Principles and Methods of UQ
A minitutorial, Part I

Houman Owhadi

IPAM Sep 13-14, 2017

DARPA EQUiPS / AFOSR award no FA9550-16-1-0054
(Computational Information Games)

Syllabus

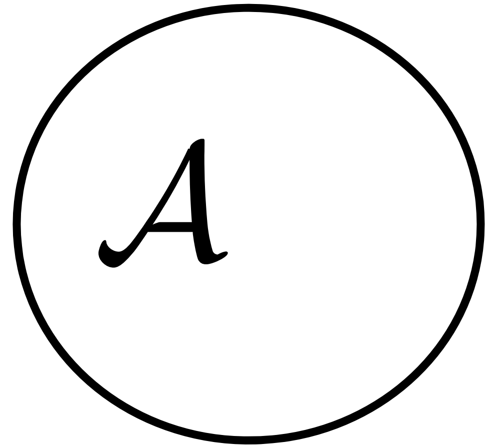
- **Principles and method of UQ**

- **Applications to computation with partial information**
  - **Complex energy landscapes (can be observed only at a finite number of points)**
  - **Coarse graining, multiresolution analysis/modeling**
  - **Fast eigensubspace projections**
  - **Wannier functions**
  - **Scalable numerical methods**

3 main approaches to UQ

The worst case approach

``When in doubt, assume the worst!"



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

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

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

$$\text{What is } \Phi(u^\dagger)?$$

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

Robust Optimization/ Min and Max approach

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.

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.

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

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

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.

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

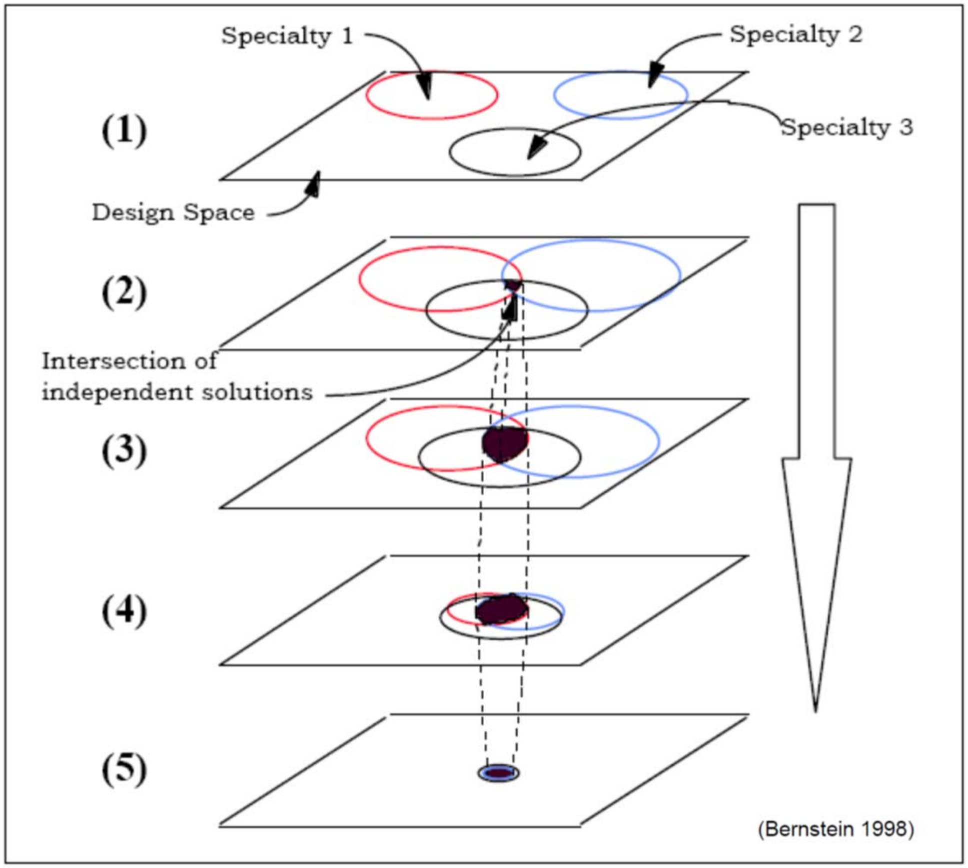
Value at Risk

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.

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.



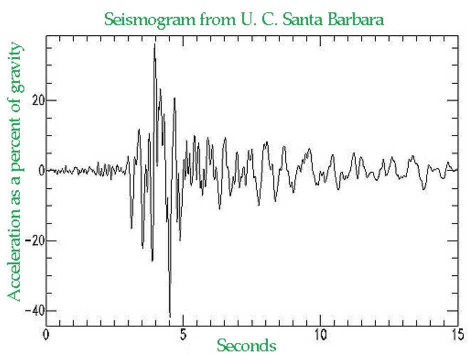
Bernstein JI (1998) *Design methods in the aerospace industry: looking for evidence of* ***set-based practices***. Master's thesis. Massachusetts Institute of Technology, Cambridge, MA.

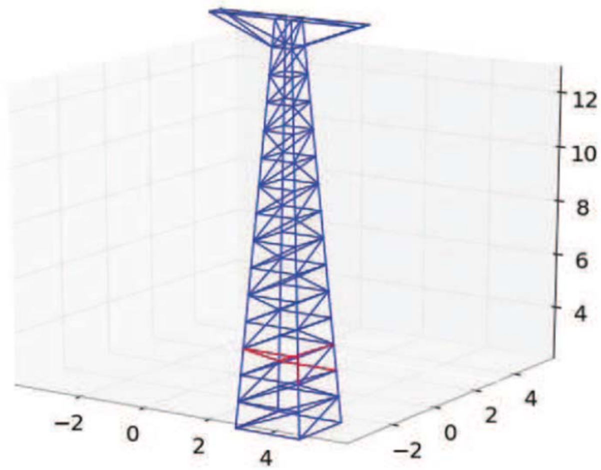
Set based design/analysis

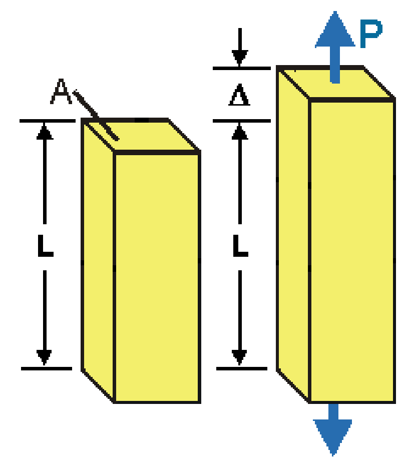
David J. Singer, PhD., Captain Norbert Doerry, PhD., and Michael E. Buckley," *What is* ***Set-Based Design***? ," Presented at ASNE DAY 2009, National Harbor, MD., April 8-9, 2009. Also published in ASNE Naval Engineers Journal, 2009 Vol 121 No 4, pp. 31-43.

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

Seismic Safety Assessment of a Truss Structure







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

Ground Acceleration

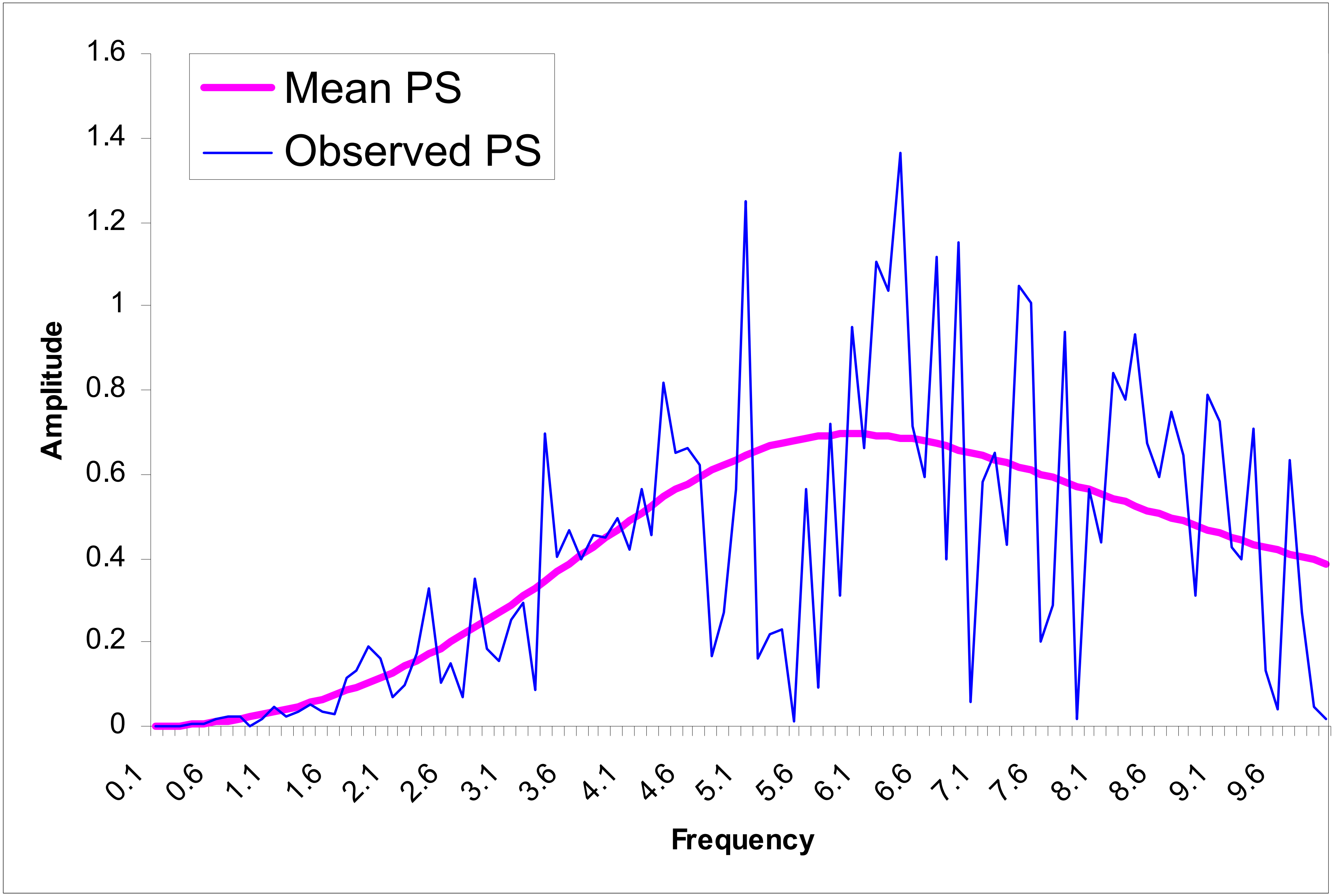
F

min(Yield Strain - Axial Strain)

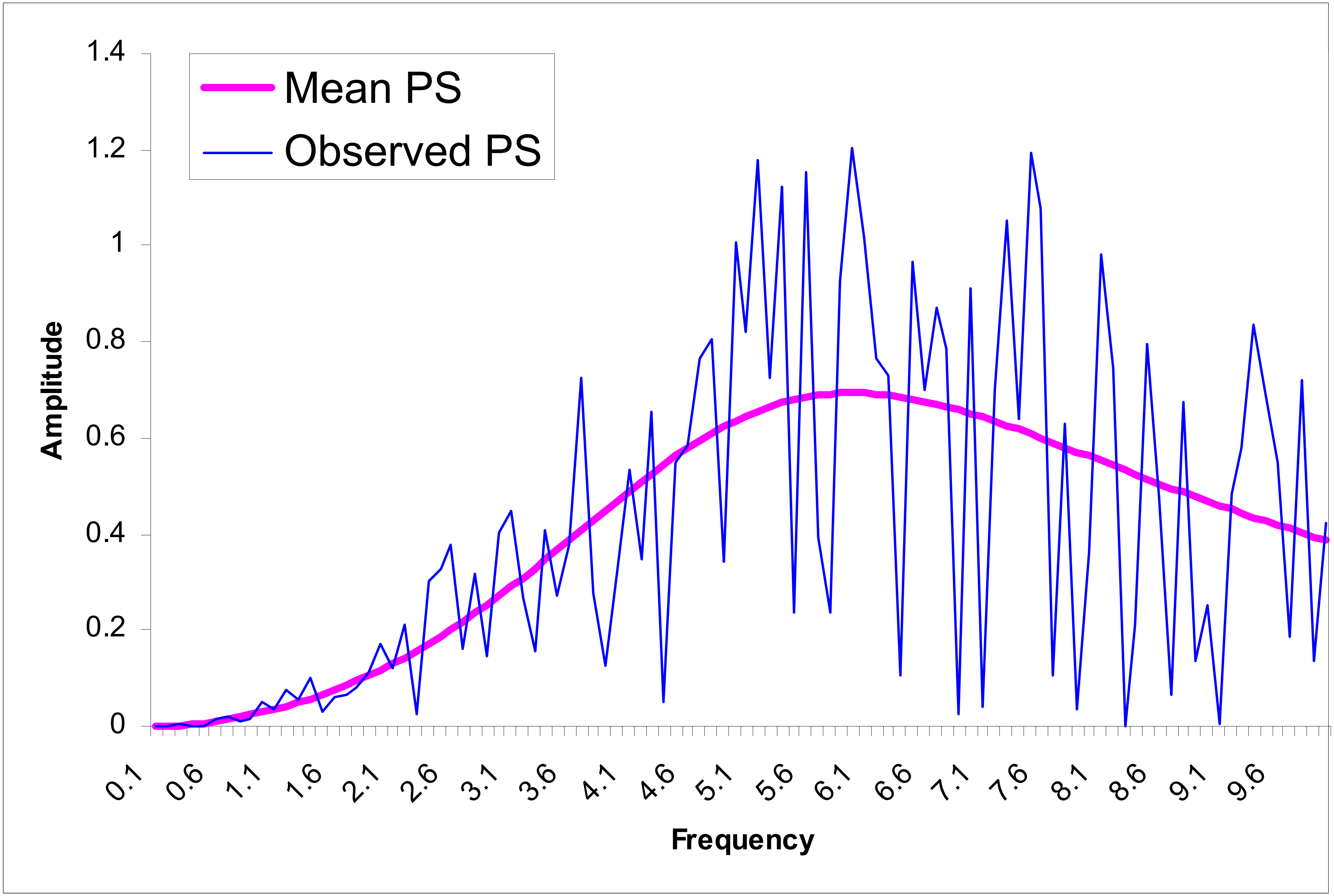
We want to certify that

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

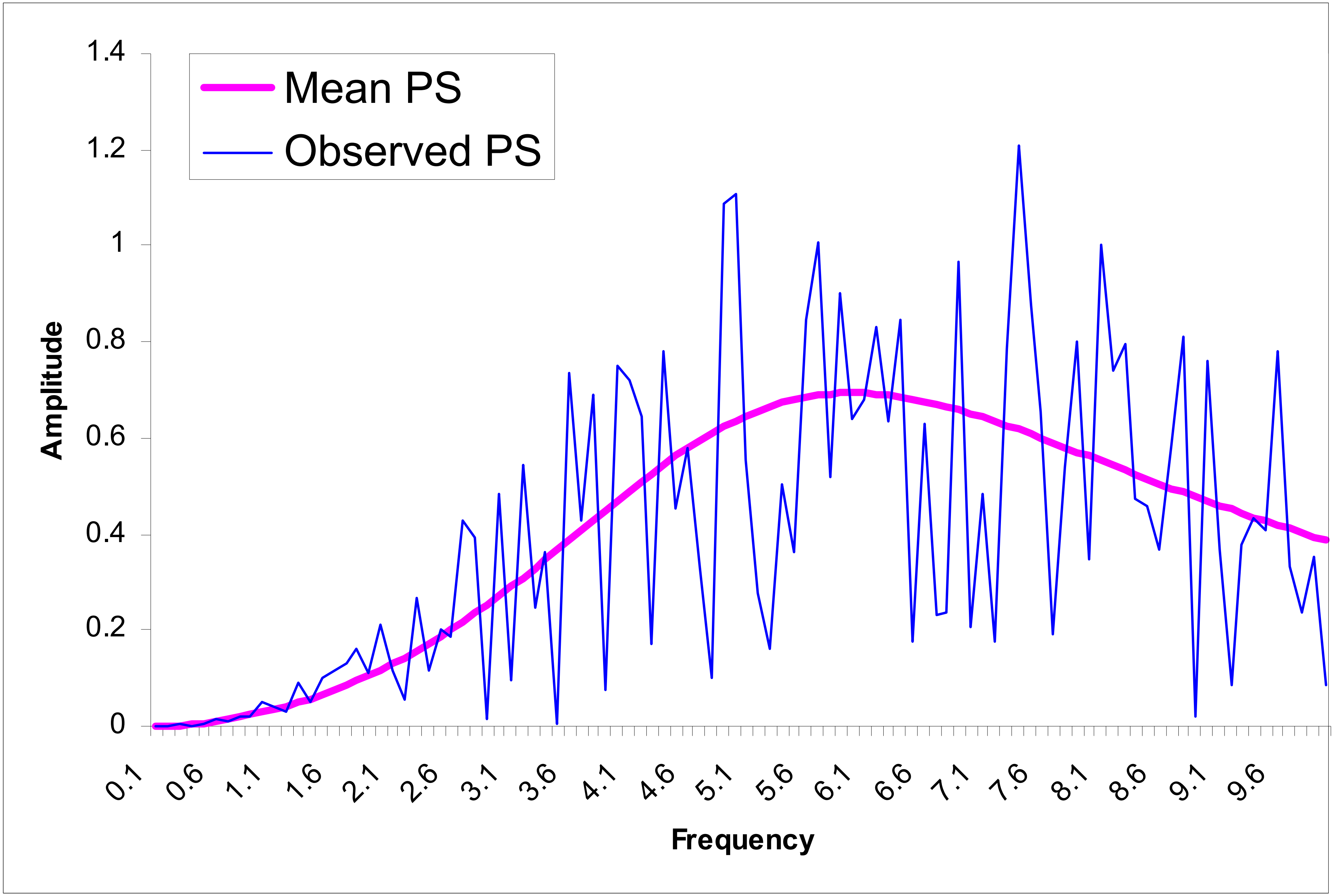
Power Spectrum



Power Spectrum



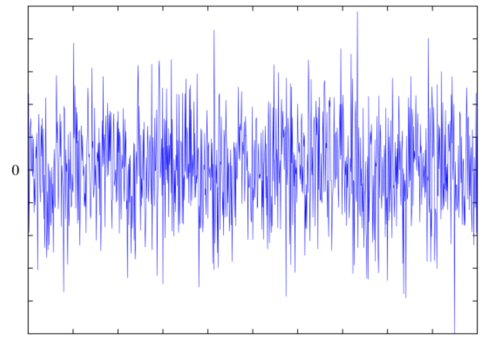
Power Spectrum



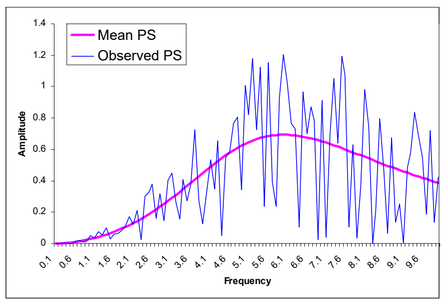
Filtered White Noise Model

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.

