AFRL-AFOSR-VA-TR-2021-0107

Games for Computation and Learning

Owhadi, Houman
CALIFORNIA INSTITUTE OF TECHNOLOGY
1200 E CALIFORNIA BLVD
PASADENA, CA, 91125
USA

08/17/2021
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
17-08-2021	Final	15 Jul 2018 - 14 Jul 2021

4. TITLE AND SUBTITLE
Games for Computation and Learning

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-18-1-0271

5c. PROGRAM ELEMENT NUMBER
61102F

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
PASADENA, CA 91125
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2021-0107

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project had two main objectives for methods emerging at the interface be-tween game theory, uncertainty quantiï¬•cation, and numerical approximation (I) their continued application to high impact problems of practical importance in compu-tational mathematics (II) their development towards machine learning. With this purpose and a dual emphasis on conceptual/theoretical advancements and algorith-mic/computational complexity advancements the accomplishments of this program are as follows. (1) We have developed general robust methods for learning kernels through (a) hyperparameter tuning via Kernel Flows (a variant of cross-validation) with applications to learning dynamical systems and to the extrapolation of weather time series, and (b) programming kernels through interpretable regression networks (kernel mode decomposition) with applications to empirical mode decomposition.(2) We have discovered a very robust and massively parallel algorithm, based on Kullback-Liebler divergence (KL) minimization that computes accurate approxima-tions of the inverse Cholesky factors of dense kernel matrices with rigorous a priori O(N log(N) log2d(N/î€•) complexity vs. accuracy guarantees

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			FARIBA FAHROO
U	U	U	UU	13	**19b. TELEPHONE NUMBER** *(Include area code)* 426-8429

Standard Form 298 (Rev.8/98)
Prescribed by ANSI Std. Z39.18

Games for Computation and Learning

Final Report

Program Manager: Dr. Fariba Fahroo AFOSR/RTA
Contract/Grant #: FA9550-18-1-0271.
Reporting Period: 7/15/2018 to 7/14/2021
PI: Houman Owhadi
Organization: California Institute of Technology

Abstract:

This project had two main objectives for methods emerging at the interface between game theory, uncertainty quantification, and numerical approximation (I) their continued application to high impact problems of practical importance in computational mathematics (II) their development towards machine learning. With this purpose and a dual emphasis on conceptual/theoretical advancements and algorithmic/computational complexity advancements the accomplishments of this program are as follows. (1) We have developed general robust methods for learning kernels through (a) hyperparameter tuning via Kernel Flows (a variant of cross-validation) with applications to learning dynamical systems and to the extrapolation of weather time series, and (b) programming kernels through interpretable regression networks (kernel mode decomposition) with applications to empirical mode decomposition. (2) We have discovered a very robust and massively parallel algorithm, based on Kullback-Liebler divergence (KL) minimization that computes accurate approximations of the inverse Cholesky factors of dense kernel matrices with rigorous a priori $\mathcal{O}(N\log(N)\log^{2d}(N/\epsilon)$ complexity vs. accuracy guarantees (this is the new state of the art) (3) We have introduced Competitive Gradient Descent, a surprisingly simple but powerful generalization of the gradient descent to the two-player setting where the update is given by the Nash equilibrium of a regularized bilinear local approximation of the underlying game. This algorithm avoids oscillatory and divergent behaviors seen in alternating gradient descent, and the ability to choose larger step-sizes furthermore allows the proposed algorithm to achieve faster convergence. (4) We have developed a rigorous framework for the analysis of artificial neural networks as discretized image registration algorithms with images replaced by high dimensional functions in high dimensional spaces. (5) We have introduced a general Gaussian Process/Kernel method approach for solving and learning arbitrary nonlinear PDEs. (6) We have introduced a new Uncertainty Quantification framework addressing the

limitations of traditional approaches (in terms of accuracy, robustness, and computational complexity).

