AFRL-AFOSR-VA-TR-2019-0028

Computational Information Games

Houman Owhadi
CALIFORNIA INSTITUTE OF TECHNOLOGY
1200 E. CALIFORNIA BLDV
PASADENA, CA 91125

01/11/2019
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2
Arlington, Virginia 22203
Air Force Material Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
11/15/2018	Final Technical	11/15/2015 - 11/14/2018

4. TITLE AND SUBTITLE
Computational Information Games

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-16-1-0054

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Houman Owhadi

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Institute of Technology
1200 E. California Blvd.,
MC: 9-94
Pasadena, CA 91125

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF OFFICE OF SCIENTIFIC RESEARCH
875 NORTH RANDOLPH STREET, RM 3112
ARLINGTON VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution unlimited, approved for public release.

13. SUPPLEMENTARY NOTES
Program Manager: Dr. Fariba Fahroo DR-04 USAF AFMC AFOSR/RTA <fariba.fahroo@us.af.mil>
Change in AFOSR program manager: Dr. Fariba Fahroo has replaced Dr. Jean-Luc Cambier

14. ABSTRACT
The objective of AFOSR/EQUIPS Grant number FA9550-16-1-0054 (Computational Information Games) was to develop a game theoretic approach to numerical approximation and algorithm design. This approach has been turned in a general framework which has lead to the discoveries of (i) a general solution to the numerical homogenization problem (ii) wavelets a (i) wavelets adapted to arbitrary linear operators (gamblets) (ii) scalable solvers with some degree of universality (iii) new tools for numerical analysis and algorithm design such as the Fast Gamblet Transform (FGT).

15. SUBJECT TERMS
Gamblets. Statistical Numerical Approximation. Fast Solvers. Operator adapted wavelets. Game theoretic approach to algorithmic design and numerical analysis.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Houman Owhadi
					19b. TELEPHONE NUMBER *(Include area code)*
UU	UU	UU	UU	22	626 395 4547

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Computational Information Games

Final Report

Program Manager: Dr. Fariba Fahroo DR-04 USAF AFMC AFOSR/RTA
Change in AFOSR program manager: Dr. Fariba Fahroo DR-04 USAF AFMC AFOSR/RTA has replaced Dr. Jean-Luc Cambier DR-04 USAF AFMC AFOSR/RTA.
Contract/Grant #: FA9550-16-1-0054.
Reporting Period: 11/15/2015 to 11/14/2018.
PI: Houman Owhadi
Organization: California Institute of Technology

Abstract: This project has explored interplays between Game Theory, Numerical Approximation and Gaussian Process Regression and developed a general theory for a game theoretic approach to numerical approximation and algorithm design. This game theoretic approach has lead to the discoveries of (i) wavelets adapted to arbitrary linear operators (gamblets) (ii) scalable solvers with some degree of universality (iii) new tools for numerical analysis and algorithm design such as the Fast Gamblet Transform (FGT). These discoveries provide new tools for numerical analysis and algorithm design and one of these tools is the Fast Gamblet Transform (FGT). The scope of potential applications of those tools is comparable to having a Fast Fourier Transform that could be applied to arbitrary linear operators. For instance (i) the FGT enables the multi-resolution analysis of arbitrary linear operators defined on Sobolev spaces in near-linear complexity (ii) The FGT leads to a solver for arbitrary continuous symmetric linear bijections mapping $H^s_0(\Omega)$ to $H^{-s}(\Omega)$ in $\mathcal{O}(N \log^{2d+1} N)$ complexity (this is the state of the art for PDEs with rough coefficients) (iii) The FGT enables near-linear complexity when using implicit solvers for hyperbolic and parabolic PDEs (thereby opening the complexity bottleneck of such solvers) (iv) We have derived a simple incomplete Cholesky factorization algorithm for the compression, inversion and approximate the PCA of dense $N \times N$ kernel matrices in $\mathcal{O}(N \log^{2d+2} N)$ complexity.

1 Accomplishments

The objective of AFOSR/EQUIPS Grant number FA9550-16-1-0054 (Computational Information Games) was to develop a game theoretic approach to numerical approximation and algorithm design [4, 3, 11, 7, 13]. This approach has been turned in a general framework [7] which has lead to the discoveries of (i) a general solution to the numerical homogenization problem [7] (ii) wavelets adapted to large classes of linear operators (gamblets) [7] (iii) scalable solvers with some degree of universality [7, 13] (e.g., of $\mathcal{O}(N \log^{2d+1} N)$ complexity for arbitrary continuous symmetric linear bijections mapping $H^s_0(\Omega)$ to $H^{-s}(\Omega)$, this is the state of the art for such operators). These discoveries provide new tools for numerical analysis and algorithm design and one of these tools is the Fast Gamblet Transform (FGT). The scope of potential applications of those tools is comparable to having a Fast Fourier Transform that could be applied to a large class of linear operators. Another major breakthrough is [13] which introduces a very simple incomplete Cholesky factorization algorithm to compress, invert and approximate the PCA of dense $N \times N$ kernel matrices in $\mathcal{O}(N \log^{2d+2} N)$ complexity. This algorithm is derived and analysed by using gamblets to represent fundamental linear algebra operations such as Schur complementation.

1.1 List of accomplishments

Major accomplishments

- Discovery of (the first) fully operator adapted wavelets (this discovery has been and is still branching into many others).
- Discovery of robust scalable solvers for general elliptic, parabolic and hyperbolic linear PDEs of rigorous a priori complexity vs accuracy estimates (of $\mathcal{O}(N \log^{2d+1} N)$ complexity to achieve grid size accuracy in energy norm, this is state of the art for PDEs with rough coefficients).
- Discovery of a very simple incomplete Cholesky factorization algorithm to compress, invert and approximate the PCA of dense $N \times N$ kernel matrices in $\mathcal{O}(N \log^{2d+2} N)$ complexity (this is state of the art for kernels defined as Green's functions of general elliptic operators). A software has been made available at

  https://github.com/f-t-s/nearLinKernel

- Development of a general game/decision theoretic framework to numerical approximation (this development has been and still branching into many others, with impacts in Kriging, and statistical estimation/computational tradeoffs, etc... the fundamental impacts are on the interplays between computational complexity and UQ).

- First rigorous a priori rigorous exponential decay estimates on the screening effect.
- Thresholding the Gamblet Transform of noisy solutions of observations of solutions of PDEs or Graph Laplacians is a near minimax recovery estimator [17].

Broader impact of transition from FA9550-12-1-0389 into FA9550-16-1-0054

- Our results on the non robustness of Bayesian inference [10, 6, 9, 5] lead us to predict that that deep/machine learning algorithms could be non robust and could lead to increased confidence in incorrect solutions (see talk by Mike McKerns at SciPy 2013, `https://www.youtube.com/watch?v=o-nwSnLC6DU&feature=youtu.be&t=74` published on July 2013). These predictions have been confirmed [16] and addressing these vulnerabilities (1) is now recognized as critical to the safety of Machine Learning Algorithms and (2) has stimulated the emergence of **"advesarial machine learning"**.
- FA9550-12-1-0389 supported the further development of the Mystic framework and the broader impact is now manifest under FA9550-16-1-0054. In particular Mystic and Pathos have been used at the Intelligence Systems Support Office (ISSO) of USAF and these codes are now in the top charts in GitHub (see https://hugovk.github.io/top-pypi-packages/ for official stats and links to code, mystic, pathos are rank 252 with 4.141M downloads per year).

