Computational Information Games
A minitutorial Part I

Houman Owhadi

ICERM June 5, 2017

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

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

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]

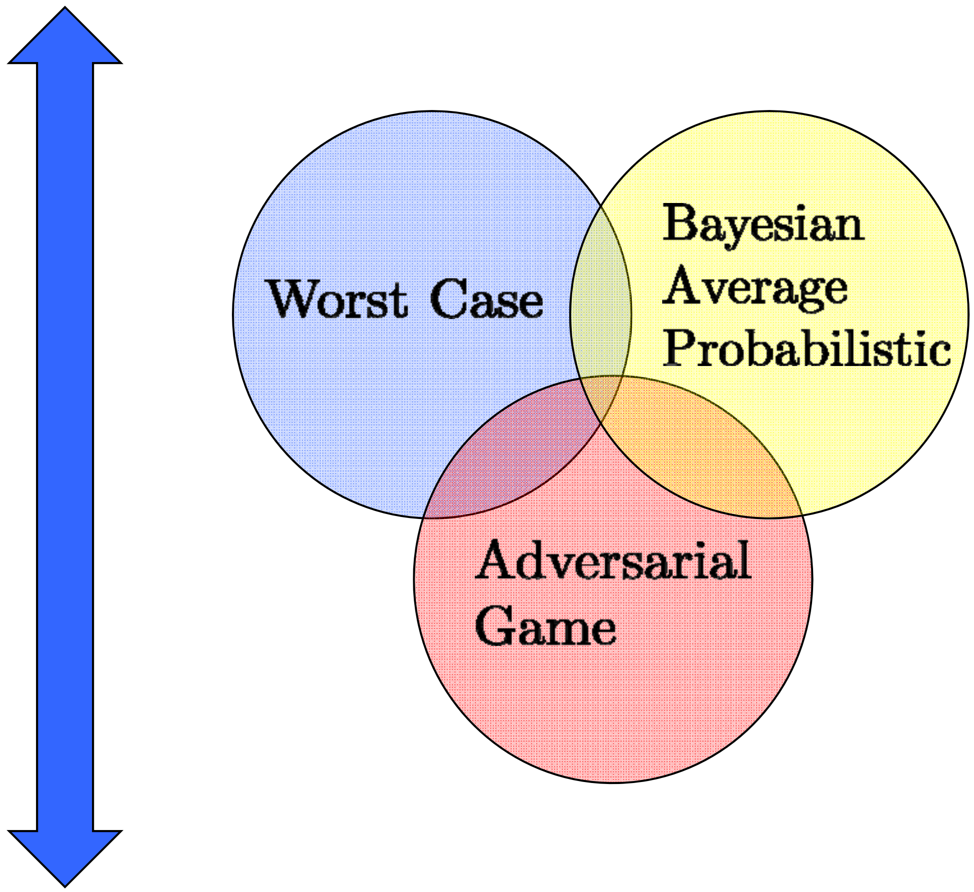
[Jon Cockayne, Chris Oates, Tim Sullivan, Mark Girolami. Bayesian Probabilistic Numerical Methods. arXiv:1702.03673, 2017]

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

http://probabilistic-numerics.org/

http://oates.work/samsi

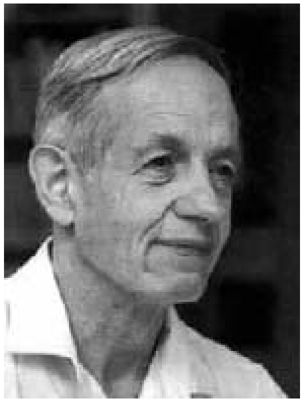
3 approaches to inference and to dealing with uncertainty



3 approaches to Numerical Approximation

Game theory

John Von Neumann



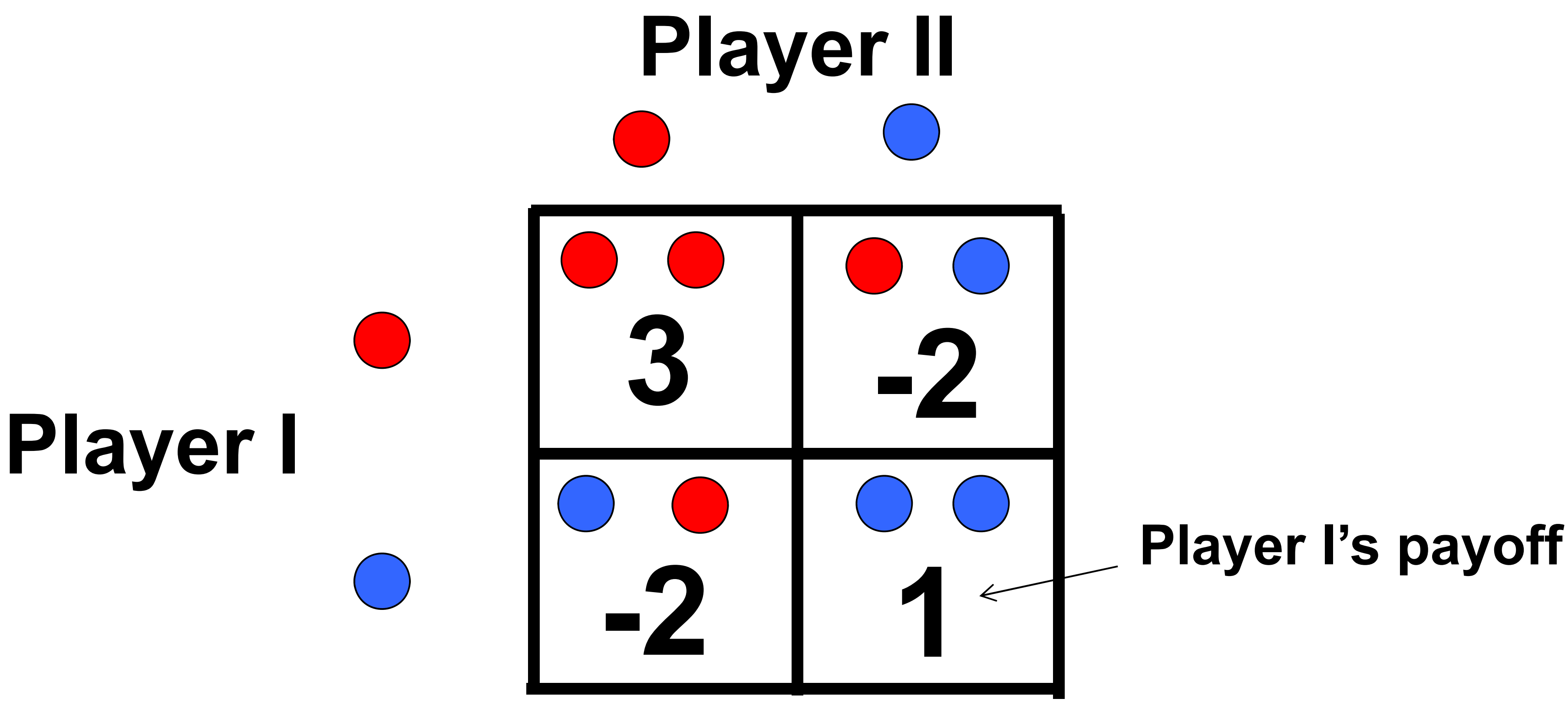
John Nash

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.