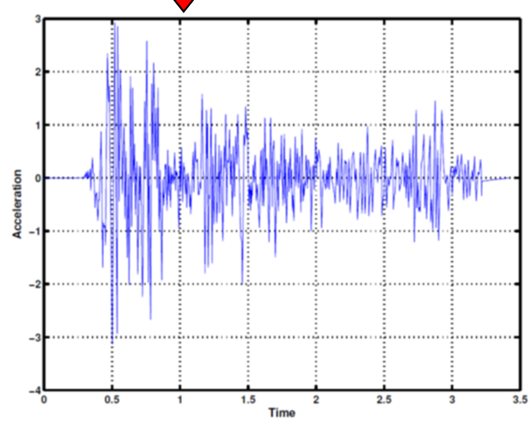
White noise

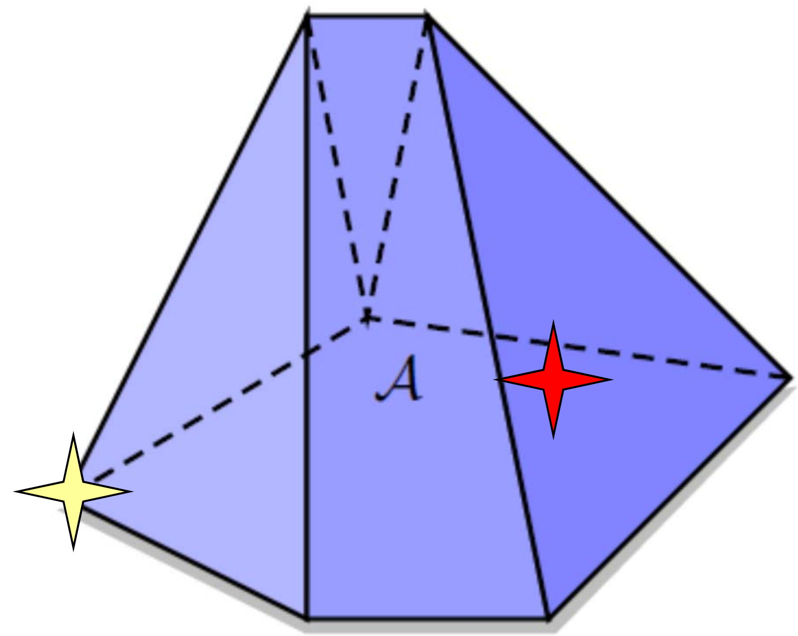


Filter

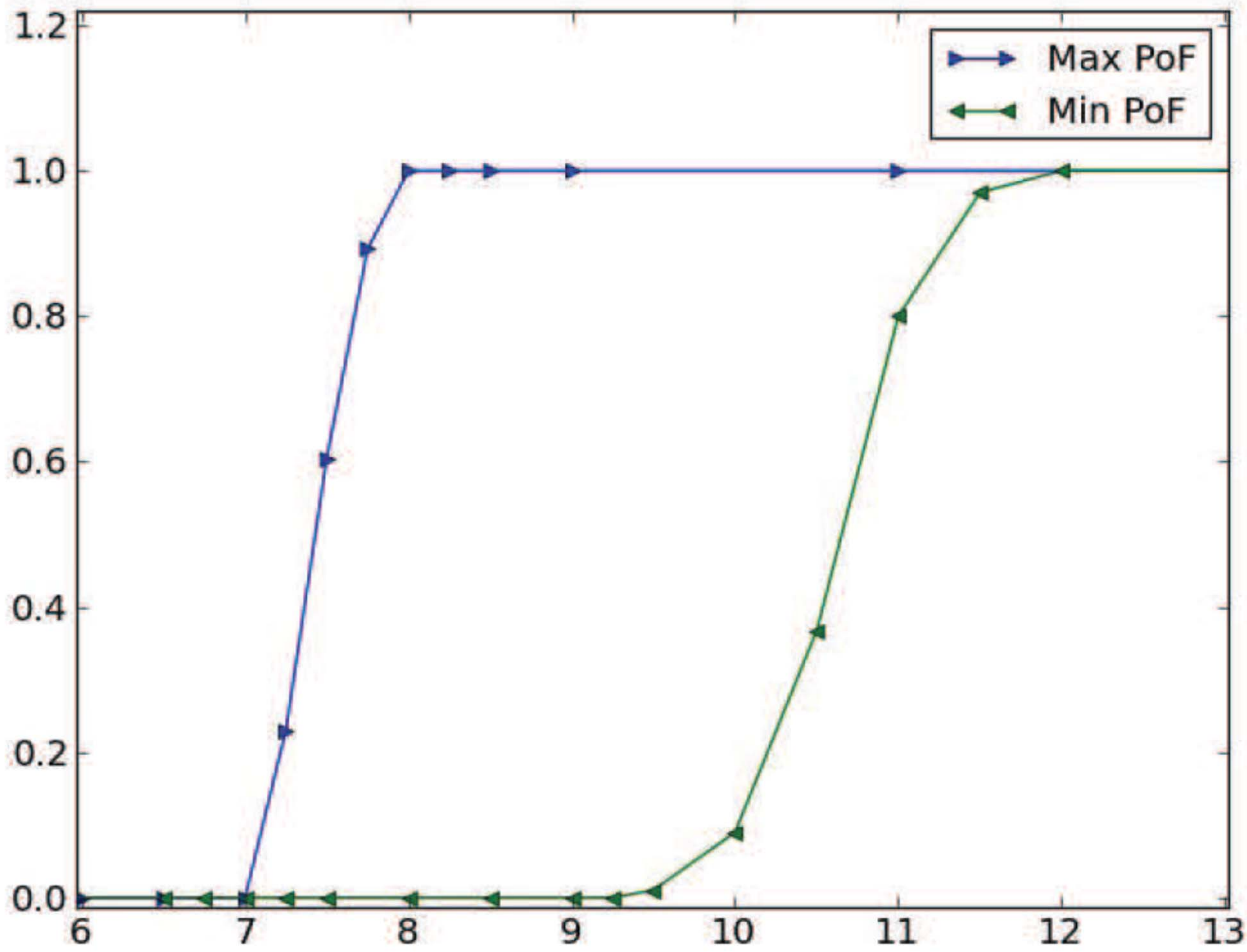


Ground acceleration

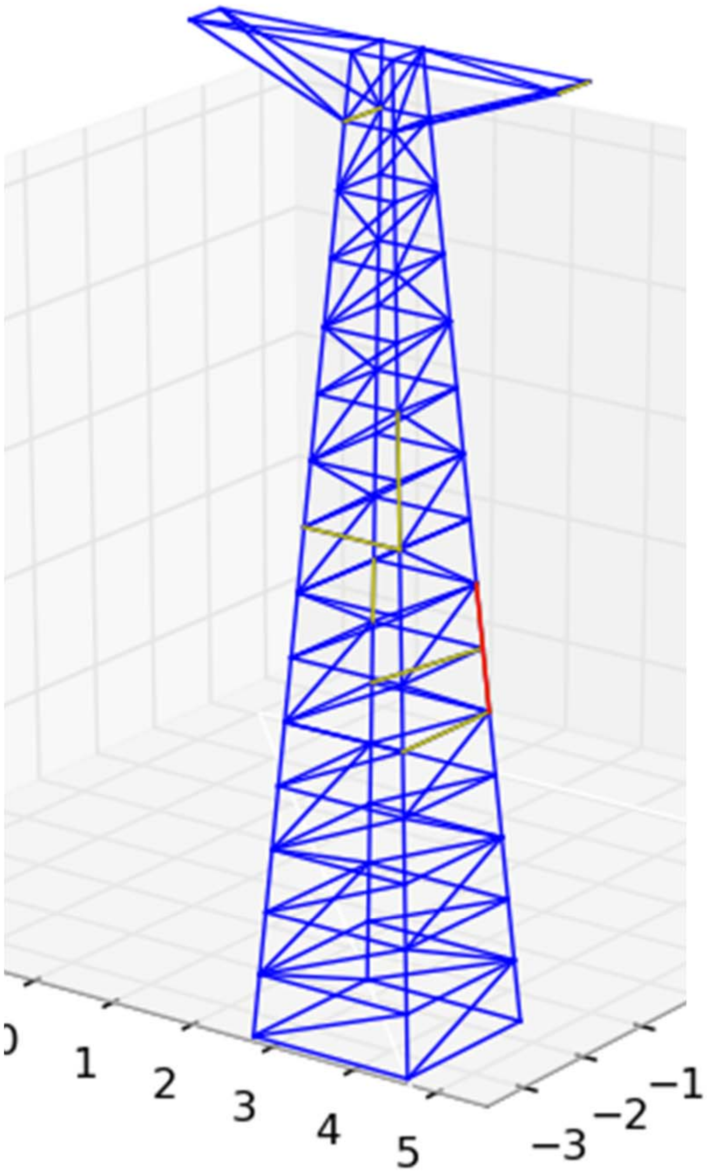


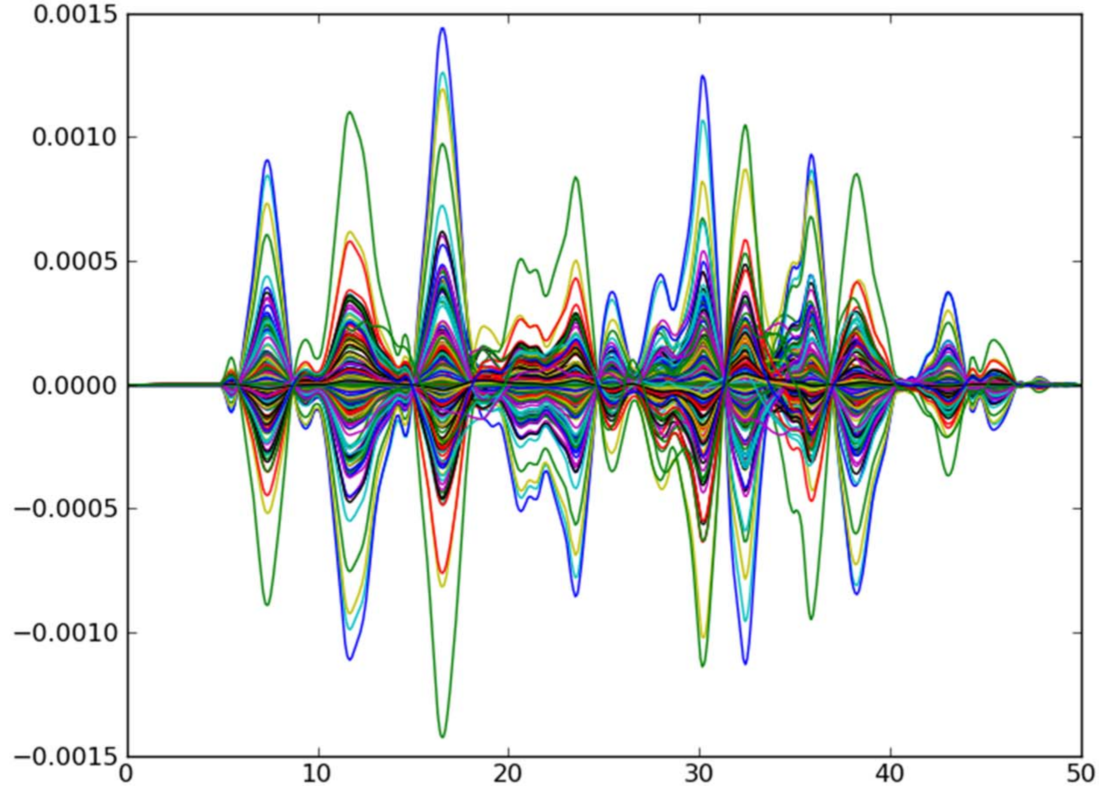


Vulnerability Curves (vs earthquake magnitude)



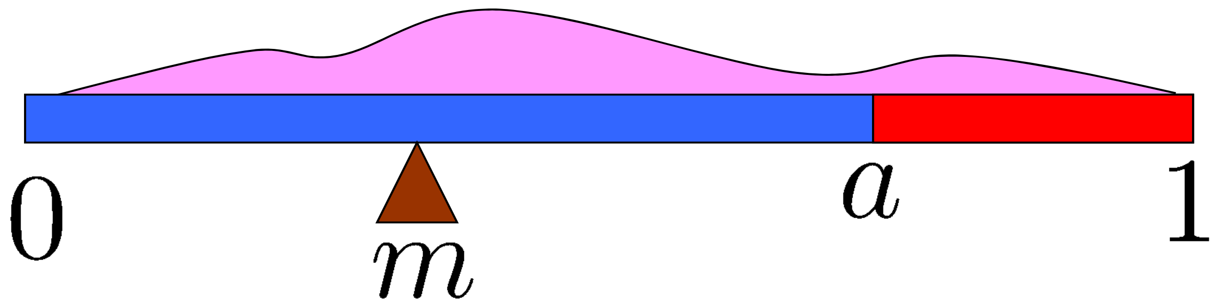
Identification of the weakest elements





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

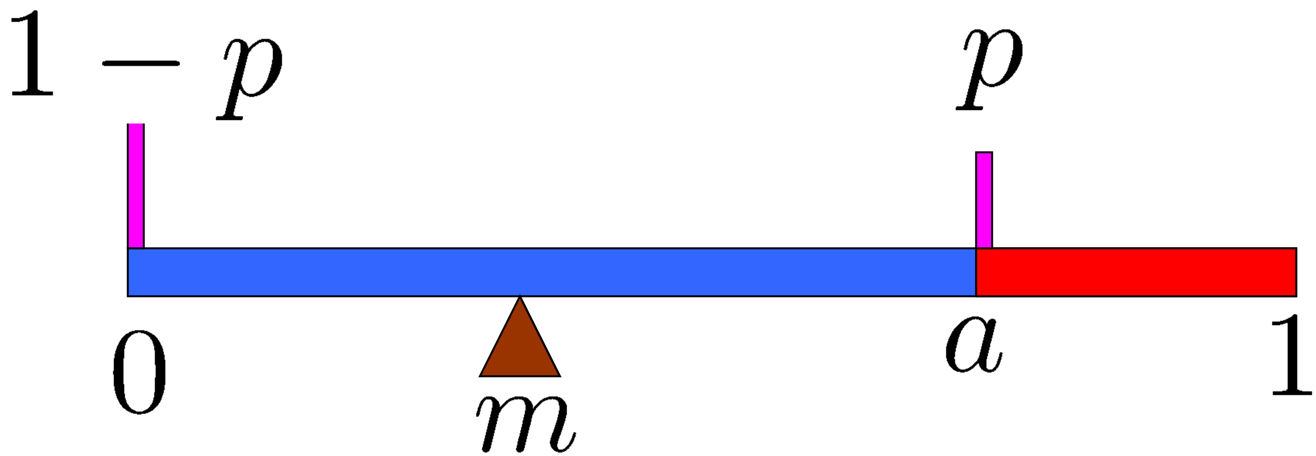
You are given one pound of playdoh,
how much mass can you put above a
while keeping the seesaw balanced around m?



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

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

Bayesian Approach

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

$$f(x) = \exp\left(\cosh\left(\frac{x^2 + \sin(x)}{3 + \cos(x^3)} \right) \right)$$

Compute

$$\int_0^1 f(x)\,dx$$

P. Diaconis

Numerical Analysis Approach

Find a good quadrature rule for the numerical integration of f

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

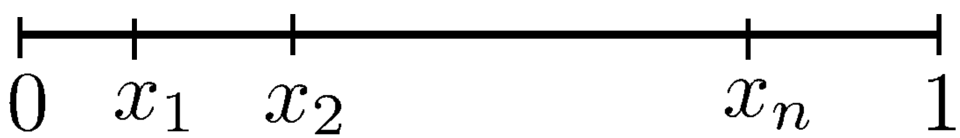
$$f(x) = \exp\left(\cosh\left(\frac{x^2 + \sin(x)}{3 + \cos(x^3)} \right) \right)$$

Compute

$$\int_0^1 f(x)\, dx$$

Bayesian Approach

- Put a prior in $\mathcal{C}([0,1])$
- Calculate f at $x_1, \ldots, x_n$



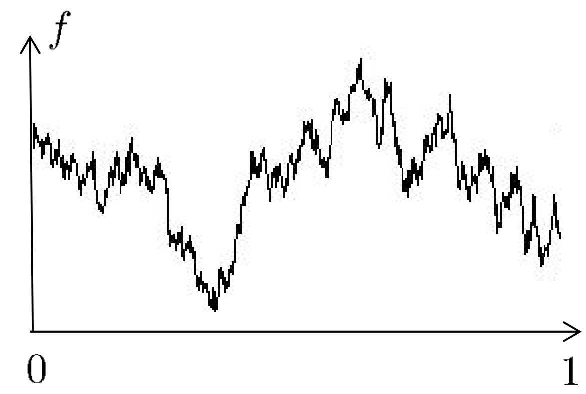
- Compute

$$\mathbb{E}\left[\int_0^1 f(x)\, dx \,\middle|\, f(x_1), \ldots, f(x_n) \right]$$

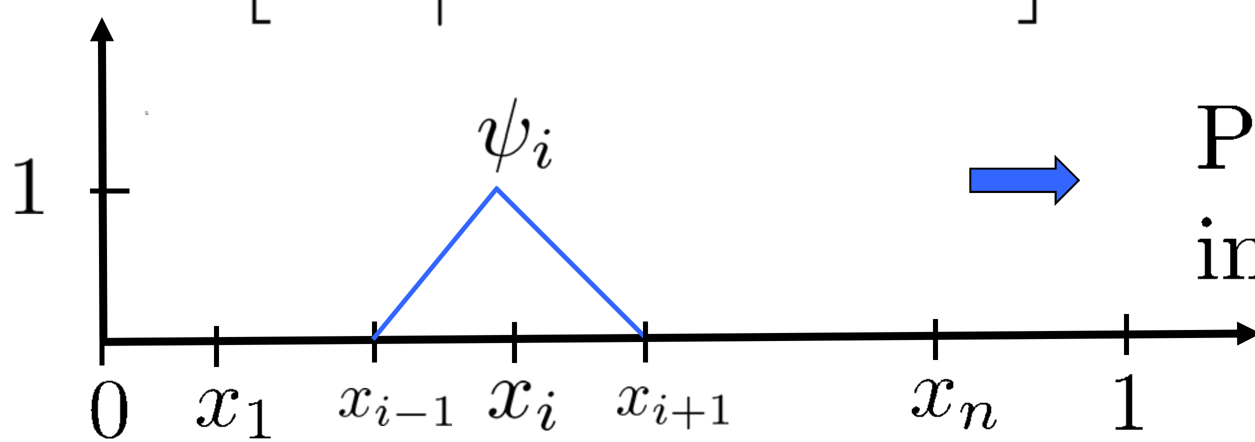
E.g.