1.2 Generalization of the gamblet transform

We have completed the monograph [7] describing the generalization of the game theoretic approach and gamblets introduced in [4] (for second order divergence operator) to large classes of operators. We are currently finishing a book [8] (to appear soon) to make these results accessible.

1.2.1 The numerical approximation/optimal recovery game

Gamblets emerge from a game theoretic approach to numerical approximation and algorithm design [7], which unfolds from the following observations: (i) to compute fast one must compute with partial information of hierarchies of levels of complexity (ii) the process of computing with partial information can be turned into an adversarial game with respect to the missing information (iii) inaccurate approximations, in repeated intermediary calculations, lead to loss in CPU time and the total CPU time required to invert a given linear operator is the sum of these losses.

One fundamental result of [7] can be described as follows. Let $\mathcal{B}$ be a Banach space endowed with a quadratic norm $\|\cdot\|$. Let $\phi_1, \ldots, \phi_m$ be m independent linear elements of $\mathcal{B}^*$. Consider an adversarial zero sum game where player I chooses $u \in \mathcal{B}$ and player II must recover u based on the incomplete information $([\phi_1, u], \ldots, [\phi_m, u])$. Then the optimal bet of player II is

$$v = \mathbb{E}\big[\xi \mid [\phi_i, \xi] = [\phi_i, u] \text{ for all } i\big] \tag{1}$$

where ξ is a centered Gaussian field defined by $[\phi, \xi] \sim \mathcal{N}(0, \|\phi\|_*^2)$ for $\phi \in \mathcal{B}^*$ and writing $\|\cdot\|_*$ for the dual norm. Furthermore this optimal bet can be written as a linear combination of elementary bets, i.e. as

$$v = \sum_i [\phi_i, u] \psi_i \tag{2}$$

with

$$\psi_i = \mathbb{E}\big[\xi \mid [\phi_j, \xi] = \delta_{i,j} \text{ for all } j\big] \tag{3}$$

Based on this observation, [7] derives wavelets adapted to the norm $\|\cdot\|$ and fast solvers for the operator defining that norm. For the sake of the clarity of this report we will present the results when $(\mathcal{B}, \|\cdot\|)$ is the Sobolev space $\mathcal{H}_0^s(\Omega)$ endowed with the operator norm

$$\|u\|^2 := \int_\Omega u \mathcal{L} u \tag{4}$$

where

$$\mathcal{L} : \mathcal{H}_0^s(\Omega) \to \mathcal{H}^{-s}(\Omega) \tag{5}$$

is an arbitrary symmetric, positive linear bijection that is assumed to be local, in that $\int_\Omega u \mathcal{L} v = 0$ if u and v have disjoint supports.

1.2.2 Numerical homogenization

The generalized purpose of *numerical homogenization*, can be described as follows.

Problem 1.1. *Given an arbitrary operator $\mathcal{L} : \mathcal{H}_0^s(\Omega) \to \mathcal{H}^{-s}(\Omega)$ and a given m, to find m basis functions $\psi_1, \ldots, \psi_m$ satisfying the following two requirements:*

1. Accuracy. *The approximation error*

$$\sup_{f\in L^2(\Omega)} \inf_{c\in\mathbb{R}^m} \frac{\|\mathcal{L}^{-1}f-\sum_{i=1}^m c_i\psi_i\|}{\|f\|_{L^2(\Omega)}} \tag{6}$$

must be as small as possible.

2. Localization. *The basis functions ψ_i must be as localized as possible (e.g. with compact support or exponentially decaying).*

These requirements are, to some degree, conflicting because the basis functions minimizing (6) (i.e. achieving the Kolmogorov n-width [2] with $n = m$) are the eigenfunctions of $\mathcal{L}$ corresponding to the m smallest eigenvalues [2, 1, 4], which are not localized.

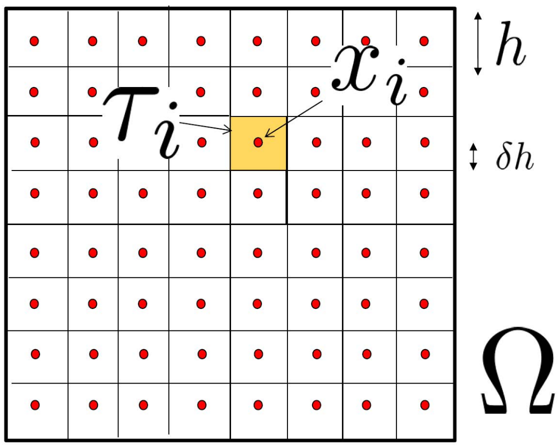

Figure 1: τ_i and x_i. Used from forthcoming book [8] with permission from Cambridge University Press.

Gamblets provide a very simple and natural solution to Problem 1.1. Let $\tau_1,\cdots,\tau_m$ form a partition of Ω of resolution h as illustrated in Figure 1. Let $x_1,\ldots,x_m$ be points centered in $\tau_1,\cdots,\tau_m$. For $i\in\{1,\ldots,m\}$ take $\phi_i = 1_{\tau_i}$ or (for $s > d/2$) $\phi_i = \delta(x-x_i)$.

Let $\psi_1,\ldots,\psi_m$ be the corresponding gamblets defined in (3). Then, [7] shows that $\psi_1,\ldots,\psi_m$ are a solution to Problem 1.1. More precisely, they achieve the accuracy of the Kolmogorov n-width [2] (the minimum of (6)) up to a multiplicative constant: for $f\in L^2(\Omega)$,

$$\inf_{\psi\in\text{span}\{\psi_1,\ldots,\psi_m\}} \|\mathcal{L}^{-1}f-\psi\|_{\mathcal{H}_0^s(\Omega)} \le Ch^s\|f\|_{L^2(\Omega)}. \tag{7}$$

Furthermore, they are exponentially localized, i.e.

$$\|\psi_i\|_{\mathcal{H}^s(\Omega/B(x_i,nh))} \leq Ce^{-n/C}, \quad i = 1, \dots, m\,. \tag{8}$$

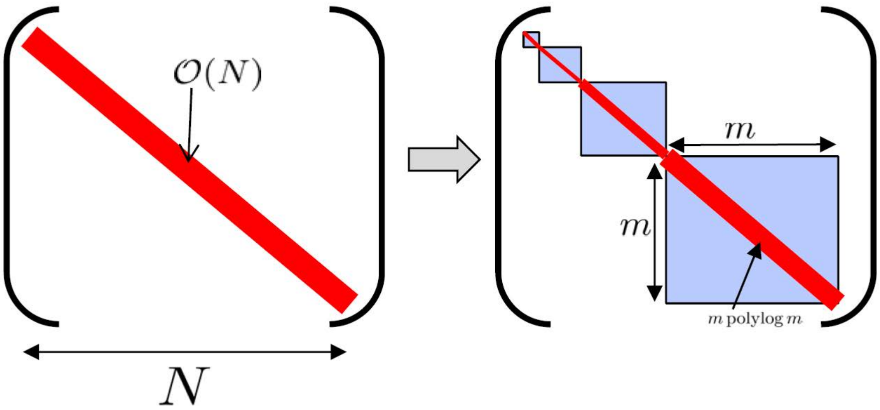

Figure 2: Matrix representation with fully adapted wavelets. Used from forthcoming book [8] with permission from Cambridge University Press.

