y)\,dy$$

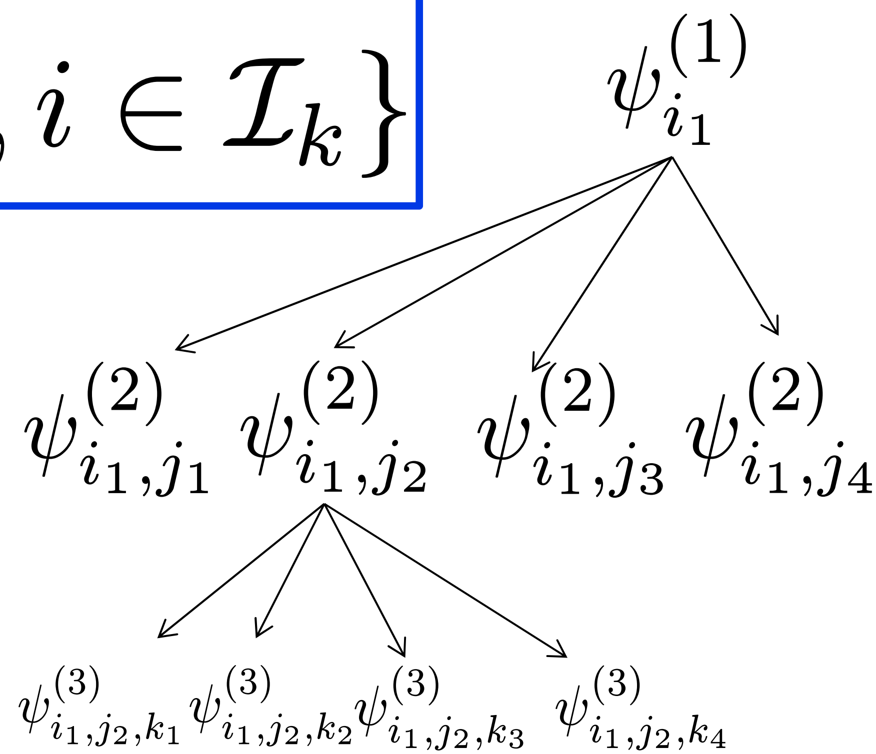
$\psi_i^{(k)}$: Elementary gambles/bets at resolution $H_k = 2^{-k}$

$$\psi_i^{(k)}(x) := \mathbb{E}\Big[v(x)\Big|\int_\Omega v(y)\phi_j^{(k)}(y)\,dy = \delta_{i,j},\, j \in \mathcal{I}_k\Big]$$



# Gamblets are nested

$$\mathfrak{V}^{(k)} := \operatorname{span}\{\psi_i^{(k)}, i \in \mathcal{I}_k\}$$




## Theorem

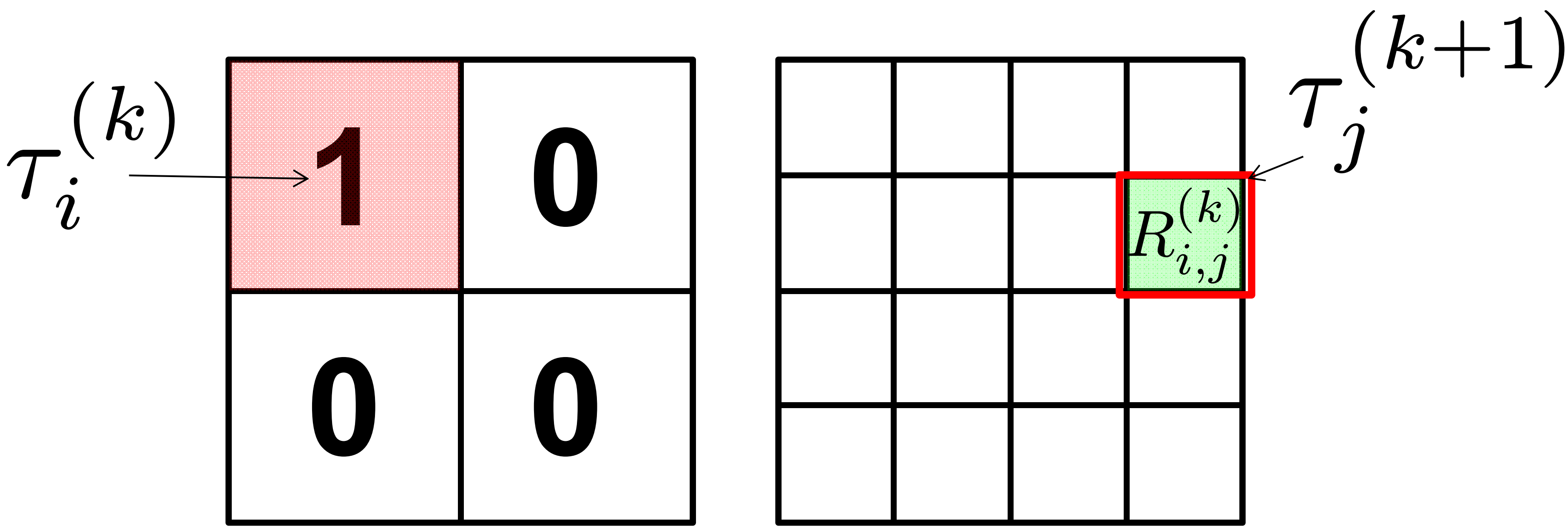
$$\mathfrak{V}^{(k)} \subset \mathfrak{V}^{(k+1)}$$

$$\psi_i^{(k)}(x) = \sum_{j\in\mathcal{I}_{k+1}} R^{(k)}_{i,j}\psi_j^{(k+1)}(x)$$

# Interpolation/Prolongation operator

$$R^{(k)}_{i,j} = \mathbb{E}\big[\int_\Omega v(y)\phi^{(k+1)}_j(y)\,dy \big| \int_\Omega v(y)\phi^{(k)}_l(y)\,dy = \delta_{i,l},\, l\in\mathcal{I}_k\big]$$

$R^{(k)}_{i,j}$ Your best bet on the value of $\int_{\tau^{(k+1)}_j} u$ given the information that $\int_{\tau^{(k)}_i} u = 1$ and $\int_{\tau_l} u = 0$ for $l \neq i$



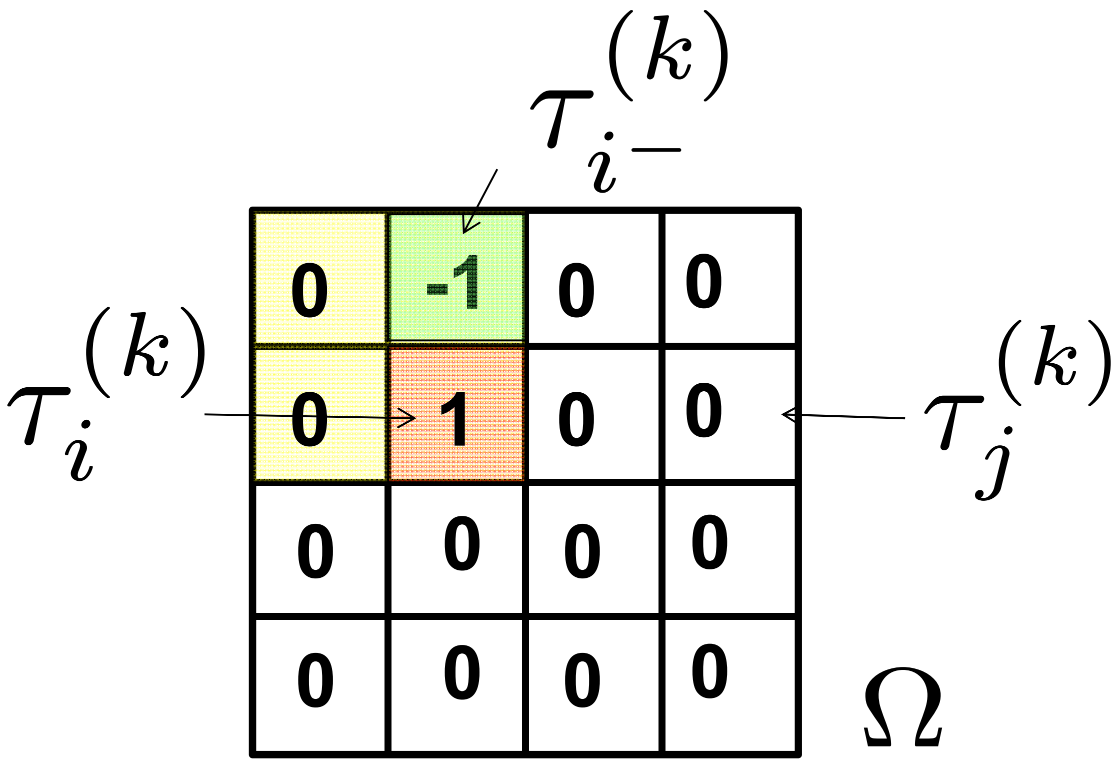
**At this stage you can finish with classical multigrid**

**But we want multiresolution decomposition**

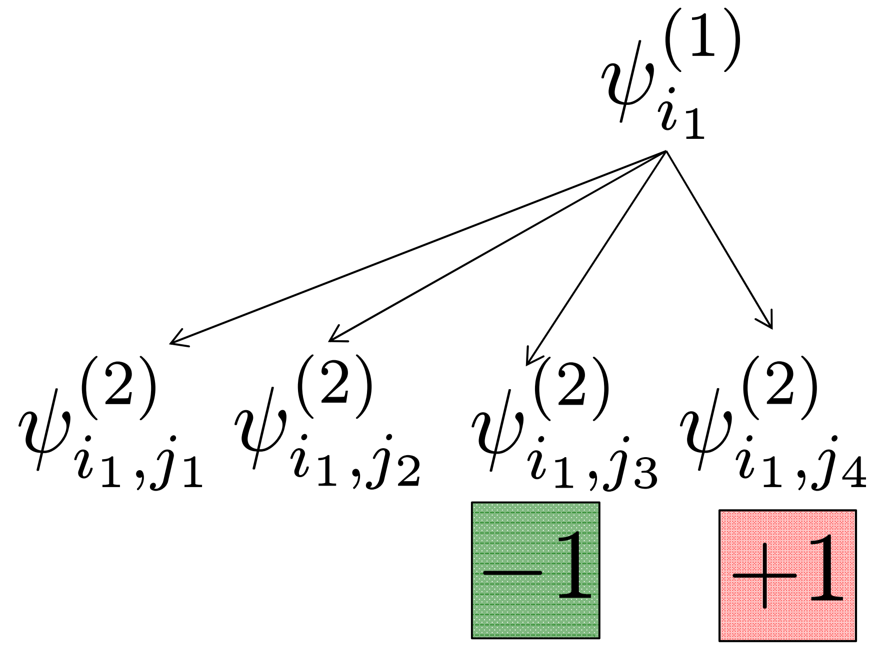
## Elementary gamble

$\chi_i^{(k)}$ Your best bet on the value of $u$ given the information that

$$\int_{\tau_i^{(k)}} u = 1,\ \int_{\tau_{i-}^{(k)}} u = -1 \text{ and } \int_{\tau_j^{(k)}} u = 0 \text{ for } j \neq i$$