1 Accomplishments in Scientific Computing

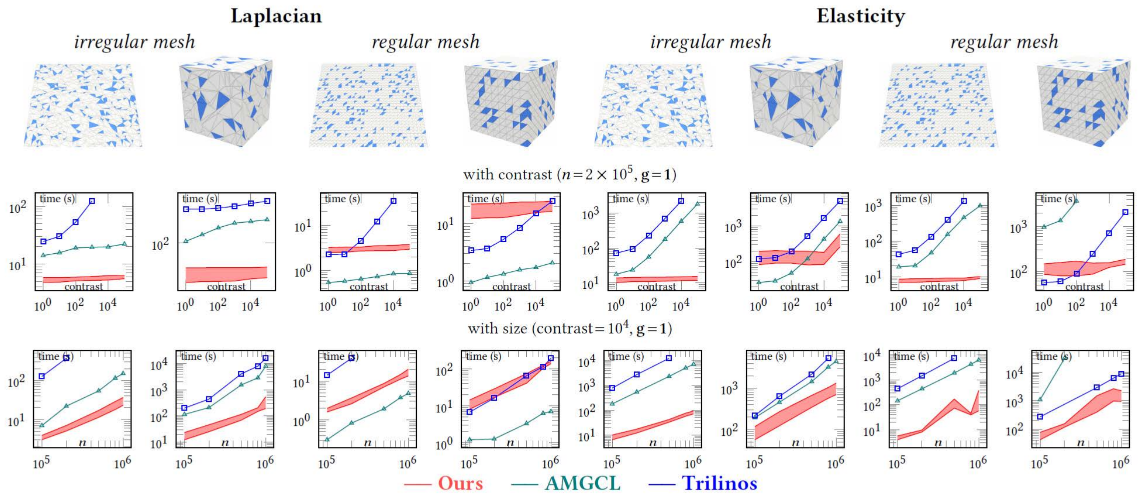

Figure 1.1: [2, Fig. 14]. Comparison between our sparse Cholesky solver and two implementations of AMG from two libraries, namely AMGCL [5] and Trilinos [1].

1.1 Cholesky factorization of dense kernel matrices and fast linear solvers

- [16] has been accepted for publication by SIAM SISC. This paper introduces a rigorous and embarrassingly parallel sparse Cholesky factorization algorithm for the inversion of the dense kernel matrices (useful for kernel methods, Gaussian process regression, the computation of integral operators, etc.). The algorithm is both rigorous (with the best-known complexity vs. accuracy guarantees) and practical.
  - A presentation is available online at https://www.youtube.com/watch?v=VAGtjw0_Mj8.
  - This work introduces an algorithm for computing the inverse Cholesky factors of an $N \times N$dense kernel matrix in (rigorous) $Nlog^{2d}(N/\epsilon)$ complexity (this is the new state of the art)
  - The algorithm is also very robust and practical (it is stable and has small constants): it achieves the factorization of a $10^6 \times 10^6$ dense kernel matrix in 10s on a single core and is also massively parallelizable.
  - This work also solves a major problem in computational statistics (the efficient factorization of dense kernel matrices with additive noise).

  - Julia codes have been released at https://github.com/f-t-s/cholesky_by_KL_minimization.
  - ARL [18] has used an older version [17] to significantly accelerate surrogate modeling for energetic materials.

- [2] has presented a concrete implementation (as a fast linear solver for elliptic PDEs) of the sparse Cholesky factorization algorithm introduced in [17]. As observed in [2], the proposed approach

  > "far exceeds the performance of existing carefully-engineered libraries for graphics problems involving bad mesh elements and/or high contrast of coefficients."

  See Fig. 1.1 (The proposed method is robust to high contrast and more efficient than AMG.

1.2 Solving and learning nonlinear PDEs with GPs

[3] has introduced a general Gaussian Process/Kernel method approach for solving and learning arbitrary nonlinear PDEs. This approach is an extension of the gamblet framework [10, 8, 11] to nonlinear PDEs. This method was co-developed under AFOSR MURI # 19 and the complexity of the proposed approach reduces to the inversion/compression of dense kernel matrices. Therefore solving and learning nonlinear PDEs inherit the (state of the art) complexity of algorithm developed in [16] under FA9550-18-1-0271.

1.3 Publications

The following book has been published by Cambridge University Press

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

The research performed under this program has also been featured in Notices of the AMS

- Statistical Numerical Approximation. H. Owhadi, C. Scovel and F. Schäfer. Notices of the American Mathematical Society, volume 66, number 10, featured article, pages 1608-1617, 2019.

We also report the following publications.