1.2.3 Operator adapted wavelets

As emphasized in [15, p. 83] ideal operator adapted wavelets should be characterized by three properties (see Figure 2) described in the following problem.

Problem 1.2. *Given an arbitrary operator $\mathcal{L}$, find wavelets simultaneously satisfying the following three properties.*

1. ***Scale-orthogonality** with respect to the operator scalar product. This property implies that the (stiffness) matrix representation of the operator in the wavelet basis is block-diagonal.*

2. ***Local support (or rapid decay)** of the basis functions. This property implies that the individual blocks are sparse or nearly sparse.*

3. ***Riesz stability in the energy norm**. This property implies that the blocks are uniformly well-conditioned.*

As discussed in [15, p. 83], although adapted wavelets achieving 2 of these properties have been constructed, "it is not known if there is a practical technique for ensuring all the three properties simultaneously in general".

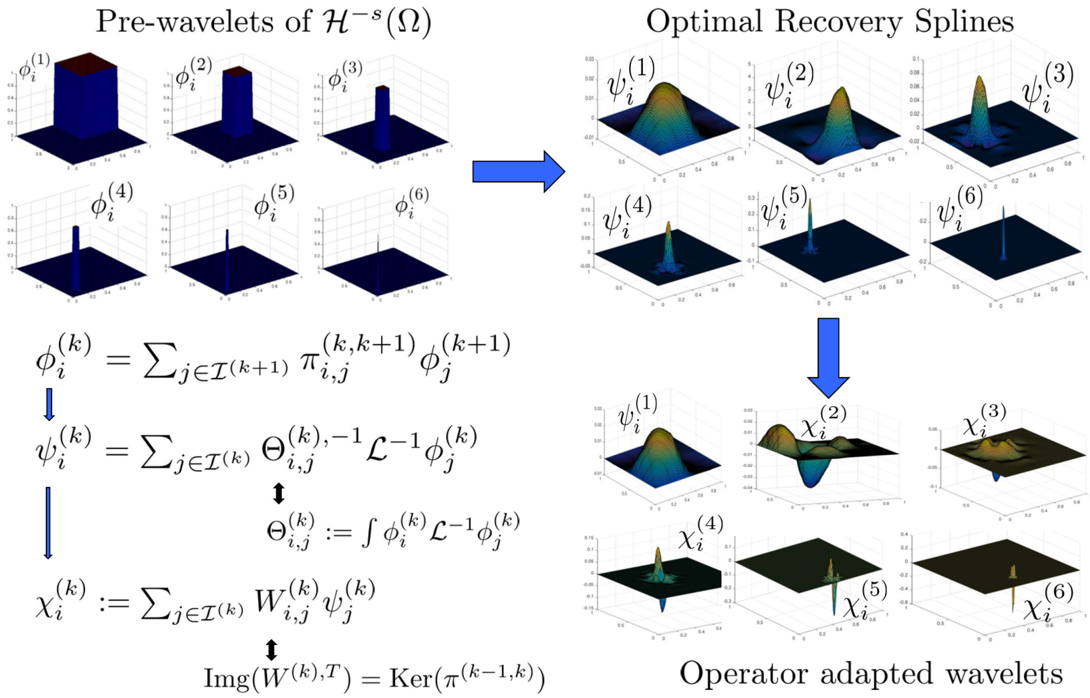

Figure 3: Overview of the construction. Used from forthcoming book [8] with permission from Cambridge University Press.

Gamblets provide a very simple solution to Problem 1.2. This solution, illustrated in Figure 3, can be described as follows.

1. Select non-operator adapted pre-wavelets for $\mathcal{H}^{-s}(\Omega)$. These pre-wavelets $\phi_i^{(k)} \in \mathcal{H}^{-s}(\Omega)$ (where heuristically k stands for scale and i for location) form a hierarchy satisfying the nesting relation

$$\phi_i^{(k)} = \sum_j \pi_{i,j}^{(k,k+1)} \phi_j^{(k+1)} . \tag{9}$$

   Prototypical examples for the choice of $\phi_i^{(k)}$ are Haar pre-wavelets (Figure 4, which induce a multi-resolution decomposition of the compact embedding of $L^2(\Omega)$ into $\mathcal{H}^{-s}(\Omega)$ that is not adapted to the operator $\mathcal{L}$) and, for $s > d/2$, sub-sampled Diracs (Figure 5).

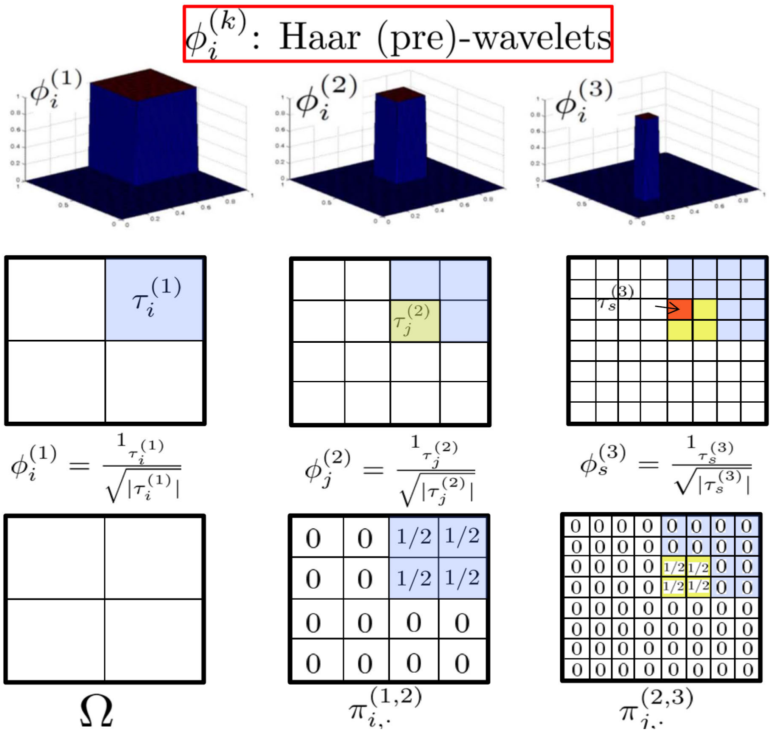

Figure 4: Haar pre-wavelets as $\phi_i^{(k)}$. Used from forthcoming book [8] with permission from Cambridge University Press.