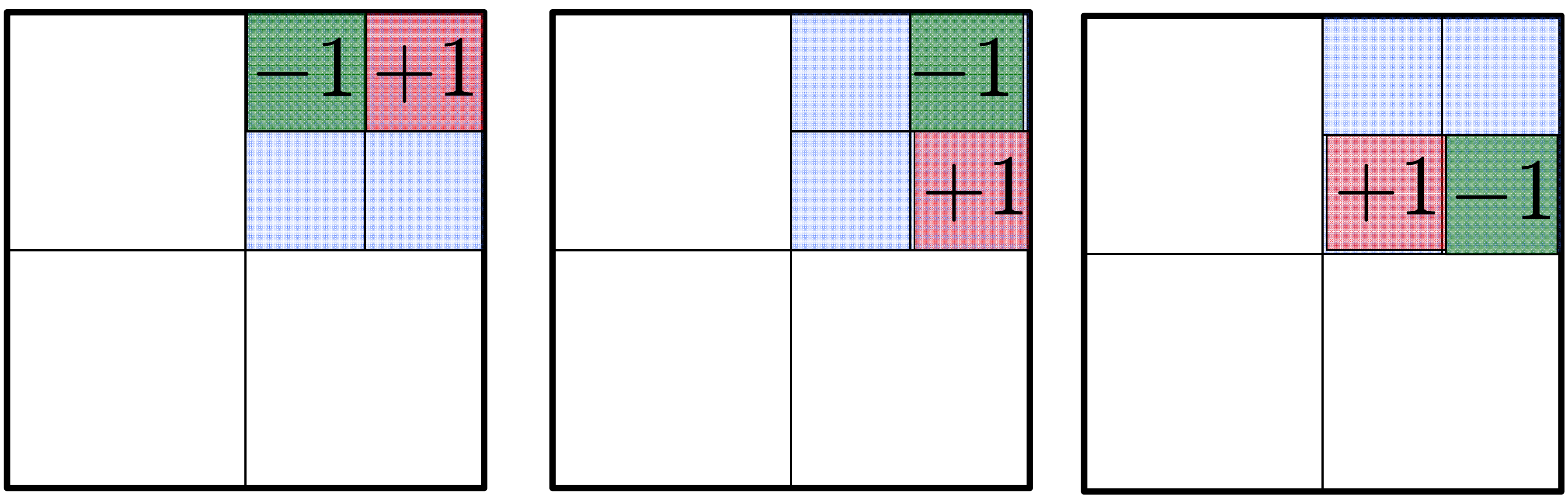


$$\chi_i^{(k)} = \psi_i^{(k)} - \psi_{i^-}^{(k)}$$

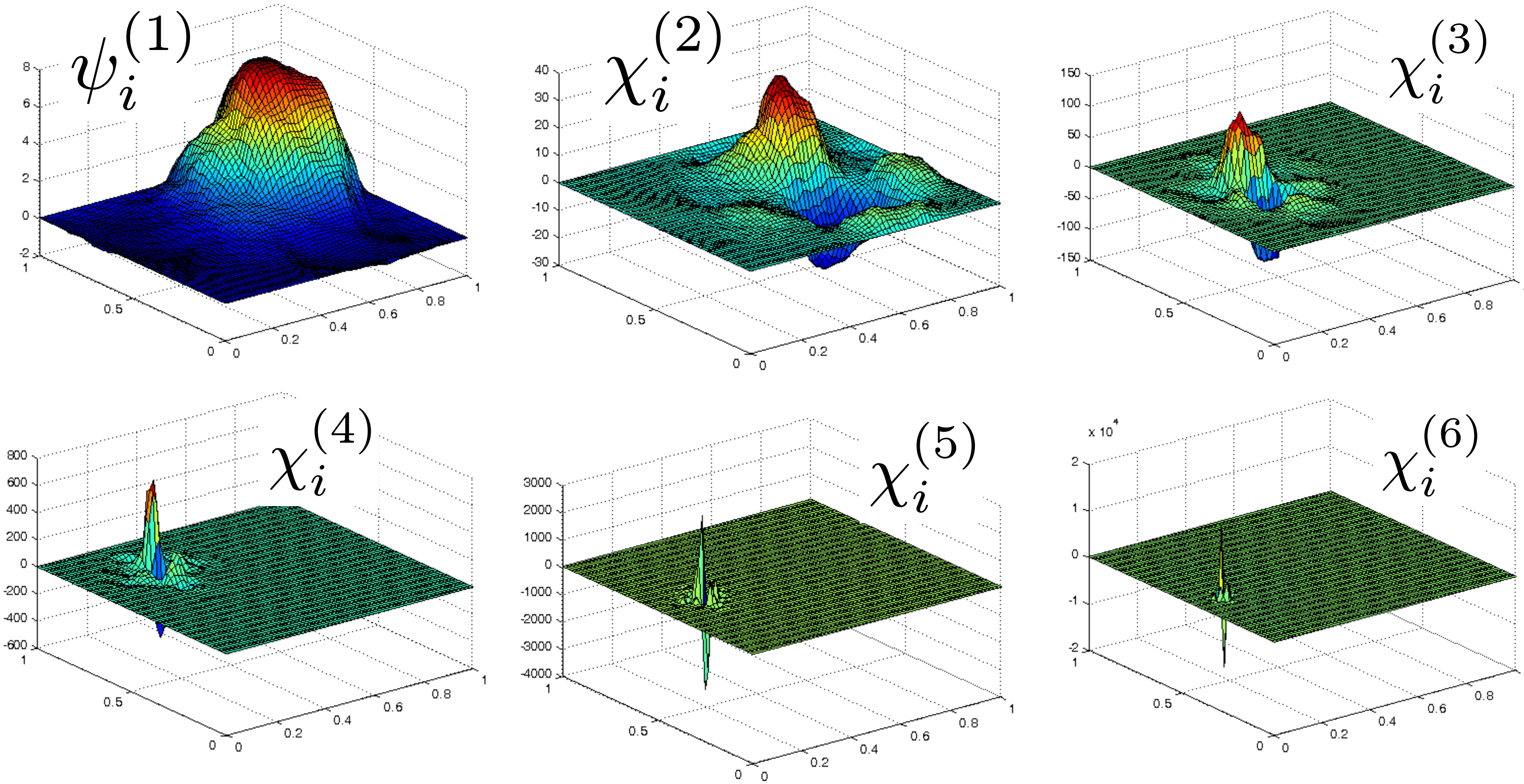



$$i = (i_1, \ldots, i_{k-1}, i_k)$$

$$i^- = (i_1, \ldots, i_{k-1}, i_k - 1)$$



$$\chi_i^{(k)} = \psi_i^{(k)} - \psi_{i^-}^{(k)}$$



## Multiresolution decomposition of the solution space

$$\mathfrak{V}^{(k)} := \text{span}\{\psi_i^{(k)}, i \in \mathcal{I}_k\}$$

$$\mathfrak{W}^{(k)} := \text{span}\{\chi_i^{(k)}, i\}$$

$\mathfrak{W}^{(k+1)}$: Orthogonal complement of $\mathfrak{V}^{(k)}$ in $\mathfrak{V}^{(k+1)}$ with respect to $< \psi, \chi >_a := \int_\Omega (\nabla \psi)^T a \nabla \chi$

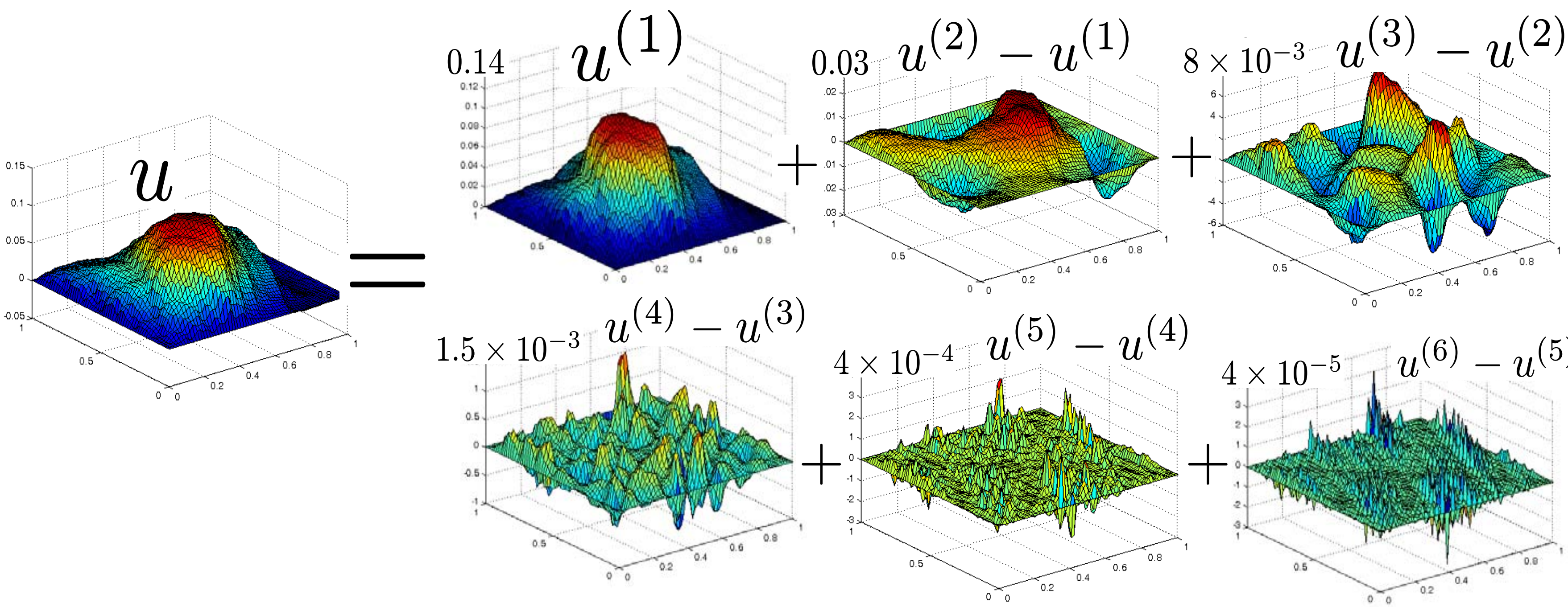
## Theorem

$$H_0^1(\Omega) = \mathfrak{V}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \cdots \oplus_a \mathfrak{W}^{(k)} \oplus_a \cdots$$