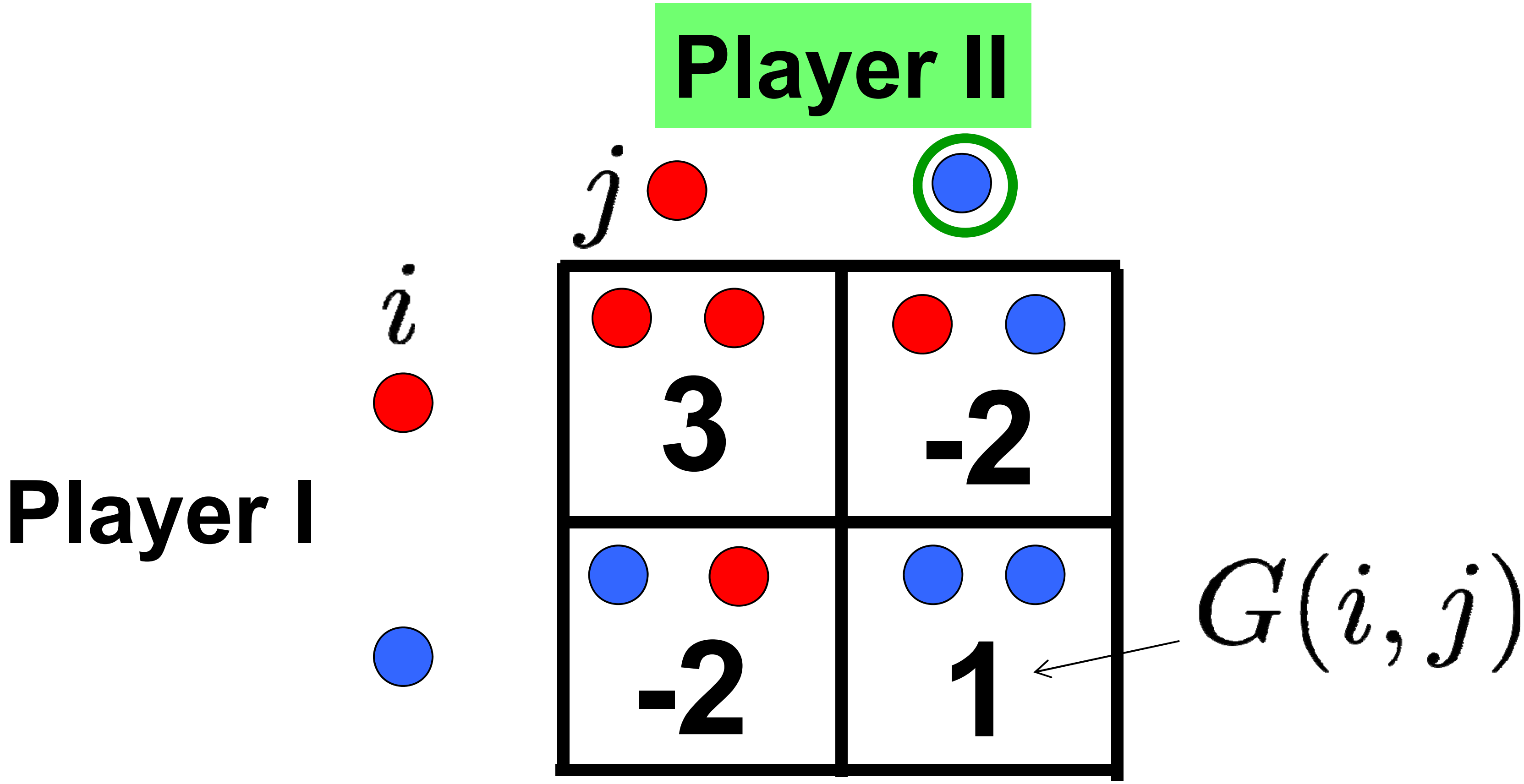
N. Nash. Non-cooperative games. *Ann. of Math.*, 54(2), 1951.