Assume $f(t) = \xi + B_t$

$\xi \sim \mathcal{N}(0,1)$, B_t: B.M.



$$\mathbb{E}\Big[f(x)\Big|f(x_1),\dots,f(x_n)\Big] = \textstyle\sum_i f(x_i)\psi_i(x)$$



→ Piecewise linear interpolation of f

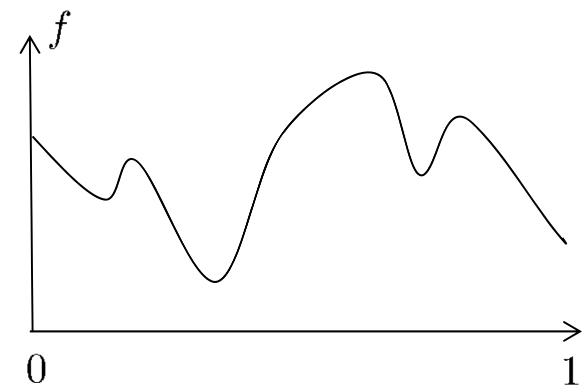
$$\mathbb{E}\Big[\int_0^1 f(x)\,dx\Big|f(x_1),\dots,f(x_n)\Big]$$

→ Trapezoidal quadrature rule

E.g.

Assume $f(t) = \xi + \int_0^t B_s \, ds$

$\xi \sim \mathcal{N}(0,1)$, B_s: B.M.



$\mathbb{E}\left[f(x) \middle| f(x_1), \ldots, f(x_n)\right]$ ➡ Cubic spline interpolant

E.g.

Integrate B.M. k times ➡ Splines of order $2k+1$

O'Hagan (1991). Bayes-Hermite quadrature

Q Similar link between coarse graining and Bayesian Inference?

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

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

a unif. ell. $a_{i,j} \in L^\infty(\Omega)$

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

Approximate the solution space of (1) with a finite dimensional space

Numerical Homogenization Approach

Work hard to find good basis functions

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] [Chung-Efendiev-Hou, JCP 2016]

Variational Multiscale Method, Orthogonal decomposition

[Hughes, Feijóo, Mazzei, Quincy. 1998]

[Malqvist-Peterseim 2012] Local Orthogonal Decomposition

Projection based method Nolen, Papanicolaou, Pironneau, 2008

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

Flux norm Berlyand, Owhadi 2010; Symes 2012

Harmonic continuation [Babuska-Lipton 2010]

Bayesian Approach

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

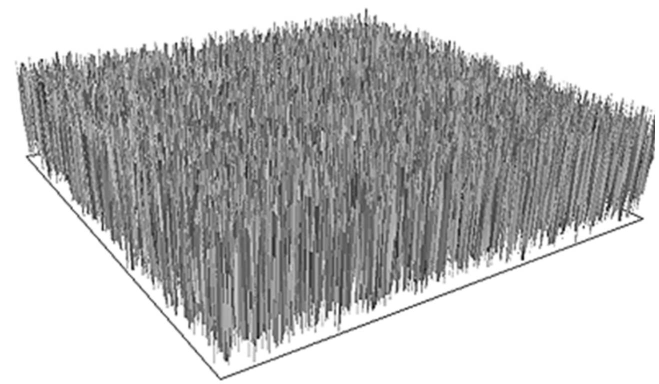
Proposition

- Put a prior on g
- Compute $\mathbb{E}\left[u(x)\middle|\text{finite no of observations}\right]$

Bayesian approach

Replace g by ξ

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



ξ: White noise

Gaussian field with covariance function $\Lambda(x,y) = \delta(x-y)$

$\Leftrightarrow \quad \forall f \in L^2(\Omega),\ \int_\Omega f(x)\xi(x)\,dx$ is $\mathcal{N}\big(0, \|f\|^2_{L^2(\Omega)}\big)$

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

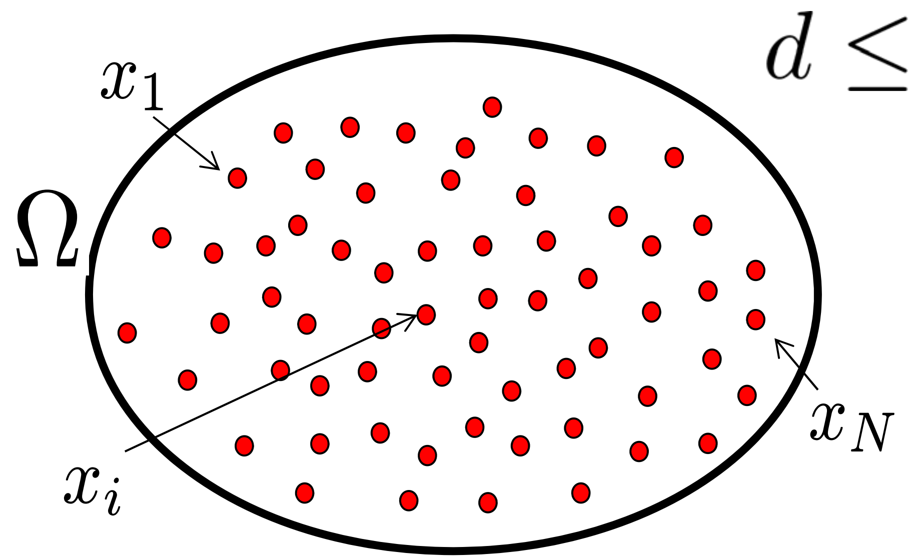
[J. Cockayne, C. J. Oates, T. Sullivan, and M. A. Girolami. Probabilistic meshless methods for partial differential equations and bayesian inverse problems. arXiv:1605.07811, 2016]

[M. Raissi, P. Perdikaris, and G. E. Karniadakis. Inferring solutions of differential equations using noisy multidelity data. JCP, 2017.]

$d \leq 3$

Let

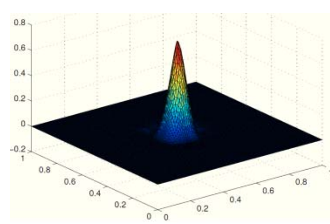
$x_1, \ldots, x_N \in \Omega$



Theorem

$$\mathbb{E}\Big[v(x)\Big|v(x_1),\ldots,v(x_N)\Big] = \sum_{i=1}^{N} v(x_i)\psi_i(x)$$

$a = I_d$ ⟷ ψ_i: Polyharmonic splines



[Harder-Desmarais, 1972]

[Duchon 1976, 1977,1978]

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

[Owhadi-Zhang-Berlyand 2013]

Standard deviation of the statistical error bounds/controls the worst case error

$$\big(v(x)\big|v(x_1),\ldots,v(x_N)\big) \sim \mathcal{N}\Big(\sum_{i=1}^N v(x_i)\psi_i(x), \sigma^2(x)\Big)$$

$\sigma^2(x)$: Kriging function (geostatistics)

D. E. Myers. Kriging, co-Kriging, radial basis functions and the role of positive definiteness. *Comput. Math. Appl.*, 24(12):139–148, 1992. Advances in the theory and applications of radial basis functions.

$\sigma^2(x)$: Power function (radial basis function interpolation)

G. E. Fasshauer. Meshfree methods. In *Handbook of Theoretical and Computational Nanotechnology*. American Scientific Publishers, 2005.

Holger Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

Z. Min Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.*, 13(1):13–27, 1993.

Theorem

$$\Big|u(x) - \sum_{i=1}^N u(x_i)\psi_i(x)\Big| \leq \sigma(x)\|g\|_{L^2(\Omega)}$$

Bayesian/probabilistic approach to numerical approximation not new but appears to have remained overlooked

Pioneering work

[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]

" These concepts and techniques have attracted little attention among numerical analysts" (Larkin, 1972)

Bayesian Numerical Analysis

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

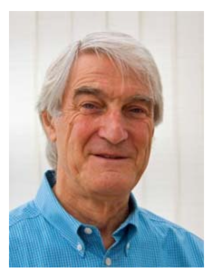
[J. E. H. Shaw. A quasirandom approach to integration in Bayesian statistics. Ann. Statist, 1988.]

[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.]

P. Diaconis



A. O' Hagan

J. E. H. Shaw

Information based complexity

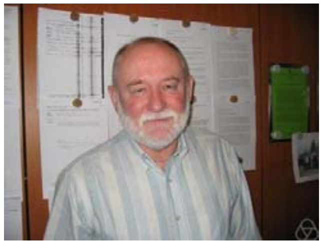
[H. Woźniakowski. Probabilistic setting of information-based complexity. J. Complexity, 1986.]

[G. W. Wasilkowski and H. Woźniakowski. Average case optimal algorithms in Hilbert spaces. Journal of Approximation Theory, 47(1):1725, 1986.]

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

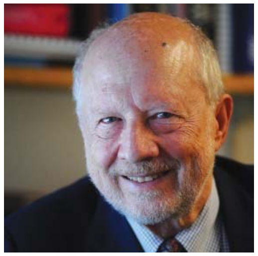
[J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. Information-based complexity. 1988]

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



H. Wozniakowski

G. W. Wasilkowski



J. F. Traub

E. Novak

Probabilistic Numerical Methods

⇕

Statistical Inference approaches to numerical approximation and algorithm design

[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. R. Conrad, M. Girolami, S. Srkk, A. Stuart, and K. Zygalakis. Probability measures for numerical solutions of differential equations. 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.]

[Owhadi 2017, Multi-grid with rough coefficients and Multiresolution PDE decomposition from Hierarchical Information Games, arXiv:1503.03467, SIREV]

[J. Cockayne, C. J. Oates, T. Sullivan, and M. A. Girolami. Probabilistic meshless methods for partial differential equations and bayesian inverse problems. arXiv:1605.07811, 2016]

[I. Bilionis. Probabilistic solvers for partial differential equations. arXiv:1607.03526, 2016]

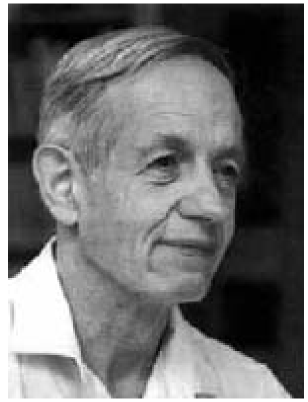
[Jon Cockayne, Chris Oates, Tim Sullivan, Mark Girolami. Bayesian Probabilistic Numerical Methods. arXiv:1702.03673, 2017]

[M. Raissi, P. Perdikaris, and G. E. Karniadakis. Inferring solutions of differential equations using noisy multidelity data. JCP, 2017.]

[H. Owhadi and C. Scovel. Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis 2017. arXiv:1703.10761]

The Game theoretic approach

John Von Neumann



John Nash

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.

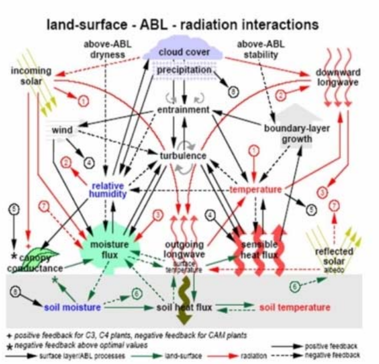
N. Nash. Non-cooperative games. *Ann. of Math.*, 54(2), 1951.

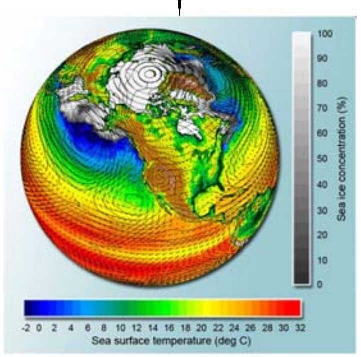
A. Wald. Statistical decision functions which minimize the maximum risk. *Ann. of Math. (2)*, 46:265–280, 1945.

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.

Find the best climate model and incorporate data





Find a 95% interval of confidence on average global temperatures in 50 years

Problem

- Incomplete information on underlying processes
- Limited computation capability
- You don't know $\mathbb{P}$
- You have limited data
- How do you use the data in an optimal way?