- Sparse Cholesky Factorization by Kullback–Leibler Minimization Schäfer, Florian, Katzfuss, Matthias, and Owhadi, Houman SIAM J. Sci. Comput. 2021.
- Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity Schäfer, Florian, Sullivan, T. J., and Owhadi, Houman Multiscale Model. Simul. 2021.
- Material-adapted Refinable Basis Functions for Elasticity Simulation Jiong Chen, Max Budninskiy, Houman Owhadi, Hujun Bao, Jin Huang, and Mathieu Desbrun. ACM Trans. Graph. (SIGGRAPH Asia), 38(6), Art. 161, 2019.
- Fast eigenpairs computation with operator adapted wavelets and hierarchical subspace correction. H. Xie, L. Zhang and H. Owhadi, SIAM Journal on Numerical Analysis, 2019.
- J. Chen, F. Schäfer, J. Huang, and M. Desbrun. Multiscale Cholesky preconditioning for ill-conditioned problems. Sci- Graph. 2021.
- Solving and Learning Nonlinear PDEs with Gaussian Processes. 2021. Y. Chen, B. Hosseini, H. Owhadi, AM. Stuart [arXiv:2103.12959]

2 Accomplishments in learning

2.1 Kernel flows

We have introduced kernel flows (KF) [12] from a numerical approximation approach to kernel design and learning. Kernel Flows (KF) offer a scalable solution to the kernel construction/selection problem based on the simple premise that a kernel K must be good if the number N of interpolation points can be halved without significant loss in accuracy (measured using the intrinsic RKHS norm $\|\cdot\|_K$ associated with the kernel).

2.1.1 Learning dynamical systems

Regressing the vector field of a dynamical system from a finite number of observed states is a natural way to learn surrogate models for such systems. In [7] we used Kernel Flows [12] and its variants as simple and effective approaches for learning the kernel used in these emulators.

2.1.2 Extrapolation of weather/climate time series

Figure 2.1 provides an example of application of [7] to the extrapolation of weather/climate time series. In that example, in collaboration with Argonne National Lab [6] (see figures 2.1 and 2.2), we used Kernel Flows [12] (an algorithm performing Kriging with a kernel learned from data) for weather time series prediction and compared those

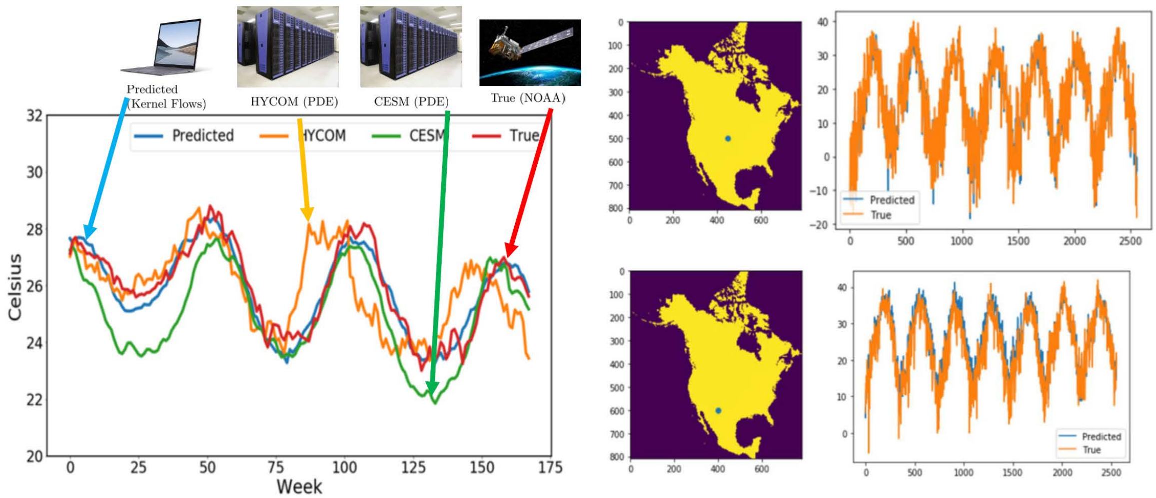

Figure 2.1: Left: Weather prediction, true satellite data vs Kernel Flows vs 2 ANL PDE based models ran on HPC. Right: Kernel Flows predictions vs true data.

predictions to true satellite data to PDE based models (HYCOM and CESM, developed by ANL) run on HPC and to LSTM neural networks (also developed by ANL and run on HPC). Our simple laptop-run algorithm ended up outperforming both ANL PDE-based models and LSTM neural networks both in terms of complexity and accuracy. Indeed, HYCOM took 800 core-hours per day of forecast on a Cray XC40 system. CESM took 17 million core-hours on Yellowstone, NCAR's high-performance computing resource. The architecture optimized LSTM took 3 hours of wall time on 128 compute nodes of the Theta supercomputer. Our method took 40 seconds to train on a single node machine (a laptop) without acceleration