# Multiresolution decomposition of the solution

## Theorem

$$u^{(k+1)} - u^{(k)} = \text{F.E. sol. of PDE in } \mathfrak{W}^{(k+1)}$$




Subband solutions $u^{(k+1)} - u^{(k)}$ can be computed independently

# Uniformly bounded condition numbers

$$A^{(k)}_{i,j} := \langle \psi^{(k)}_i, \psi^{(k)}_j \rangle_a$$

$$B^{(k)}_{i,j} := \langle \chi^{(k)}_i, \chi^{(k)}_j \rangle_a$$

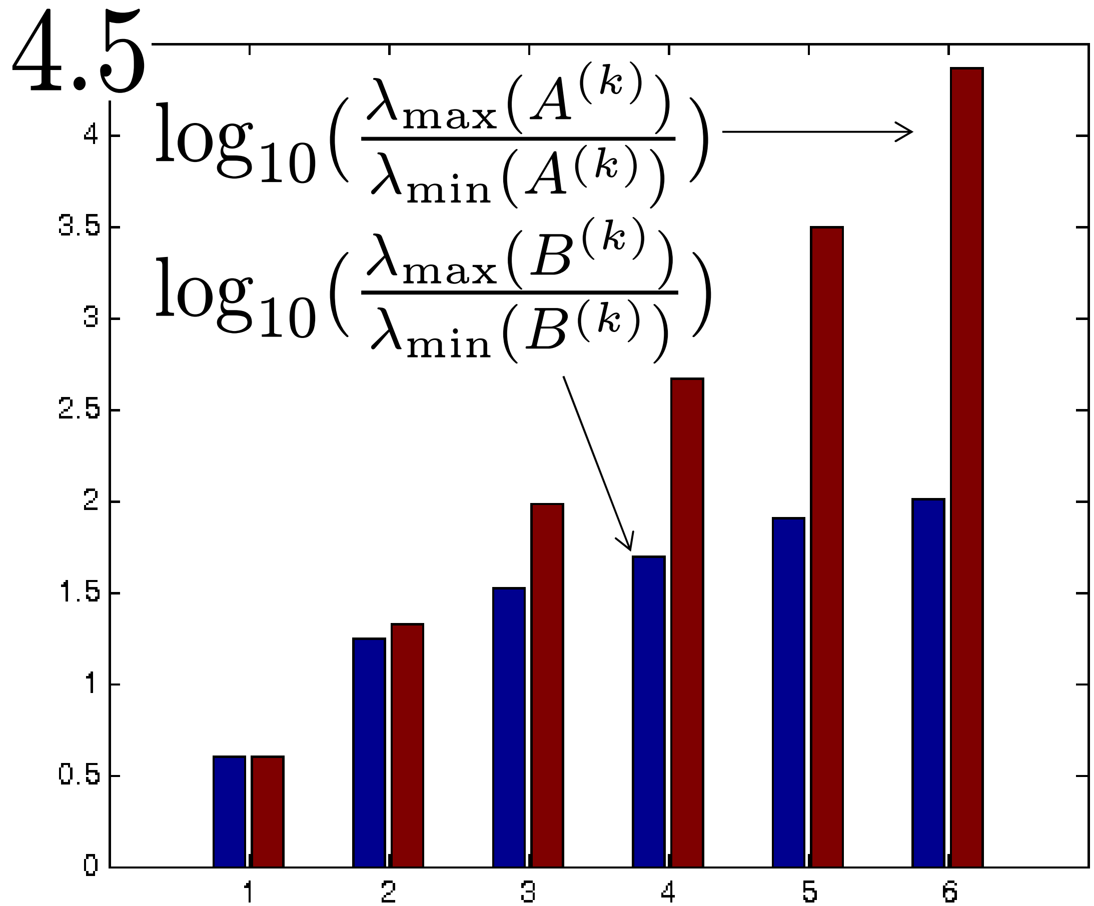
## Theorem

$$\frac{\lambda_{\max}(B^{(k)})}{\lambda_{\min}(B^{(k)})} \leq C$$

⇓

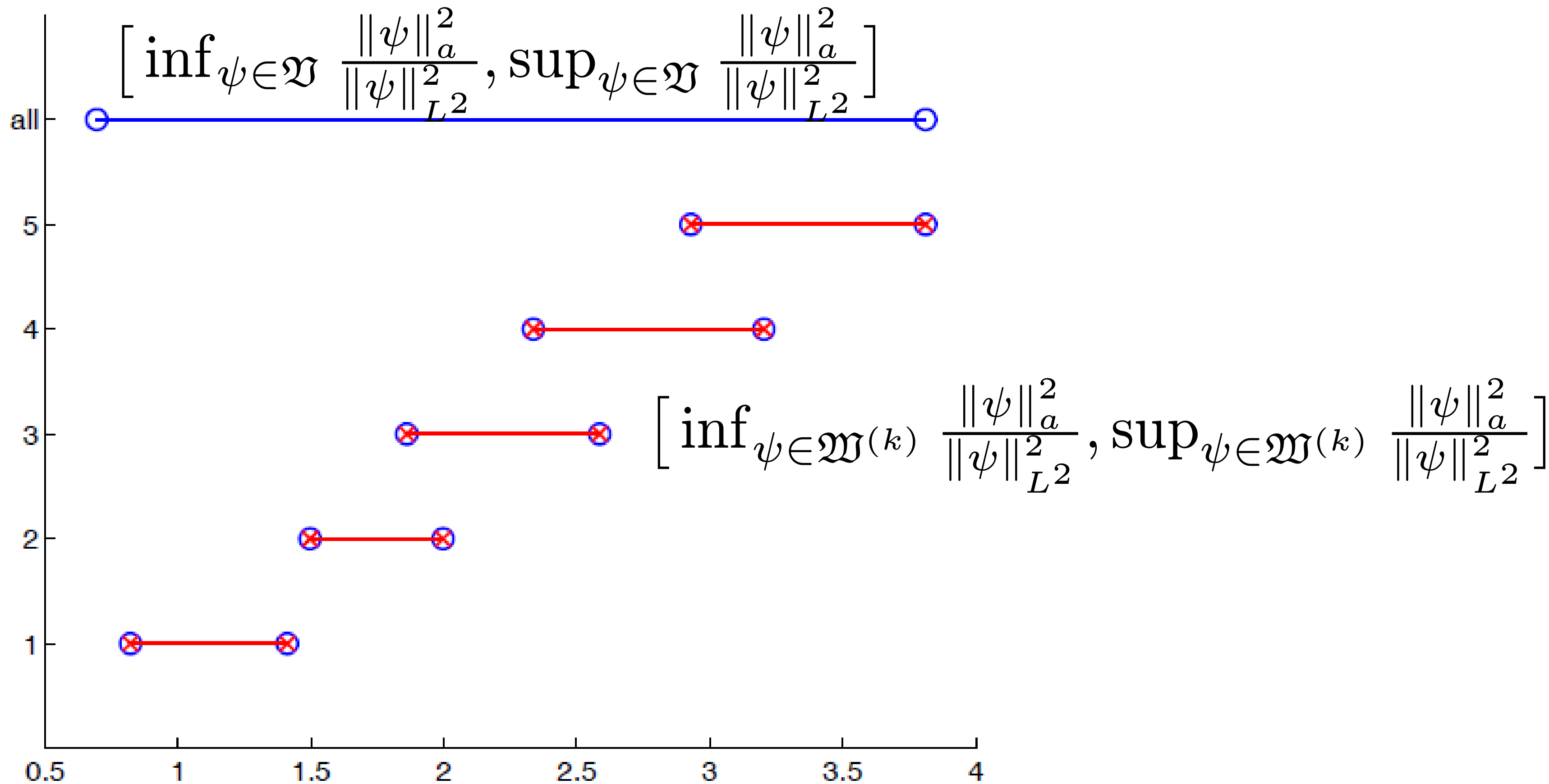
**Just relax!**

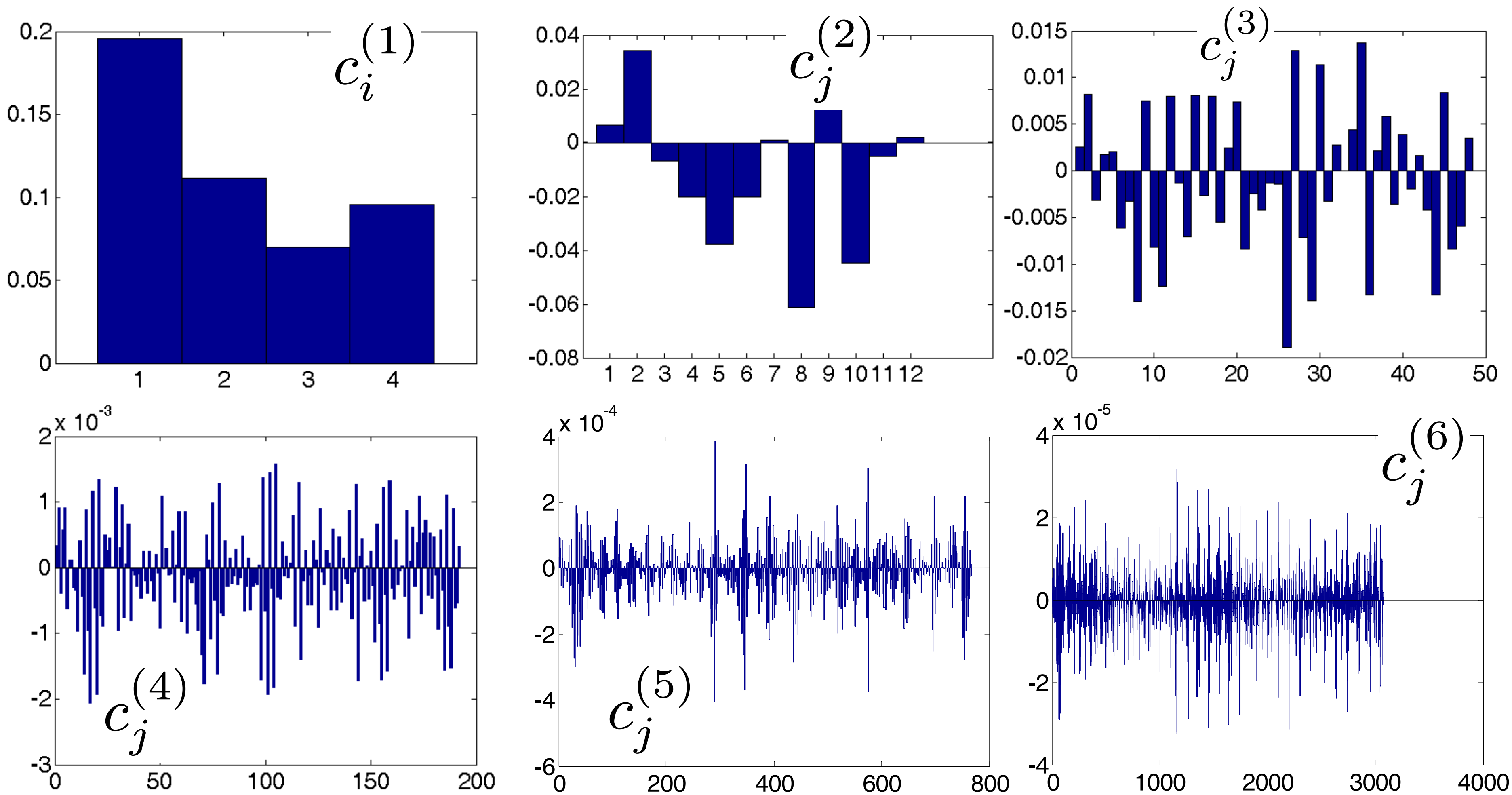
In $v \in \mathfrak{W}^{(k)}$
to get
$u^{(k)} - u^{(k-1)}$



$$\mathfrak{V} = \mathfrak{W}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \cdots \oplus_a \mathfrak{W}^{(k)}$$

Ranges of eigenvalues in $\mathfrak{V}$ and $\mathfrak{W}^{(k)}$ $(k = 1, \ldots, 5)$ in log scale



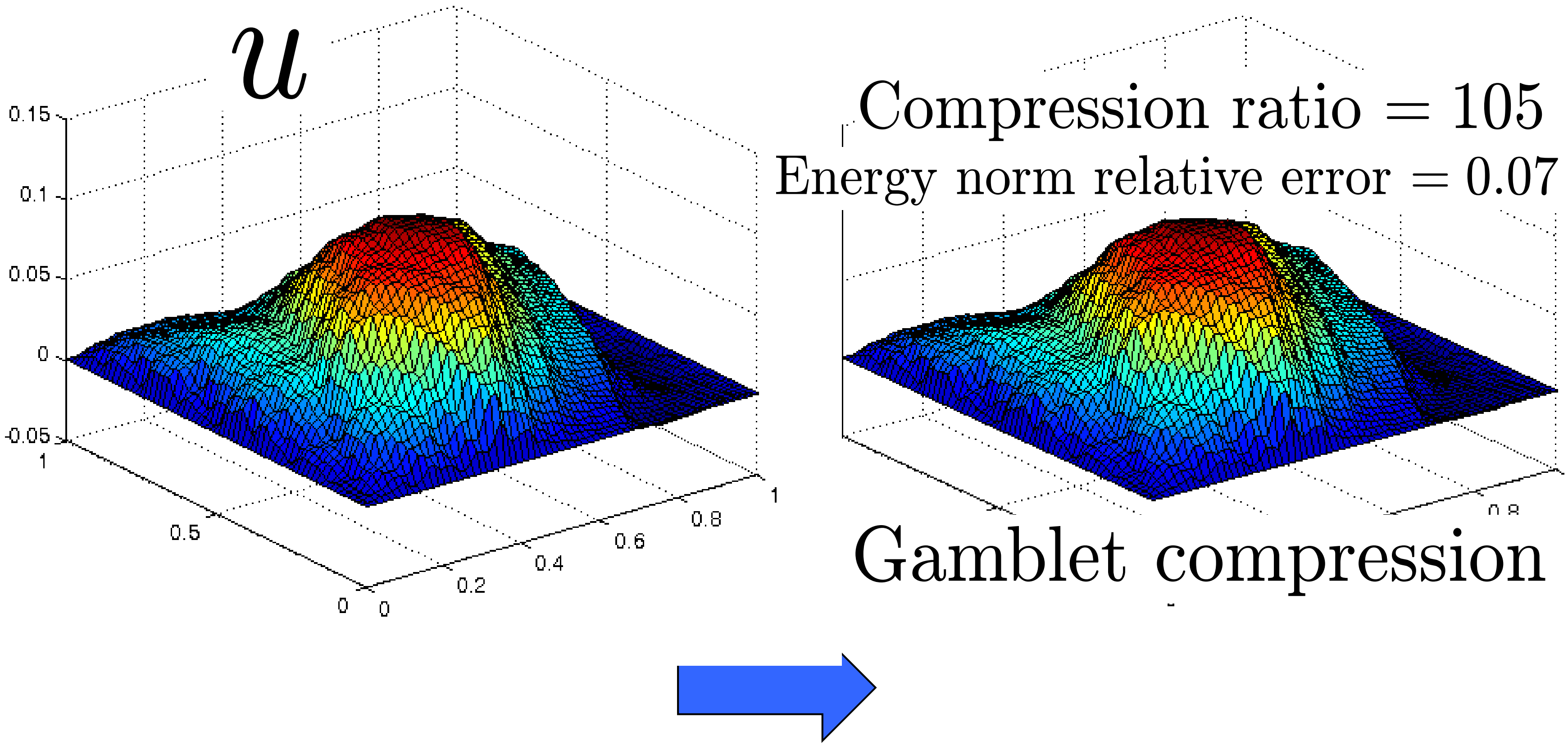



$$u = \sum_i c_i^{(1)} \frac{\psi_i^{(1)}}{\|\psi_i^{(1)}\|_a} + \sum_{k=2}^{q} \sum_j c_j^{(k)} \frac{\chi_j^{(k)}}{\|\chi_j^{(k)}\|_a}$$