Deterministic zero sum game



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

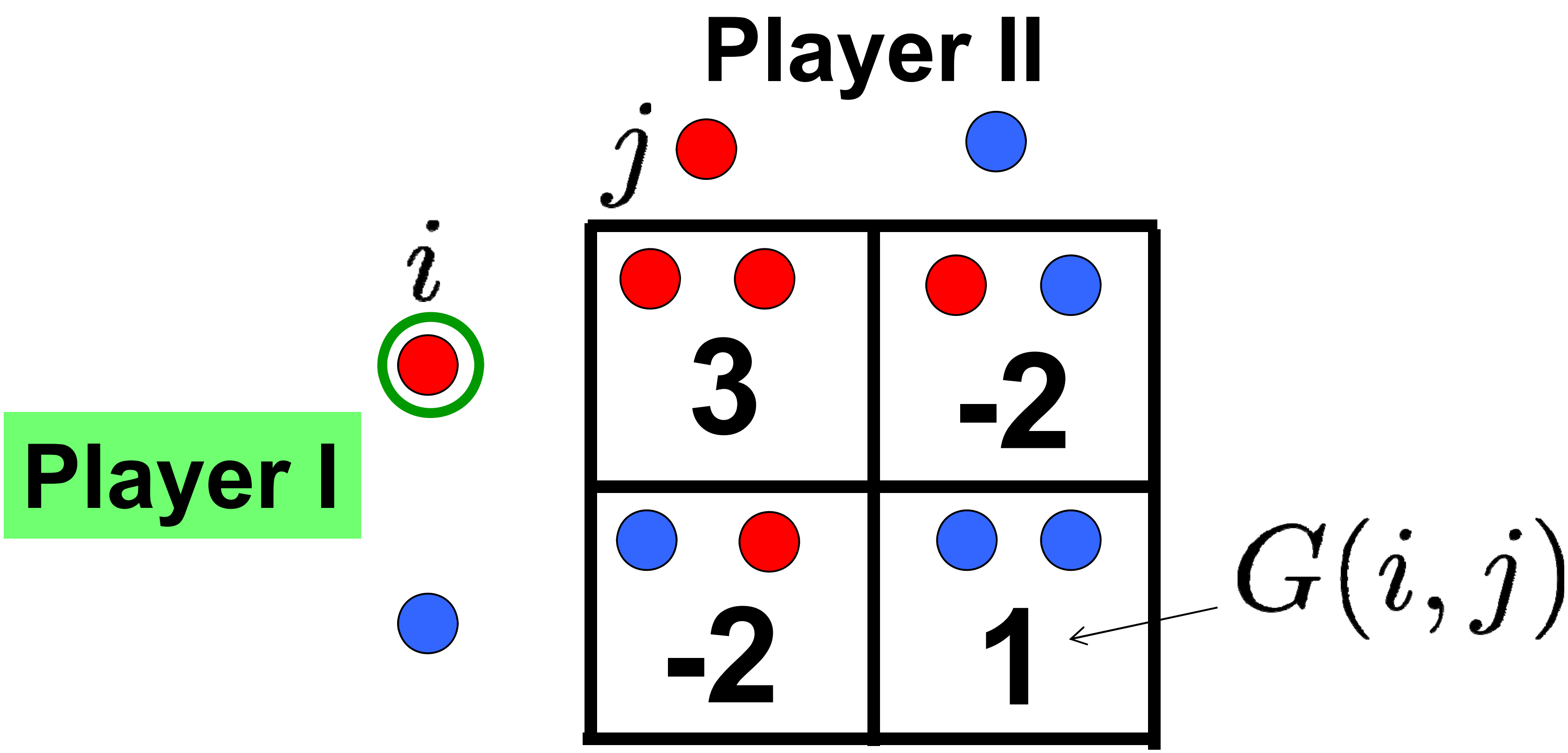
Worst case approach



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

II should play blue and lose 1 in the worst case

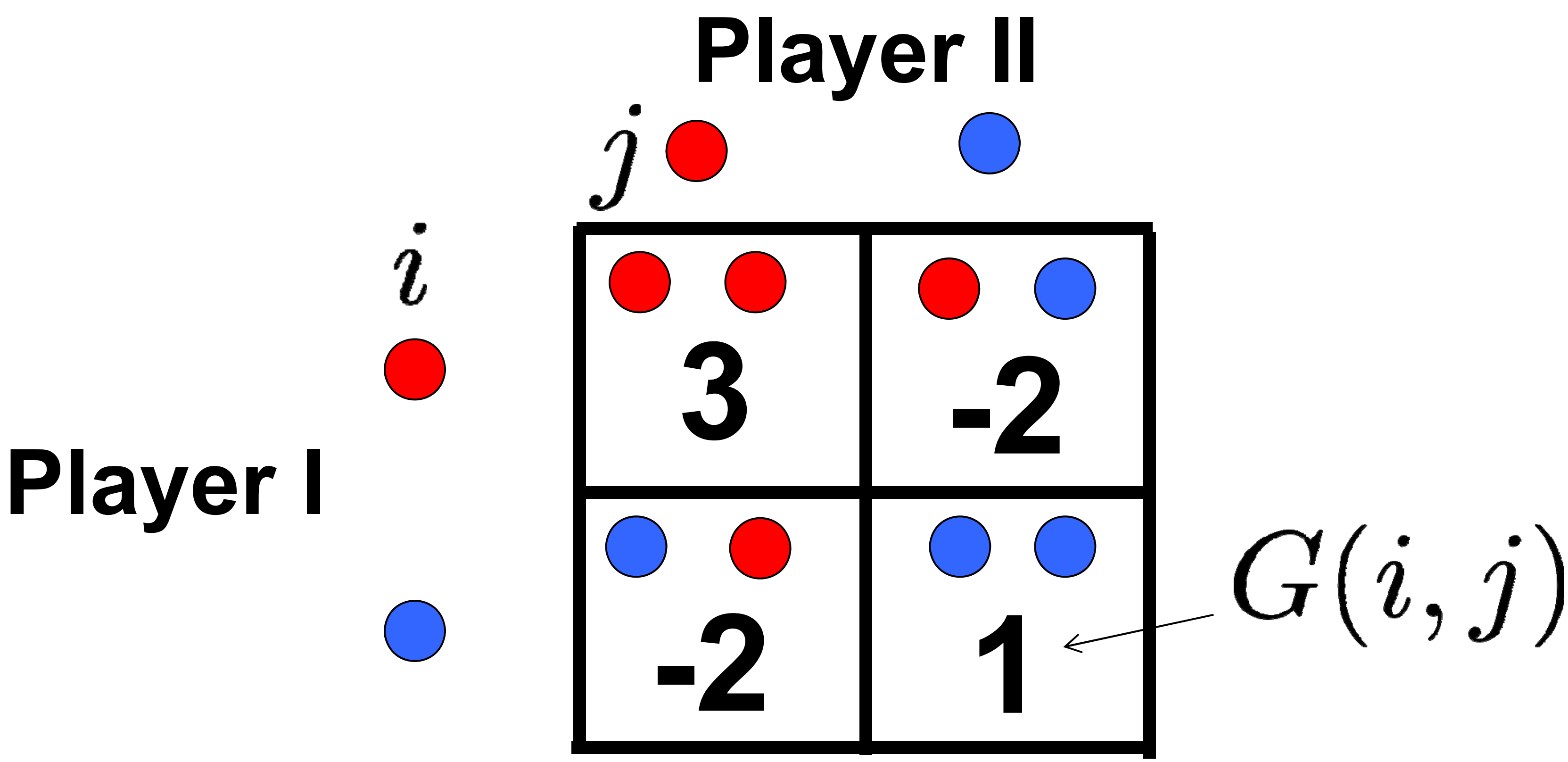
Worst case approach



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

I should play red and lose 2 in the worst case

No saddle point

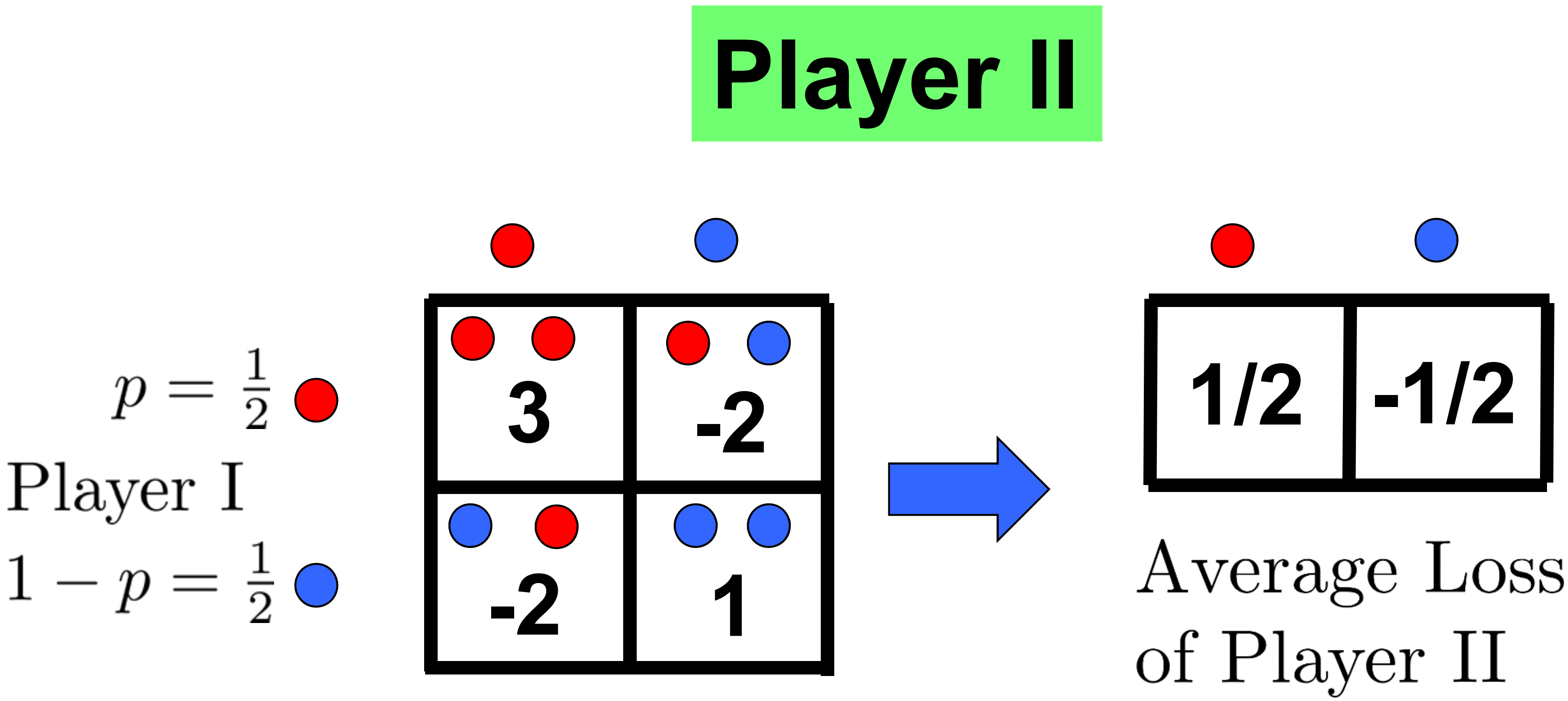


$$\max_i \min_j G(i,j) \neq \min_j \max_i G(i,j)$$

Not an equilibrium for a repeated game

Average case (Bayesian) approach

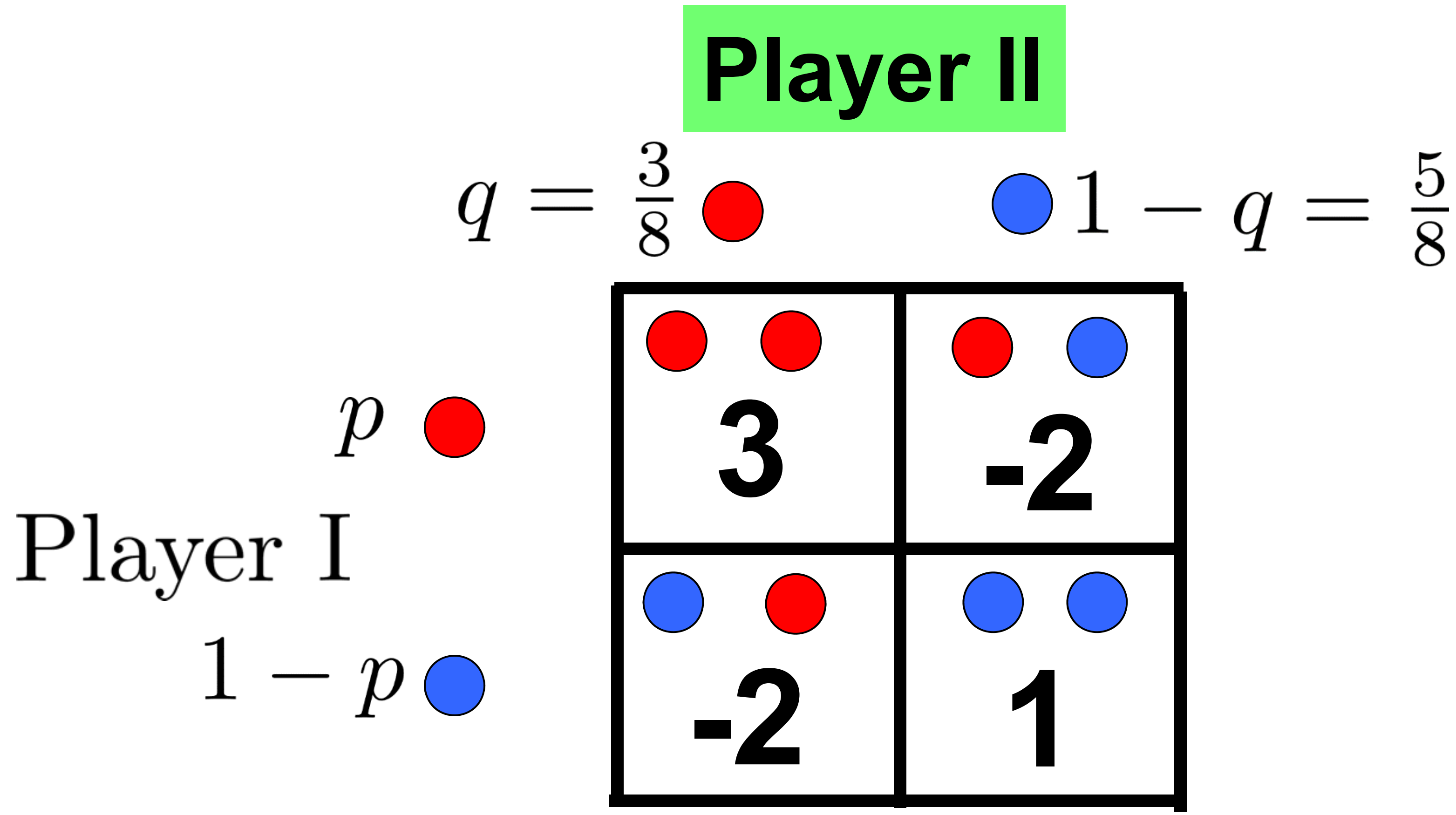
Place a uniform prior on the choice of Player I



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 II

q $\qquad\qquad$ $1-q$

Player I

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

Player I

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

	3	**-2**
	-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

Player II

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

John Von Neumann

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

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

Player II

$q = \frac{3}{10}$ $\quad$ $1 - q = \frac{7}{10}$

Player I

	$q = \frac{3}{10}$	$1 - q = \frac{7}{10}$
p	**5**	**-2**
$1 - p$	**-2**	**1**

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