2.1.3 Analysis

Hierarchical modeling and learning has proven very powerful in the field of Gaussian process regression and kernel methods, especially for machine learning applications and, increasingly, within the field of inverse problems more generally. The classical approach to learning hierarchical information is through Bayesian formulations of the problem, implying a posterior distribution on the hierarchical parameters or, in the case of empirical Bayes, providing an optimization criterion for them. In [4] we have compared the empirical Bayesian and the Kernel Flows [12] approach to hierarchical learning, in terms of large data consistency, variance of estimators, robustness of the estimators to model misspecification, and computational cost. Our analysis shows the consistency, near optimal accuracy and robustness of Kernel Flows.

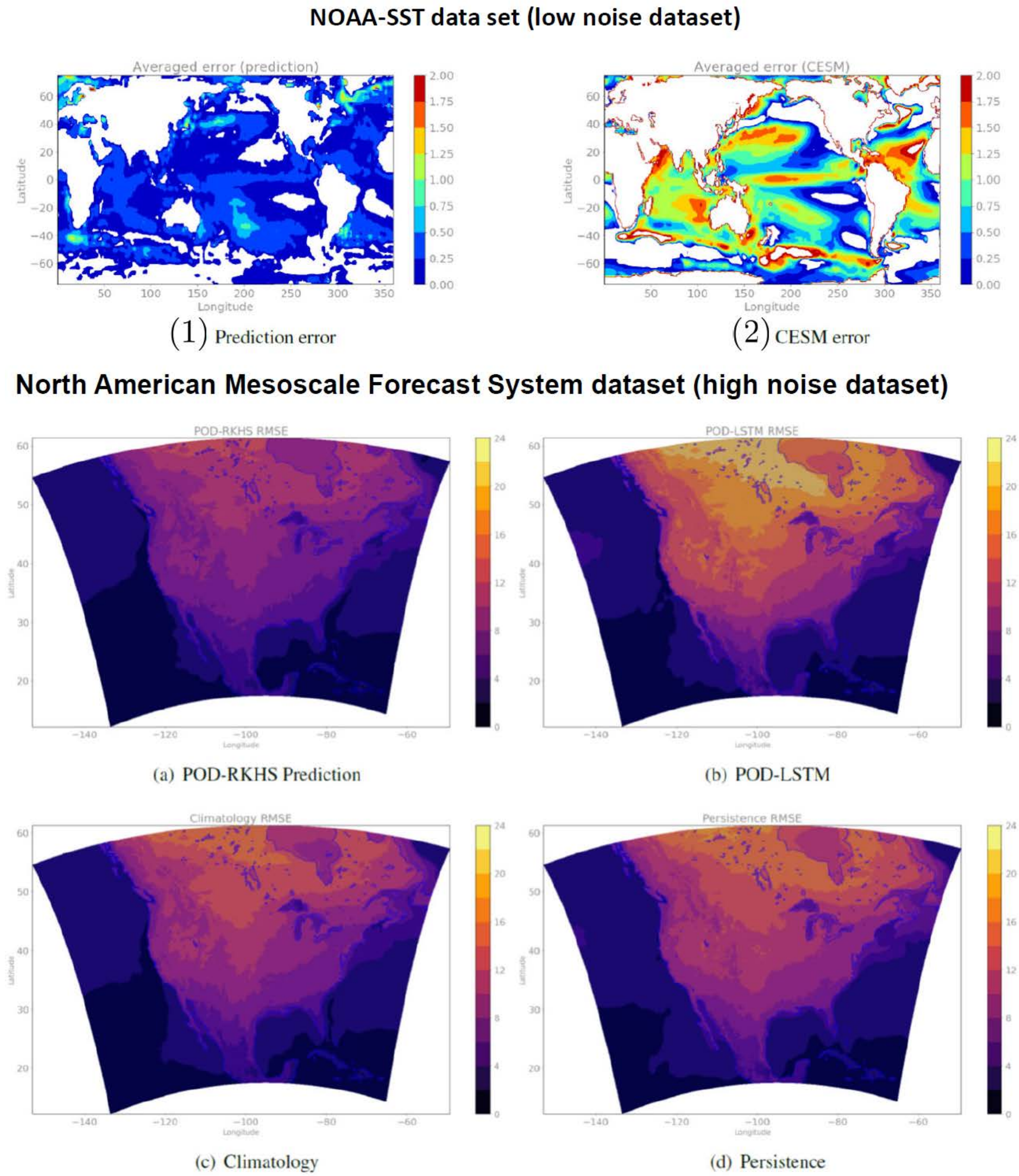

Figure 2.2: Top row: Comparison between (1) our method and (2) CESM on the NOAA-SST dataset. Bottom rows: comparisons between (a) our method (b) an architecture optimized LSTM (c) persistence and (d) climatology.

2.2 Kernel mode decomposition and the programming of kernels

In [13] we introduced a new approach to the classical mode decomposition problem through nonlinear regression models, which achieve near-machine precision in the recovery of the modes. Although kernel methods have strong theoretical foundations, they require the prior selection of a good kernel. While the usual approach to this kernel selection problem is hyperparameter tuning, we discovered an alternative (programming) approach to the kernel selection problem while using mode decomposition as a prototypical pattern recognition problem. In this approach, kernels are programmed for the task at hand through the programming of interpretable regression networks in the context of additive Gaussian processes. As a prototypical application we have programmed regression networks approximating the modes $v_i = a_i(t)y_i\big(\theta_i(t)\big)$ of a (possibly noisy) signal $\sum_i v_i$ when the