2. View the $\phi_i^{(k)}$ as measurement functions for numerical approximation games and define $\psi_i^{(k)}$ as the corresponding gamblets, i.e. using (3),

$$\psi_i^{(k)} := \sum_j \Theta_{ij}^{(k),-1} \mathcal{L}^{-1} \phi_j^{(k)}, \tag{10}$$

where $\Theta^{(k),-1}$ is the inverse of the Gramian matrix $\Theta^{(k)}$ defined by $\Theta_{i,j}^{(k)} := [\phi_i^{(k)}, \mathcal{L}^{-1}\phi_j^{(k)}]$. The elements $\psi_i^{(k)}$ are then pre-wavelets, adapted to the operator $\mathcal{L}$, forming a nested hierarchy of $\mathcal{H}_0^s(\Omega)$, i.e. $\psi_i^{(k)} = \sum_j R_{i,j}^{(k,k+1)} \psi_j^{(k+1)}$.

3. For $k \geq 2$ orthogonalize the pre-wavelets $\psi_i^{(k)}$ through local linear combinations/differences with coefficients spanning the kernel of $\pi^{(k-1,k)}$, i.e. introducing $W^{(k)}$ as a sparse matrix such that $\mathrm{Im}(W^{(k),T}) = \mathrm{Ker}(\pi^{(k-1,k)})$ and

$$\chi_i^{(k)} := \sum_j W_{ij}^{(k)} \psi_j^{(k)}, \tag{11}$$

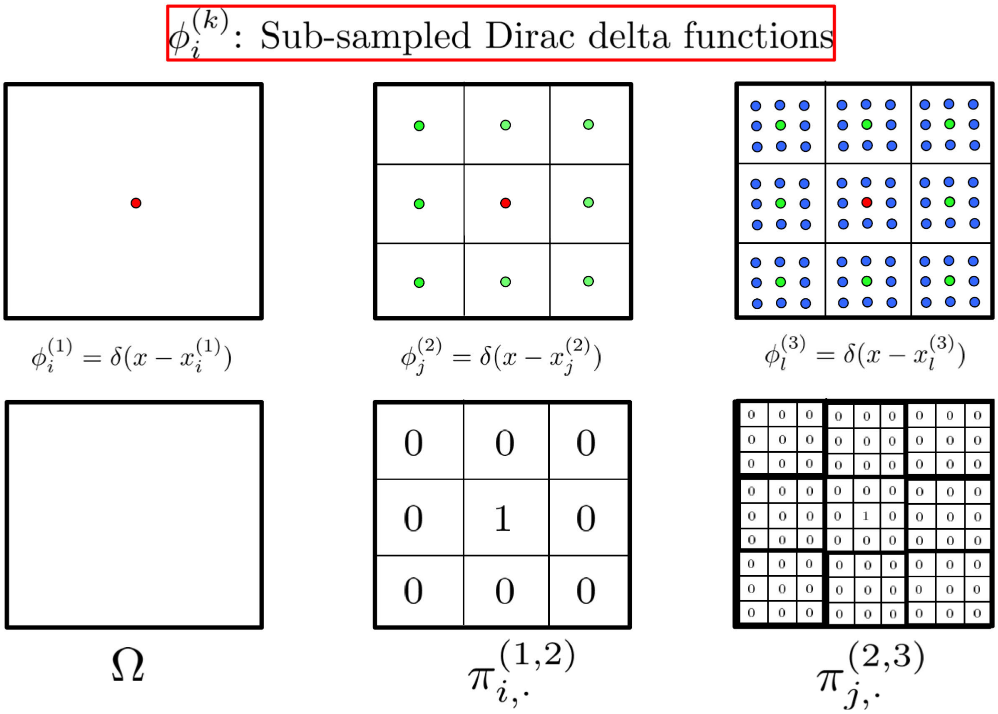

Figure 5: Sub-sampled Diracs as $\phi_i^{(k)}$. Used from forthcoming book [8] with permission from Cambridge University Press.

the $\chi_i^{(k)}$ are operator adapted wavelets (inducing a multi-resolution decomposition of $\mathcal{H}_0^s(\Omega)$ adapted to $\mathcal{L}$ in the sense of Problem 1.2).

Algorithm 1 The Gamblet Transform.

$\psi_i^{(q)} = \varphi_i$
$A_{i,j}^{(q)} = \langle \psi_i^{(q)}, \psi_j^{(q)} \rangle$
for $k = q$ to 2 **do**
  $B^{(k)} = W^{(k)} A^{(k)} W^{(k),T}$
  $\chi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$
  $R^{(k-1,k)} = \pi^{(k-1,k)} (I^{(k)} - A^{(k)} W^{(k),T} B^{(k),-1} W^{(k)})$
  $A^{(k-1)} = R^{(k-1,k)} A^{(k)} R^{(k,k-1)}$
  $\psi_i^{(k-1)} = \sum_{j \in \mathcal{I}^{(k)}} R_{i,j}^{(k-1,k)} \psi_j^{(k)}$
end for

The resulting algorithm for computing those operator adapted wavelets is presented in Algorithm 1. In this algorithm the operator adapted pre-wavelets are computing in hierarchical nested fashion, i.e.

$$\psi_i^{(k)} = \sum_j R_{i,j}^{(k,k+1)} \phi_j^{(k+1)} , \tag{12}$$

and this computation can be done fast because the interpolation matrices $R^{(k,k+1)}$ are sparse (since the gamblets are exponentially decaying). As a consequence the overall complexity of this algorithm is $\mathcal{O}(N \ln^{2d+1} N)$.

1.2.4 Fast solvers

Is it possible to identify/design a scalable solver that could be applied to a large class of linear operators? One incentive to ask this question is the vast and increasing literature on the numerical approximation of linear operators where the number of linear solvers seems to trail the number of possible linear systems. Paraphrasing Sard's assertion, one reason not to ask this question is the historical presupposition that [12, pg. 223] "of course no one method of approximation of a linear operator can be universal." Using gamblets to answer this question in the setting of linear operators leads to a fast solver with some degree of universality.

For instance, consider the problem of solving

$$\mathcal{L}u = f \tag{13}$$

as fast as possible to a given accuracy, where $\mathcal{L}$, defined in (5), is arbitrary. Gamblets provide a very simple solver based on the decomposition of the solution space $\mathcal{H}_0^s(\Omega)$ into sub-bands $\mathfrak{W}^{(k)} = \text{span}\{\chi_i^{(k)} | i \in \mathcal{I}^{(k)}\}$ that are orthogonal in operator scalar product, i.e.

$$\mathcal{H}_0^s(\Omega) = \mathfrak{V}^{(1)} \oplus \mathfrak{W}^{(2)} \oplus \cdots \oplus \mathfrak{W}^{(q)} \oplus \mathfrak{W}^{(q+1)} \tag{14}$$

and such that $\mathcal{L}$ is uniformly well-conditioned within each sub-band.