Framework: Game/Decision Theory

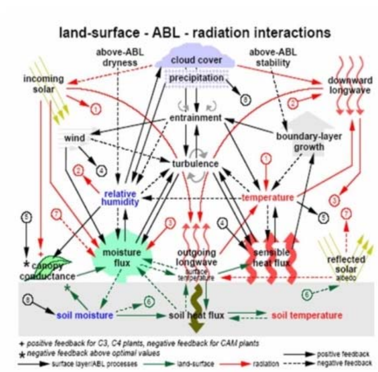
Player I

Chooses candidate

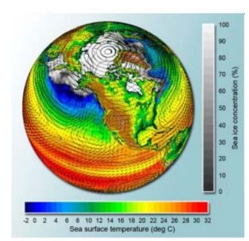
Player II

Sees data

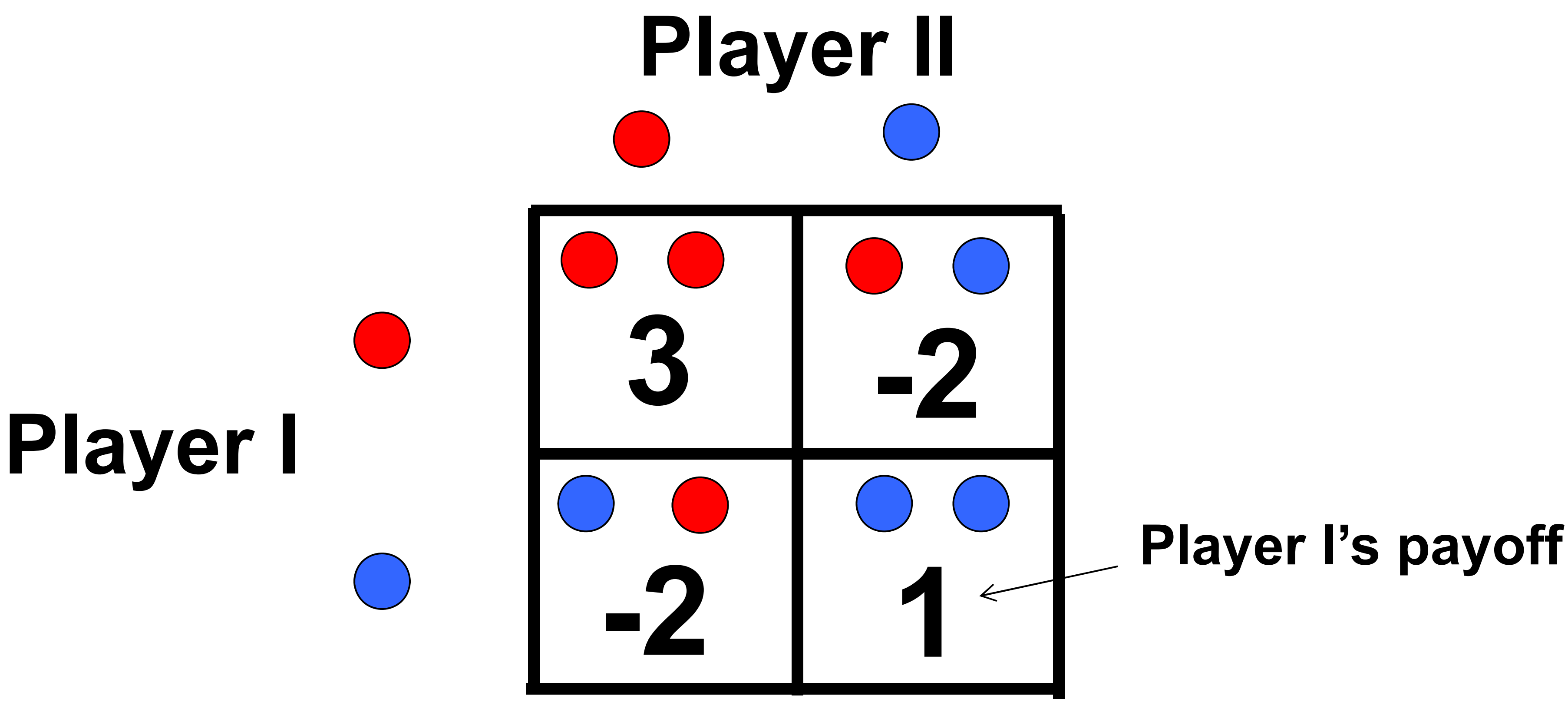
Chooses model



$$\mathcal{E}(\text{candidate}, \text{model}(\text{data}))$$

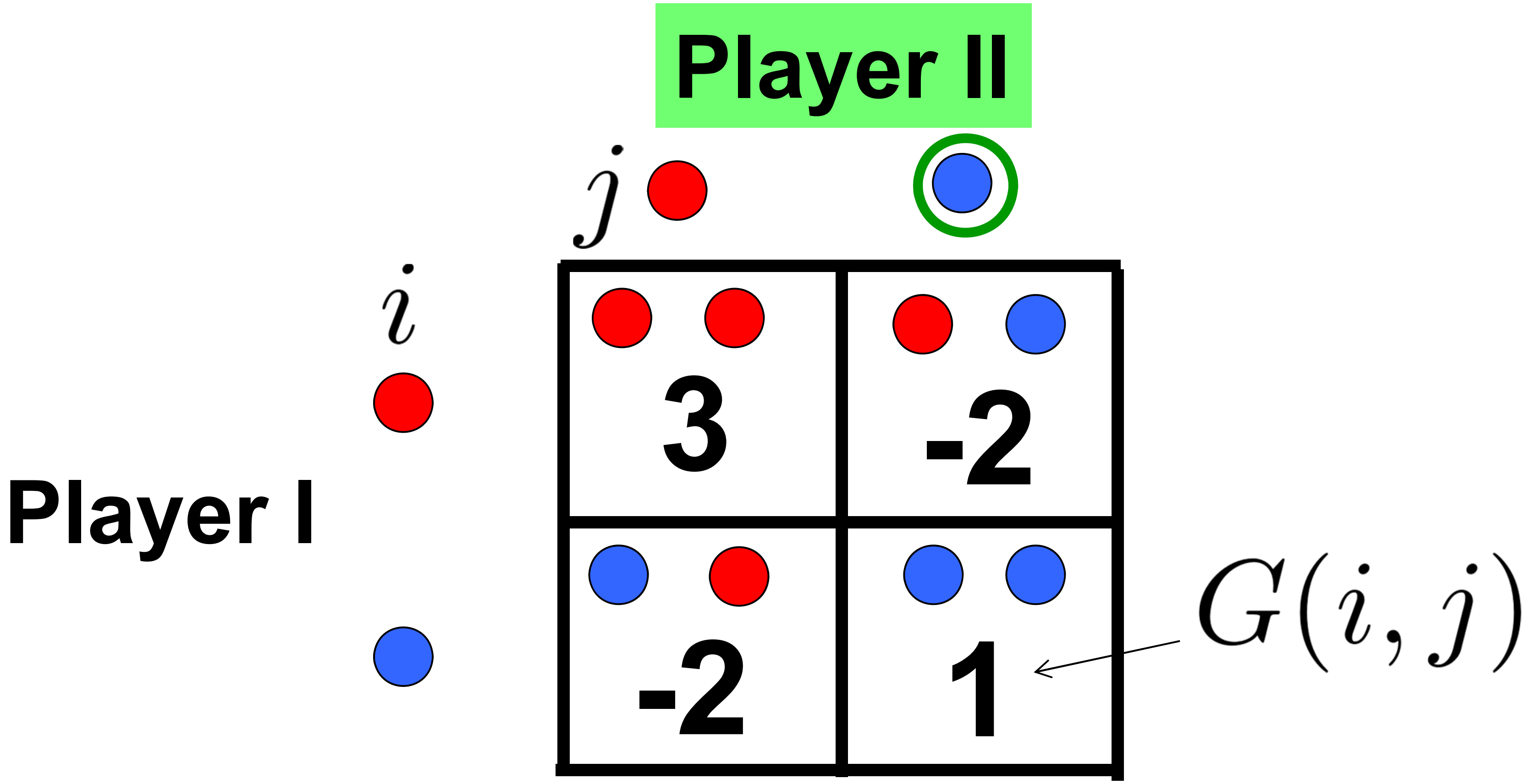


Deterministic zero sum game



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

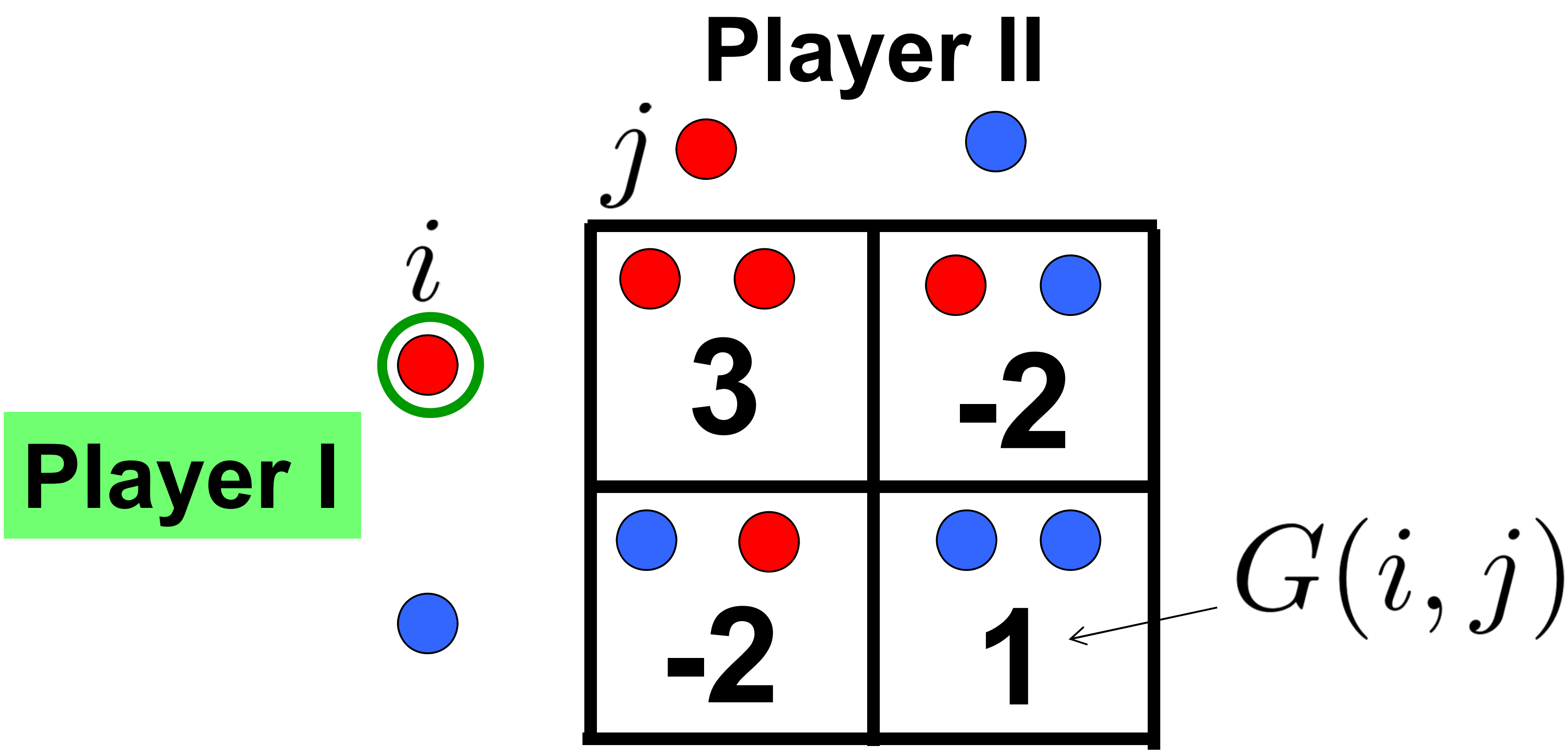
Worst case approach



$$\min_j \max_i G(i,j)$$

II should play blue and lose 1 in the worst case

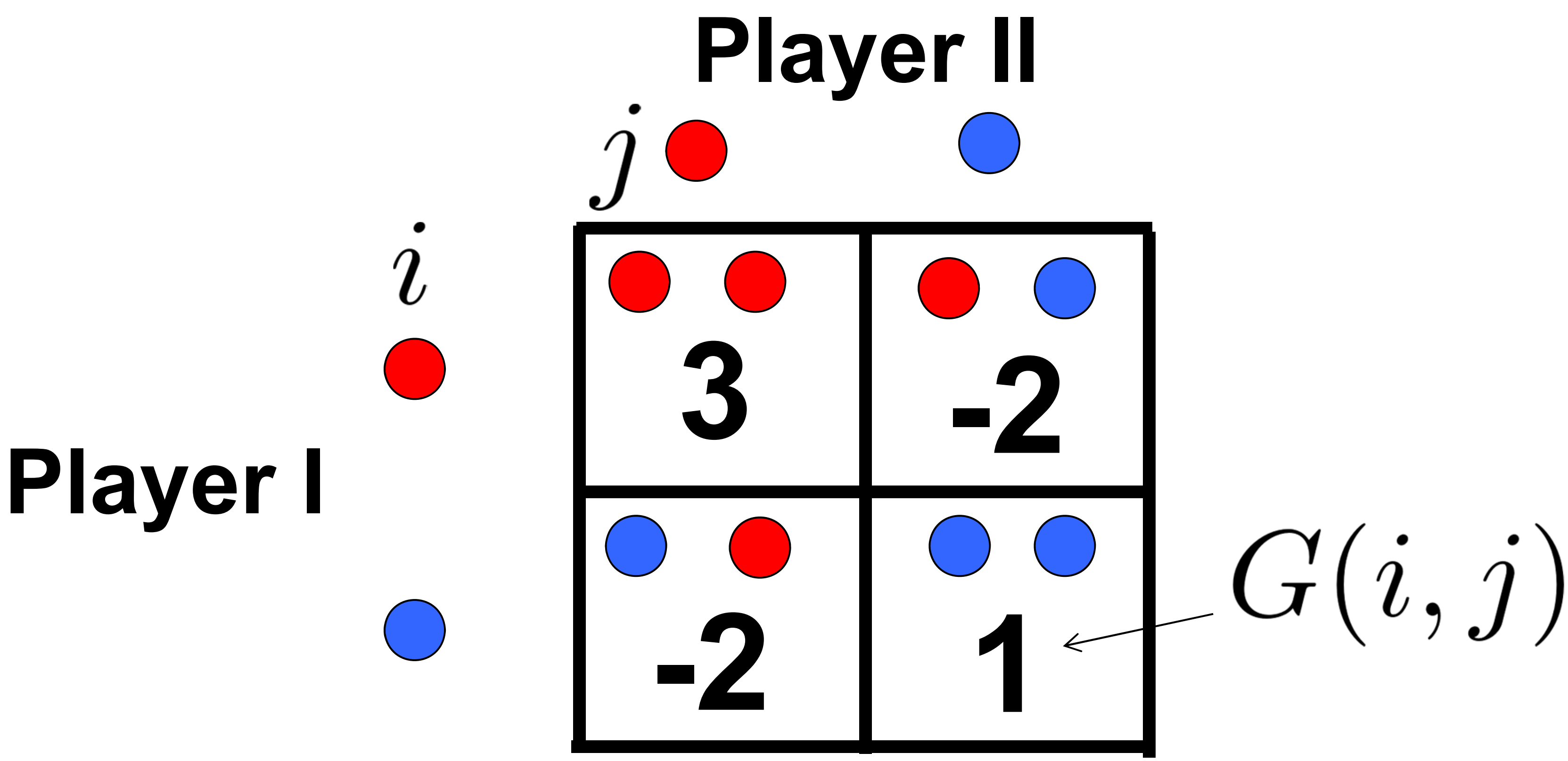
Worst case approach



$$\max_i \min_j G(i, j)$$

I should play red and lose 2 in the worst case

No saddle point



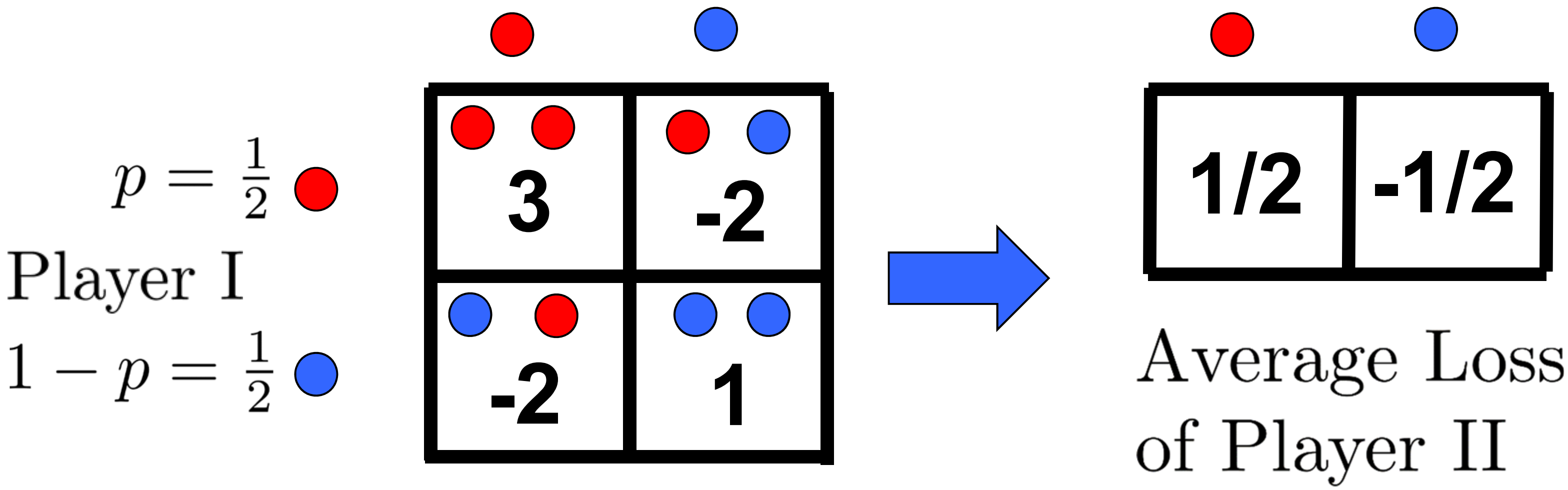
$$\max_i \min_j G(i,j) \neq \min_j \max_i G(i,j)$$

Not an equilibrium for a repeated game

Average case (Bayesian) approach

Place a uniform prior on the choice of Player I

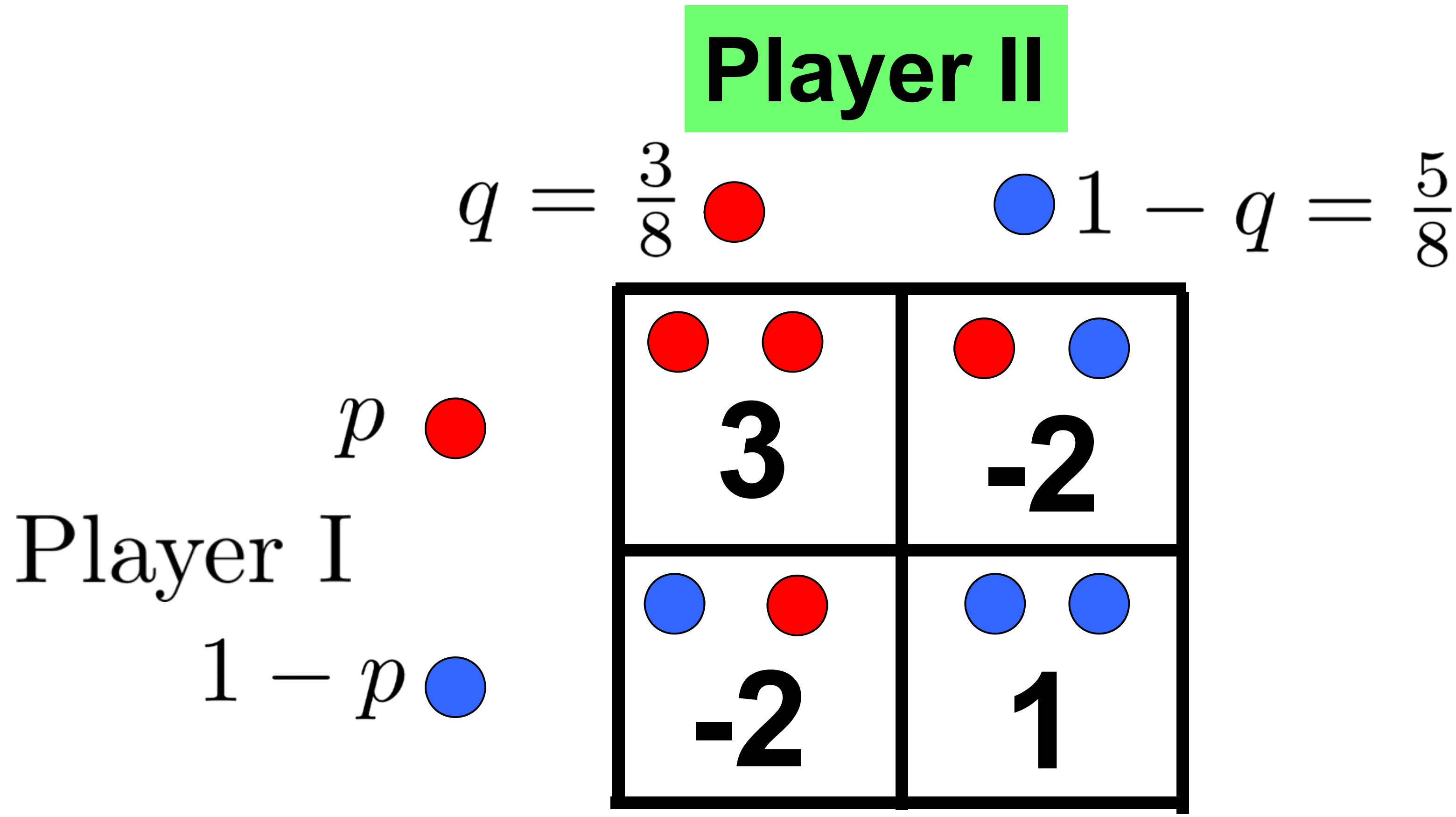
Player II



II Should always play blue

Not an equilibrium for a repeated game

Mixed strategy (repeated game) solution



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

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

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

Mixed strategy (repeated game) solution

Player I

		Player II	
		q (red)	$1-q$ (blue)
Player I	$p = \frac{3}{8}$ (red)	3	-2
	$1-p = \frac{5}{8}$ (blue)	-2	1

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

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

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

Game theory

Optimal strategies are mixed strategies

Optimal way to play is at random

John Von Neumann

$\min \max = \max \min$

Player II: q (red), $1-q$ (blue)