amplitudes a_i, instantaneous phases θ_i and periodic waveforms y_i may all be unknown and shown near machine precision recovery under regularity and separation assumptions on the instantaneous amplitudes a_i and frequencies $\dot{\theta}_i$. Python codes have been released at https://github.com/kernel-enthusiasts/Kernel-Mode-Decomposition-1D.

2.3 Competitive gradient descent

We have introduced Competitive Gradient Descent [14], a surprising simple but powerful generalization of the gradient descent to the two-player setting where the update is given by the Nash equilibrium of a regularized bilinear local approximation of the underlying game. This algorithm avoids oscillatory and divergent behaviors seen in alternating gradient descent and the ability to choose larger stepsizes furthermore allows the proposed algorithm to achieve faster convergence.

In [15] we have introduced competitive mirror descent (CMD, a generalization of competitive gradient descent [14] completed during year 1), a general-purpose algorithm for solving constrained competitive problems, as a counterpart of gradient descent in unconstrained single-agent optimization.

2.4 Idea registration

We have shown that artificial neural networks are discretizations of image registration problems with images replaced by high dimensional RKHS spaces [9]. Whereas image registration compares images through deformations of their coordinate systems, idea registration compares abstractions (ideas) through deformations of their feature space representations (Fig. 2.3). The proposed theory provides principled solutions to many questions in deep learning such as: (a) how to design good architectures? (b) how to design neural networks that are equivariant to an arbitrary group of transformation? (c) what is the stochastic process underlying deep learning? (d) How to make Deep Learning rigorously robust to attacks?

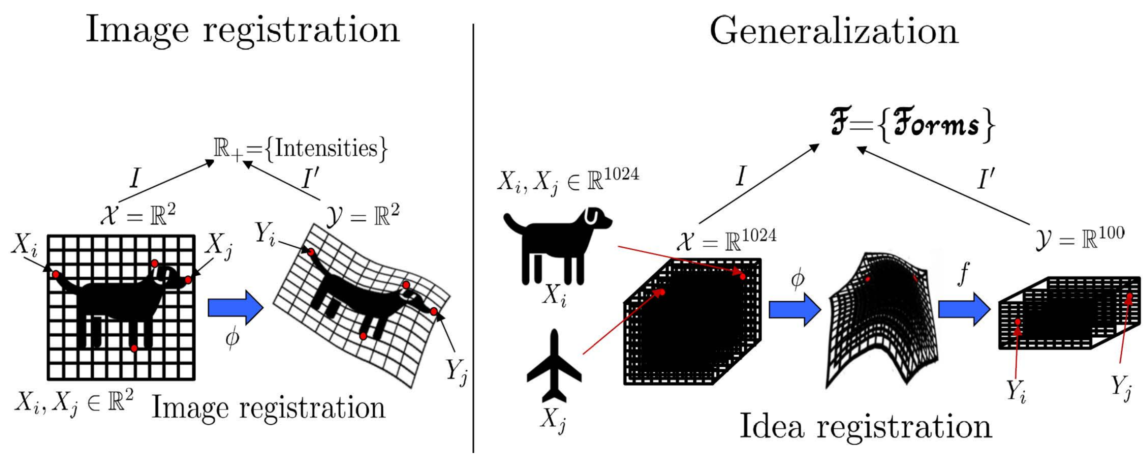

Figure 2.3: The continuous limit of ANNs is a registration problem with images replaced by high dimensional functions.