Therefore, the gamblet transform turns the inverse problem $\mathcal{L}u = f$ into a sequence of independent, uniformly well-conditioned, sparse linear systems. Each sub-band solution (illustrated in Figure 6) can be computed independently. This fast gamblet solve is presented in Algorithm 2.

Since the matrices $B^{(k)}$ of Algorithm 2 are uniformly well-conditioned and sparse, the complexity of the algorithm is $\mathcal{O}(N \log^{d+1} N)$ for each linear solve.

1.3 Dense kernel matrices

The fast solver described in Subsection 1.2 is based on an explicit computation of the gamblets. By representing elementary linear algebra operations, such as computing Schur complements, as operations on Gamblets (i.e. on wavelets adapted

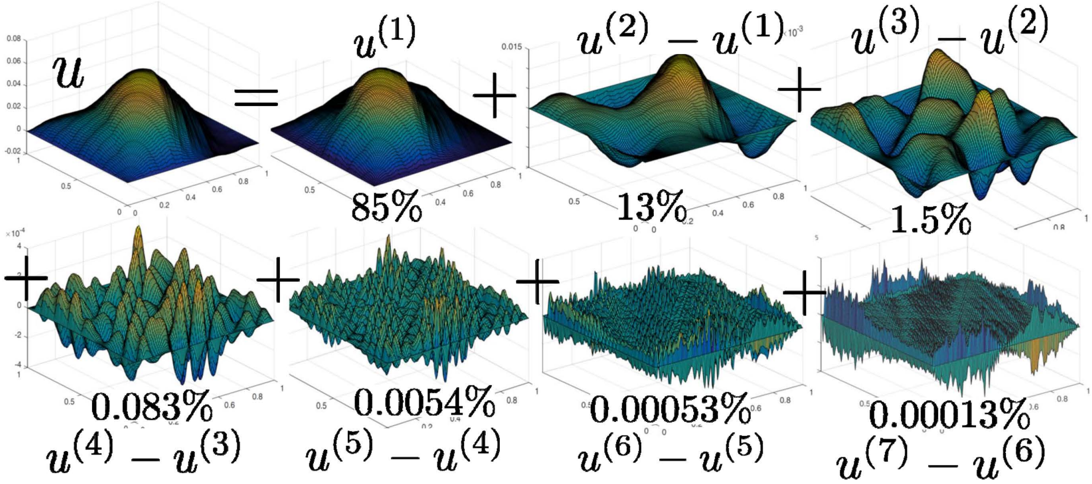

Figure 6: Multi-resolution decomposition of the solution of $\mathcal{L}u = f$. Used from forthcoming book [8] with permission from Cambridge University Press.

Algorithm 2 The Gamblet Solve.

1: $f_i^{(q)} = \int_\Omega f \psi_i^{(q)}$
2: **for** $k = q$ to 2 **do**
3: $\quad w^{(k)} = B^{(k),-1} W^{(k)} f^{(k)}$
4: $\quad u^{(k)} - u^{(k-1)} = \sum_{i \in \mathcal{J}^{(k)}} w_i^{(k)} \chi_i^{(k)}$
5: $\quad f^{(k-1)} = R^{(k-1,k)} f^{(k)}$
6: **end for**
7: $U^{(1)} = A^{(1),-1} f^{(1)}$
8: $u^{(1)} = \sum_{i \in \mathcal{I}^{(1)}} U_i^{(1)} \psi_i^{(1)}$
9: $u = u^{(1)} + (u^{(2)} - u^{(1)}) + \cdots + (u^{(q)} - u^{(q-1)})$

to the underlying operator) one can derive simple algorithms that can be analysed through gamblets. This approach is illustrated in the recent preprint [13], which introduces a simple incomplete Cholesky factorization algorithm (rigorously) achieving $\mathcal{O}(N \log^{2d+2} N)$ complexity for the inversion, compression, and approximate PCA of an $N \times N$ dense kernel matrix $\Theta \in \mathbb{R}^{N \times N}$ obtained from point evaluations of the Green's function G at locations $\{x_1, \ldots, x_N\}$ $(\Theta_{i,j} = G(x_i, x_j))$ of a local continuous symmetric linear operator $\mathcal{L}$ mapping $H^s(\Omega)$ to $H^{-s}(\Omega)$.

Such dense kernel matrices arise in computational physics, numerical analysis, statistics and machine learning, which can suffer from the computational complexity bottleneck associated with simple operations such as (i) storing Θ (a naive approach costs $\mathcal{O}(N^2)$) (ii) computing the matrix vector product Θv (a naive approach costs $\mathcal{O}(N^2)$) (iii) computing the inverse $\Theta^{-1}v$ (a naive approach costs $\mathcal{O}(N^3)$) (iv) computing the determinant of Θ (a naive approach costs $\mathcal{O}(N^3)$) (v) approximation the PCA of Θ (a naive approach costs $\mathcal{O}(N^4)$).

The overall algorithm introduced in [13] performs an ordering of $\{x_i\}_{1\leq i\leq N}$, with corresponding permutation matrix P, and computes, in complexity $N\operatorname{polylog}(N)\operatorname{polylog}(\frac{1}{\epsilon})$ in time and space, a sparse lower triangular matrix L (whose number of non zero entries is $\mathcal{O}(N\operatorname{polylog}(N)\operatorname{polylog}(\frac{1}{\epsilon}))$) such that

$$\|\Theta - PLL^TP^T\| \leq \mathcal{O}(\epsilon).$$

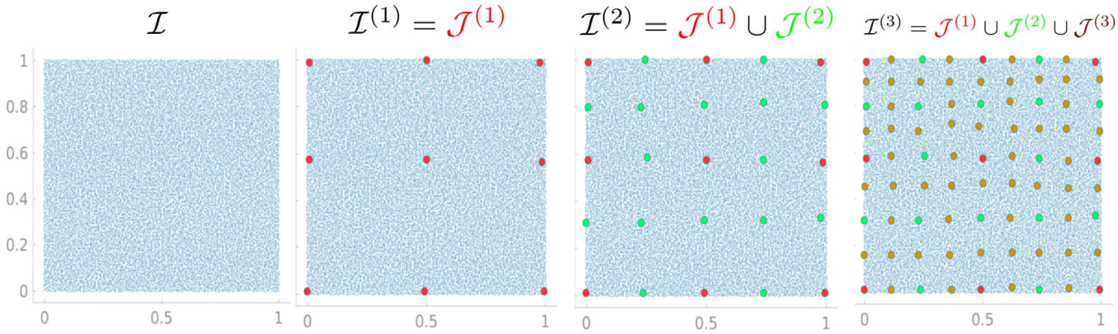

Figure 7: Decompose $\{x_i\}_{i\in\mathcal{I}}$ (fine blue dots in the left figure) into a nested (subsampled) hierarchy $\{x_i\}_{i\in\mathcal{I}^{(1)}} \subset \{x_i\}_{i\in\mathcal{I}^{(2)}} \subset \{x_i\}_{i\in\mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i\in\mathcal{I}^{(q)}}$. Used from forthcoming book [8] with permission from Cambridge University Press.