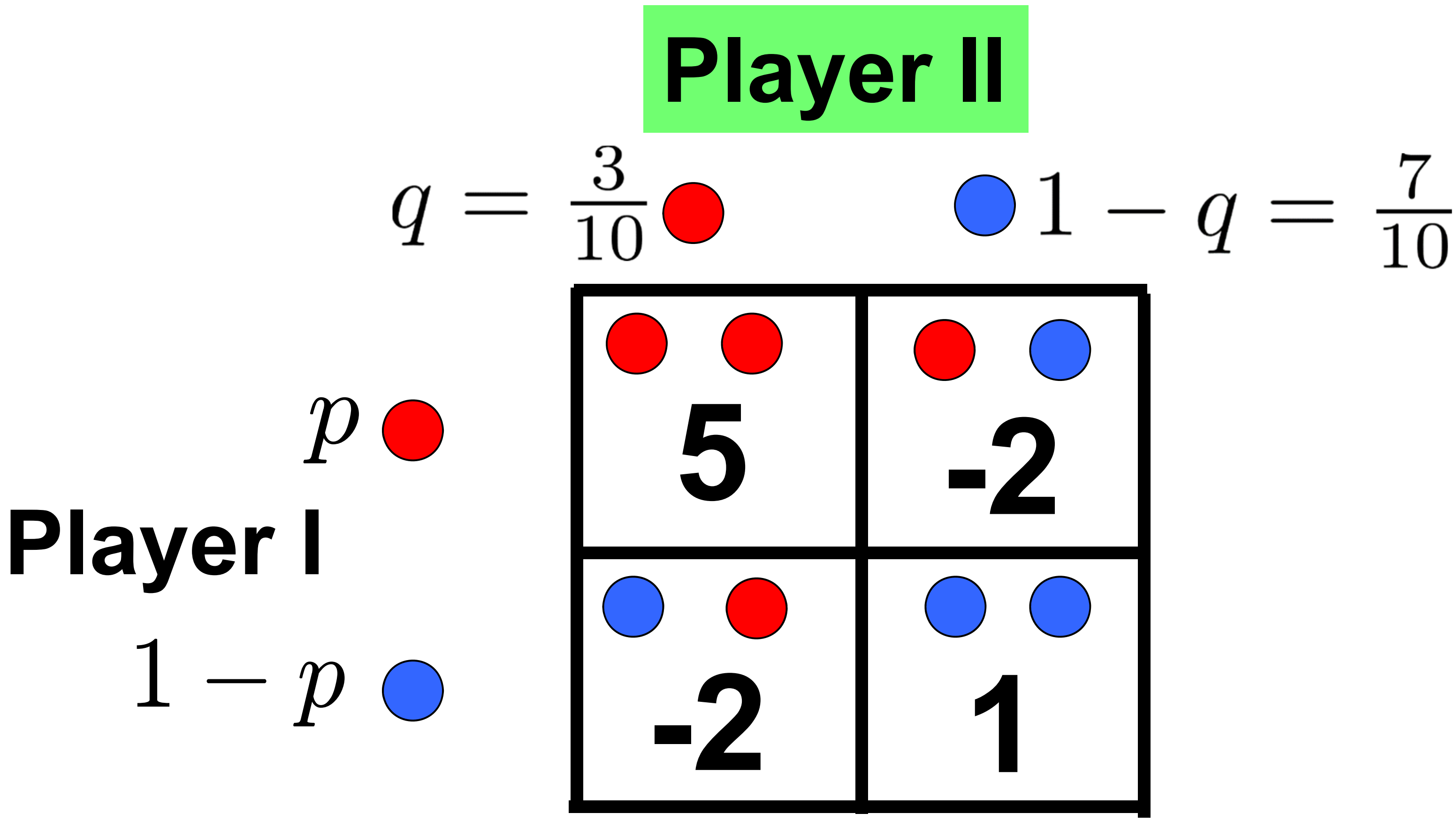
Player I: p (red), $1-p$ (blue)

	q (red)	$1-q$ (blue)
p (red)	3	-2
$1-p$ (blue)	-2	1

Saddle point

$$\min_q \max_p q_j p_i G(i,j) = \max_p \min_q q_j p_i G(i,j)$$

The optimal mixed strategy is determined by the loss matrix



II should play red with probability 3/10 and win 1/8 on average

$$\text{Player } I\text{'s expected payoff} = 5pq + (1-p)(1-q) - 2p(1-q) - 2q(1-p)$$

$$= 1 - 3q + p(10q - 3) = -\tfrac{1}{8} \text{ for } q = \tfrac{3}{10}$$

Framework: Game/Decision Theory

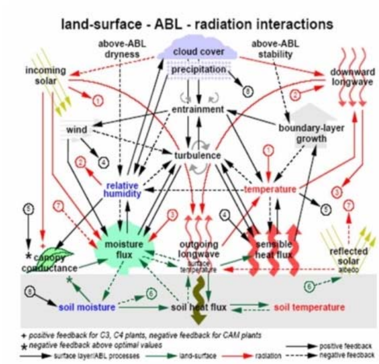
Player I

Chooses candidate

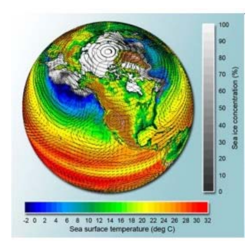
Player II

Sees data

Chooses model



$$\mathcal{E}(\text{candidate}, \text{model}(\text{data}))$$



statistical decision theory

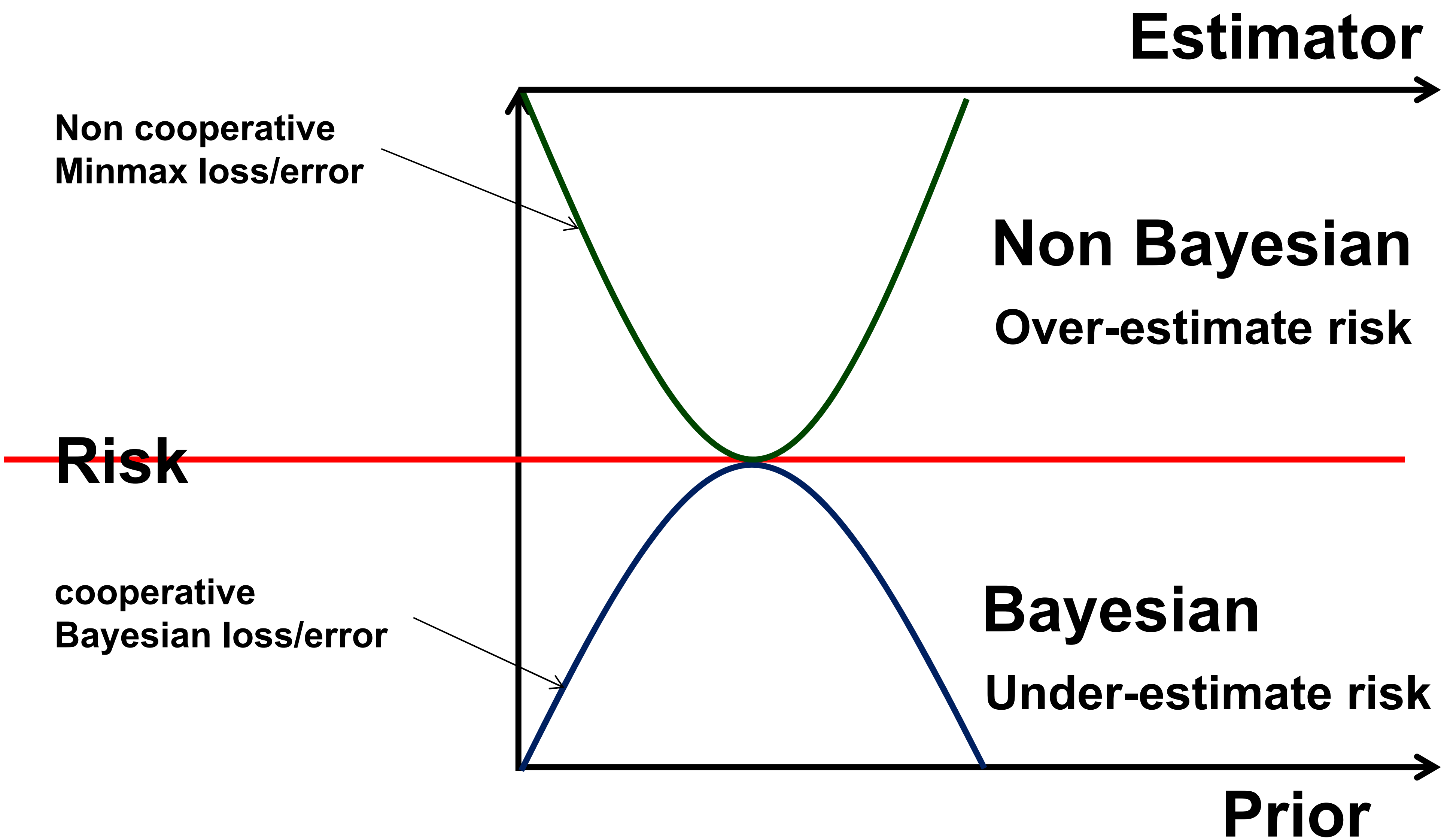
Abraham Wald

The best strategy is to play at random

Obtained by finding the worst prior in the Bayesian class of estimators

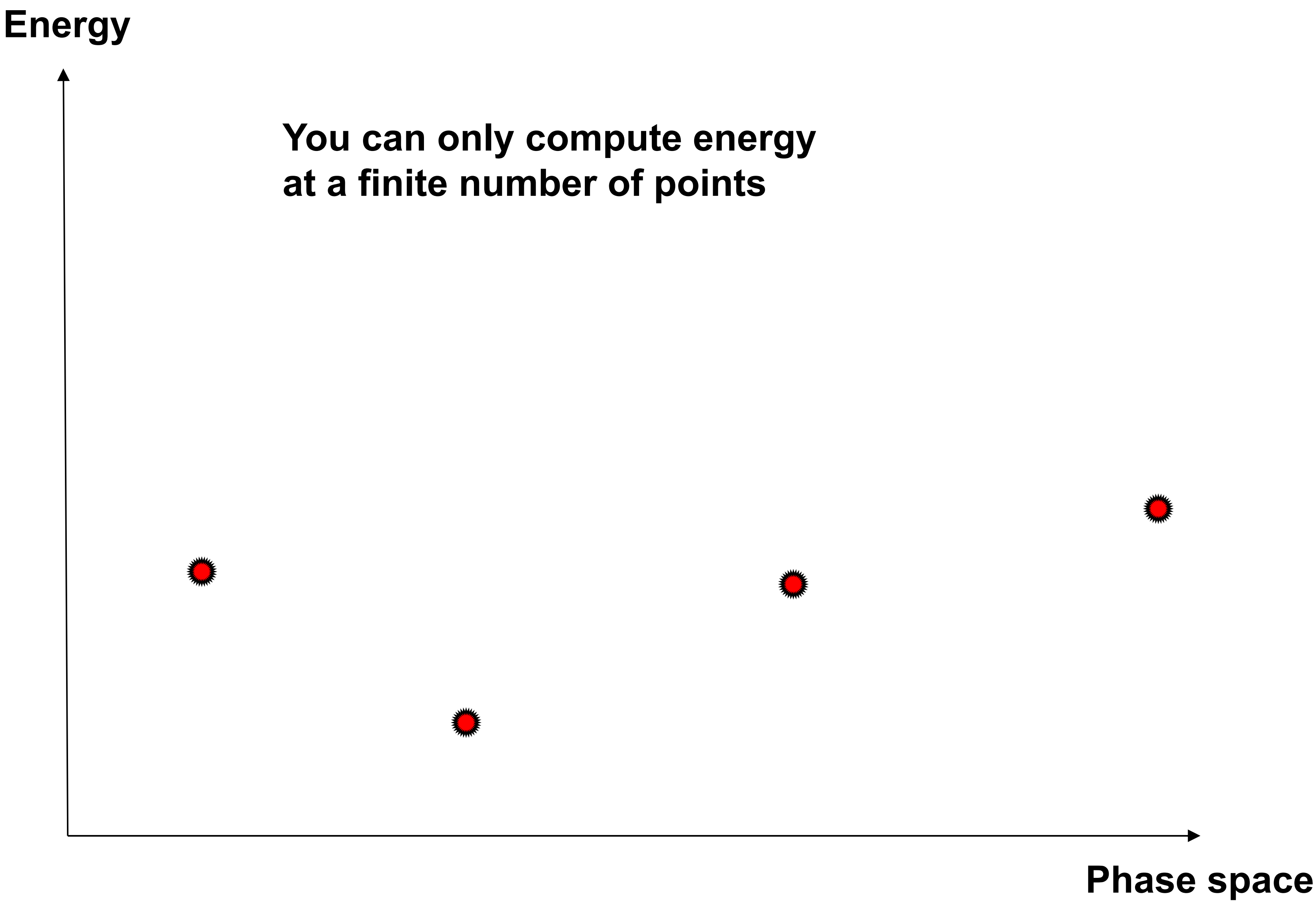
Incorporation of data naturally done via conditioning

Complete class theorem



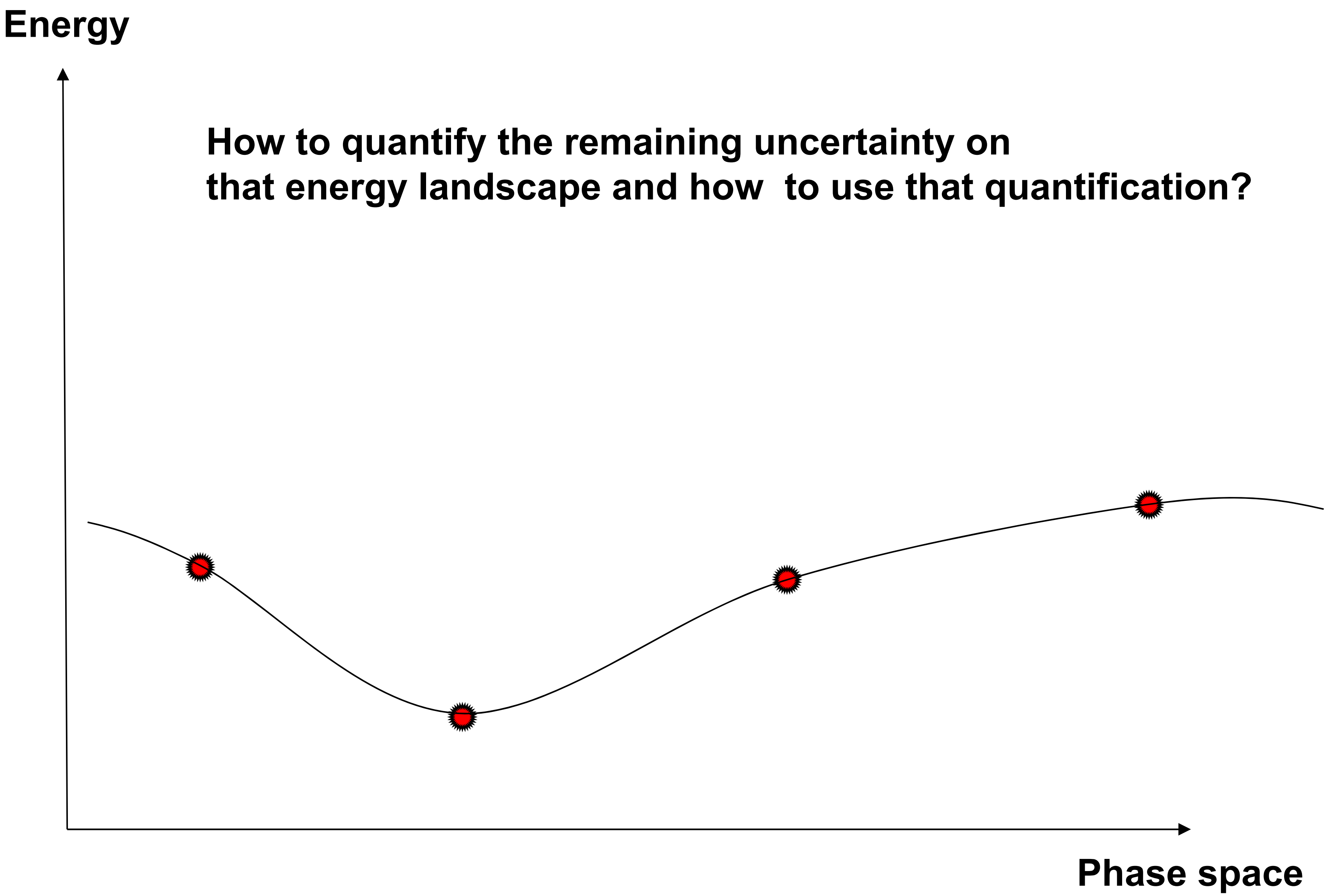
Application to energy landscapes

You can only compute energy
at a finite number of points

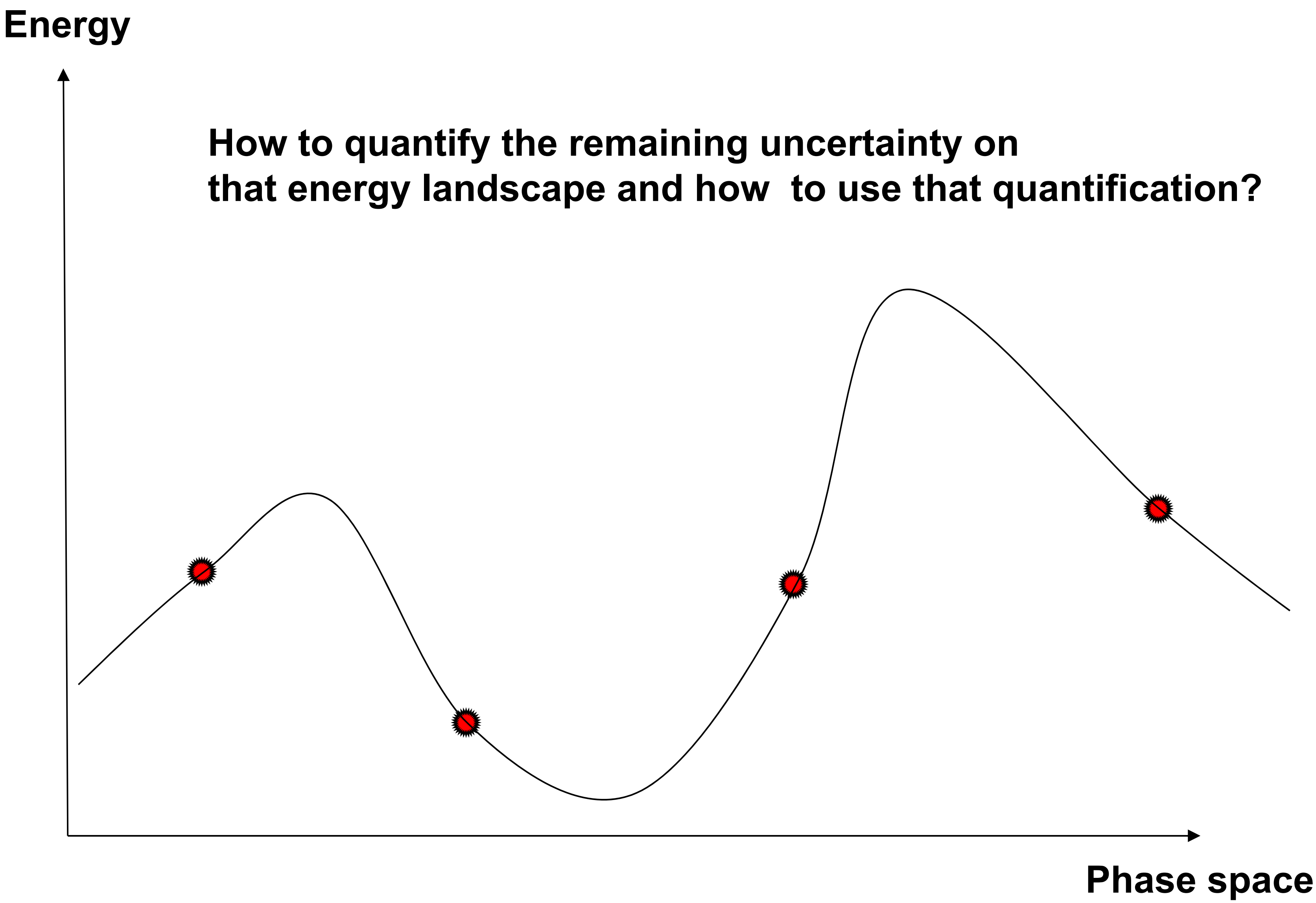


Application to energy landscapes

How to quantify the remaining uncertainty on that energy landscape and how to use that quantification?