**Coefficients of the solution in the gamblet basis**

# Operator Compression

**Gamblets behave like wavelets but they are adapted to the PDE and can compress its solution space**




# Throw 99% of the coefficients

**Fast gamblet transform**

$\mathcal{O}(N \ln^{3d} N)$ complexity

**Nesting**

$$A^{(k)} = (R^{(k,k+1)})^T A^{(k+1)} R^{(k,k+1)}$$

**Level(k) gamblets and stiffness matrices can be computed from level(k+1) gamblets and stiffness matrices**

**Well conditioned linear systems**

**Underlying linear systems have uniformly bounded condition numbers**

$$\psi_i^{(k)} = \psi_{(i,1)}^{(k+1)} + \sum_j C_{i,j}^{(k+1),\chi} \chi_j^{(k+1)}$$

$$C^{(k+1),\chi} = (B^{(k+1)})^{-1} Z^{(k+1)}$$

$$Z_{j,i}^{(k+1)} := -(e_j^{(k+1)} - e_{j^-}^{(k+1)})^T A^{(k+1)} e_{(i,1)}^{(k+1)}$$

**Localization**

**The nested computation can be localized without compromising accuracy or condition numbers**

**Theorem**

Localizing $(\psi_i^{(k)})_{i\in\mathcal{I}_k}$ and $(\chi_i^{(k)})_i$ to subdomains of size

$\geq C H_k \ln \frac{1}{H_k}$ ➡ $\text{Cond. No } (B^{(k),\text{loc}}) \leq C$

$\geq C H_k (\ln \frac{1}{H_k} + \ln \frac{1}{\epsilon})$ ➡

$$\left\| u - u^{(1),\text{loc}} - \sum_{k=1}^{q-1} (u^{(k+1),\text{loc}} - u^{(k),\text{loc}}) \right\|_a \leq \epsilon$$

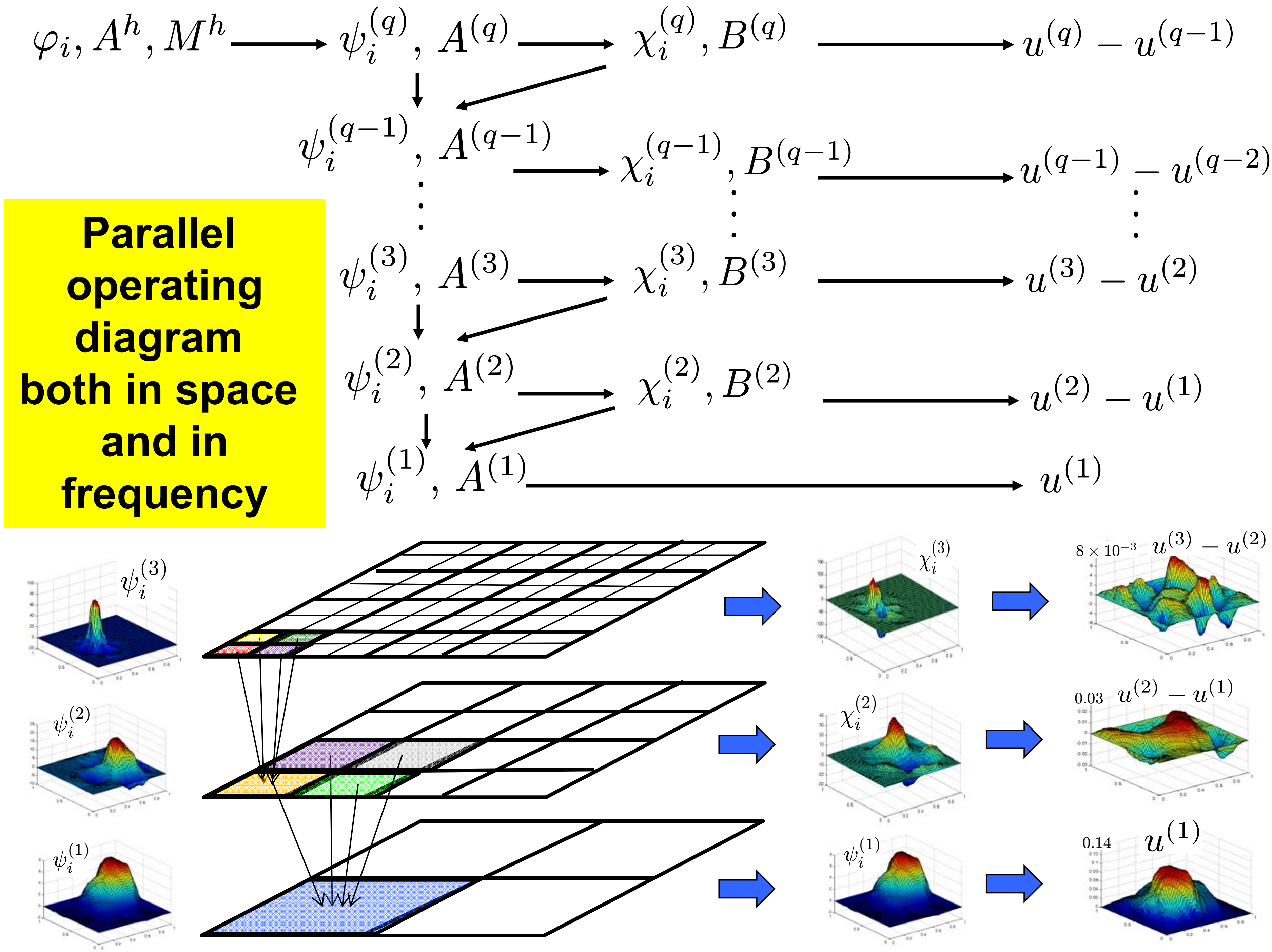
**Theorem**

The number of operations to compute gamblets and achieve accuracy $\epsilon$ is $\mathcal{O}\big(N \ln^{3d}\big(\max(\frac{1}{\epsilon}, N^{1/d})\big)\big)$

(and $\mathcal{O}\big(N \ln^{d}(N^{1/d}) \ln \frac{1}{\epsilon}\big)$ for subsequent solves)

**Complexity**

| | | | |
|---|---|---|---|
| Gamblet Transform | $\mathcal{O}(N \ln^{3d} N)$ | Linear Solve | $\mathcal{O}(N \ln^{d+1} N)$ |


Parallel operating diagram both in space and in frequency
$\varphi_i, A^h, M^h$
$\psi_i^{(q)}, A^{(q)}$
$\chi_i^{(q)}, B^{(q)}$
$u^{(q)} - u^{(q-1)}$
$\psi_i^{(q-1)}, A^{(q-1)}$
$\chi_i^{(q-1)}, B^{(q-1)}$
$u^{(q-1)} - u^{(q-2)}$
$\psi_i^{(3)}, A^{(3)}$
$\chi_i^{(3)}, B^{(3)}$
$u^{(3)} - u^{(2)}$
$\psi_i^{(2)}, A^{(2)}$
$\chi_i^{(2)}, B^{(2)}$
$u^{(2)} - u^{(1)}$
$\psi_i^{(1)}, A^{(1)}$
$u^{(1)}$
$\psi_i^{(3)}$
$\psi_i^{(2)}$
$\psi_i^{(1)}$
$\chi_i^{(3)}$
$\chi_i^{(2)}$
$\psi_i^{(1)}$
$8 \times 10^{-3}$ $u^{(3)} - u^{(2)}$
0.03 $u^{(2)} - u^{(1)}$
0.14 $u^{(1)}$

## Numerical Homogenization

**Harmonic Coordinates** Babuska, Caloz, Osborn, 1994

Kozlov, 1979 Allaire Brizzi 2005; Owhadi, Zhang 2005

**MsFEM** [Hou, Wu: 1997]; [Efendiev, Hou, Wu: 1999]

[Fish - Wagiman, 1993] [Gloria 2010] Arbogast, 2011: Mixed MsFEM

**Projection based method** Nolen, Papanicolaou, Pironneau, 2008

**HMM** Engquist, E, Abdulle, Runborg, Schwab, et Al. 2003-...

**Flux norm** Berlyand, Owhadi 2010; Symes 2012

**Localization** [Chu-Graham-Hou-2010] (limited inclusions)

[Efendiev-Galvis-Wu-2010] (limited inclusions or mask)

[Babuska-Lipton 2010] (local boundary eigenvectors)

[Owhadi-Zhang 2011] (localized transfer property)

[Malqvist-Peterseim 2012] Volume averaged interpolation

[Owhadi-Zhang-Berlyand 2013] (Rough Polyharmonic Splines)

[Owhadi, SIAM MMS, 2015]
Bayesian Numerical Homogenization

[Hou and Liu,DCDS-A, 2016]

## Statistical approach to numerical approximation

[ Henri Poincaré. Calcul des probabilités. 1896. ]

[ A. V. Sul'din, Wiener measure and its applications to approximation methods. Matematika 1959 ]

[ A. Sard. Linear approximation. 1963. ]

[ G. S. Kimeldorf and G. Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. 1970 ]

[ F.M. Larkin. Gaussian measure in Hilbert space and applications in numerical analysis. Rocky Mountain J. Math, 1972 ]

[ H. Woźniakowski. Probabilistic setting of information-based complexity. J. Complexity, 1986.]

