The game theoretic approach to Uncertainty Quantification, reduced order modeling and numerical analysis

Houman Owhadi[*]

We discuss the development of Uncertainty Quantification framework founded upon a combination of game/decision theory and information based complexity. We suggest that such a framework could be used not only to guide decisions in presence of epistemic uncertainties and complexity management capabilities constraints but also to automate the process of discovery in (1) model form uncertainty quantification and design (2) model reduction (3) the design of fast, robust and scalable numerical solvers. Although these applications appear dissimilar, they are all based on the efficient processing of incomplete information with limited computational resources: (1) model form UQ and design require the management and processing of epistemic uncertainties and limited data (2) model reduction requires the approximation of the full state of a complex system through operations performed on a few (coarse/reduced) variables (3) fast and robust computation requires computation with partial information. The core idea of the proposed framework is to reformulate the process of computing with partial information and limited resources as that of playing underlying hierarchies of adversarial information games characterizing the adversarial and nested processing of hierarchies of partial/missing information.

I.   Introduction

The purpose of this paper is to review and discuss the development of the game the game/decison theoretic approach to UQ and numerical analysis presented in.[15,16] The proposed approach is built upon a generalization of Optimal Uncertainty Quantification,[17] Von Neumann's Game Theory,[25] Nash's non-cooperative games[9,10] and Wald's Decision Theory.[26]

We will discuss the practical applications of this framework to (1) the design of fast, robust and scalable numerical solvers and the analysis of partial differential equations (2) model form uncertainty quantification and design (3) model reduction. Although these applications may appear, at first glance, dissimilar, they are all based on the efficient processing of incomplete information with limited computational resources (i.e., complexity management capabilities constraints): (1) fast and robust computation requires computation with partial information (this concept forms the core of Information Based Complexity[24] where it is also augmented by concepts of contaminated and priced information associated with, for example, truncation errors and the cost of numerical operations) (2) model form UQ and design require the management and processing of epistemic uncertainties and limited data (3) model reduction requires the approximation of the full state of a complex system through operations performed on a few (coarse/reduced) variables. The core idea of this framework is to guide this process through its reformulation as an adversarial game with respect to the missing information. For instance, in the model reduction application, this game would oppose Player I who must choose the full state of the system (in the admissible set defined by what the missing information could be) and Player II who must approximate that state based on the observed partial information (e.g., the values of the coarse/reduced variables). In the proposed adversarial game Player I would try to maximize the approximation error, while Player II would try to minimize it. One aspect of this framework is its application to the identification of hierarchies of discrete and finite operators enabling the approximation of arbitrary models/functions by nested hierarchies of discrete and finite (reduced) models/functions acting at different

---
[*]Professor of Applied and Computational Mathematics and Control and Dynamical Systems, California Institute of Technology, Computing & Mathematical Sciences , MC 9-94 Pasadena, CA 91125, owhadi@caltech.edu.

levels of complexity (partial information). These concepts are implicitly present in multigrid/multiresolution methods where a given PDE must be approximated at different levels of complexity (scales) and fast and accurate operators enabling the exchange of information between successive levels of complexity (scales) must be identified. An example of application of this framework, that will be discussed, can be found in[15] where the problem of finding fast scalable solvers for PDEs with rough coefficients (i.e. those associated with transport in porous media) is turned into an algorithm by reformulating the underlying process of discovery as non cooperative hierarchical information games. This application could be seen as part of the (re)emerging field of statistical approaches to numerical approximation.[1–8, 11–14, 19–23, 27, 28]