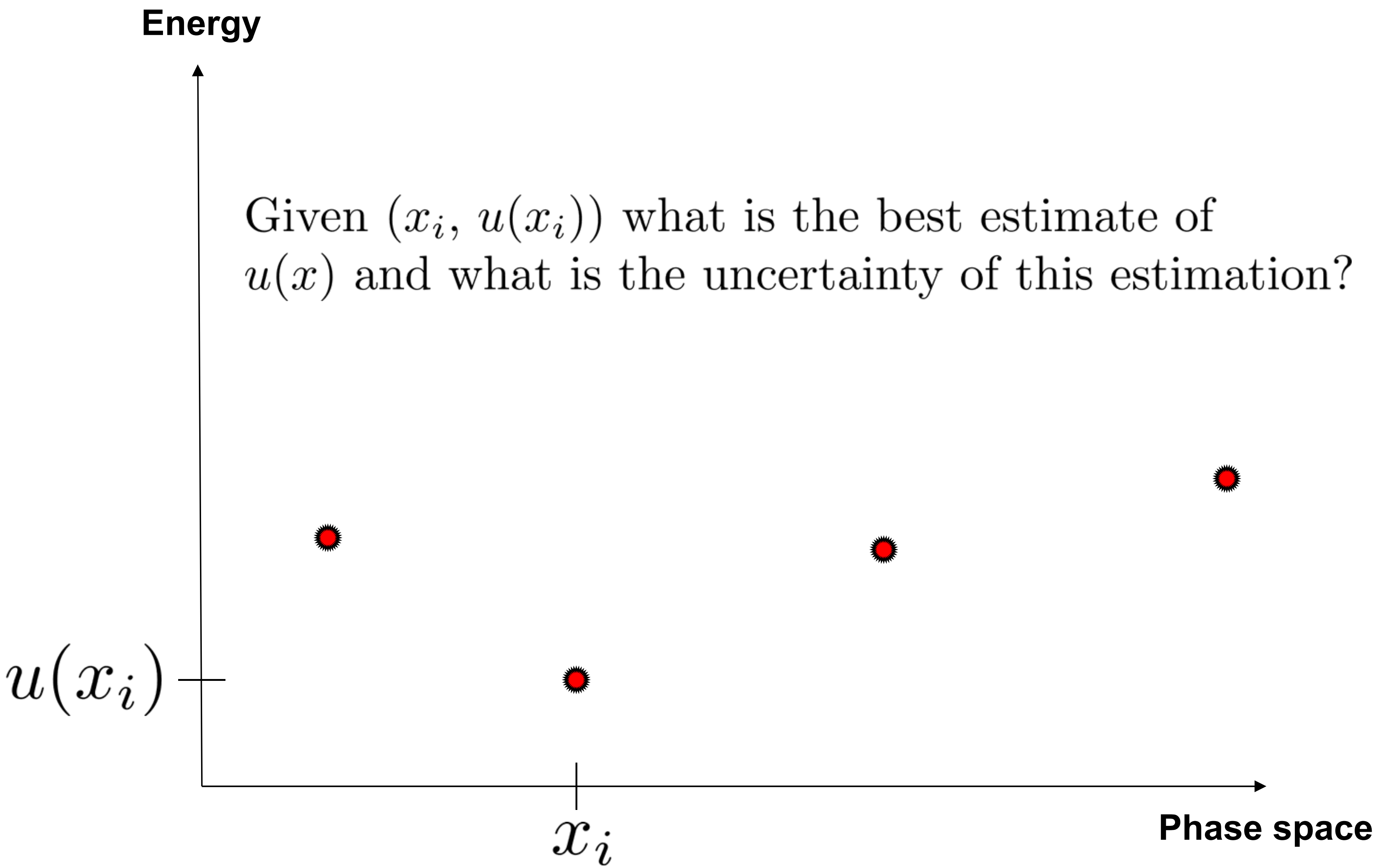


Application to energy landscapes



Application to energy landscapes

Given $(x_i, u(x_i))$ what is the best estimate of $u(x)$ and what is the uncertainty of this estimation?



A simple approximation problem

Approximate

$$x \in \mathbb{R}^n$$

Based on the information that

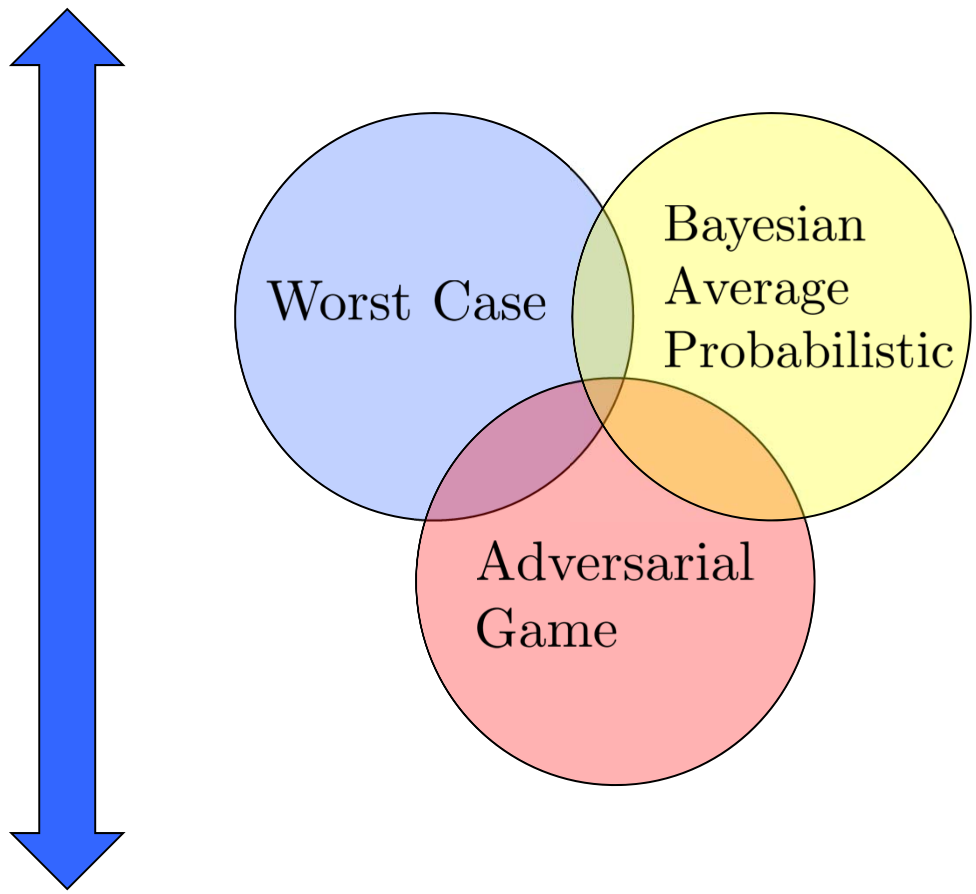
$$\Phi x = y$$

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

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

$$v(y)$$ Your approximation

3 approaches to inference and to dealing with uncertainty



3 approaches to Numerical Approximation

Worst case approach (Optimal Recovery)

Problem $\|\cdot\|$: Quadratic norm on $\mathbb{R}^n$

Find $v : \mathbb{R}^m \to \mathbb{R}^n$ minimizing worst case error

$$\inf_{v} \sup_{x \in \mathbb{R}^n} \frac{\left\| x - v(\Phi x) \right\|}{\|x\|}$$

C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.

C. A. Micchelli. Orthogonal projections are optimal algorithms. *Journal of Approximation Theory*, 40(2):101–110, 1984.

Solution

$v(y)$ is the minimizer of

$$\begin{cases} \text{Minimize } \|w\| \\ \text{Subject to } w \in \mathbb{R}^n \text{ and } \Phi w = y \end{cases}$$

$$v(y) = \sum_i y_i \psi_i \qquad \psi_i\text{: Optimal recovery splines}$$

C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.

C. A. Micchelli. Orthogonal projections are optimal algorithms. *Journal of Approximation Theory*, 40(2):101–110, 1984.

Average case approach (IBC)

$\|\cdot\|$: Quadratic norm on $\mathbb{R}^n$

μ: Measure of probability on $\mathbb{R}^n$ s.t. $\int \|x\|^2 \mu(dx) < \infty$

$$\mathcal{E}(v) = \int \|x - v(\Phi x)\|^2 \mu(dx)\text{: Average error}$$

Problem

Find $v : \mathbb{R}^m \to \mathbb{R}^n$ minimizing average error

G. W. Wasilkowski and H. Woźniakowski. Average case optimal algorithms in Hilbert spaces. *Journal of Approximation Theory*, 47(1):17–25, 1986.

J. B. Kadane and G. W. Wasilkowski. Average case ϵ-complexity in computer science: A Bayesian view. 1983. Columbia Univ. Report CUSC-6583.

Solution

Has a natural Bayesian interpretation

$$\mu = \mathcal{N}(0, C) \iff v(y) = \mathbb{E}_{x\sim\mu}[x|\Phi x = y]$$

$v(y)$ is the minimizer of

$$\begin{cases} \text{Minimize } w^T C^{-1} w \\ \text{Subject to } w \in \mathbb{R}^n \text{ and } \Phi w = y \end{cases}$$

If $\|x\|^2 = x^T C^{-1} x$ and $\mu = \mathcal{N}(0, C)$ then average case solution $=$ worst case solution

E. W. Packel. The algorithm designer versus nature: a game-theoretic approach to information-based complexity. *J. Complexity*, 3(3):244–257, 1987.

Adversarial game approach

Player I

Chooses $x \in \mathbb{R}^n$

Player II

Sees $y = \Phi x$

Chooses $v(y)$

Max

Min

$$\frac{\left\| x - v(\Phi x) \right\|^2}{\|x\|^2}$$

[H. Owhadi and C. Scovel. Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis 2017. arXiv:1703.10761]

Loss function

$$\mathcal{E}(x,v) = \frac{\left\| x - v(\Phi x) \right\|^2}{\|x\|^2}$$

Player I

$$\max_x \min_v \mathcal{E}(x,v) = 0$$

Player II

$$\min_v \max_x \mathcal{E}(x,v) \neq 0$$

No saddle point of pure strategies

Randomized strategy for player I

Player I

Chooses $\mu \in \mathcal{P}(\mathbb{R}^n)$

Samples $x \sim \mu$

Player II

Sees $y = \Phi x$

Chooses $v(y)$

Max

Min

$$\frac{\int_{\mathbb{R}^n} \|x - v(\Phi x)\|^2 \mu(dx)}{\int_{\mathbb{R}^n} \|x\|^2 \mu(dx)}$$

[H. Owhadi and C. Scovel. Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis 2017. arXiv:1703.10761]

Loss function

$$\mathcal{E}(\mu, v) = \frac{\int_{\mathbb{R}^n} \|x - v(\Phi x)\|^2 \mu(dx)}{\int_{\mathbb{R}^n} \|x\|^2 \mu(dx)}$$

Saddle point

$$\max_\mu \min_v \mathcal{E}(\mu, v) = \min_v \max_\mu \mathcal{E}(\mu, v)$$

$$\exists \mu^\dagger, v^\dagger$$

$$\mathcal{E}(\mu, v^\dagger) \leq \mathcal{E}(\mu^\dagger, v^\dagger) \text{ for all } \mu$$

$$\mathcal{E}(\mu^\dagger, v) \geq \mathcal{E}(\mu^\dagger, v^\dagger) \text{ for all } v$$

Canonical Gaussian field

$$\|x\|^2 := x^T A x$$

A: $n \times n$ symmetric postive definite matrix

ξ: Canonical Gaussian field on $(\mathbb{R}^n, \|\cdot\|)$

Density function

$$f(x) = \frac{e^{-\frac{\|x\|^2}{2}}}{C}$$

$$\xi \sim \mathcal{N}(0, A^{-1})$$

For $z \in \mathbb{R}^n$, $z^T \xi \sim \mathcal{N}(0, \|z\|_*^2)$

$$\|z\|_* = \sup_{x \in \mathbb{R}^n} \frac{z^T x}{\|x\|} = z^T A^{-1} z$$

Equilibrium saddle point

Player I

$$\mu^{\dagger} \quad \Longleftrightarrow \quad x \sim \xi - \mathbb{E}\big[\xi \mid \Phi\xi\big]$$

Player II

$$v^{\dagger}(y) = \mathbb{E}[\xi \mid \Phi\xi = y]$$

The optimal bet of Player II is Bayesian

⇕

Complete Class Theorem

Statistical decision theory

Abraham Wald

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

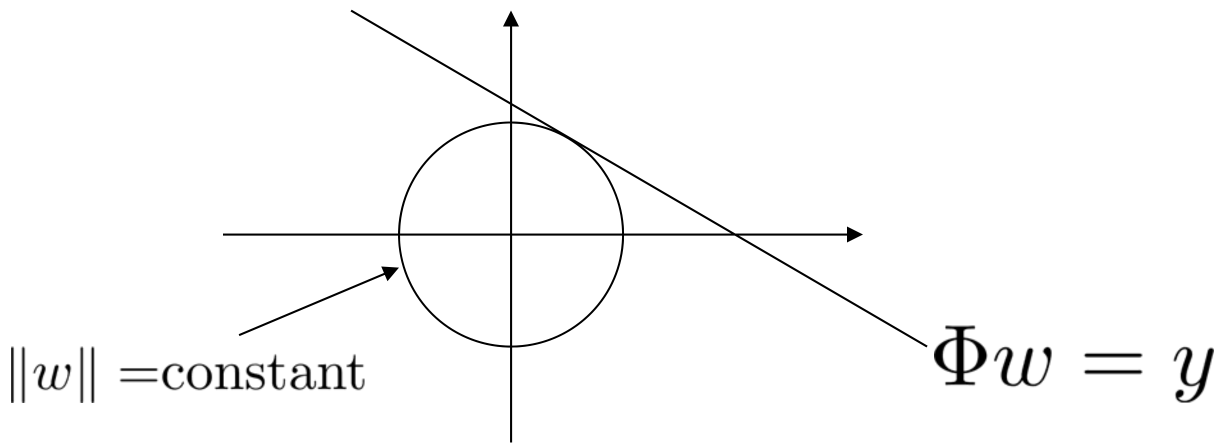
A. Wald. An essentially complete class of admissible decision functions. *Ann. Math. Statistics*, 18:549–555, 1947.

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

The game theoretic solution is equal to the worst case solution

$$v^\dagger(y) = \mathbb{E}[\xi \mid \Phi\xi = y]$$

ξ has density $\dfrac{e^{-\frac{\|w\|^2}{2}}}{C}$



$v^\dagger(y)$ is the minimizer of

$$\begin{cases} \text{Minimize } \|w\| \\ \text{Subject to } w \in \mathbb{R}^n \text{ and } \Phi w = y \end{cases}$$

Generalization

$(\mathcal{B}, \|\cdot\|)$: separable Banach space

$\|\cdot\|$: Quadratic norm

$\Updownarrow$

$$\|u\|^2 := [\mathcal{T}u, u]$$

$\mathcal{T}$: Symmetric positive continuous linear bijection

$$\mathcal{B} \xrightarrow{\quad \mathcal{T} \quad} \mathcal{B}^*$$

For $u, v \in \mathcal{B}$,

- $[\mathcal{T}u, v] = [\mathcal{T}v, u]$,
- $[\mathcal{T}u, u] \geq 0$

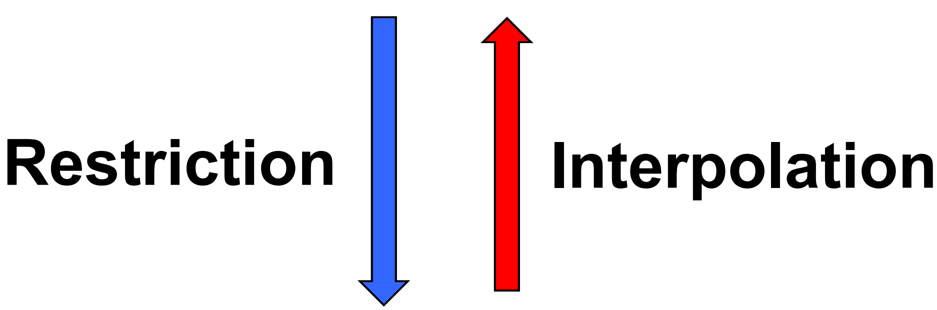
$$\phi_1, \ldots, \phi_m \in \mathcal{B}^*$$