[ E. W. Packel. The algorithm designer versus nature: a game-theoretic approach to information-based complexity. J. Complexity, 1987]

[ J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. Information-based complexity. 1988]

# Statistical approach to numerical approximation


[ P. Diaconis. Bayesian numerical analysis. In Statistical decision theory and related topics, 1988 ]

[ J. E. H. Shaw. A quasirandom approach to integration in Bayesian statistics. Ann. Statist, 1988. ]

[ A. O'Hagan. Bayes-Hermite quadrature. J. Statist. Plann. Inference, 29(3):245 260, 1991. ]

[ A. O'Hagan. Some Bayesian numerical analysis. Bayesian statistics, 1992. ]

[ Skilling, J. Bayesian solution of ordinary differential equations. 1992. ]

[ Erich Novak and Henryk Woźniakowski, Tractability of Multivariate Problems, 2008-2010 ]

[ Chkrebtii, O. A., Campbell, D. A., Girolami, M. A. and Calderhead, B. Bayesian uncertainty quantification for differential equations. arXiv:1306.2365. 2013 ]

[ H. Owhadi. Bayesian Numerical Homogenization. SIAM MMS, 2015 ]

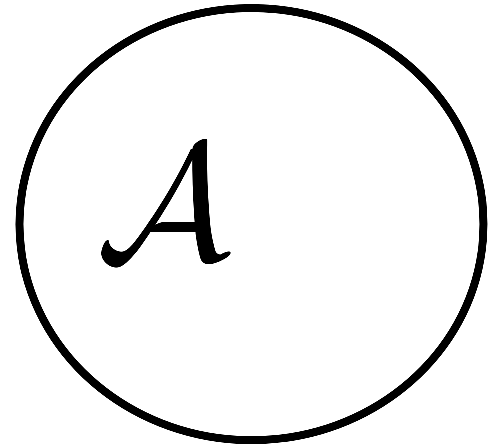
[ P. Hennig. Probabilistic interpretation of linear solvers. SIAM Journal on Optimization, 2015. ]

[ P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. Journal of the Royal Society A, 2015. ]

# Some high level remarks

## What is the worst?

$u^\dagger$: Unknown element of $\mathcal{A}$



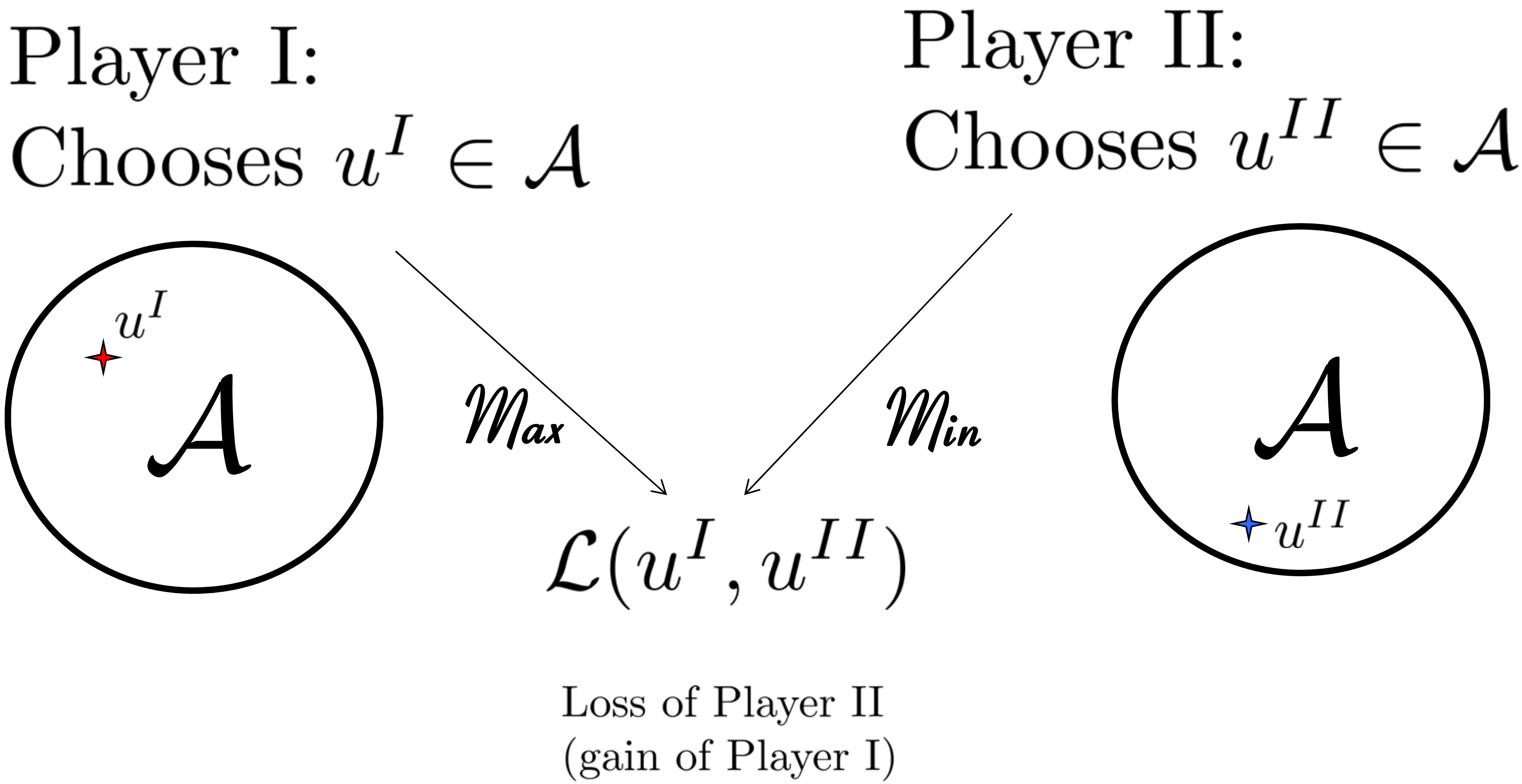

$$\Phi : \mathcal{A} \longrightarrow \mathbb{R}$$

$$u \longrightarrow \Phi(u) \quad \text{Quantity of Interest}$$

What is $\Phi(u^\dagger)$?

$$\inf_{u\in\mathcal{A}} \Phi(u) \leq \Phi(u^\dagger) \leq \sup_{u\in\mathcal{A}} \Phi(u)$$

**Robust Optimization worst case**




**Game theoretic worst case**

## Robust Optimization worst case

Failure is not an option.
You want to always be right.

## Game theoretic worst case

Interpretation depends on the choice of loss function.

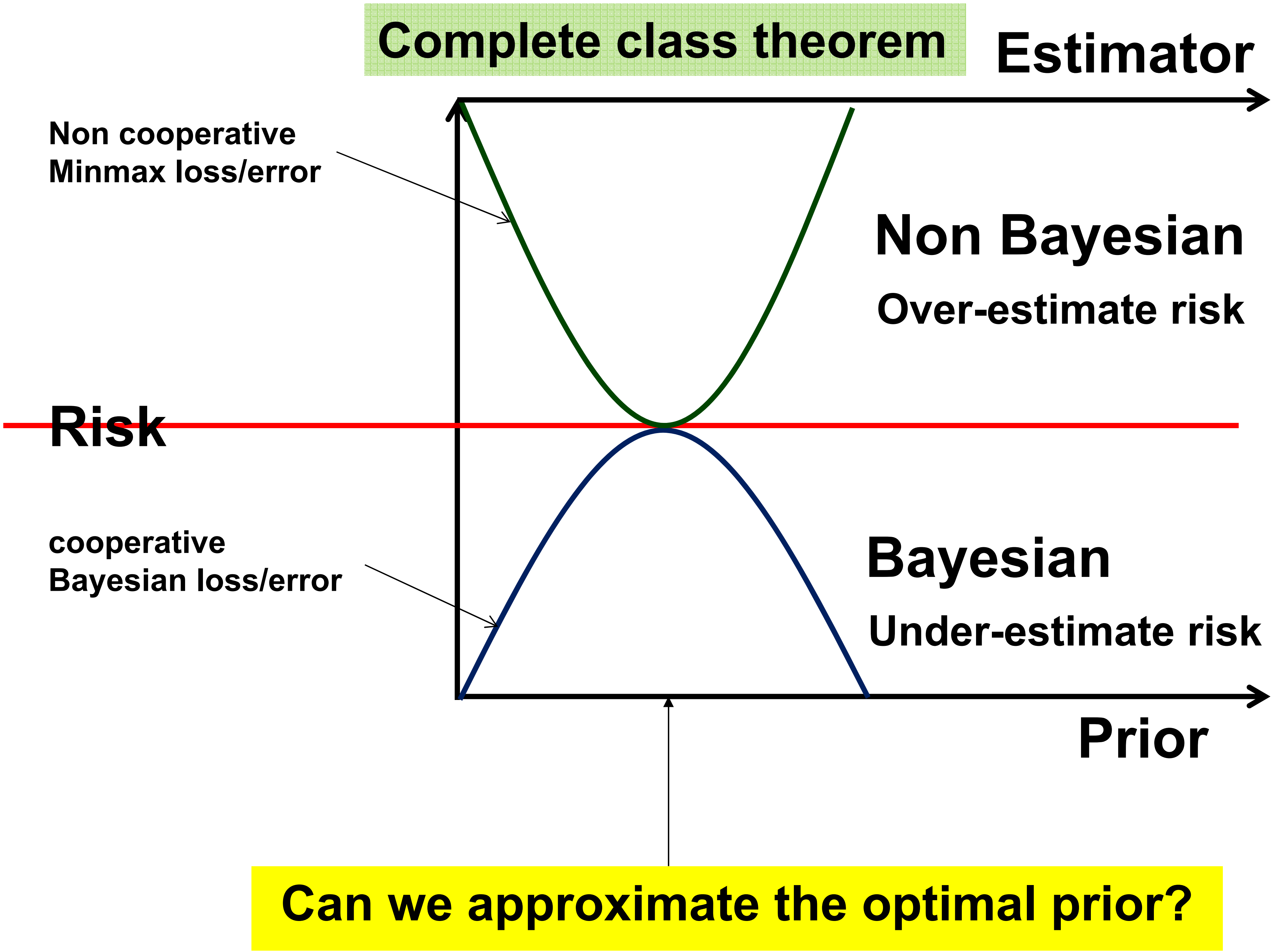
### Confidence error

You want to be right with high probability.

### Quadratic error

You want to be right on average.

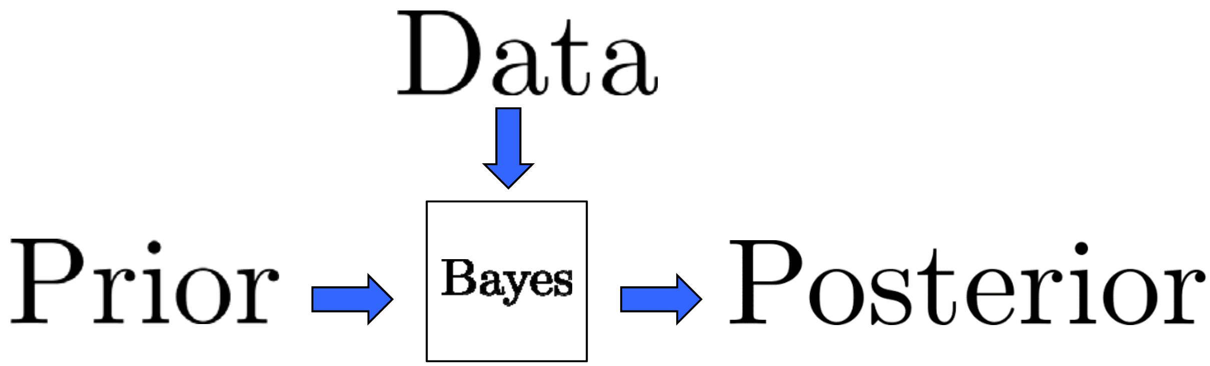
Well suited for numerical computation where you need to keep computing with partial information (e.g. invert a 1,000,000 by 1,000,000 matrix)