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

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

Bayesian/probabilistic approach 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]

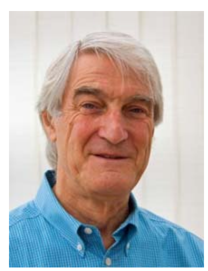
[J. E. H. Shaw. A quasirandom approach to integration in Bayesian statistics. Ann. Statist, 1988.]

[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

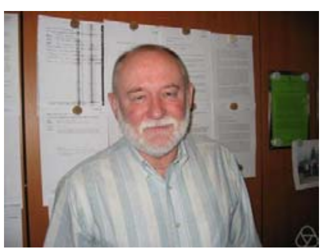
[H. Woźniakowski. Probabilistic setting of information-based complexity. J. Complexity, 1986.]

[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]

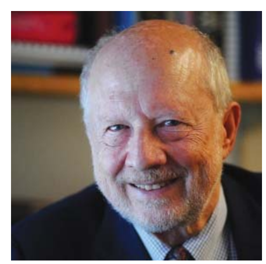
[J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. Information-based complexity. 1988]

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



H. Wozniakowski

G. W. Wasilkowski



J. F. Traub

E. Novak

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

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]

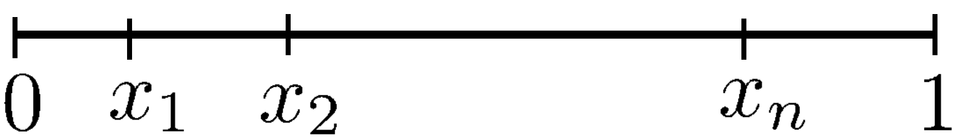
$$f(x) = \exp\left(\cosh\left(\frac{x^2+\sin(x)}{3+\cos(x^3)}\right)\right)$$

Compute

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

Bayesian Approach

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



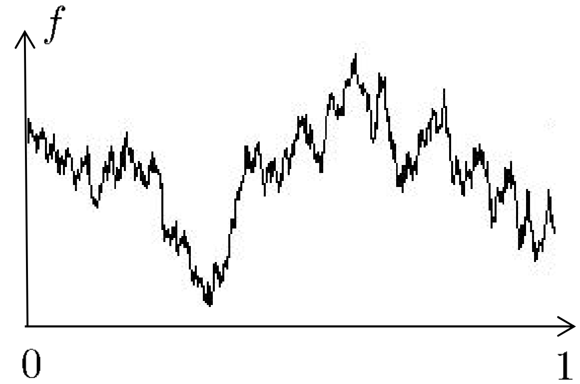
- Compute

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

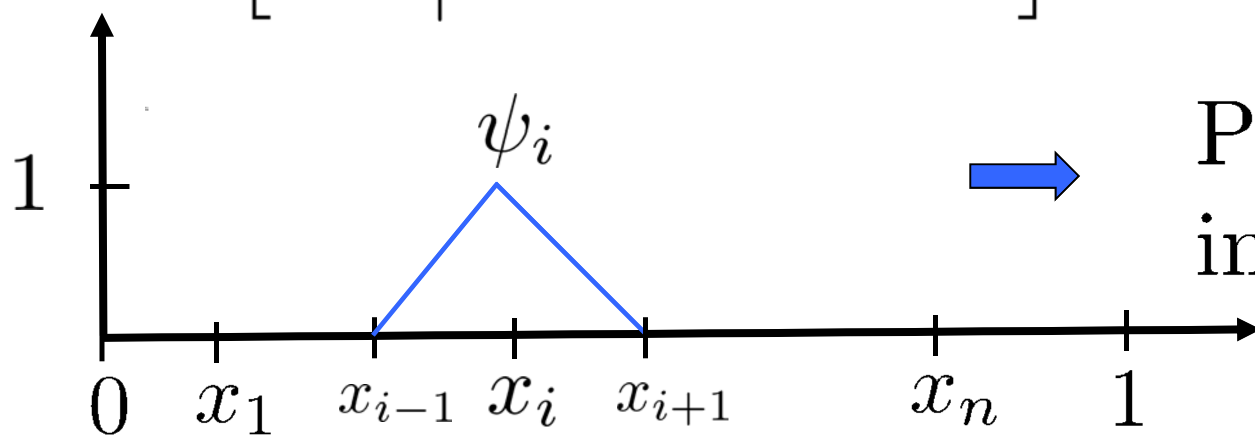
E.g.