II.   The worst case approach to UQ

As understood in the popular adage, "When in doubt, assume the worst!" the method of reasoning with "the worst that may befall" offers a rational framework to making decisions under uncertainty. Let us observe that there are essentially two worst case approaches to UQ. The first one is the Robust Optimization approach in which $u^\dagger$ is an unknown element of an admissible/ambiguity set $\mathcal{A}$, Φ is some quantity of interest mapping elements $u \in \mathcal{A}$ onto $\Phi(u) \in \mathbb{R}$, and one is interested in estimating $\Phi(u^\dagger)$. The Robust Optimization approach (which is essentially a min and max approach and the framework employed in Optimal Uncertainty Quantification[17]) then seeks to compute best and worst case scenarios corresponding to optimal bounds on $\Phi(u^\dagger)$, i.e.

$$\inf_{u \in \mathcal{A}} \Phi(u) \leq \Phi(u^\dagger) \leq \sup_{u \in \mathcal{A}} \Phi(u) \tag{1}$$

We refer to Subsection A of the appendix for an illustration of the Optimal Uncertainty Quantification method applied to the seismic safety study of a truss structure.

The second worst case approach is the game theoretic approach illustrated in Figure 1. The problem is to estimate some quantity of interest $\Phi(\mu^\dagger)$ depending on an unknown probability distribution $\mu^\dagger$ based on the information that $\mu^\dagger$ belongs to the admissible/ambiguity set $\mathcal{A}$ and the observation of $d = (d_1, \ldots, d_n) \sim (\mu^\dagger)^n$ (i.e. n i.i.d. data points sampled from $\mu^\dagger$).

As illustrated in Figure 1 the game theoretic approach provides a rational framework to the problem of making designing an optimal model or making an optimal prediction. In this approach we have two players (I and II) in an adversarial game. Player I chooses $\mu \in \mathcal{A}$ and Player II chooses a function θ of the data d (a model whose value $\theta(d)$ against the data provides an estimation of the quantity of interest $\Phi(\mu)$) and receives a loss $\mathcal{E}(\mu, \theta)$ (e.g. a mean squared error, a confidence error or another statistical error). The objective of Player I is to maximize Player II's loss and the objective of Player II is to minimize it. Although Player I
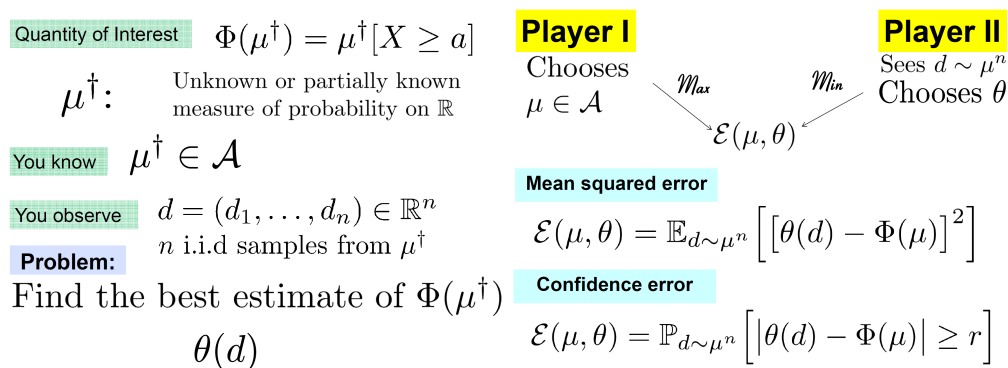

Figure 1. The game theoretic approach to finding the reservoir model and incorporating historical data.

can be a real player (e.g. a group of engineers trying to quantify the robustness/confidence of a model that another group has designed), it can also be a conceptual one used to formalize the process of playing against one's lack of information and finding an optimal model.