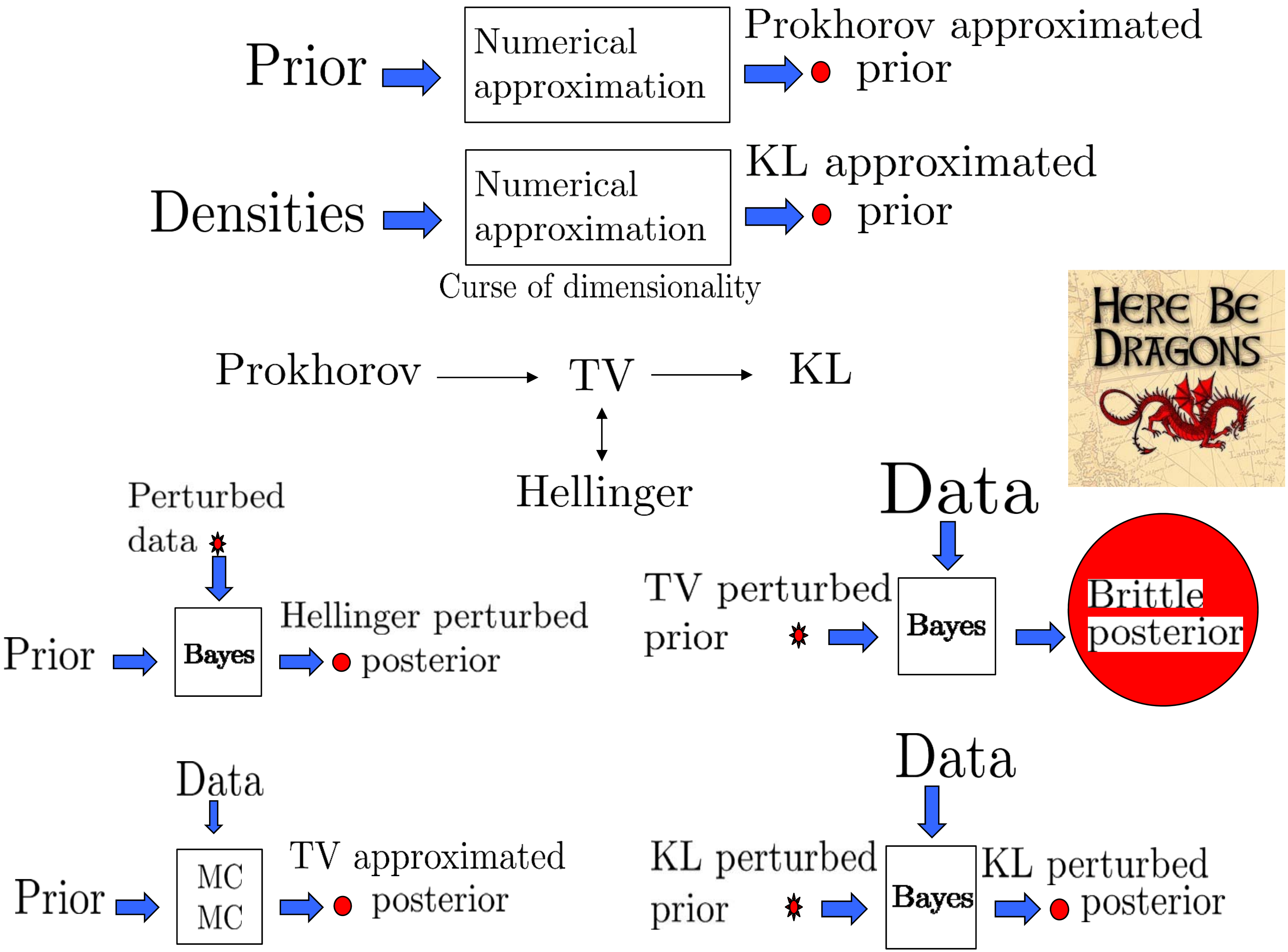
Complete class theorem
Estimator
Non cooperative
Minmax loss/error
Non Bayesian
Over-estimate risk
Risk
cooperative
Bayesian loss/error
Bayesian
Under-estimate risk
Prior
Can we approximate the optimal prior?

# Numerical robustness of Bayesian inference

$$\theta_\pi(d) = \mathbb{E}_{\mu\sim\pi,\, d\sim\mu^n}\left[\Phi(\mu)\middle| d\right]$$




**Can we numerically approximate the prior when closed form expressions are not available for posterior values?**


Prior
Numerical approximation
Prokhorov approximated prior
Densities
Numerical approximation
KL approximated prior
Curse of dimensionality
HERE BE DRAGONS
Prokhorov
TV
KL
Hellinger
Perturbed data
Prior
Bayes
Hellinger perturbed posterior
Data
TV perturbed prior
Bayes
Brittle posterior
Data
Prior
MC MC
TV approximated posterior
Data
KL perturbed prior
Bayes
KL perturbed posterior

# Robustness of Bayesian conditioning in continuous spaces

- Brittleness of Bayesian Inference under Finite Information in a Continuous World. H. Owhadi, C. Scovel and T. Sullivan. Electronic Journal of Statistics, vol 9, pp 1-79, 2015. arXiv:1304.6772
- Brittleness of Bayesian inference and new Selberg formulas. H. Owhadi and C. Scovel. Communications in Mathematical Sciences (2015). arXiv:1304.7046
- On the Brittleness of Bayesian Inference. H. Owhadi, C. Scovel and T. Sullivan. SIAM Review, 57(4), 566-582, 2015, arXiv:1308.6306
- Qualitative Robustness in Bayesian Inference (2015). H. Owhadi and C. Scovel. arXiv:1411.3984

## Positive

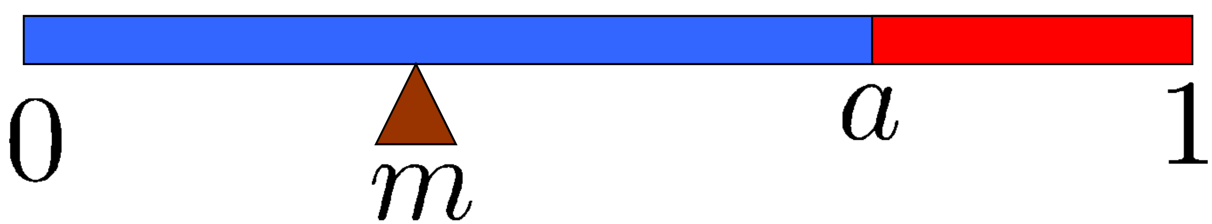
- Classical Bernstein Von Mises
- Wasserman, Lavine, Wolpert (1993)
- P Gustafson & L Wasserman (1995)
- Castillo and Rousseau (2013)
- Castillo and Nickl (2013)
- Stuart & Al (2010+).
- ….

## Negative

- Freedman (1963, 1965)
- P Gustafson & L Wasserman (1995)
- Diaconis & Freedman 1998
- Johnstone 2010
- Leahu 2011
- Belot 2013

Brittleness of Bayesian Inference under Finite Information in a Continuous World. H. Owhadi, C. Scovel and T. Sullivan. Electronic Journal of Statistics, vol 9, pp 1-79, 2015. arXiv:1304.6772

**10.000 children are given one pound of play-doh. On average, how much mass can they put above a while, on average, keeping the seesaw balanced around m?**

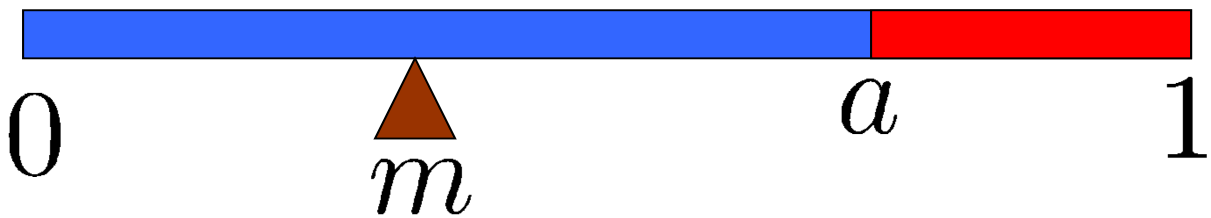



**Paul is given one pound of play-doh.
What can you say about how much mass he is putting above a if all you have is the belief that he is keeping the seesaw balanced around m?**

**What is the least upper bound on**

$$\mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\right]$$

**If all you know is** $\mathbb{E}_{\mu\sim\pi}\left[\mathbb{E}_{\mu}[X]\right] = m$ **?**




$$\mu \in \mathcal{A} := \mathcal{M}\big([0,1]\big)$$

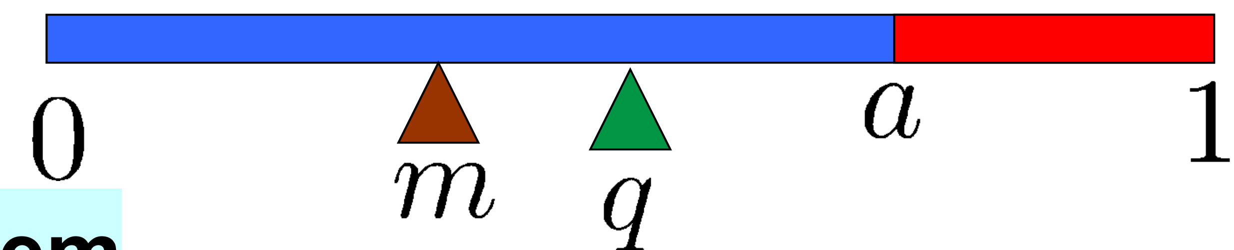
**Answer**

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\right]$$

$$\Pi := \left\{\pi \in \mathcal{M}(\mathcal{A}) \,:\, \mathbb{E}_{\mu\sim\pi}\left[\mathbb{E}_{\mu}[X]\right] = m\right\}$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right]$$

$$\Pi := \left\{ \pi \in \mathcal{M}(\mathcal{M}([0,1])) \; : \; \mathbb{E}_{\mu \sim \pi}\left[\mathbb{E}_{\mu}[X]\right] = m \right\}$$

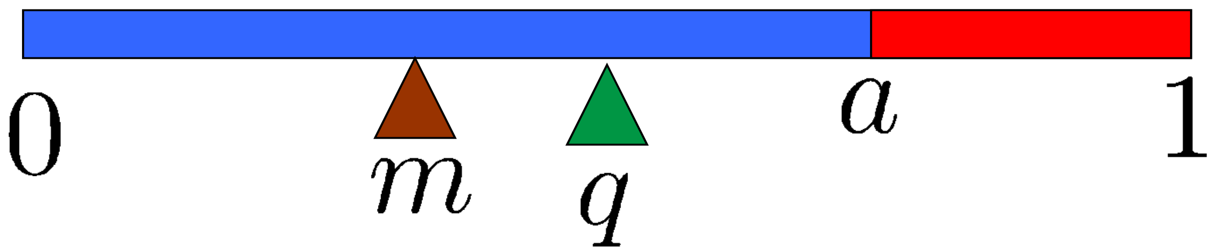



**Theorem**

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right] = \sup_{\mathbb{Q} \in \mathcal{M}([0,1]) \,:\, \mathbb{E}_{\mathbb{Q}}[q] = m} \mathbb{E}_{q \sim \mathbb{Q}} \left[ \sup_{\mu \in \mathcal{M}([0,1]) \,:\, \mathbb{E}_{\mu}[X] = q} \mu[X \geq a] \right]$$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \ge a]\right]$$

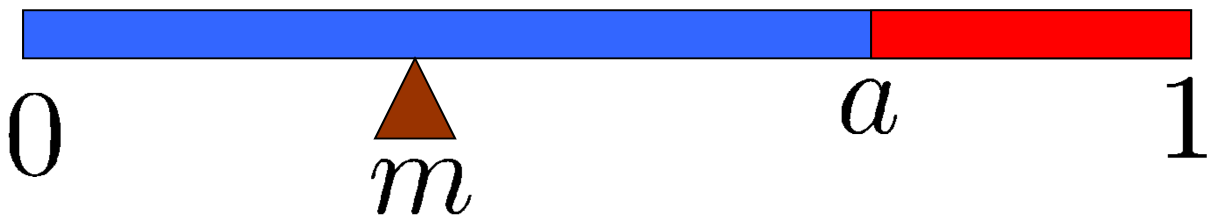
$$\Pi := \left\{\pi \in \mathcal{M}\big(\mathcal{M}([0,1])\big) \;:\; \mathbb{E}_{\mu\sim\pi}\left[\mathbb{E}_{\mu}[X]\right] = m\right\}$$




$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \ge a]\right] = \sup_{\mathbb{Q}\in\mathcal{M}([0,1])\,:\,\mathbb{E}_{\mathbb{Q}}[q]=m} \mathbb{E}_{q\sim\mathbb{Q}}\left[\min\left(\frac{q}{a}, 1\right)\right]$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right]$$

$$\Pi := \left\{ \pi \in \mathcal{M}(\mathcal{M}([0,1])) \ : \ \mathbb{E}_{\mu \sim \pi} \left[ \mathbb{E}_{\mu}[X] \right] = m \right\}$$




$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[ \mu[X \geq a] \right] = \frac{m}{a}$$

# Reduction calculus with measures over measures

$$\mathcal{M}(\mathcal{X}) \supset \mathcal{A} \xrightarrow{\ \Psi\ } \mathcal{Q}$$

Polish space

$$\mathcal{M}(\mathcal{A}) \supset \Pi \xleftarrow{\ \Psi^{-1}\ } \mathfrak{Q} \subset \mathcal{M}(\mathcal{Q})$$