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

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



$$\mathbb{E}\left[f(x) \middle| f(x_1), \ldots, f(x_n) \right] = \sum_i f(x_i)\psi_i(x)$$



→ Piecewise linear interpolation of f

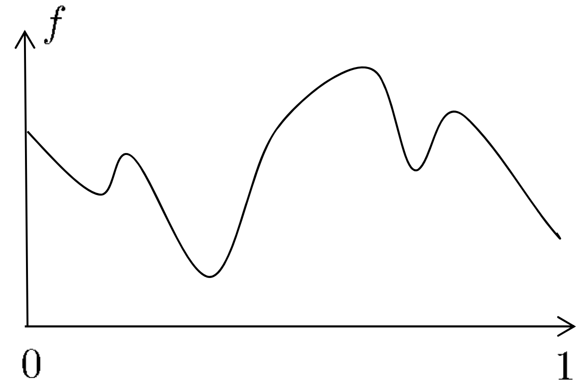
$$\mathbb{E}\left[\int_0^1 f(x)\,dx \middle| f(x_1), \ldots, f(x_n) \right]$$

→ 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 PDEs and Bayesian Inference?

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

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

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

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

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

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

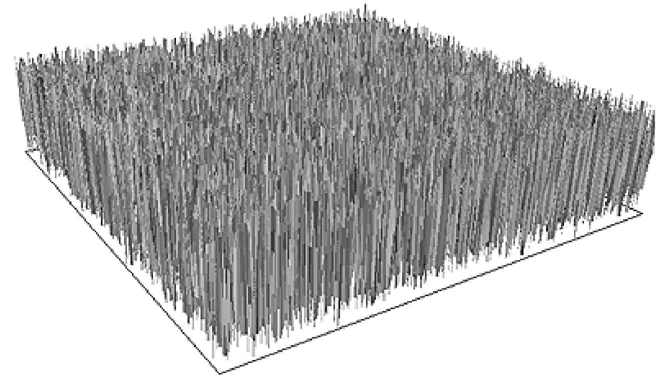
Proposition

- 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 $\mathcal{N}\left(0, \|f\|^2_{L^2(\Omega)}\right)$

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

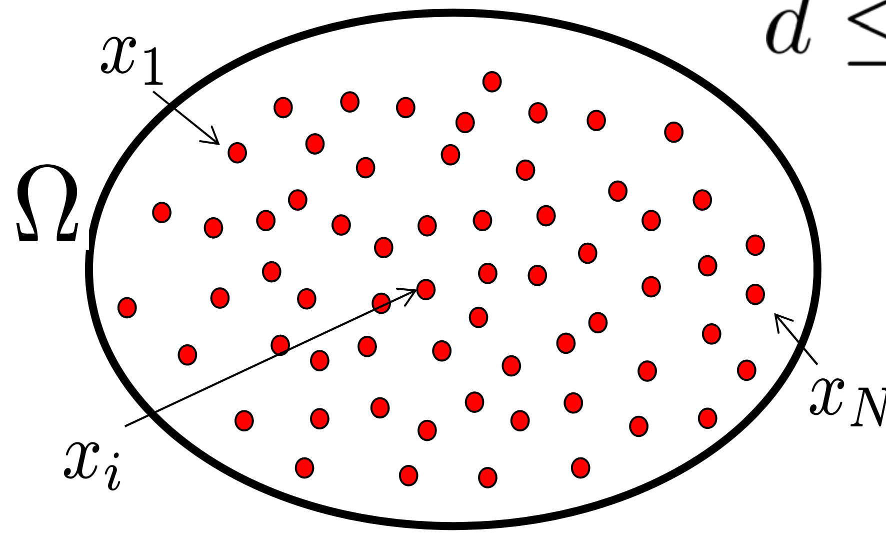
[J. Cockayne, C. J. Oates, T. Sullivan, and M. A. Girolami. Probabilistic meshless methods for partial differential equations and bayesian inverse problems. arXiv:1605.07811, 2016]

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

Let

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

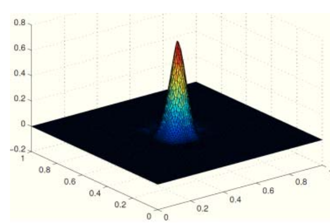
$d \leq 3$



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

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

$\sigma^2(x)$: 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

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

Summary

The Bayesian approach leads to old and new quadrature rules.

Statistical errors seem to imply/control deterministic worst case errors

Questions

- **Why does it work?**
- **How far can we push it?**
- **What are its limitations?**
- **How can we make sense of the process of randomizing a known function?**

$$\mathcal{B}_1 \xrightarrow{\mathcal{L}} \mathcal{B}_2$$

$$u \longrightarrow g$$

Direct Problem Given u find g

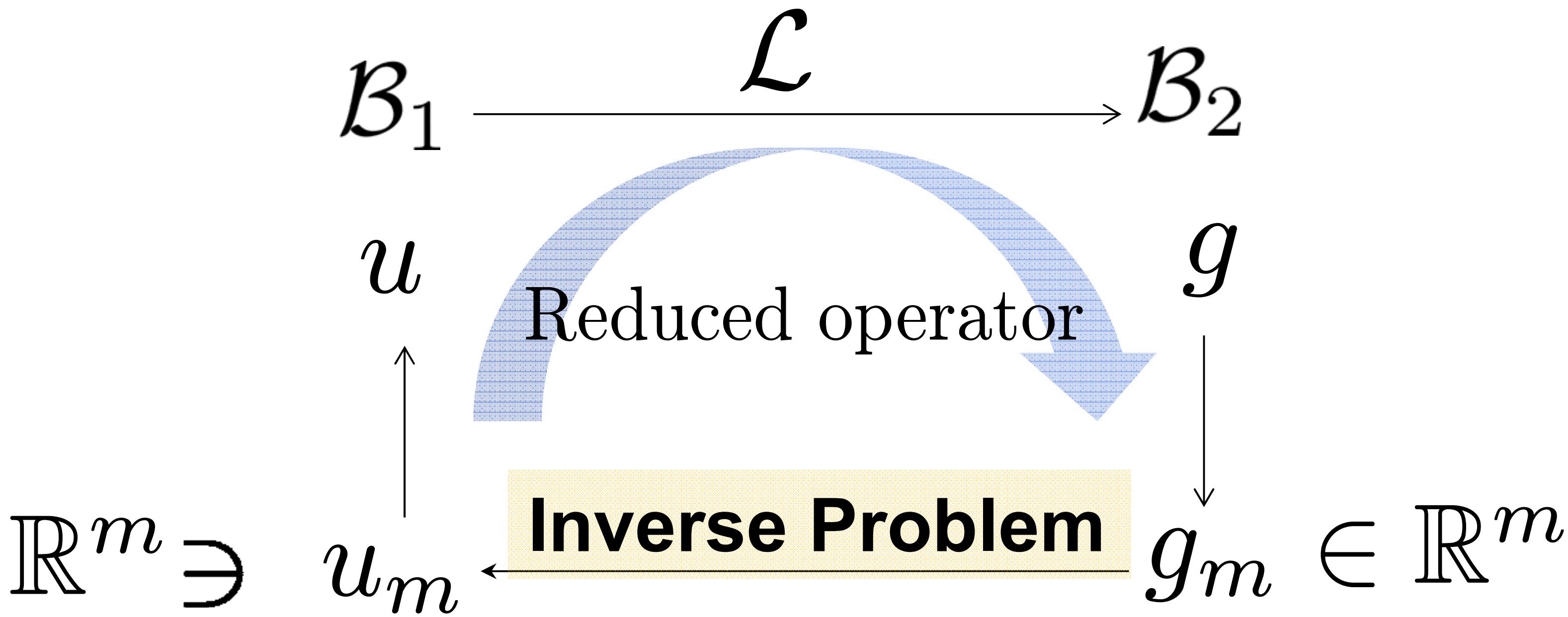
Inverse Problem Given g find u

u and g live in infinite dimensional spaces

⇓

Direct computation is not possible

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



Numerical implementation requires computation with partial information.

$$\phi_1, \ldots, \phi_m \in \mathcal{B}_1^*$$

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

$$u_m \in \mathbb{R}^m \xrightarrow{\text{Missing information}} u \in \mathcal{B}_1$$

Fast Solvers

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]