Problem

Let u be an unknown element of $\mathcal{B}$

Given $([\phi_1, u], \ldots, [\phi_m, u])$ recover u

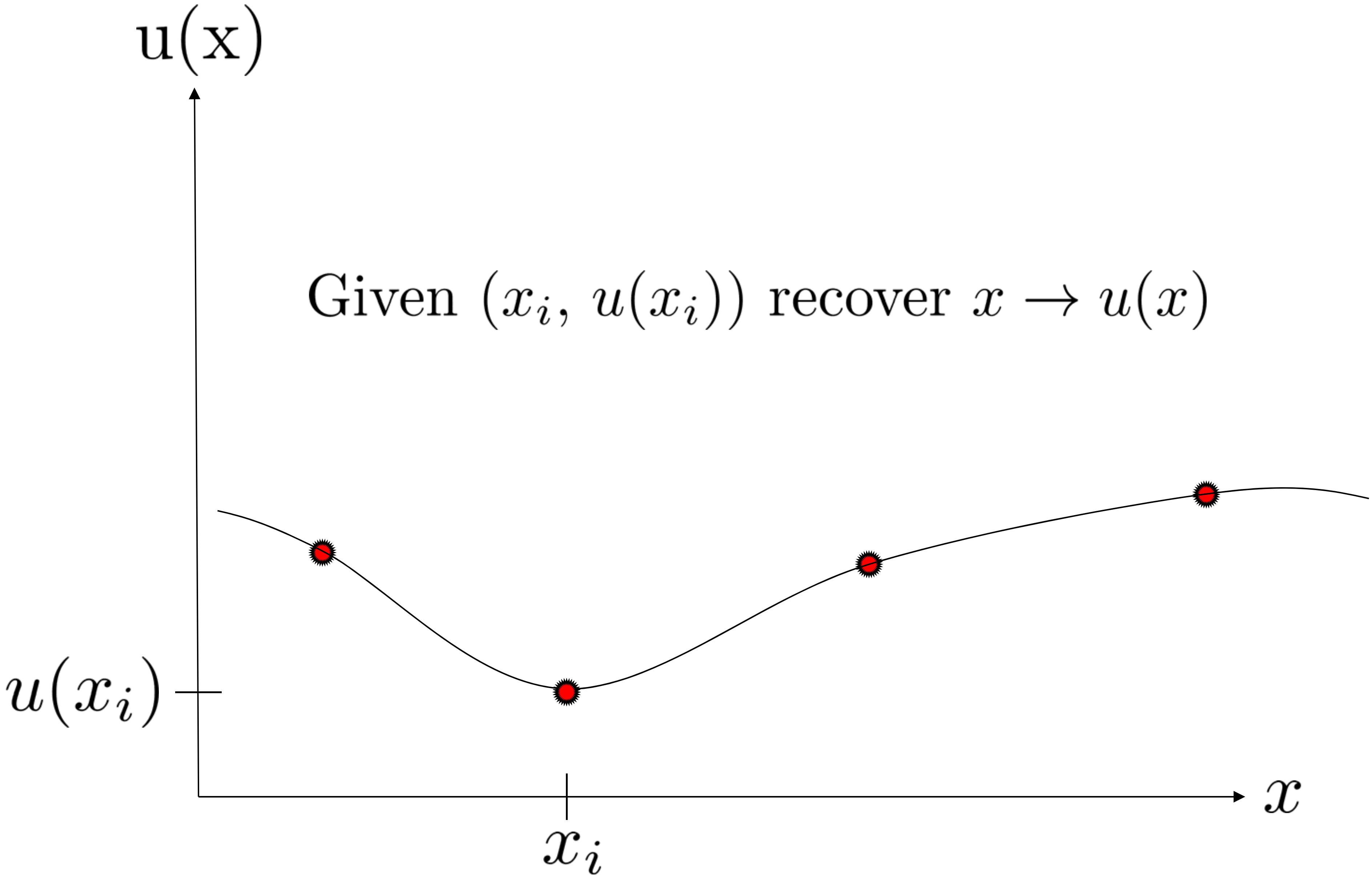
$$u \in \mathcal{B}$$



$$u_m = ([\phi_1, u], \ldots, [\phi_m, u])$$

Application to energy landscapes

Given $(x_i,\, u(x_i))$ recover $x \to u(x)$



Application to energy landscapes

$$\mathcal{B} = H^s(\Omega)$$

Ω: Phase space $\qquad s > d/2$

$\|\cdot\|$: Quadratic norm on $H^s(\Omega)$

$$\|u\|^2 = \int_\Omega u \mathcal{L} u$$

$\mathcal{L}$: symmetric, positive, continuous linear bijection

$$H^s(\Omega) \xrightarrow{\quad \mathcal{L} \quad} H^{-s}(\Omega)$$

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

Choose the norm in which you want optimality

$$\phi_1, \ldots, \phi_m \in \mathcal{B}^*$$

Player I

Chooses $u \in \mathcal{B}$

Player II

Sees $y = \big([\phi_1, u], \ldots, [\phi_m, u]\big)$

Chooses $v(y) \in \mathcal{B}$

Max

Min

$$\frac{\Big\| u - v([\phi_1, u], \ldots, [\phi_m, u]) \Big\|^2}{\|u\|^2}$$

Examples

$$\mathcal{B} := \mathbb{R}^N$$

Player I

Chooses $x \in \mathbb{R}^N$

Player II

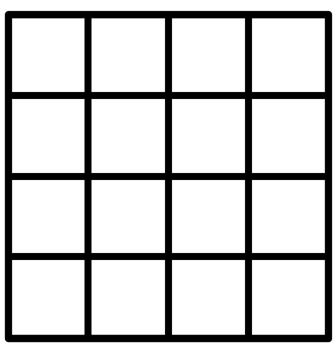
Sees $(\phi_1^T x, \ldots, \phi_m^T x)$

Chooses $v(\phi_1^T x, \ldots, \phi_m^T x) \in \mathbb{R}^N$

$$\mathcal{B} := H_0^s(\Omega)$$

Player I

Chooses $u \in H_0^s(\Omega)$



Player II

Sees $(\int_{\tau_1} u, \ldots, \int_{\tau_m} u)$

Chooses $v(\int_{\tau_1} u, \ldots, \int_{\tau_m} u) \in H_0^s(\Omega)$

Application to energy landscapes

Player I

Chooses $u \in H^s(\Omega)$

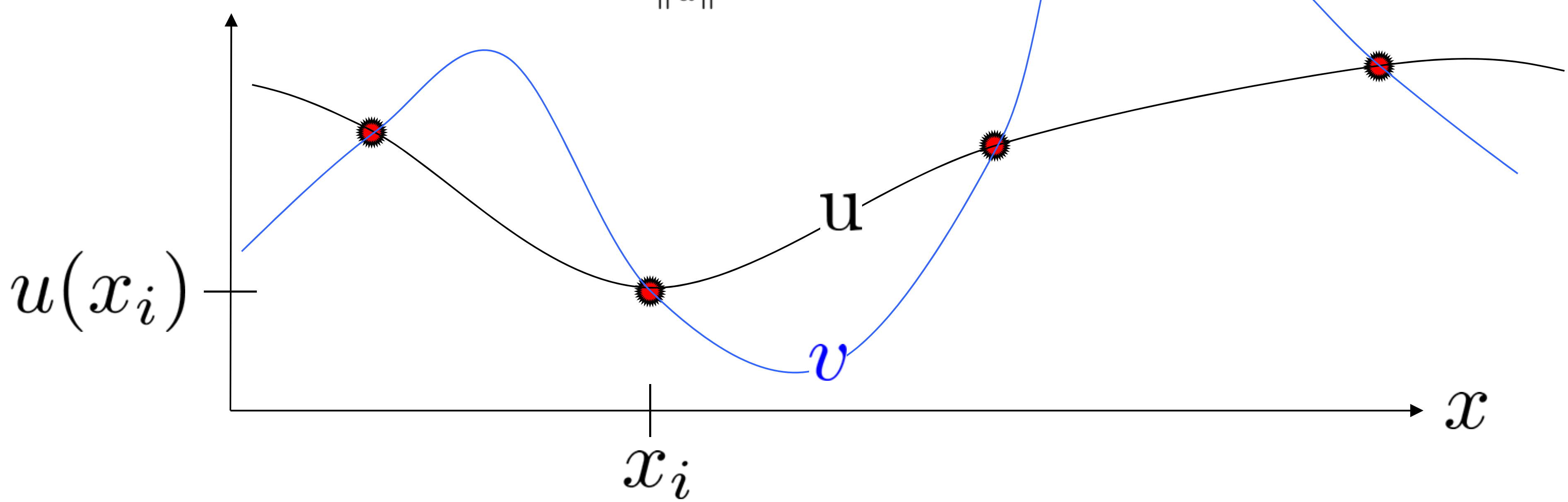
Player II

Sees $u(x_1), \ldots, u(x_m)$

Chooses $v(u(x_1), \ldots, u(x_m)) \in H^s(\Omega)$

Max / Min

$$\frac{\left\| u - v(u(x_1), \ldots, u(x_m)) \right\|^2}{\|u\|^2}$$



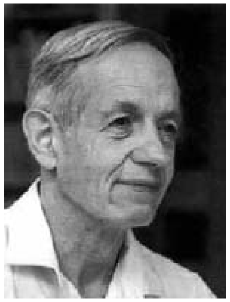
Loss function

$$\mathcal{E}(u,v) = \frac{\left\| u - v([\phi_1,u],\ldots,[\phi_m,u]) \right\|^2}{\|u\|^2}$$

Player I $\max_u \min_v \mathcal{E}(u,v) = 0$

Player II $\min_v \max_u \mathcal{E}(u,v) \neq 0$

No saddle point of pure strategies



Need to lift the minimax to mixed strategies

Randomized strategy for player I

$$\phi_1, \ldots, \phi_m \in \mathcal{B}^*$$

Player I

Chooses $\mu \in \mathcal{P}(\mathcal{B})$

Samples $u \sim \mu$

Player II

Sees $y = ([\phi_1, u], \ldots, [\phi_m, u])$

Chooses $v(y) \in \mathcal{B}$

Max

Min

$$\frac{\int \left\| u - v([\phi_1, u], \ldots, [\phi_m, u]) \right\|^2 \mu(du)}{\int \|u\|^2 \mu(du)}$$

Loss function

$$\mathcal{E}(\mu, v) = \frac{\int \left\| u - v([\phi_1, u], \ldots, [\phi_m, u]) \right\|^2 \mu(du)}{\int \|u\|^2 \mu(du)}$$

Theorem

$$\sup_\mu \inf_v \mathcal{E}(\mu, v) = \inf_v \sup_\mu \mathcal{E}(\mu, v)$$

What is the saddle point?

Can we find $\mu^\dagger, v^\dagger$ such that for all μ, v

$$\mathcal{E}(\mu, v^\dagger) \leq \mathcal{E}(\mu^\dagger, v^\dagger) \leq \mathcal{E}(\mu^\dagger, v)$$

$$\mathcal{B} := \mathbb{R}^N \qquad \|x\|^2 := x^T A x$$

A: $N \times N$ symmetric postive definite matrix

ξ: Canonical Gaussian field on $(\mathbb{R}^n, \|\cdot\|)$

Density function of ξ: $f(x) = \frac{e^{-\frac{\|x\|^2}{2}}}{C}$

$$\xi = \mathcal{N}(0, A^{-1})$$

Theorem

$$\mu^\dagger \stackrel{\text{dist}}{=} \xi - \mathbb{E}\big[\xi \mid ([\phi_i, \xi])_{i \in \mathcal{I}}\big]$$

$$v^\dagger\big([\phi_1, u], \ldots, [\phi_m, u]\big) = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\big]$$

Generalization $(\mathcal{B}, \|\cdot\|)$: separable Banach space

$$\|u\|^2 := [\mathcal{T}u, u]$$

$\mathcal{T}$: Symmetric positive continuous linear bijection

$$\mathcal{B} \xrightarrow{\quad\mathcal{T}\quad} \mathcal{B}^*$$

Canonical Gaussian field

ξ: Linear isometry mapping $\mathcal{B}^*$ to a Gaussian Space

$$\xi\colon\ \mathcal{B}^* \longrightarrow \mathcal{H}$$
$$\phi \longrightarrow [\phi, \xi] \sim \mathcal{N}(0, \|\phi\|_*^2)$$

$$\|\phi\|_* := \sup_{v \in \mathcal{B}} \frac{[\phi, v]}{\|v\|}$$

$$\xi \sim \mathcal{N}(0, \mathcal{T}^{-1})$$

For $\varphi, \phi \in \mathcal{B}^*$ $\quad \mathbb{E}\big[[\varphi, \xi][\phi, \xi]\big] = \langle \varphi, \phi \rangle_* = [\varphi, \mathcal{T}^{-1}\phi]$

Examples

$$\mathcal{B} := \mathbb{R}^N \qquad \|x\|^2 := x^T A x$$

A: $N \times N$ symmetric postive definite matrix

$$\xi = \mathcal{N}(0, A^{-1})$$

$$\mathcal{B} := H_0^s(\Omega) \qquad \|u\|^2 := [\mathcal{L}u, u]$$

$\mathcal{L}$: arbitrary symmetric, positive, continuous linear bijection

$$(H_0^s(\Omega), \|\cdot\|_{H_0^s(\Omega)}) \xrightarrow{\mathcal{L}} (H^{-s}(\Omega), \|\cdot\|_{H^{-s}(\Omega)})$$

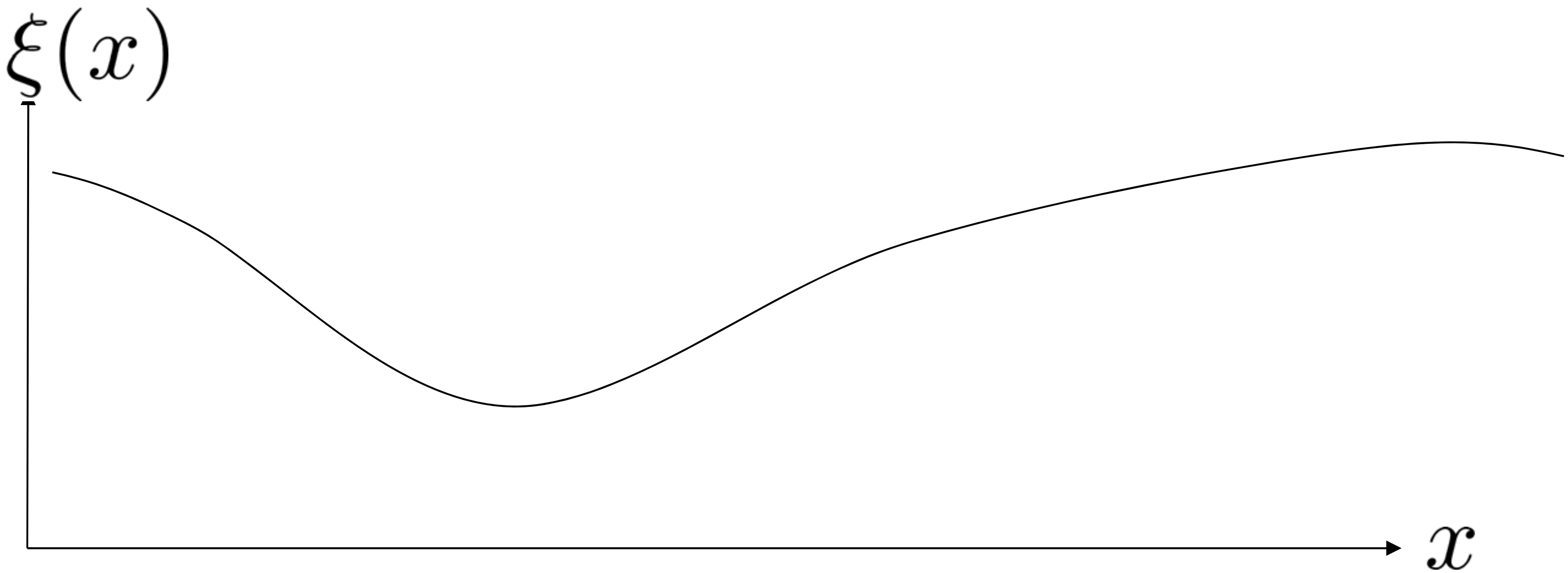
$$\int_\Omega \phi(x)\xi(x)\,dx \sim \mathcal{N}(0, \int_{\Omega^2} \phi(x)G(x,y)\phi(y)\,dx\,dy)$$

Application to energy landscapes

$$G := \mathcal{L}^{-1} \text{ Green's function of } \mathcal{L}$$

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

$$\mathbb{E}[\xi(x)] = 0 \qquad \mathrm{Cov}\left(\xi(x), \xi(y)\right) = G(x, y)$$



Let

$$\mu^\dagger \overset{\text{dist}}{=} \xi - \mathbb{E}\big[\xi \mid ([\phi_i, \xi])_{i \in \mathcal{I}}\big]$$

$$v^\dagger(y) = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = y_i \text{ for } i \in \mathcal{I}\big]$$

Theorem $\exists\, \mu_n \in \mathcal{P}(\mathcal{B})$ indexed by $n \in \mathbb{N}^*$

$$\text{s.t.} \quad \mu_n \xrightarrow[n\to\infty]{w} \mu^\dagger$$

i.e. for all $k \geq 1$ and $\varphi_1, \ldots, \varphi_k \in \mathcal{B}^*$

$$\mathbb{E}_{u\sim\mu_n}\Big[\big([\varphi_1, u], \ldots, [\varphi_k, u]\big)\Big] \underset{n\to\infty}{\longrightarrow} \mathbb{E}_{u\sim\mu^\dagger}\Big[\big([\varphi_1, u], \ldots, [\varphi_k, u]\big)\Big]$$

and s.t. for all μ, v

$$\mathcal{E}(\mu, v^\dagger) - \frac{1}{n} \leq \mathcal{E}(\mu_n, v^\dagger) \leq \mathcal{E}(\mu_n, v) + \frac{1}{n}$$

Optimal bet of player II

$$u^* = \mathbb{E}\Big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\Big]$$

The bet is both repeated game and worst case optimal

Theorem

$$v^\dagger(y) = \mathbb{E}\Big[\xi \mid [\phi_i, \xi] = y_i \text{ for } i \in \mathcal{I}\Big]$$

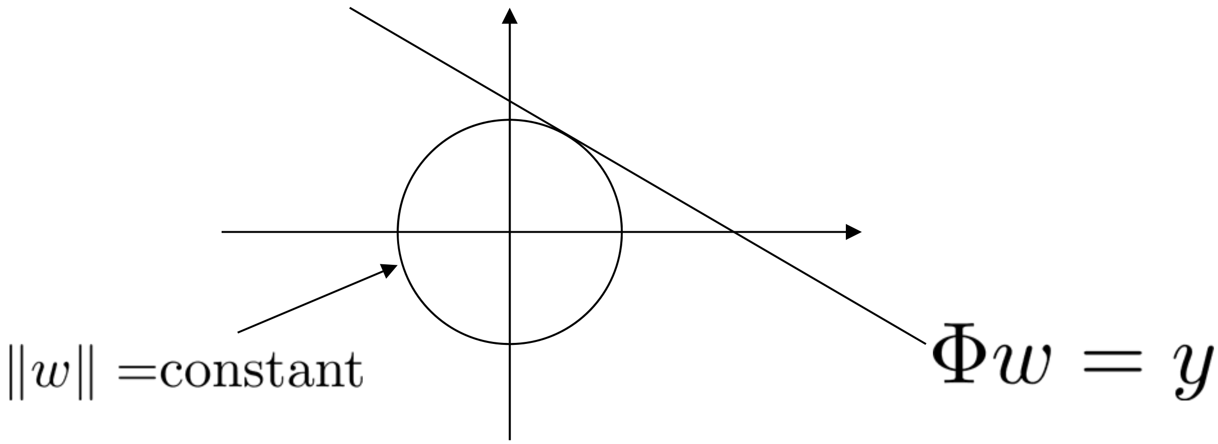
minimizer of

$$\inf_v \sup_{u \in \mathcal{B}} \frac{\Big\|u - v([\phi_1, u], \ldots, [\phi_m, u])\Big\|^2}{\|u\|^2}$$