**Theorem** Brittleness of Bayesian Inference under Finite Information in a Continuous World. H. Owhadi, C. Scovel and T. Sullivan. Electronic Journal of Statistics, vol 9, pp 1-79, 2015. arXiv:1304.6772

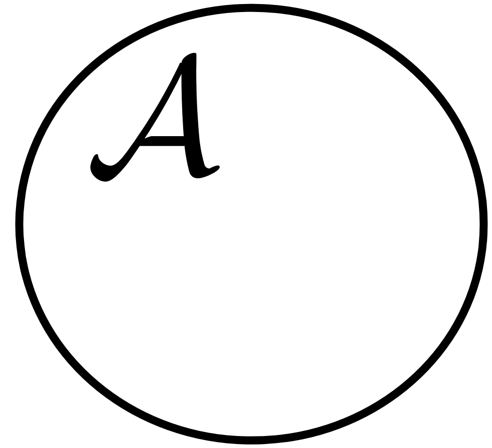
$$\sup_{\pi \in \Psi^{-1}\mathfrak{Q}} \mathbb{E}_{\mu\sim\pi}\big[\Phi(\mu)\big]$$

$$\Big\|$$

$$\sup_{\mathbb{Q}\in\mathfrak{Q}} \Big[\mathbb{E}_{q\sim\mathbb{Q}}\big[\sup_{\mu\in\Psi^{-1}(q)} \Phi(\mu)\big]\Big]$$

## What is the worst with random data?

$\mu^\dagger$: Unknown element of $\mathcal{A}$




$$\Phi : \mathcal{A} \longrightarrow \mathbb{R}$$

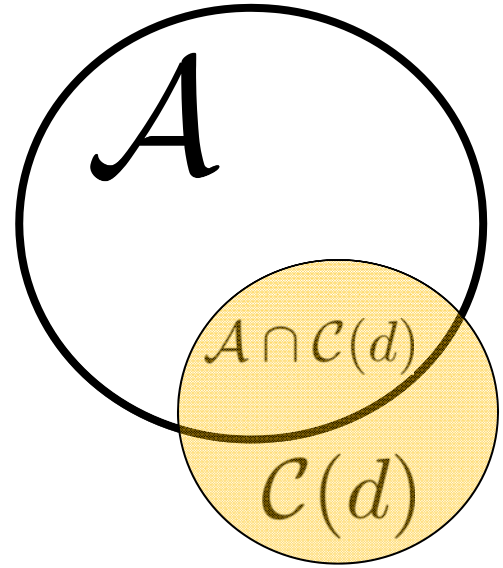
$$\mu \longrightarrow \Phi(\mu) \quad \text{Quantity of Interest}$$

You observe data $d \sim (\mu^\dagger)^n$

What is $\Phi(\mu^\dagger)$?

Find a (confidence) set $\mathcal{C}(d)$ such that with probability $1-\epsilon$

$$\mu^\dagger \in \mathcal{C}(d)$$




With probability $1-\epsilon$

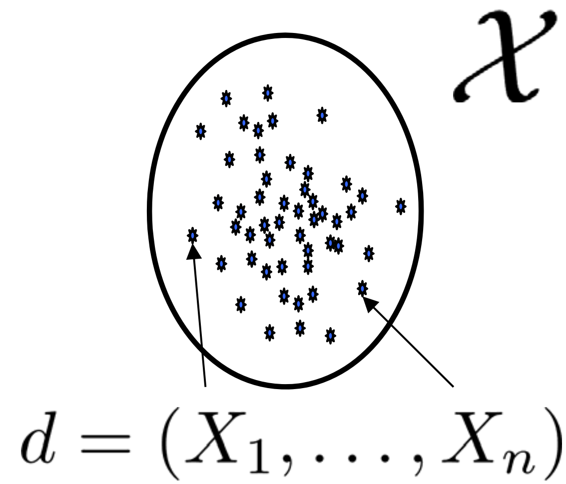
$$\inf_{\mu \in \mathcal{A} \cap \mathcal{C}(d)} \Phi(\mu) \leq \Phi(\mu^\dagger) \leq \sup_{\mu \in \mathcal{A} \cap \mathcal{C}(d)} \Phi(\mu)$$

Notion of worst/sharpest depends on the particular choice of $\mathcal{C}(d)$

**Frequentist/Concentration of measure worst case**

$$\mu \in \mathcal{M}(\mathcal{X})$$

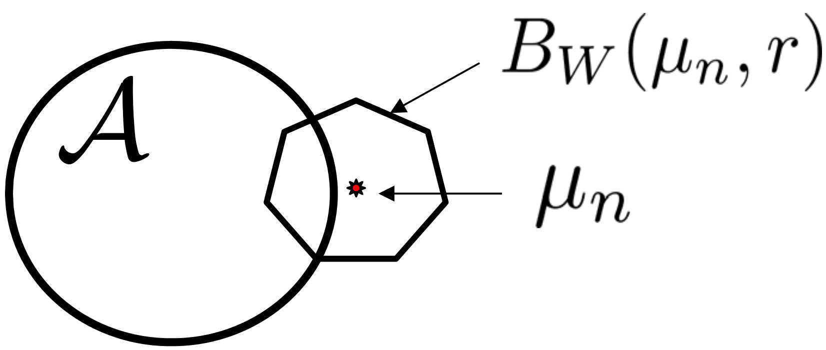
$$\mu_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \in \mathcal{M}(\mathcal{X})$$



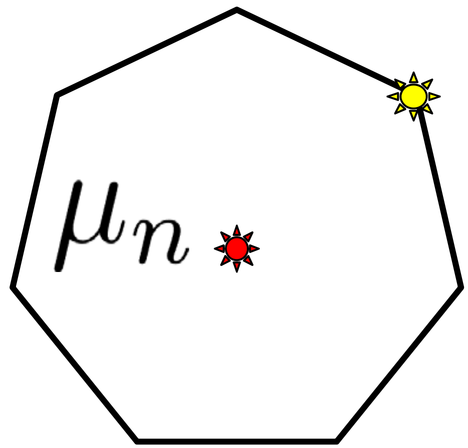

**Theorem** $d_{\text{Wasserstein}}(\mu, \mu_n) = \sup_{f \in \text{Lip}_1}\left(\mathbb{E}_{\mu}[f] - \mathbb{E}_{\mu_n}[f]\right)$ $\mathbb{E}_{X\sim\mu}\left[e^{\|X\|^a}\right] < \infty$

$$\mu^n\left[d_{\text{W}}(\mu, \mu_n) \geq r\right] \leq C\left(e^{-cnr^m}1_{r\leq 1} + e^{-cnr^a}1_{r>1}\right)$$

N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probability Theory and Related Fields, (2014), pp. 1-32.



## Reduction calculus of the ball about the empirical distribution




- D. Wozabal. A framework for optimization under ambiguity. Annals of Operations Research, 193(1):21–47, 2012.
- P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. arXiv:1505.05116, 2015.
- Extreme points of a ball about a measure with finite support (2015). H. Owhadi and Clint Scovel. arXiv:1504.06745

**The extreme points of the Prokhorov, Monge-Wasserstein and Kantorovich metric balls about a measure whose support has at most n points, consist of measures whose supports have at most n+2 points.**

## Question

**Game/Decision Theory + Information Based Complexity**

↓

**Turn the process of discovery of scalable numerical solvers into an algorithm**

**Worst case calculus**

↓

**?**

# The truncated moment problem

$$\mathcal{M}[0,1] \xrightarrow{\quad\Psi\quad} \mathbb{R}^k$$

$$\mu \longrightarrow \Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big)$$