What are the optimal strategies for such games? A remarkable result of decision theory,[26] stemming from Von Neumann's Game Theory[25] (see also Nash's non-cooperative games[9, 10] is the complete class theorem[26] (see Figure 2) which states that optimal strategies are randomized strategies: The best strategy for Player

I is to play select a candidate for the real system at random by placing a prior probability measure π_I on $\mathcal{A}$ and the optimal strategy of Player II lives in the Bayesian class of estimators and obtained by identifying an optimal prior (as the solution of a minimax problem).
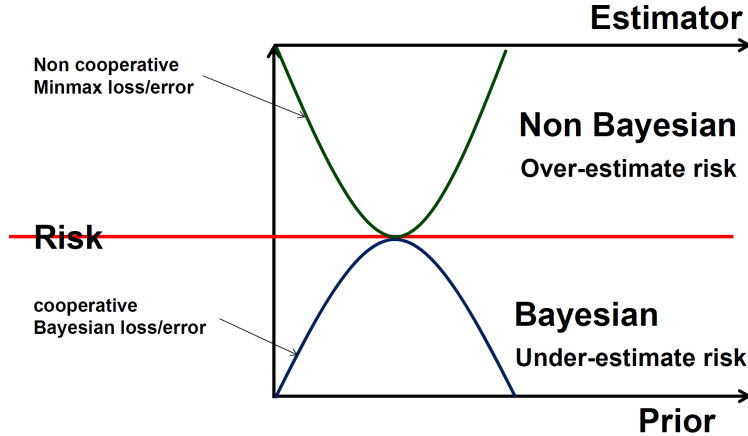

Figure 2. The complete class theorem. Bayesian strategies underestimate risk (the value of the minimax game), non Bayesian strategies overestimate risk. Optimal strategies are at the interface and obtained by identifying an optimal prior (obtained as the prior whose statistical loss is maximal in a cooperative game).

III. Illustration of the proposed approach in a simple numerical approximation problem.

A proof of concept of the efficacy of the game theoretic framework can be found in[15] where the problem of finding fast scalable solvers for PDEs with rough coefficients is turned into an algorithm by reformulating the underlying process of discovery as non cooperative hierarchical information games and produces the fastest known linear solver for PDEs with rough coefficients with a rigorous bound on the complexity.

To illustrate the proposed approach, we will describe here its application to a simple numerical approximation problem. In this application we want to approximate the solution x of the linear system

$$Ax = b \tag{2}$$

where A is a known $n \times n$ positive definite matrix and b is an unknown element of $\mathbb{R}^n$. We are given the information that

$$\Phi x = y \tag{3}$$

where Φ is a known $m \times n$ rank m matrix with $m < n$ and y is a known element of $\mathbb{R}^m$. We also given the compactness condition

$$b^T Q^{-1} b \leq 1 \tag{4}$$

where Q is a known $n \times n$ symmetric positive definite matrix. Observe that since $m < n$ we do not have sufficient measurements to recover x. Instead the measurements (3) define an ambiguity set or a set of candidates for x, illustrated in Figure 3 and defined by

$$\mathcal{A} = \left\{ z \in \mathbb{R}^n \mid \Phi z = y, \text{ and } |Az|_{Q^{-1}} \leq 1 \right\} \tag{5}$$

where we have used the notation $|b|_{Q^{-1}} := \sqrt{b^T Q^{-1} b}$

The classical numerical analysis minimax solution of the problem of recovering x is to approximate x with the minimizer z^* of

$$\min_{z^* \in \mathcal{A}} \max_{z \in \mathcal{A}} \|z - z^*\|. \tag{6}$$

where $\|\cdot\|$ is some norm on $\mathbb{R}^n$. Evidently the formulation (6) looks like that of an adversarial zero sum game (illustrated in Figure 3) where Player I chooses $z \in \mathcal{A}$ and Player II chooses $z^* \in \mathcal{A}$ and the loss of Player
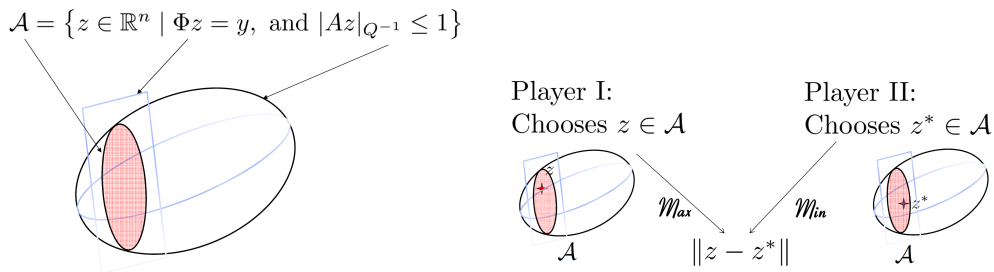
$$\mathcal{A} = \left\{ z \in \mathbb{R}^n \mid \Phi z = y, \text{ and } |Az|_{Q^{-1}} \le 1 \right\}$$