The simplicity of this algorithm is remarkable: First, decompose $\{x_i\}_{i\in\mathcal{I}}$ into a nested hierarchy $\{x_i\}_{i\in\mathcal{I}^{(1)}} \subset \{x_i\}_{i\in\mathcal{I}^{(2)}} \subset \{x_i\}_{i\in\mathcal{I}^{(3)}} \subset \cdots \subset \{x_i\}_{i\in\mathcal{I}^{(q)}}$ and define $\mathcal{I}^{(1)} = \mathcal{J}^{(1)}$,. . . ,$\mathcal{I}^{(k)} = \mathcal{J}^{(1)} \cup \cdots \cup \mathcal{J}^{(k)}$ as in Figure 7 (see the $\mathcal{J}^{(k)}$ are the hierarchical stratification defined below).

Next, order the degrees of freedom (elements of $\mathcal{I}$) from $\mathcal{J}^{(1)}$ to $\mathcal{J}^{(q)}$ (i.e. the nesting/inclusion of the set of points $\mathcal{I}^{(k)}$ defines a stratification leading to an ordering of the points of $\mathcal{I}$) and for $\epsilon \in (0, e^{-1})$ define the sparsity pattern, $S := \big\{(i,j) \in \mathcal{I}\times\mathcal{I} \big| i \in \mathcal{J}^{(k)}, j \in \mathcal{J}^{(l)}, \operatorname{dist}(x_i,x_j) \leq 2\ln\frac{1}{\epsilon} \times 2^{-\min(k,l)}\big\}$. S contains $\mathcal{O}(N(\operatorname{polylog} N \operatorname{polylog}\frac{1}{\epsilon}))$ elements and all the entries of Θ lying outside of this sparsity pattern are ignored (i.e. the resulting $G(x_i,x_j)$ do not even need to be evaluated).

Algorithm 1: Cholesky factorisation

```
for i ← 1 to N do
    A_{i,i} ← √A_{i,i};
    for j ← i + 1 to N do
        for k ← j to N do
            A_{k,j} ← A_{k,j} − A_{k,i} A_{j,i} / A_{i,i};
    A_{:,i} ← A_{:,i} / √A_{i,i};
return LowerTriang(A)
```

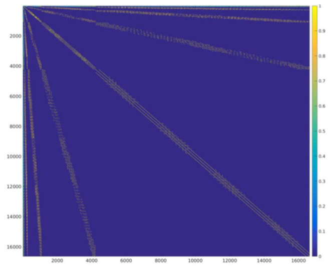

Figure 8: The algorithm. Skip entries of A (computational surrogate for Θ) outside of the sparsity pattern (i.e. skip all operations for which (k, j), (k, i) or (j, i) are outside of the sparsity pattern illustrated in the right figure).

The Cholesky factorization $A = LL^T$ (computational surrogate for Θ) is then computed as in Figure 8 with one small tweak: skip all operations for which (k, j), (k, i) or (j, i) are outside of the sparsity pattern (right hand side of Figure 8) (the Cholesky factorization is not applied to Θ but only to a subset of entries of Θ defined by the sparsity pattern).

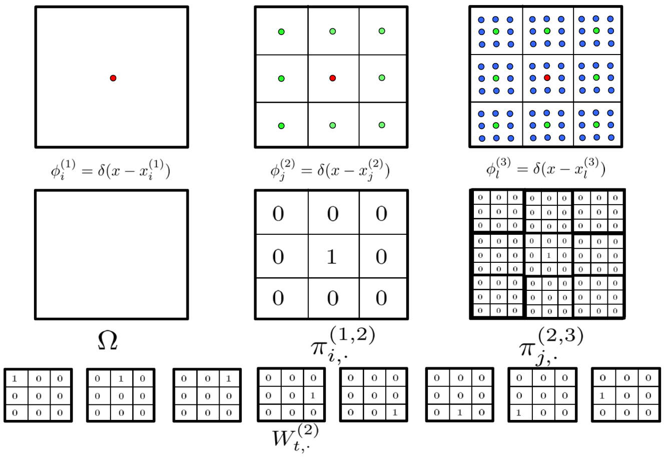

Figure 9: The matrices $\pi^{(k-1,k)}$ and $W^{(k)}$ for sub-sampled masses of Diracs. Used from forthcoming book [8] with permission from Cambridge University Press.

Why does it work?

The analysis of the algorithm is performed using gamblets. By choosing Dirac delta functions at the locations x_i as measurement functions at the finest scale (i.e.

$\phi_i^{(q)} = \delta(\cdot - x_i)$) one can represent Θ as the Gram matrix of these measurement functions, i.e. $\Theta_{i,j} = [\phi_i^{(q)}, \mathcal{L}^{-1}\phi_j^{(q)}]$, and its inverse Θ^{-1} is the stiffness matrix of the corresponding gamblets, i.e. $\Theta^{-1}_{i,j} = \langle \psi_i^{(q)}, \psi_j^{(q)} \rangle := [\mathcal{L}\psi_i^{(q)}, \psi_j^{(q)}]$.

Next by constructing the hierarchy of measurement functions $\phi_i^{(k)}$ via sub-sampling (as illustrated in Figure 9) and ordering the points x_i according to this hierarchy one obtains that the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k)}$ Gram matrix $\Theta^{(k)}_{i,j} = [\phi_i^{(k)}, \mathcal{L}^{-1}\phi_j^{(k)}]$ corresponds to the upper left block of Θ as illustrated in Figure 10.

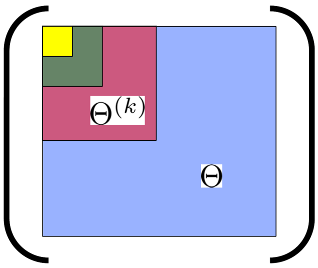

Figure 10: Θ and its sub-matrices $\Theta^{(k)}$. Used from forthcoming book [8] with permission from Cambridge University Press.

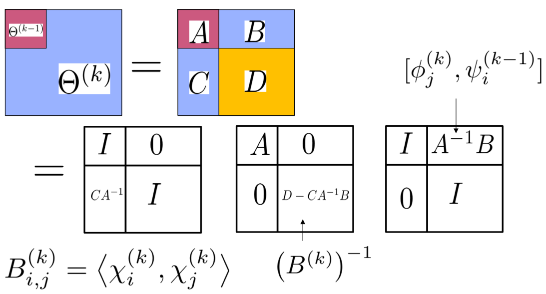

Figure 11: Correspondence between Schur complements and gamblets. Used from forthcoming book [8] with permission from Cambridge University Press.