C. A. Micchelli. Orthogonal projections are optimal algorithms. *Journal of Approximation Theory*, 40(2):101–110, 1984.

C. A. Micchelli and T. J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pages 1–54. Springer, 1977.

$$u^* = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\big]$$



Variational formulation of the optimal bet

u^*: Minimizer of

$$\begin{cases} \text{Minimize } \|w\| \\ \text{Subject to } w \in \mathcal{B} \text{ and } [\phi_i, w] = [\phi_i, u] \text{ for } i \in \mathcal{I} \end{cases}$$

$$u^* = \mathbb{E}\Big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\Big]$$

Kernel formulation of the optimal bet

$$u^* = \sum_{i,j} [\phi_i, u] \Theta^{-1}_{i,j} \mathcal{T}^{-1} \phi_j$$

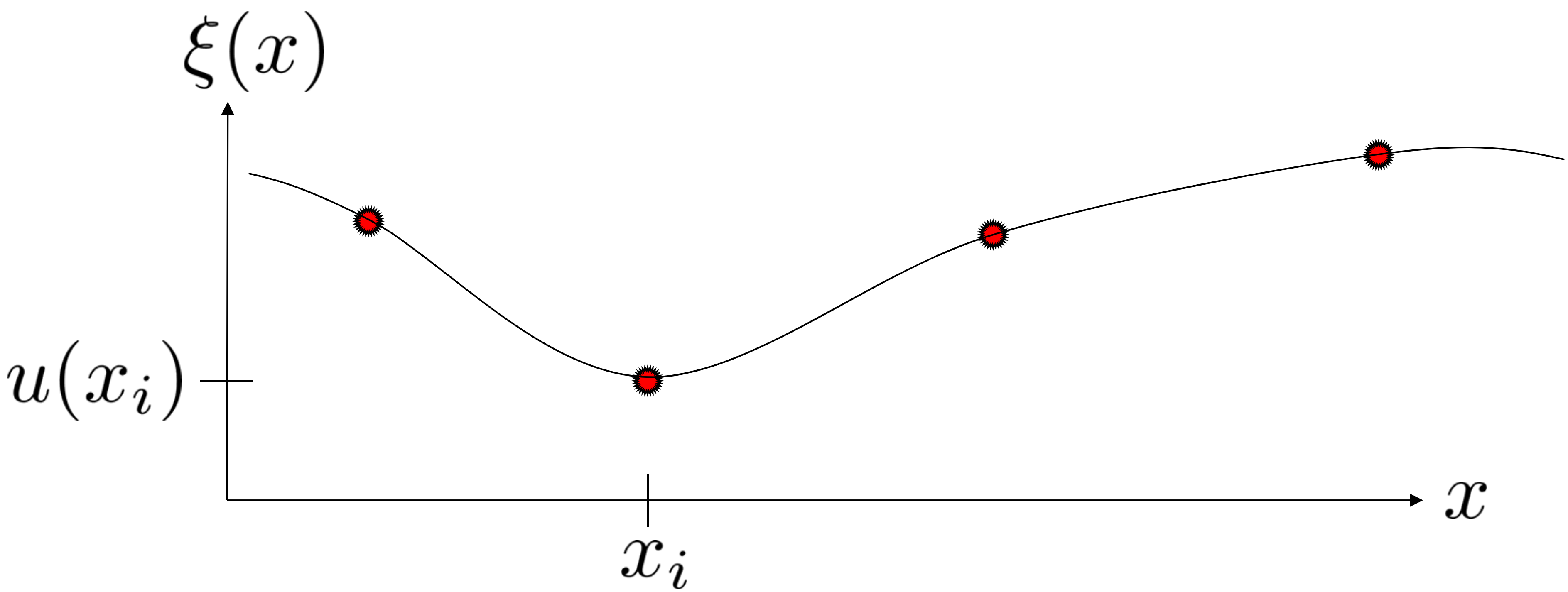
$$\Theta_{i,j} = [\phi_i, \mathcal{T}^{-1} \phi_j]$$

Application to energy landscapes

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

Best guess

$$u^*(x) = \mathbb{E}\big[\xi(x)\big|\xi(x_i) = u(x_i),\ i \in \{1, \ldots, m\}\big]$$



Best guess

$G := \mathcal{L}^{-1}$ Green's function of $\mathcal{L}$

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

$$\mathbb{E}[\xi(x)] = 0 \qquad \mathrm{Cov}\big(\xi(x), \xi(y)\big) = G(x, y)$$

Best guess

$$u^*(x) = \mathbb{E}\big[\xi(x) \big| \xi(x_i) = u(x_i),\ i \in \{1, \ldots, m\}\big]$$

Variational and Kernel interpretation

u^*: Minimizer of

$$\begin{cases} \text{Minimize } \|w\| \\ \text{Subject to } w \in H^s(\Omega) \text{ and } w(x_i) = u(x_i) \text{ for } i \in \{1, \ldots, m\} \end{cases}$$

$$u^*(x) = \sum_{i=1}^{m} c_i G(x_i, x)$$

$$c_i = \sum_j \Theta^{-1}_{i,j} u(x_j)$$

$$\Theta_{i,j} = G(x_i, x_j)$$

Optimal bet of player II

$$u^* = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\big]$$

Error of the recovery

Theorem

$$\|u - u^*\| = \min_{\psi \in \mathcal{T}^{-1} \operatorname{span}\{\phi_i | i \in \mathcal{I}\}} \|u - \psi\|$$

$$\|u - u^*\| = \min_{\phi \in \operatorname{span}\{\phi_i | i \in \mathcal{I}\}} \|\mathcal{T} u - \phi\|_*$$

Application to energy landscapes

$$\mathcal{B} = H^s(\Omega)$$

$$\|u\|^2 = \int_\Omega u \mathcal{L} u$$

$\mathcal{L}$: symmetric, positive, continuous linear bijection

$$H^s(\Omega) \xrightarrow{\quad\mathcal{L}\quad} H^{-s}(\Omega)$$

Theorem

$$\|u - u^*\|_{H^s} \leq C \min_{\phi \in \text{span}\{\delta(x - x_i) | i \in \mathcal{I}\}} \|\mathcal{L}u - \phi\|_{H^{-s}(\Omega)}$$

$$\|u - u^*\|_{H^s} \leq C h^s \|\mathcal{L}u\|_{L^2(\Omega)}$$

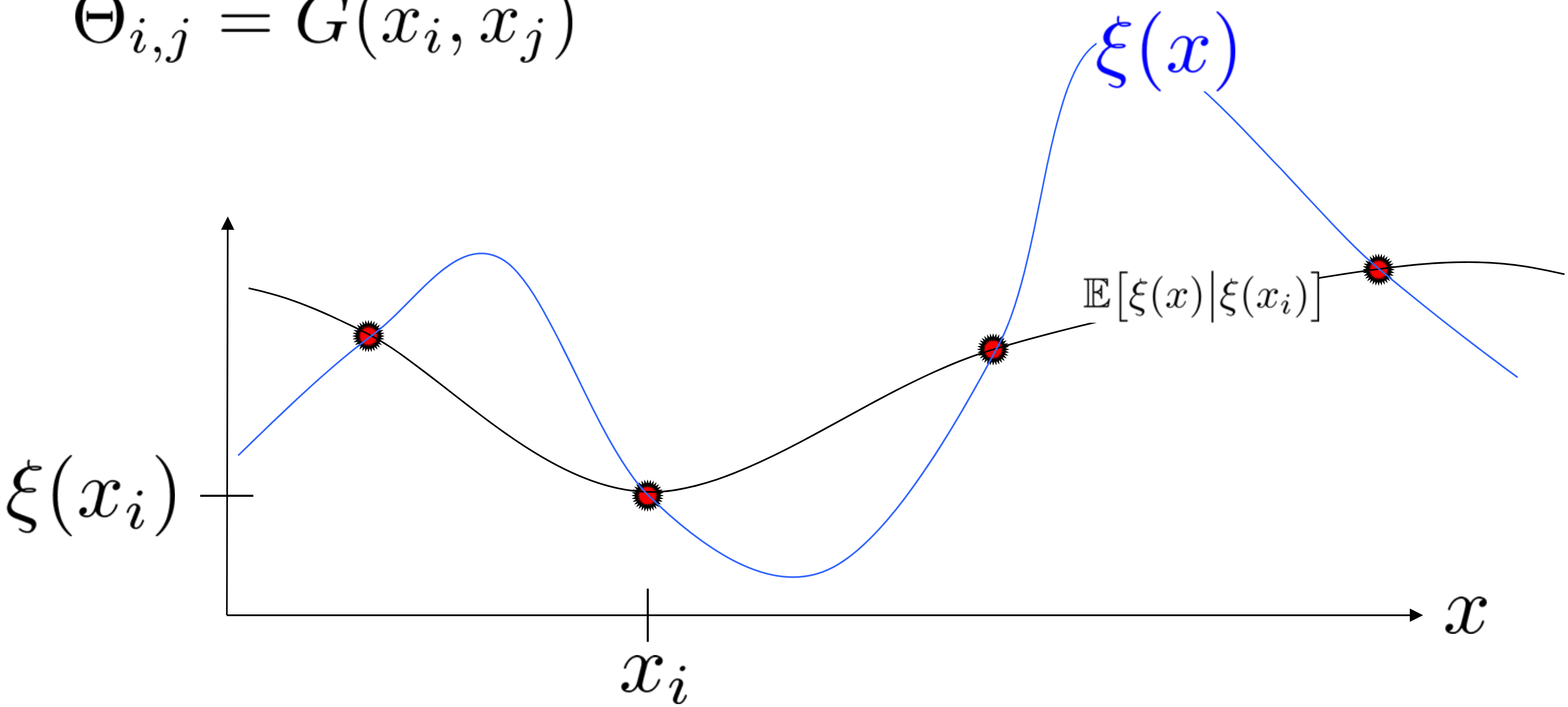
$$h = \sup_{x \in \Omega} \min_i |x - x_i|$$

Uncertainty Quantification

$$\xi(x) - \mathbb{E}\big[\xi(x)\big|\xi(x_i)\big] \sim \mathcal{N}(0, K)$$

$$K(x,y) = G(x,y) - \sum_{i,j} \Theta^{-1}_{i,j} G(x,x_i) G(y,x_j)$$

$$\Theta_{i,j} = G(x_i, x_j)$$



The standard deviations controls the deterministic error

$$\sigma^2(x) = K(x,x) = \mathbb{E}\Big[\big|\xi(x) - \mathbb{E}\big[\xi(x)\big|\xi(x_i)\big]\big|^2\Big]$$

Theorem

$$\big|u(x) - u^*(x)\big| \leq \sigma(x)\|u\|$$

Proof

$$u(x) - u^*(x) = \int_\Omega \mathcal{L}u(y)\left(G(y,x) - \sum_{i,j} G(y,x_i)\Theta^{-1}_{i,j}G(x_j,x)\right)dy$$

$$\big|u(x) - u^*(x)\big| \leq \|u\|\,\Big\|G(y,x) - \sum_{i,j} G(y,x_i)\Theta^{-1}_{i,j}G(x_j,x)\Big\|$$

Optimal bet of player II

$$u^* = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for } i \in \mathcal{I}\big]$$

Gamblets

$$u^* = \sum_i [\phi_i, u]\psi_i$$

$$\psi_i \in \mathcal{B}$$

$$\psi_i = \mathbb{E}\big[\xi \mid [\phi_j, \xi] = \delta_{i,j} \text{ for } j \in \mathcal{I}\big]$$

Optimal recovery splines

ψ_i: Unique minimizer of

$$\begin{cases} \text{Minimize} & \|\psi\| \\ \text{Subject to} & \psi \in \mathcal{B} \text{ and } [\phi_j, \psi] = \delta_{i,j}, \quad j = 1, \ldots, m \end{cases}$$

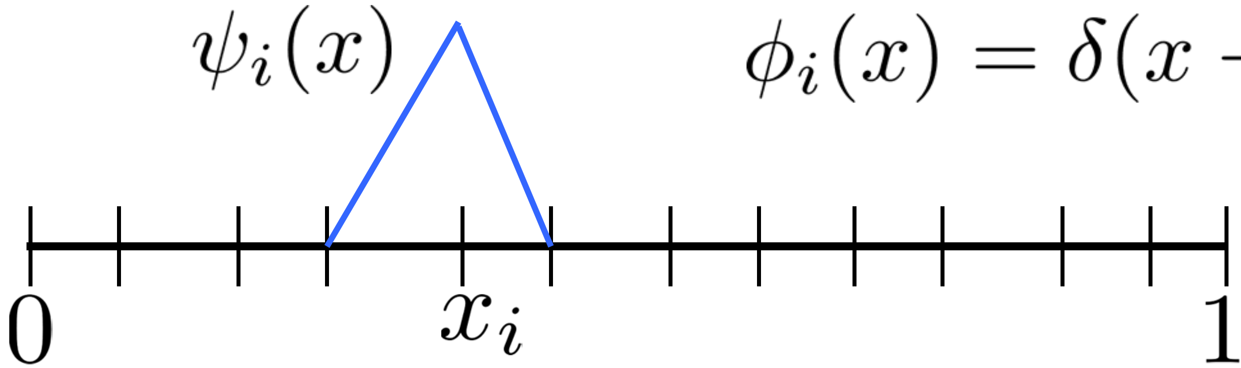
Example

$$\mathcal{B} := H^1(0,1)$$

$$\|u\|^2 = a\big(u(0)\big)^2 + b\int_0^1 \left(\frac{du}{dx}\right)^2 dx$$

$$\xi_t = \alpha\mathcal{N}(0,1) + \beta B_t$$

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



$$\mathbb{E}\Big[\xi(x)\Big|\xi(x_1) = f(x_1), \ldots, \xi(x_n) = f(x_n)\Big]$$

Piecewise linear interpolation of f

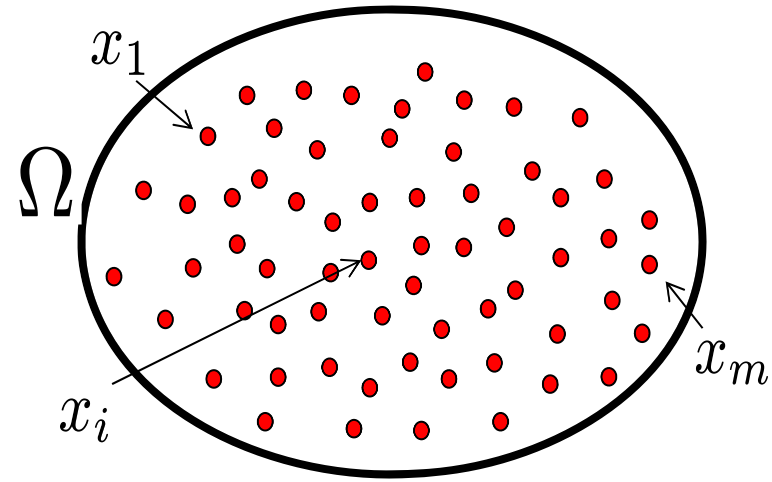
Example

$$\mathcal{B} := H_0^1(\Omega) \cap H^2(\Omega)$$

$$\|u\|^2 := \int_\Omega (\Delta u)^2$$

$$\mathcal{T} = \Delta^2 \qquad \xi \sim \mathcal{N}(0, \mathcal{T}^{-1})$$

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



ψ_i: Polyharmonic splines

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

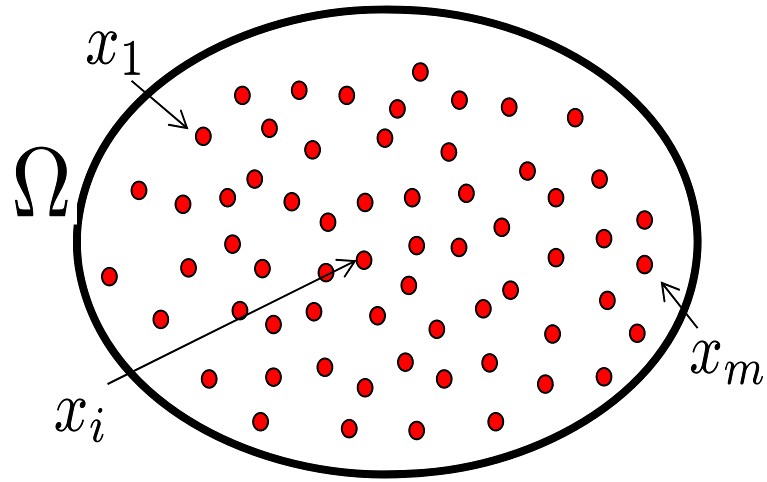
Example

$$\mathcal{B} := \{u \in H^1_0(\Omega) \mid \int_\Omega \big|\operatorname{div}(a\nabla u)\big|^2 < \infty\}$$

$$\|u\|^2 := \int_\Omega \big|\operatorname{div}(a\nabla u)\big|^2 \qquad a_{i,j} \in L^\infty(\Omega)$$

$$\mathcal{T} = (-\operatorname{div}(a\nabla\cdot))^2$$

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



ψ_i: Rough Polyharmonic splines

[Owhadi-Zhang-Berlyand 2013]

Thank you

- **Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis, 2017. arXiv:1703.10761. H. Owhadi and C. Scovel.**
- Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity, Schäfer, Sullivan, Owhadi. 2017.
- Multigrid with gamblets. L. Zhang and H. Owhadi, 2017
- Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients, 2016. H. Owhadi and L. Zhang. arXiv:1606.07686
- Multigrid with rough coefficients and Multiresolution operator decomposition from Hierarchical Information Games. H. Owhadi. SIAM Review, 59(1), 99149, 2017. arXiv:1503.03467
- Towards Machine Wald (book chapter). Houman Owhadi and Clint Scovel. Springer Handbook of Uncertainty Quantification, 2016, arXiv:1508.02449.
- Bayesian Numerical Homogenization. H. Owhadi. SIAM Multiscale Modeling & Simulation, 13(3), 812828, 2015. arXiv:1406.6668

DARPA EQUiPS / AFOSR award no FA9550-16-1-0054
(Computational Information Games)

Florian Schäfer

Clint Scovel

Tim Sullivan

Lei Zhang