Player I:
Chooses $z \in \mathcal{A}$

Player II:
Chooses $z^* \in \mathcal{A}$

$\mathcal{M}_{ax}$ $\mathcal{M}_{in}$

$\mathcal{A}$ $\|z - z^*\|$ $\mathcal{A}$

Figure 3. The ambiguity set and the numerical analysis minimax problem.

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

$\min\max \ne \max\min$

Player II

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

Player I

J. Von Neumann

Player I's payoff

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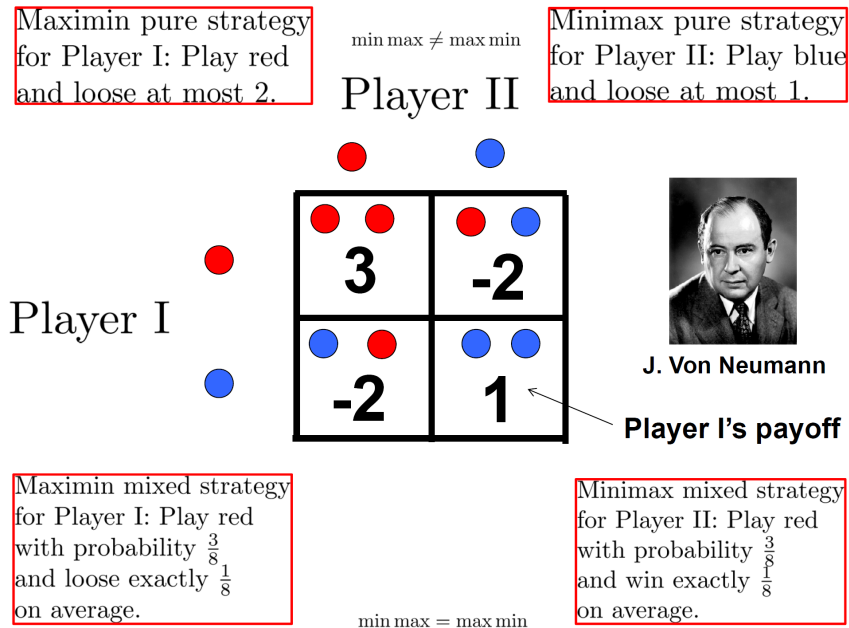
$\min\max = \max\min$

Figure 4. A simple adversarial zero sum game.

II is $\|z - z^*\|$. Observe that the solution z^* of (6) is not a saddle point since $\max_{z \in \mathcal{A}} \min_{z^* \in \mathcal{A}} \|z - z^*\| = 0$ and therefore $\min_{z^* \in \mathcal{A}} \max_{z \in \mathcal{A}} \|z - z^*\| \ne \max_{z \in \mathcal{A}} \min_{z^* \in \mathcal{A}} \|z - z^*\|$. Why is it important to have a saddle point? How should we play such a game?