The elementary steps of the algorithm correspond to the computation of the Schur complement of $\Theta^{(k)}$ in Θ. As illustrated in Figure 11, this Schur complement is equal to $(B^{(k)})^{-1}$, where $B^{(k)}$ is the stiffness matrix of the orthogonalized gamblets, i.e. $B^{(k)}_{i,j} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle$. Using the fact that $B^{(k)}$ is uniformly well-conditioned and sparse (i.e. exponentially decaying away from the diagonal) we deduce the

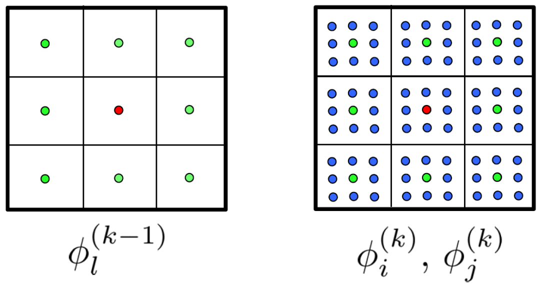

Figure 12: Conditional sparsity and the screening effect. Used from forthcoming book [8] with permission from Cambridge University Press.

sparsity of $(B^{(k)})^{-1}$ and therefore, of the Schur complement (which allows us to skip the steps involving negligible entries of $(B^{(k)})^{-1}$).

The screening effect.

Writing ξ for the Gaussian field with covariance matrix G (emerging as the optimal mixed strategy in the game theoretic approach), as illustrated in Figure 12, $(B^{(k)})^{-1}_{i,j}$ is the covariance between $[\phi_i^{(k)}, \xi]$ and $[\phi_j^{(k)}, \xi]$ given $[\phi_l^{(k-1)}, \xi]$, $\forall l$, i.e.

$$(B^{(k)})^{-1}_{i,j} = \operatorname{Cov}\Big([\phi_i^{(k)}, \xi], [\phi_j^{(k)}, \xi] \Big| [\phi_l^{(k-1)}, \xi], \forall l\Big)$$

Therefore, the sparsity of $(B^{(k)})^{-1}$ corresponds to the de-correlation of the Gaussian field on fine nodes after conditioning on coarse nodes (a phenomenon known as the *screening effect* in kriging [14], proved and used in [7, 13]).

2 Broader impact of the work accomplished

2.1 Brittleness/Robustness of Bayesian Inference and Machine Learning Algorithms

One of the results (initiated under FA9550-12-1-0389, completed under FA9550-16-1-0054 and whose impact is now manifest), was the derivation of Brittleness (non robustness) results for Bayesian estimators [10, 6, 9, 5].

Since this lack of robustness is caused by a mechanism that (1) is inherent to doing inference in a continuous world with finite-information (2) implies that consistency and robustness are conflicting requirements, we predicted that deep/machine learning algorithms could be non robust and could lead to increased confidence in incorrect solutions (see talk by Mike McKerns at SciPy 2013, https://www.youtube.com/watch?v=o-nwSnLC6DU&feature=youtu.be&t=74 published on July 2013).

Google engineers who were present at the talk tested these predictions for neural networks and observed in [16] (Szegedy et al, Dec 2013, Intriguing properties of neural networks) the non robustness of these algorithms to adversarial examples. This area has grown into the field known as **"advesarial machine learning"** which studies (1) how those algorithms could be attacked by exploiting their brittleness and (2) how to protect those algorithms against those attacks. Addressing these vulnerabilities is now recognized as critical to the safety of Machine Learning Algorithms.

2.2 The Mystic framework

FA9550-12-1-0389 supported the further development of the Mystic framework and the broader impact is now manifest under FA9550-16-1-0054. In particular Mystic and Pathos have been used at the Intelligence Systems Support Office (ISSO) of USAF and these codes are now in the top charts in GitHub (see https://hugovk.github.io/top-pypi-packages/ for official stats and links to code, mystic, pathos are rank 252 with 4.141M downloads per year).

Mystic is a highly-configurable framework for highly constrained non convex optimization and uncertainty quantification. Its environment includes Pathos, a parallel graph execution framework providing a high-level programmatic interface to high-performance computing. Both Mystic and Pathos were publicly released with planned long term support and a large number of individual downloads.

Mystic and Pathos have their own webpages at http://trac.mystic.cacr.caltech.edu/project/mystic/wiki.html and http://trac.mystic.cacr.caltech.edu/project/pathos/wiki.html They were initially developed under 12$M NSF IMR-MIP DANSE software project (for neutron scattering and optimization problems in material science) and upgraded for UQ calculations under the Caltech PSAAP, the LANL/LLNL ExMatEx and the AFOSR (FA9550-12-1-0389) projects. In addition to adding UQ components, under ExMatEx and FA9550-12-1-0389 asynchronous computing capabilities have been added to pathos. The klepto package (http://trac.mystic.cacr.caltech.edu/project/pathos/wiki/klepto.html) has been created to provide an abstraction for storage and retrieval of objects in a database, in memory, or on disk. Under ExMatEx and FA9550-12-1-0389 the majority of Mystic has also been converted to asynchronous computing, thus enabling optimization to dramatically scale in size and complexity.

3 List of publications

3.1 Published or accepted for publications

1. De-noising by thresholding operator adapted wavelets. G. R. Yoo and H. Owhadi, 2018. arXiv:1805.10736. To appear in Statistics and Computing.

2. Conditioning Gaussian measure on Hilbert space. H. Owhadi and C. Scovel. To appear in Journal of Mathematical and Statistical Analysis. 2018.

3. Qualitative Robustness in Bayesian Inference. ESAIM: Probability and Statistics, 2017. H. Owhadi and C. Scovel. arXiv:1411.3984

4. Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients. Journal of Computational Physics, Vol 347, pages 99-128, 2017. Houman Owhadi and Lei Zhang. arXiv:1606.07686

5. Separability of reproducing kernel spaces. H. Owhadi and C. Scovel. Proceedings of the AMS. Volume 145, Number 5, Pages 2131-2138, 2017. arXiv:1506.04288

6. Extreme points of a ball about a measure with finite support. H. Owhadi and C. Scovel. Communications in Mathematical Sciences. Vol. 15, No. 1, pp. 77–96, 2017. arXiv:1504.06745

7. Multigrid with rough coefficients and Multiresolution operator decomposition from Hierarchical Information Games. H. Owhadi. SIAM Review (Research spotlights), 59(1), 99-149, 2017. arXiv:1503.03467.

8. Towards Machine Wald (book chapter). H. Owhadi and C. Scovel. Springer Handbook of Uncertainty Quantification, 2017, pages 157–191, arXiv:1508.02449.

9. Brittleness of Bayesian inference and new Selberg formulas. H. Owhadi and C. Scovel. Communications in Mathematical Sciences, vol. 14, n. 1, pp. 83-145, 2016. arXiv:1304.7046

10. On the Brittleness of Bayesian Inference. H. Owhadi, C. Scovel and T. Sullivan. SIAM Review (Research spotlights), 57(4), 566-582, 2015. arXiv:1308.6306

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

3.2 Conference publications

1. The game theoretic approach to Uncertainty Quantification, reduced order modeling and numerical analysis. H. Owhadi. 19th AIAA Non-Deterministic Approaches Conference Grapevine, Texas, 2017.