Robust/Algebraic multigrid

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

Stabilized Hierarchical bases, Multilevel preconditioners

[Vassilevski - Wang, 1997, 1998] [Chow - Vassilevski, 2003]
[Panayot - Vassilevski, 1997] [Aksoylu- Holst, 2010]

Low rank matrix decomposition methods

Fast Multipole Method: [Greengard and Rokhlin, 1987]

Hierarchical Matrix Method: [Hackbusch et al., 2002] [Bebendorf, 2008]:

Common theme between these methods

Computation is done with partial information over hierarchies of levels of complexity

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

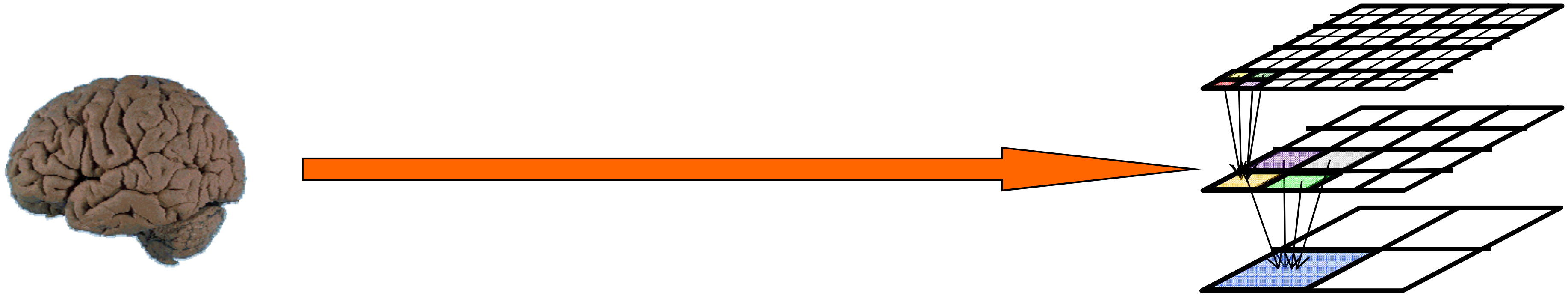
Restriction ↓ ↑ **Interpolation**

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

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

To compute fast we need to compute with partial information

The process of discovery of interpolation operators is based on intuition, brilliant insight, and guesswork



$$u_m \xrightarrow{\boxed{\text{Missing information}}} u$$

Problem

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

This is one entry point for statistical inference into Numerical analysis and algorithm design

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

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. $\int \|x\|^2 \mu(dx) < \infty$

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

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

$v(y)$ 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

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

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

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

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

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

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

Player II

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

$$\mathbb{R}^m \longrightarrow \mathbb{R}^n$$

$$y \longrightarrow x$$

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.

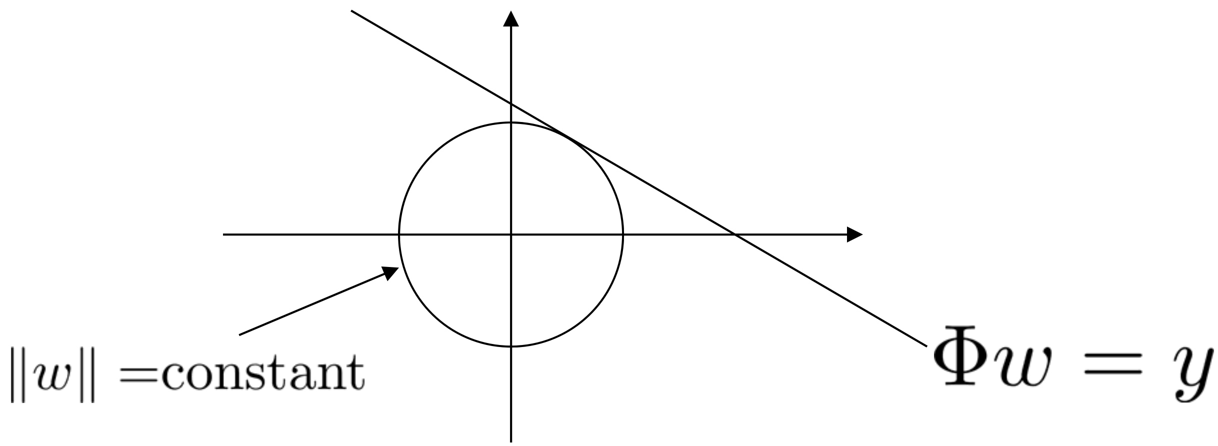
A. Wald. An essentially complete class of admissible decision functions. *Ann. Math. Statistics*, 18:549–555, 1947.

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 $\frac{e^{-\frac{\|w\|^2}{2}}}{C}$



$v^\dagger(y)$ 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 $u, v \in \mathcal{B}$,

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

Examples

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

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

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

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

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

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

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

Canonical Gaussian field

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

$\|\cdot\|$: Quadratic norm

$$\Updownarrow$$

$$\xi\colon\ \mathcal{B}^* \longrightarrow \mathcal{H}$$

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

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

$$\mathbb{E}\big[[\varphi, \xi][\phi, \xi]\big] = \big\langle \varphi, \phi \big\rangle_* \qquad \|\phi\|_* := \sup_{v \in \mathcal{B}} \frac{[\phi, v]}{\|v\|}$$

Canonical Gaussian field

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

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

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

$\Updownarrow$

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

For $\varphi, \phi \in \mathcal{B}^*$

$$\mathbb{E}\big[[\varphi, \xi][\phi, \xi]\big] = [\varphi, \mathcal{T}^{-1}\phi]$$

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

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

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

The recovery problem at the core of Algorithm Design and Numerical Analysis

To compute fast we need to compute with partial information

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

Restriction ↓ ↑ **Interpolation**

$$u_m = ([\phi_1, u], \ldots, [\phi_m, u])$$
$$\phi_1, \ldots, \phi_m \in \mathcal{B}^*$$

$$u_m \xrightarrow{\boxed{\text{Missing information}}} u$$

Problem

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

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

Player I

Chooses $u \in \mathcal{B}$

Max

Player II

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

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

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

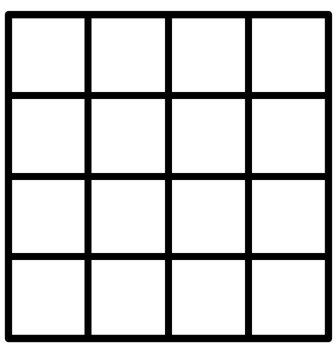
Sees $(\phi_1^T x, \ldots, \phi_m^T x)$

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

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

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

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

Theorem

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

But

$$\text{But no saddle point if } \dim(\mathcal{B}) = \infty$$

$$\text{No } \mu^* \in \mathcal{P}(\mathcal{B}) \text{ is achieving the } \sup_\mu$$

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

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

Definition

μ^*, v^* is a ϵ saddle point iff

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

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

Theorem

$$\exists\, v^\dagger : \mathbb{R}^m \to \mathcal{B}$$

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

$$\exists\, \mu^\dagger \text{ cylinder measure on } \mathcal{B}$$

s.t. for all μ, v

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

and

$$\mu_n \xrightarrow[n\to\infty]{w} \mu^\dagger$$

⇕

For all s

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

Theorem The optimal mixed strategy for Player I is

$$\mu^\dagger \Longleftrightarrow u \sim \xi - \mathbb{E}\big[\xi \mid ([\phi_i,\xi])_{i\in\mathcal{I}}\big]$$