To answer these questions let us consider the simpler deterministic adversarial zero sum game illustrated in Figure 4. In that game Player I and II both have a red and a blue marble and at the count of three they show each other a marble. If both are red Player I wins 3 points, if both are blue Player I wins 1 point and if the colors do not match then Player II wins 2 points (and Player I looses 2 points, i.e. since this is a zero sum game the loss of Player I is the gain of Player II and viceversa). How should I and II play the game? The pure strategy answer/solution is for Player I to play red and loose at most 2 (in the worst case) and the pure strategy solution for Player II is to play blue and loose at most 1. This is essentially the numerical analysis approach to obtaining a solution and it is easy to see that it is not a saddle point $\min\max \ne \max\min$. On the other hand, optimal solutions to the repeated game are mixed/randomized strategies and obtained by lifting the minimax problem to measures of probability over pure strategies. Henceforth, Player II's optimal mixed strategy is to play red with probability 3/8 and win exactly 1/8 on average (independently from the choice of Player I). Player I's optimal mixed strategy is to play red with probability 3/8 and loose 1/8 on average (independently from the choice of Player II). This can be seen by computing the average gain of Player I which, assuming that Player II plays red with probability q and blue with probability $1 - q$ and

Player I plays red with probability p and blue with probability $1 - p$, is $1 - 3q + p(8q - 3)$ (which is equal to $-1/8$ from $q = 3/8$ independently from the value of p). Note that these optimal mixed strategies form a saddle point ($\min \max \neq \max \min$).
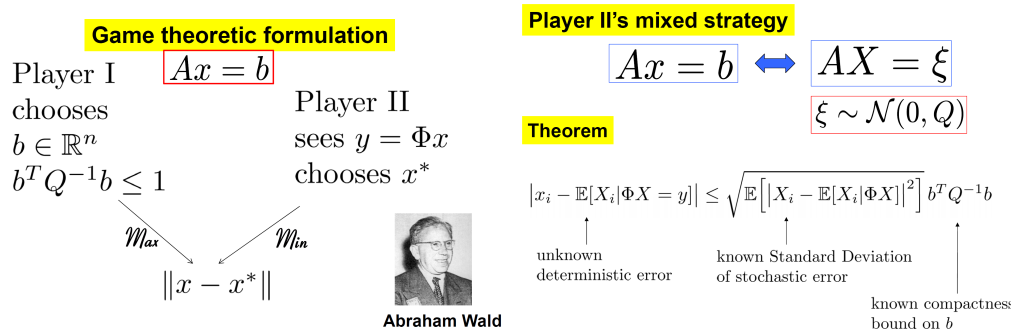

Figure 5. Optimal mixed strategy for the numerical approximation problem.

Let us now return to the numerical approximation game, illustrated in Figure 5, in which Player I chooses $b \in \mathbb{R}^n$ such that $b^T Q^{-1} b \leq 1$ and Player II sees $y = \Phi x$ and must choose x^* (to approximate the solution of $Ax = b$) and receives the loss $\|x - x^*\|$. Although this is a continuous game as in decision theory under compactness it can be approximated by a finite game and one identifies an analogous solution. More precisely to identify an optimal solution one has to lift the minimax to measures, the best strategy for Player I is to play at random and Player II's best strategy lives in the Bayesian class of estimators. This simply means that instead of considering the deterministic linear system (2), Player II should consider the stochastic linear system

$$AX = \xi \tag{7}$$

obtained by replacing the unknown right hand side b by a centered Gaussian vector ξ of covariance matrix Q (the distribution of ξ is not arbitrary or a function of Player II's belief as in Bayesian inference but the solution of a minimax problem over mixed strategies). Player II's best bet is then obtained by taking the expectation of the solution of the stochastic system conditioned on measurements of the deterministic system, i.e.

$$x^* = \mathbb{E}[X | \Phi X = \Phi x]. \tag{8}$$

Observe that Player II's recovery error on x_i, i.e. $|x_i - \mathbb{E}[X_i | \Phi X = \Phi x]|$ is unknown and that Player II's stochastic error (assuming that Player I is selecting b at random with the same prior distribution) is the random variable $|X_i - \mathbb{E}[X_i | \Phi X]|$ with known distribution. And a surprising result[15] is that the (optimal) bound on the unknown deterministic error is the known standard deviation of the stochastic error, i.e.