3.2.1 Preprints

1. Statistical Numerical Approximation. H. Owhadi, F. Schäfer and C. Scovel. Under Review in Notices of the AMS.
2. Kernel Flows: from learning kernels from data into the abyss. H. Owhadi, G. R. Yoo, 2018. arXiv:1808.04475
3. Fast eigenpairs computation with operator adapted wavelets and hierarchical subspace correction. H. Xie, L. Zhang and H. Owhadi, 2018. arXiv:1806.00565
4. Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity. F. Schäfer, T. J. Sullivan and H. Owhadi. 2017. arXiv:1706.02205
5. Universal Scalable Robust Solvers from Computational Information Games and fast eigenspace adapted Multiresolution Analysis. H. Owhadi and C. Scovel. 2017. arXiv:1703.10761

3.2.2 Book

1. Operator adapted wavelets, fast solvers, and numerical homogenization from a game theoretic approach to numerical approximation and algorithm design. H. Owhadi and C. Scovel. The Cambridge Monographs on Applied and Computational Mathematics series 2019 (Cambridge University Press).

4 List of presentations

1. November 6-8, 2015. University of Texas at Dallas. Texas Analysis and Mathematical Physics Symposium 2015.
2. January 19 - 22, 2016. IPAM. Uncertainty Quantification for Multiscale Stochastic Systems and Applications.
3. April 5-8, 2016. EPFL. SIAM UQ 2016. Mini-tutorial.
4. May 30 June 2, 2016. City University of Hong Kong. International Conference on Applied Mathematics.

5. June 18-20, 2016. 4th CAM-ICCM Workshop, Multiscale and Large-scale Scientific Computing, Chinese University of Hong Kong (CUHK).

6. August 1-5, 2016. RWTH Aachen University (Germany). XVI International Conference on Hyperbolic Problems: Theory, Numerics, Applications. Plenary.

7. January 4, 2017. AIAA SciTech. DARPA Efficient Quantification of Uncertainty in Physical Systems.

8. February 27-March 3, 2017. SIAM CSE 2017, EQUiPS minisymposia.

9. April 3-7, 2017, IPAM workshop "Multiphysics, Multiscale, and Coupled Problems in Subsurface Physics".

10. April 10-14, 2017, "Multiscale Problems: Algorithms, Numerical Analysis and Computation" Hausdorff Trimester Program.

11. June 5-10, 2017. ICERM. Probabilistic Scientific Computing: Statistical inference approaches to numerical analysis and algorithm design (co-organizer).

12. June 19-23, 2017. Dynamics, aging and universality in complex systems.

13. September 12-15, 2017. Complex High-Dimensional Energy Landscapes Tutorials.

14. October 30, 2017. Computing@PNNL seminar series.

15. Januray 24, 2018. Stanford applied math seminar.

16. March 26, 2018: University of Notre Dame. Center for Informatics and Computational Science Colloquium.

17. April 5, 2018: John Hopkins University, seminar.

18. April 6, 2018: Applied Math Colloquium, UMBC.

19. April 10, 2018: Jussieu (Paris VI). Seminaire du Laboratoire de Probabilités, Statistique et Modélisation.

20. April 11-13, 2018. The Alan Turing Institute, London. SAMSI Workshop on Probabilistic Numerics.

21. April 16, 2018. SIAM UQ 2018, MS17 Probabilistic Numerical Methods for Quantification of Discretisation Error.

22. April 20, 2018: Institut de Mathematiques de Marseille. Séminaire Probabilités et Statistique.

23. April 23-27, 2018: BIRS, Numerical Analysis and Approximation Theory meets Data Science.

24. July 10, 2018. SIAM AN 2018, MS100 Machine Learning for Scientific Computing.

25. July 24, 2018. WCCM 2018, MS104.

26. November 5-7, 2018, RICAM (Linz), Multivariate Algorithms and Information-Based Complexity.

5 Collaborations

1. Joel Tropp (Caltech). Collaboration on fast universal solvers with Gamblets.

2. Peter Schröder (Caltech). Collaboration on the fast inversion of complex connection Laplacian with gamblets.

3. Mathieu Desbrun (Caltech). Collaboration on geometric integration and model reduction in fluid dynamics with gamblets.

4. Vikram Gavini (University of Michigan). Collaboration on Large Scale Electronic Structure Calculations with Gamblets.

5. Lei Zhang (Shanghai Jiaotong University). Collaboration on fast eigen-subspace decompositions and stochastic PDEs with Gamblets.

6. Hehu Xie (Chinese Academy of Sciences). Collaboration on fast eigen-subspace decompositions with Gamblets.

7. Qingfu Zhang (China University of Petroleum). Collaboration on the application of gamblets to transport in fractured media.

8. Bruce Suter (AFRL). Collaboration on gamblets.

References

[1] L. Berlyand and H. Owhadi. Flux norm approach to finite dimensional homogenization approximations with non-separated scales and high contrast. *Archives for Rational Mechanics and Analysis*, 198(2):677–721, 2010.

[2] J. M. Melenk. On n-widths for elliptic problems. *J. Math. Anal. Appl.*, 247(1):272–289, 2000.

[3] H. Owhadi. The game theoretic approach to uncertainty quantification, reduced order modeling and numerical analysis. In *19th AIAA Non-Deterministic Approaches Conference*, page 1092, 2017.

[4] H. Owhadi. Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games. *SIAM Rev.*, 59(1):99–149, 2017.

[5] H. Owhadi and C. Scovel. Qualitative Robustness in Bayesian Inference. 2014. arXiv:1411.3984.

[6] H. Owhadi and C. Scovel. Brittleness of Bayesian inference and new Selberg formulas. *Communications in Mathematical Sciences*, 14:83–145, 2016. arXiv:1304.7046.

[7] H. Owhadi and C. Scovel. Universal scalable robust solvers from computational information games and fast eigenspace adapted multiresolution analysis. *arXiv:1703.10761*, 2017.

[8] H. Owhadi and C. Scovel. *Operator adapted wavelets, fast solvers, and numerical homogenization from a game theoretic approach to numerical approximation and algorithm design*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2019.

[9] H. Owhadi, C. Scovel, and T. Sullivan. On the brittleness of Bayesian inference. *SIAM Review (Research Spotlights)*, 57(4):566–582, 2015.

[10] H. Owhadi, C. Scovel, and T. J. Sullivan. Brittleness of Bayesian inference under finite information in a continuous world. *Electron. J. Statist.*, 9:1–79, 2015.

[11] H. Owhadi and L. Zhang. Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients. *Journal of Computational Physics*, 2017.

[12] A. Sard. Optimal approximation. *Journal of Functional Analysis*, 1(2):222–244, 1967.

[13] F. Schäfer, T. J. Sullivan, and H. Owhadi. Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity. *arXiv:1706.02205*, 2017.

[14] M. L. Stein. The screening effect in Kriging. *Annals of statistics*, pages 298–323, 2002.

[15] R. Sudarshan. *Operator-adapted finite element wavelets: Theory and applications to a posteriori error estimation and adaptive computational modeling.* ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)–Massachusetts Institute of Technology.

[16] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[17] Gene Ryan Yoo and Houman Owhadi. De-noising by thresholding operator adapted wavelets. *arXiv preprint arXiv:1805.10736, to appear in Statistics and Computing*, 2018.