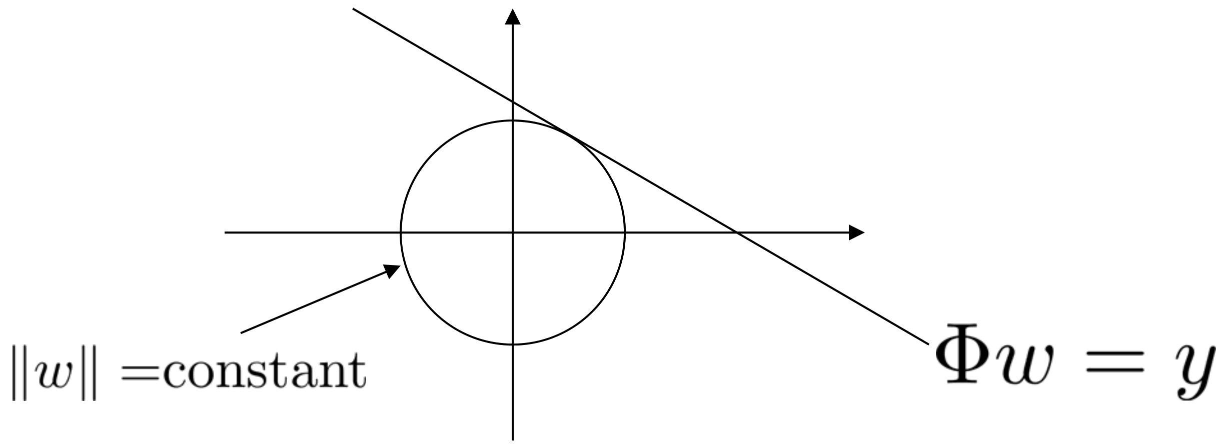
The optimal strategy for Player II is

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

$$\mathbb{R}^m \longrightarrow \mathcal{B}$$

$$([\phi_1,u],\ldots,[\phi_m,u]) \longrightarrow u$$

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



$v(y)$: Minimizer of

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

Game theoretic solution = Worst case solution

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

Optimal Recovery Solution

$v(y)$: 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.

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

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

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

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

Gamblets = Optimal Recovery Splines

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

Optimal Recovery Splines

ψ_i is the minimizer of

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

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.

Dual bases

ψ_i is the minimizer of

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

ϕ_i is the minimizer of

$$\begin{cases} \text{Minimize } \|\phi\|_* \\ \text{Subject to } \phi \in \mathcal{B}^* \text{ and } [\phi, \psi_j] = \delta_{i,j} \text{ for } j \in \mathcal{I} \end{cases}$$

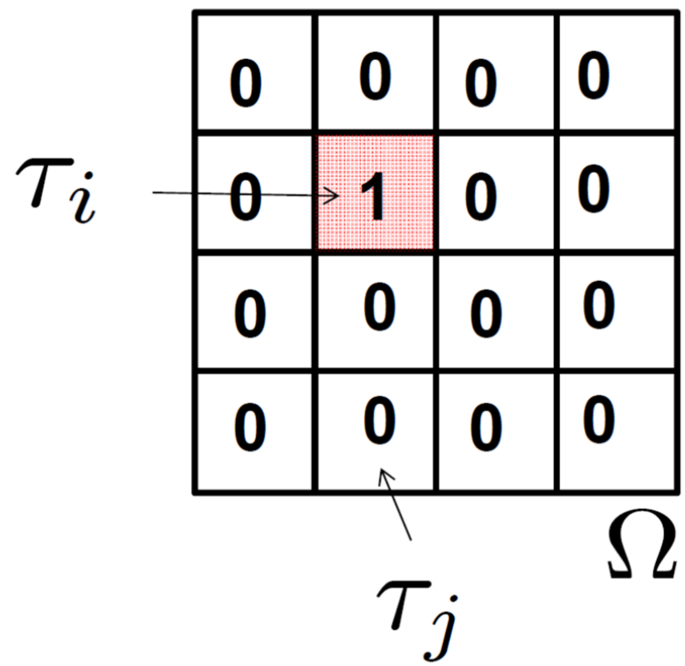
Example

$$\mathcal{B} := H_0^1(\Omega) \qquad \|u\|^2 := \int_\Omega (\nabla u)^T a \nabla u$$

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

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

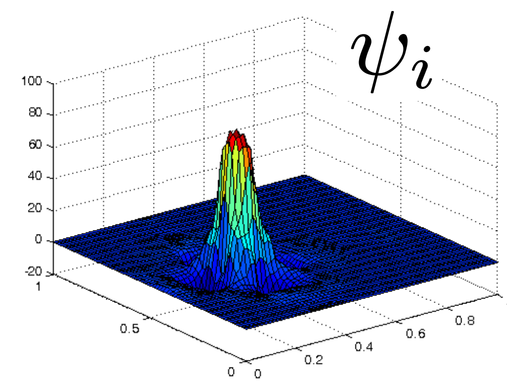
$$\begin{cases} -\operatorname{div}(a\nabla u) = g, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases}$$



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

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

ψ_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 $j \neq i$

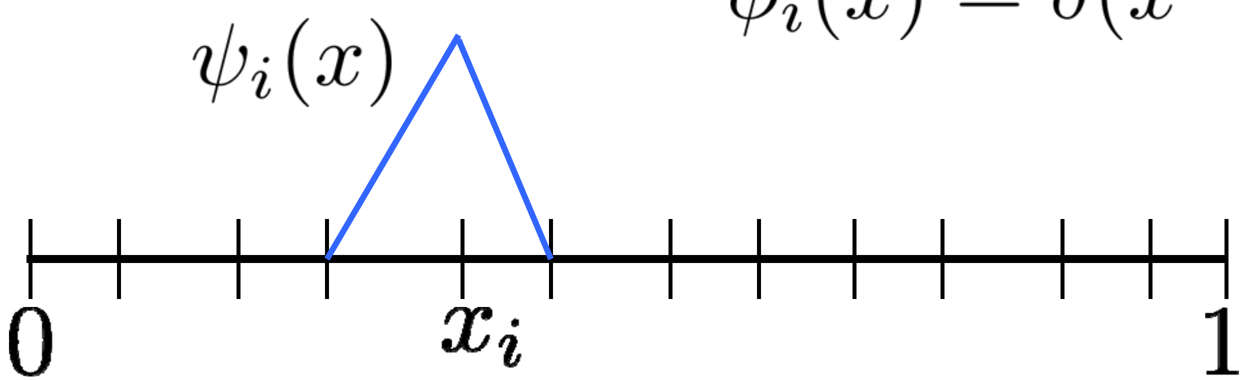


Example

$$\mathcal{B} := H_0^1(0,1) \qquad \|u\|^2 = \int_0^1 \left(\frac{du}{dx}\right)^2 dx$$