$$|x_i - \mathbb{E}[X_i | \Phi X = \Phi x]| \leq \sqrt{\mathbb{E}\left[|X_i - \mathbb{E}[X_i | \Phi X]|^2\right]} b^T Q^{-1} b. \tag{9}$$

Although this application may appear, at first glance, dissimilar to that of model selection, the game theory approach remains the same since these applications are all based on the efficient processing of incomplete information with limited computational resources (i.e., complexity management capabilities constraints): (1) fast and robust computation requires computation with partial information (this concept forms the core of Information Based Complexity[24] where it is also augmented by concepts of contaminated and priced information associated with, for example, truncation errors and the cost of numerical operations) (2) model selection and design requires the management and processing of epistemic uncertainties and limited data and model reduction requires the approximation of the full state of a complex system through operations performed on a few (coarse/reduced) variables.

IV. Application to hierarchical reduced order modelling

Results obtained in[15] suggest that concepts at the core of Information Based Complexity[11, 18, 24, 27, 28] can be generalized to the identification of hierarchies of discrete and finite operators enabling the approximation

American Institute of Aeronautics and Astronautics

of arbitrary models/functions by hierarchies of discrete and finite (reduced) models/functions acting at different levels of complexity (partial information). These generalized concepts are implicitly present in multigrid methods where a differential operator must be approximated at different levels of complexity (scales) and fast and accurate operators enabling the exchange of information between successive levels of complexity (scales) must be identified.

The core idea of the application of the game theoretic framework to model reduction is to turn the process of computing with hierarchies of partial information to that of playing hierarchies of adversarial (zero-sum) games with respect to the missing information.[15, 16] To illustrate this idea consider the approximation of a map $S : G \to U$ (discussed in[16]). Introduce a hierarchy of nested measurement functions $\Phi_i^{(k)} : U \to \mathbb{R}$ (corresponding to a hierarchies of reduced variables characterizing $u \in U$) such that each $\Phi_i^{(k)}$ is a function of $(\Phi_j^{(k+1)}(u))_{j \in \mathcal{I}^{(k+1)}}$. Write $\mathcal{F}_k$ the σ-algebra formed by level k measurements. Then since the measurements are nested, $(\mathcal{F}_k)_{k \geq 1}$ is a filtration and $\mathbb{E}[S(g)|\mathcal{F}_k]$ is a martingale under that filtration and under the measure of probability emerging from the adversarial game where Player I chooses $g \in G$ and Player II must gamble on the value of $u = S(g)$ and $(\Phi_j^{(k+1)}(u))_{j \in \mathcal{I}^{(k+1)}}$ based on level k measurements. $\mathbb{E}[S(g)|\mathcal{F}_k]$ is by construction the best function of the reduced variables $(\Phi_j^{(k)}(g))_{j \in \mathcal{I}^{(k)}}$ approximating $S(g)$ and its convergence (with respect to the level of complexity k) is naturally ensured by the martingale convergence theorem. As in,[15] by reordering measurement functions (reduced variables) into independent components (across complexity levels k) $S(g)$ can be decomposed into uncorrelated components $S(g) = \sum_{k \geq 1} \big(\mathbb{E}[S(g)|\mathcal{F}_{k+1}] - \mathbb{E}[S(g)|\mathcal{F}_k]\big)$ corresponding to a multiresolution decomposition (in complexity) of the operator S.