Study of the geometry of $M_k := \Psi\big(\mathcal{M}([0,1])\big)$

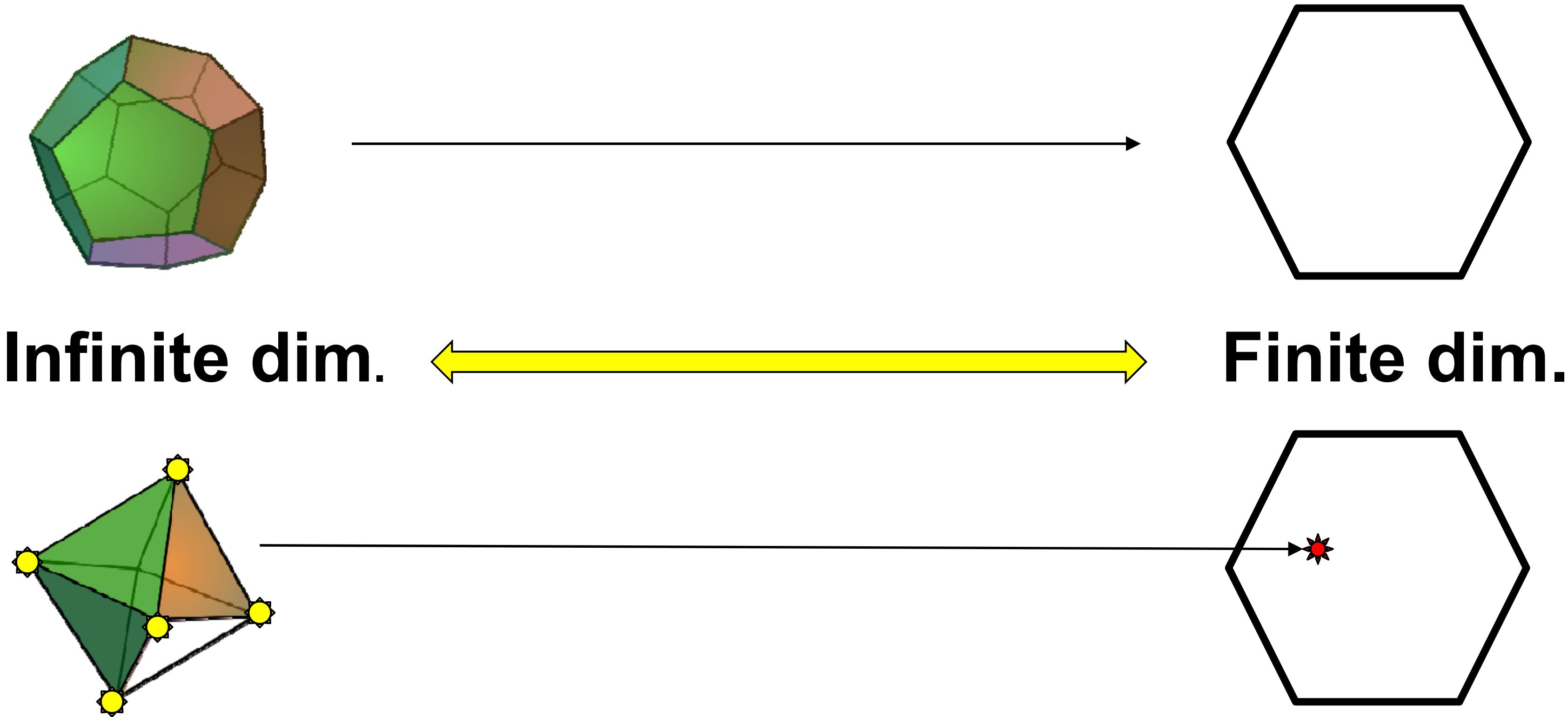
P. L. Chebyshev
1821-1894

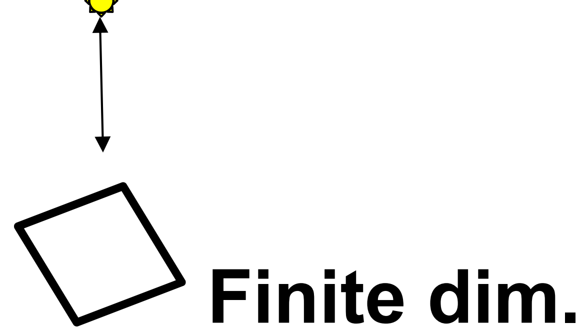
A. A. Markov
1856-1922

M. G. Krein
1907-1989

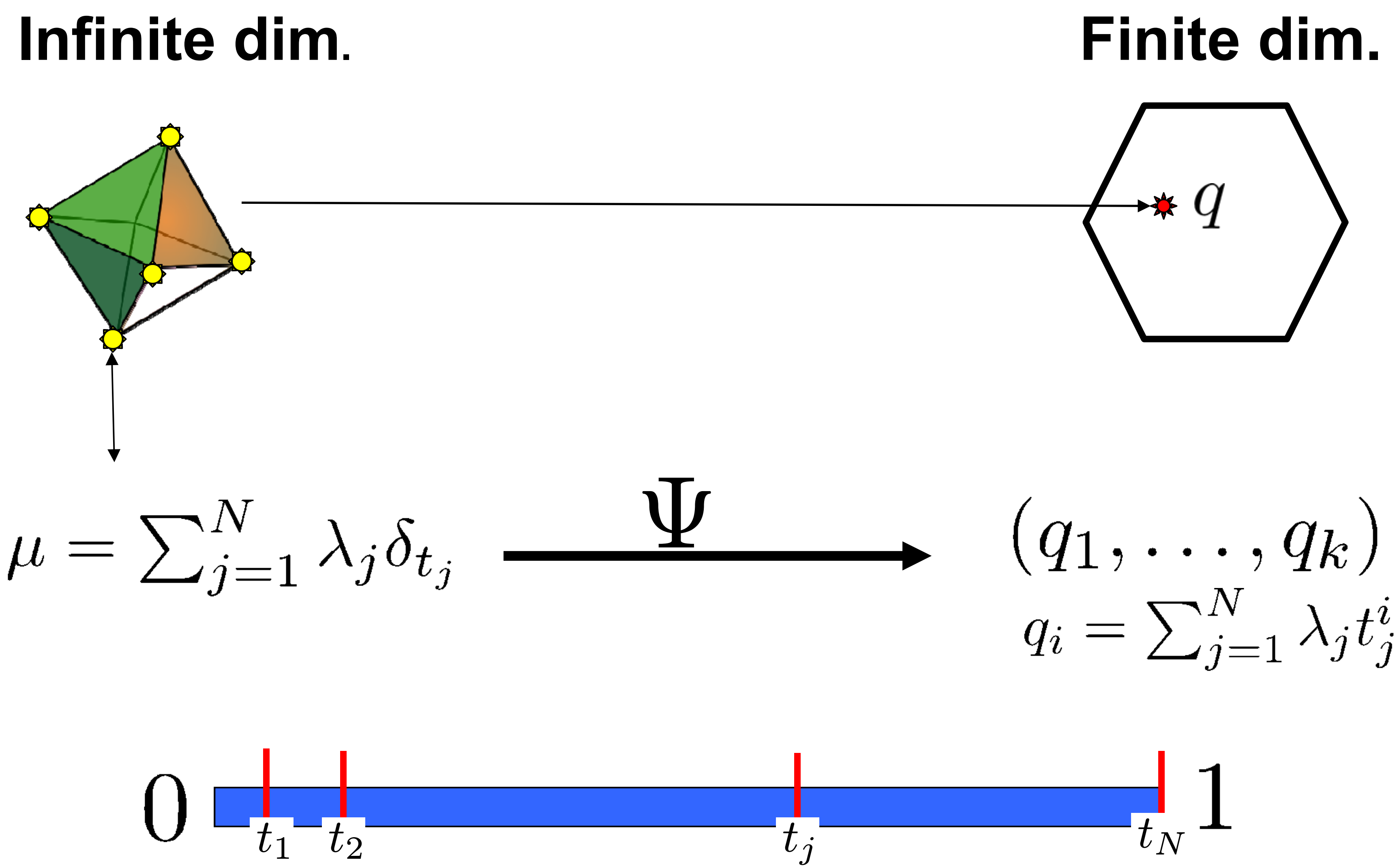
$$\mathcal{M}[0,1] \xrightarrow{\quad \Psi \quad} \mathbb{R}^k \qquad \boxed{M_k := \Psi\big(\mathcal{M}([0,1])\big)}$$

$$\mu \longrightarrow \Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big)$$




Let us compute $\mathrm{Vol}(M_k)$ using different extreme points representations.


Infinite dim.
Finite dim.
q
Ψ
μ = \sum_{j=1}^N λ_j δ_{t_j}
(q_1, . . . , q_k)
q_i = \sum_{j=1}^N λ_j t_j^i
0
1
t_1
t_2
t_j
t_N

$$\mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j}$$

**Index** $i(\mu)$: Number of support points of $\mu$

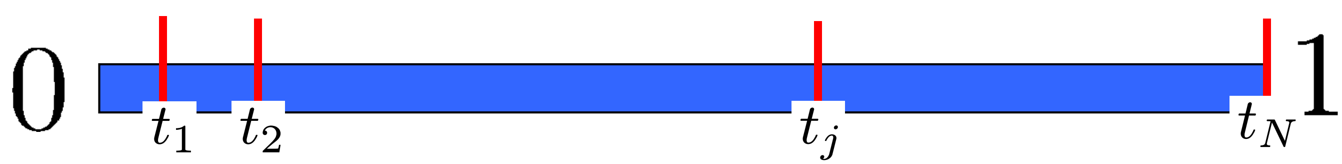
Counting interior points with weight 1 and boundary points with weight $\frac{1}{2}$

$\mu$ is called
- principal if $i(\mu) = \frac{k+1}{2}$
- canonical if $i(\mu) = \frac{k+2}{2}$
- upper if support points include 1
- lower if support points do not include 1

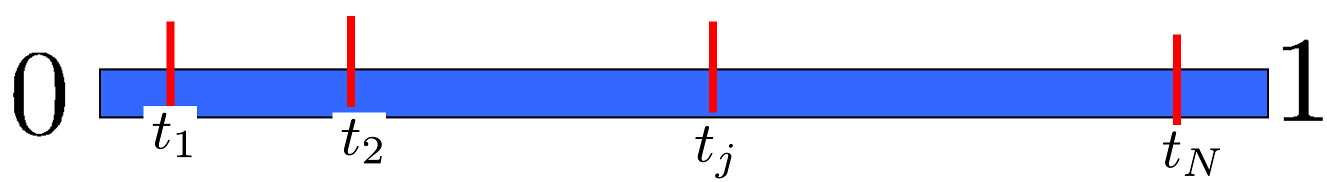
**Theorem**

Every point $q \in \mathrm{Int}(M_k)$ has a unique upper and lower principal representation.

**Upper**




**Lower**



Vol($M_{2m-1}$) using Upper Rep. = Vol($M_{2m-1}$) using Lower Rep.

$$\frac{1}{(m-1)!} S_{m-1}(3,3,2) = \frac{1}{m!} S_m(1,1,2)$$

Vol($M_{2m}$) using Upper Rep. = Vol($M_{2m}$) using Lower Rep.

$$S_m(1,3,2) = S_m(3,1,2)$$

## Selberg Identities

$$S_n(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}$$

$$S_n(\alpha,\beta,\gamma) := \int_{[0,1]^n} \prod_{j=1}^{n} t_j^{\alpha-1}(1-t_j)^{\beta-1}|\Delta(t)|^{2\gamma} dt\,.$$

$$\Delta(t) := \prod_{j<k} (t_k - t_j)$$

Brittleness of Bayesian inference and new Selberg formulas. H. Owhadi and C. Scovel. Communications in Mathematical Sciences (2015). arXiv:1304.7046

## Forrester and Warnaar 2008

## The importance of the Selberg integral

**Used to prove outstanding conjectures in Random matrix theory and cases of the Macdonald conjectures**

**Central role in random matrix theory, Calogero-Sutherland quantum many-body systems, Knizhnik-Zamolodchikov equations, and multivariable orthogonal polynomial theory**

$$\mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j}$$

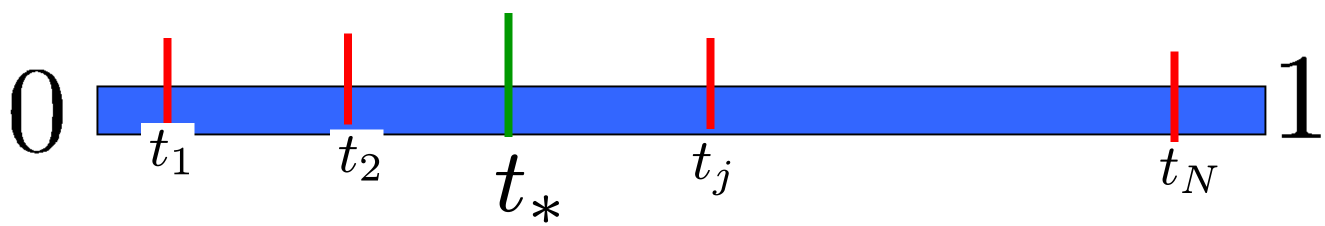
**Index** $i(\mu)$: Number of support points of $\mu$

Counting interior points with weight 1 and boundary points with weight $\frac{1}{2}$

$\mu$ is called
- principal if $i(\mu) = \frac{k+1}{2}$
- canonical if $i(\mu) = \frac{k+2}{2}$
- upper if support points include 1
- lower if support points do not include 1

**Theorem**

For $t_* \in (0,1)$, every point $q \in \mathrm{Int}(M_k)$ has a unique canonical representation whose support contains $t_*$. When $t_* = 0$ or $1$, there exists a unique principal representation whose support contains $t_*$.



**New Reproducing Kernel Hilbert Spaces and Selberg Integral formulas related to the Markov-Krein representations of moment spaces.**

$$\mathcal{M}[0,1] \xrightarrow{\quad \Psi \quad} [0,1]^k$$

$$\mu \longrightarrow \left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\right)$$

$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2 (1-t_j)^2 \Delta_m^4(t) dt = \frac{S_m(5,1,2) - S_m(3,3,2)}{2}$$

$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2 \cdot \Delta_m^4(t) dt = \frac{m}{2} S_{m-1}(5,3,2)$$

$$\Delta_m(t) := \prod_{j<k} (t_k - t_j) \quad I := [0,1]$$

$$(\Sigma\phi)(t) := \sum_{j=1}^m \phi(t_j), \quad t \in I^m$$

$$S_n(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}$$

$$e_j(t) := \sum_{i_1 < \cdots < i_j} t_{i_1} \cdots t_{i_j}$$

$\Pi_0^n$: $n$-th degree polynomials which vanish on the boundary of $[0,1]$

$M_n \subset \mathbb{R}^n$: set of $q = (q_1, \ldots, q_n) \in \mathbb{R}^n$ such that there exists a probability measure $\mu$ on $[0,1]$ with $\mathbb{E}_\mu[X^i] = q_i$ with $i \in \{1, \ldots, n\}$.

## Theorem

**Bi-orthogonal systems of Selberg Integral formulas**

Consider the basis of $\Pi_0^{2m-1}$ consisting of the associated Legendre polynomials $Q_j, j = 2, .., 2m-1$ of order 2 translated to the unit interval $I$. For $k = 2, .., 2m-1$ define

$$a_{jk} := \frac{(j+k+k^2)\Gamma(j+2)\Gamma(j)}{\Gamma(j+k+2)\Gamma(j-k+1)}, \quad k \le j \le 2m-1$$

$$\tilde{h}_k(t) := \sum_{j=k}^{2m-1} (-1)^{j+1} a_{jk} e_{2m-1-j}(t,t)\,.$$

Then for $j = k \bmod 2$, $j,k = 2, .., 2m-1$, we have

$$\int_{I^{m-1}} \tilde{h}_k(t) \Sigma Q_j(t) \prod_{j'=1}^{m-1} t_{j'}^2 \cdot \Delta_{m-1}^4(t) dt = Vol(M_{2m-1})(2m-1)!(m-1)! \frac{(k+2)!}{(8k+4)(k-2)!} \delta_{jk}\,.$$

# Collaborators

Clint Scovel (Caltech), Tim Sullivan (Warwick), Mike McKerns (Caltech), Michael Ortiz (Caltech), Lei Zhang (Jiaotong), Leonid Berlyand (PSU), Lan Huong Nguyen, Paul Herve Tamogoue Kamga.


# Research supported by

Air Force Office of Scientific Research

DARPA EQUiPS Program (Enabling Quantification of Uncertainty in Physical Systems)

U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, through the Exascale Co-Design Center for Materials in Extreme Environments

National Nuclear Security Administration