$$\mathcal{T} = -\frac{d^2}{(dx)^2} \qquad \xi\text{: Brownian bridge}$$

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



Example

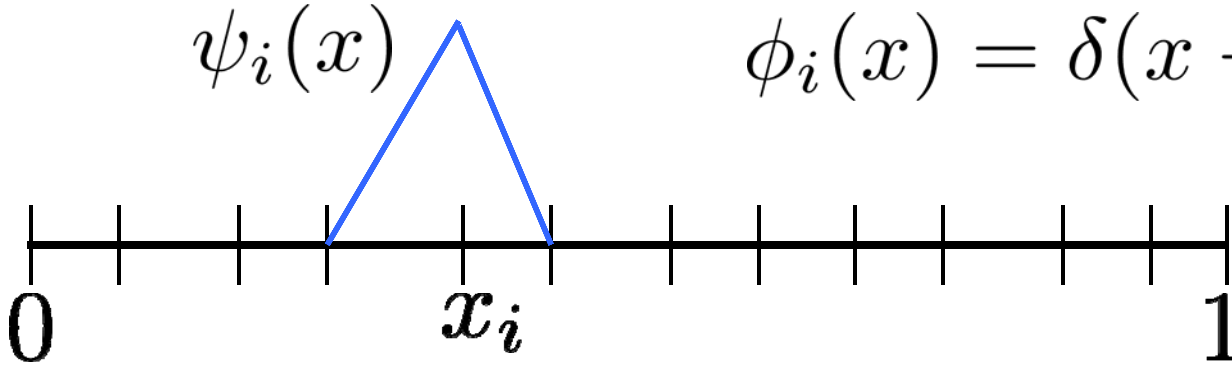
$$\mathcal{B} := H^1(0,1)$$

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

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

$\psi_i(x)$

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



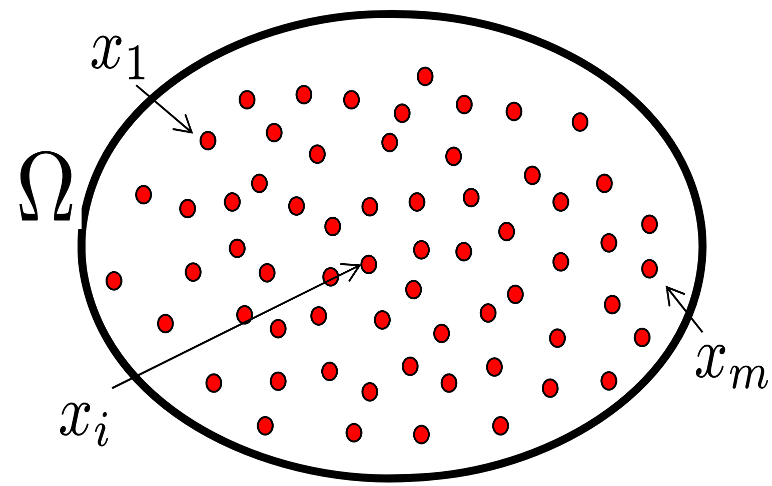
$$\mathbb{E}\Big[\xi(x)\Big|\xi(x_1) = f(x_1), \ldots, \xi(x_n) = f(x_n)\Big]$$ ➡ Piecewise linear interpolation of f

Example

$$\mathcal{B} := H_0^1(\Omega) \cap H^2(\Omega) \qquad \|u\|^2 := \int_\Omega (\Delta u)^2$$

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

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



ψ_i: Polyharmonic splines

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

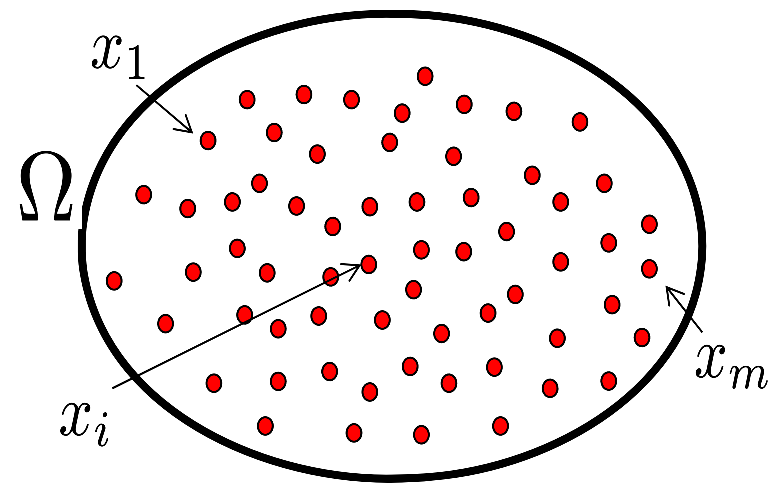
Example

$$\mathcal{B} := \{u \in H_0^1(\Omega) \mid \int_\Omega \left|\operatorname{div}(a\nabla u)\right|^2 < \infty\}$$

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

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

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



ψ_i: Rough Polyharmonic splines

[Owhadi-Zhang-Berlyand 2013]

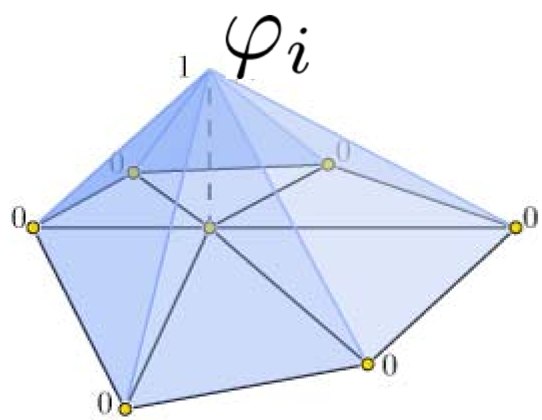
Example

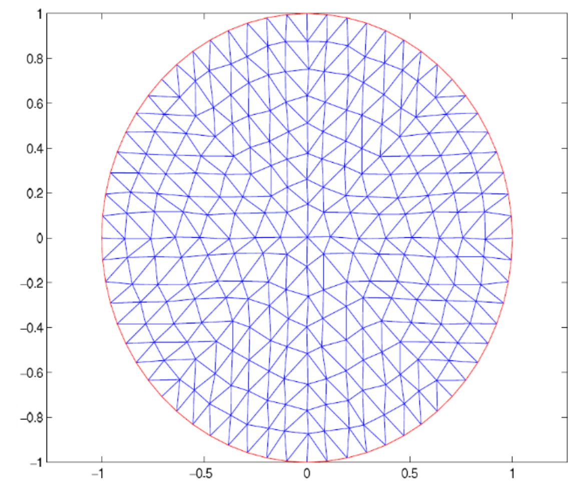
$$\mathcal{B} := H_0^1(\Omega)$$

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

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

$$\phi_i = \sum_j M_{i,j}^{-1} \varphi_j$$





Ω

ψ_i: LOD basis

A. Målqvist and D. Peterseim. Localization of elliptic multiscale problems. *Mathematics of Computation*, 83(290):2583–2603, 2014.

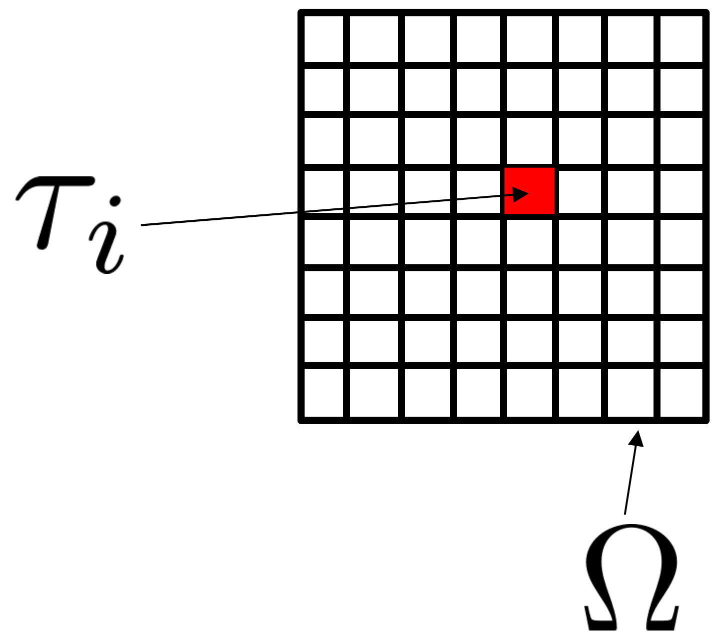
Example

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

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

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

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



H. Owhadi. Multigrid with Rough Coefficients and Multiresolution Operator Decomposition from Hierarchical Information Games. *SIAM Rev.*, 59(1):99–149, 2017. arXiv:1503.03467, 2015.

Summary

- Bayesian numerical analysis ``works” because
Gaussian priors form the optimal class of priors when losses are defined using quadratic norms and measurements are linear
- The game theoretic solution is equal to the classical worst case optimal recovery solution under above questions
- The canonical Gaussian field contains all the required information to bridge scales/levels of complexity in numerical approximation
and it does not depend on the linear measurements.

Questions

- Does the canonical Gaussian field remain optimal (or near optimal) beyond average relative errors (e.g. rare events/large deviations) or when measurements are not linear. This is a fundamental question if probabilistic numerical errors are to be merged with model errors in a unified Bayesian framework.
- What are the properties of gamblets?
- Can the game theoretic approach help us solve known open problems in numerical analysis and algorithm design?

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

Florian Schäfer

Clint Scovel

Tim Sullivan

Lei Zhang

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