V.   Appendix

A.   The Optimal Uncertainty Quantification approach to the seismic safety assessment of a truss structure.
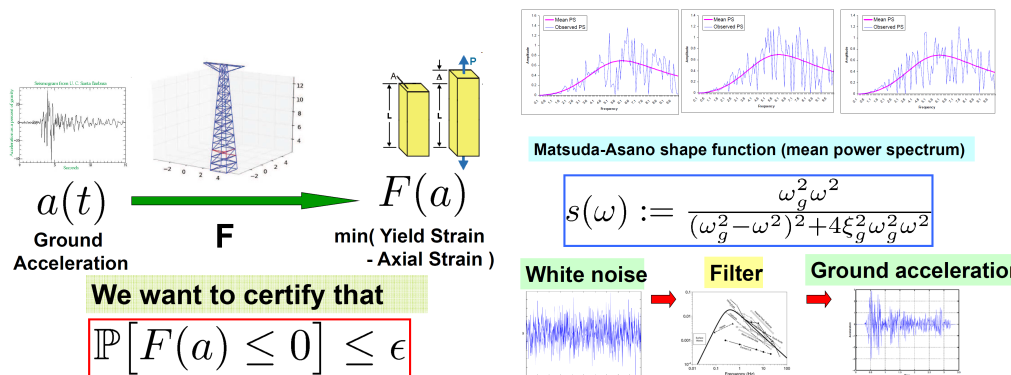

Figure 6.   Seismic safety assessment of a truss structure.

An example of the robust optimization worst case approach can in the seismic safety assessment of a truss structure[17] illustrated in Figure 6. In this example we consider the response function F of an electric tower mapping ground motion acceleration a to the minimum over all members i of the structure of the difference between the yield strain and the axial strain of member i. Therefore if $F(a)$ then some members become plastic and the structure fails under the ground motion acceleration $F(a)$ and one is interested in certifying the probability that the structure fails is smaller than the threshold ϵ, i.e. $\mathbb{P}[F(a) \leq 0] \leq \epsilon$. The major difficulty lies in the fact that the measure of probability $\mathbb{P}$ is unknown. One used by seismic engineers for assessing the probability of failure is the filtered white noise method illustrated in Figure 6. This method is based on the observation that although different earthquakes have different power spectra, their mean power spectrum has a known/given shape (typically a Lorentzian function characterizing the resonant frequency and damping factor of the site where the tower is built). The filtered white noise method is then to generate white noise and filter in the Fourier domain with that characteristic shape function (corresponding to the mean power spectrum) and return to the physical domain to obtain a stochastic

process under which one assesses the safety of the truss structure. The problem with the filtered white noise
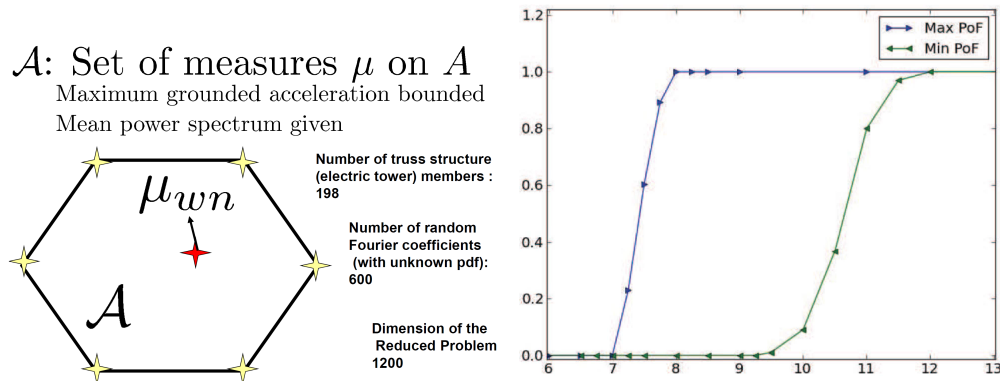

Figure 7. The admissible set and vulnerability curves.

method is that, as illustrated in Figure 7, the set $\mathcal{A}$ of probability distributions that are compatible with the given mean power spectrum is an infinite dimension polytope of the set of probability distributions on $\mathcal{A}$ and the filtered white distribution is only one point in that set. The conservative point of view is to compute a best case and a worst case scenario with respect to all distributions in that set. Although these are infinite dimensional problems, it is shown in[17] that as in linear programming they can be reduced to finite dimensional optimazation problems by considering finite-dimensional families of extremal scenarios of $\mathcal{A}$. This reduction then enables us to compute the vulnerability curves illustrated in Figure 7 corresponding to the minim and maximum probability of failure vs earthquake magnitude. Here one can see that in the worst case scenario the structure will start failing with an earthquake of magnitude 7.2 and in the best case scenario the structure starts failure with an earthquake of magnitude 9.4. Moreover for an earthquake of magnitude 8.5 the optimal bounds on the probability of failure are 0 and 1 and this simply means that the information contained in the mean power spectrum is not sufficient to improve trivial bounds (i.e. one needs more information if one wishes to improve those bounds).

References

[1] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic integration. *arXiv:1512.00933*, 2015.

[2] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami. Bayesian solution uncertainty quantification for differential equations. *arXiv:1306.2365*, 2015.

[3] P. R. Conrad, M. Girolami, S. Srkk, A. Stuart, and K. Zygalakis. Probability measures for numerical solutions of differential equations. *arXiv:1512.00933*, 2015.

[4] P. Diaconis. Bayesian numerical analysis. In *Statistical decision theory and related topics, IV, Vol. 1 (West Lafayette, Ind., 1986)*, pages 163–175. Springer, New York, 1988.

[5] P. Hennig. Probabilistic interpretation of linear solvers. *SIAM Journal on Optimization*, 25(1):234–260, 2015.

[6] P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. *Proc. A.*, 471(2179):20150142, 17, 2015.

[7] G. S. Kimeldorf and G. Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, 41:495–502, 1970.

[8] F. M. Larkin. Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mountain J. Math.*, 2(3):379–421, 1972.

[9] J. F. Nash, Jr. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. U. S. A.*, 36:48–49, 1950.

[10] John Nash. Non-cooperative games. *Ann. of Math. (2)*, 54:286–295, 1951.

[11] A. S. Nemirovsky. Information-based complexity of linear operator equations. *J. Complexity*, 8(2):153–175, 1992.

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

[13] A. O'Hagan. Some Bayesian numerical analysis. In *Bayesian statistics, 4 (Peñíscola, 1991)*, pages 345–363. Oxford Univ. Press, New York, 1992.

[14] H. Owhadi. Bayesian numerical homogenization. *SIAM Journal on Multiscale Modeling and Simulation*, 13(3):812–828, 2015. arXiv:1406.6668.

[15] H. Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. *SIAM Review*, 2016. arXiv:1503.03467.

[16] H. Owhadi and C. Scovel. Towards machine Wald. *Hanbook for Uncertainty Quantification*, 2016. arXiv:1508.02449.

[17] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *SIAM Rev.*, 55(2):271–345, 2013.

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

[19] H. Poincaré. *Calcul des probabilités*. Georges Carrés, Paris, 1896.

[20] A. Sard. *Linear approximation*. American Mathematical Society, Providence, R.I., 1963.

[21] M. Schober, D. K. Duvenaud, and P. Hennig. Probabilistic ODE solvers with Runge-Kutta means. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 739–747. Curran Associates, Inc., 2014.

[22] J. Skilling. *Maximum Entropy and Bayesian Methods: Seattle, 1991*, chapter Bayesian Solution of Ordinary Differential Equations. Springer Netherlands, Dordrecht, 1992.

[23] A. V. Sul′din. Wiener measure and its applications to approximation methods. I. *Izv. Vysš. Učebn. Zaved. Matematika*, 1959(6 (13)):145–158, 1959.

[24] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-based complexity*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boult.

[25] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1944.

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

[27] H. Woźniakowski. Probabilistic setting of information-based complexity. *J. Complexity*, 2(3):255–269, 1986.

[28] H. Woźniakowski. What is information-based complexity? In *Essays on the complexity of continuous problems*, pages 89–95. Eur. Math. Soc., Zürich, 2009.