Technical approach. Learning can be seen as approximating an unknown function $f^\dagger : \mathcal{X} \to \mathcal{Y}$ by regressing the training data (X, Y). Given a kernel K, ridge regression approximates $f^\dagger$ with the minimizer of $\lambda\|f\|_K^2 + \|f(X) - Y\|_{\mathcal{Y}}^2$. Given another kernel Γ defining an RKHS of functions mapping $\mathcal{X}$ to $\mathcal{X}$, consider the mechanical regression (MR) problem of approximating f with a minimizer of $(\nu L/2)\sum_{i=1}^L \|v_i\|_\Gamma^2 + \lambda\|f\|_K^2 + \|f \circ \phi(X) - Y\|_{\mathcal{Y}}^2$ where $\phi = (I + v_L) \circ \cdots \circ (I + v_1)$ is a deformation of the input space obtained from the composition of L small deformations $v_k : \mathcal{X} \to \mathcal{X}$. In the limit where $L \to \infty$ minimizers of (MR) converge to minimizers of the idea registration (IR) problem $(\nu/2)\int_0^1 \|v_t\|_\Gamma^2\, dt + \lambda\|f\|_K^2 + \|f \circ \phi(X) - Y\|_{\mathcal{Y}}^2$ with $\dot{\phi} = v(\phi(\cdot, t), t)$. Minimizers of (IR) have the representation $\dot{\phi} = \Gamma(\phi, q)p$ where (q, p) follows a Hamiltonian dynamic with energy $p^T\Gamma(q, q)p/2$. If the kernels Γ and K are obtained from feature maps of the form $\varphi(x) = A\mathbf{a}(x) + c$ where $\mathbf{a}$ is an activation function defined as an elementwise nonlinearity, then minimizers of (MR) are minimizers of (AF) $(\nu L/2)\sum_{i=1}^L(\|w^k\|^2 + \|b^k\|^2) + \lambda(\|\tilde{w}^k\|^2 + \|\tilde{b}^k\|^2) + \|f \circ \phi(X) - Y\|_{\mathcal{Y}}^2$ with $f = \tilde{w}\mathbf{a}(\cdot) + \tilde{b}$ and $v_k = w^k\mathbf{a}(\cdot) + b^k$. $f \circ \phi$ obtained from (AF) has the exact structure of one block of a residual neural network (ResNet). Iterating (AF) over a hierarchy of spaces (layered in between $\mathcal{X}$ and $\mathcal{Y}$) produces input-output functions that have the exact structure of Artificial Neural Networks (ANNs) and ResNets. If K and Γ are projected (reduced) and averaged with respect to the action of a group (e.g., translations), the input-output functions obtained from the hierarchical version of (AF) is a convolutional neural network (CNN). The convergence of (MR) towards (IR) implies that ANNs, ResNets and CNNs are discretized idea registration problems (they converge towards (IR) in the continuous/infinite depth limit). Whereas image registration compares two images by creating alignments via deformations of their coordinate systems, idea registration compares abstractions (ideas) by creating alignments via hierarchical deformations of their feature spaces. This identification of ANNs as discretized idea registration problems has several immediate consequences such as: (1) Energy preservation implies that, at a minimum, the Euclidean norm of the weights and biases of a ResNet must be preserved across layers. (2) The search for good architectures for ANNs is equivalent to the search of good kernels for (IR), in particular, we identify architectures generalizing CNNs that are equivariant (preserving the relative pose information) to the action of arbitrary groups of transformations. (3) Trained ANNs can be identified as MAP estimators of deep residual Gaussian processes (introduced here). (4) L_2 regularized ANNs satisfy a neural least action principle and can be represented via the Hamiltonian dynamic generated by their feature maps. (5) Trained ResNets are discretized integrators for these Hamiltonian ODEs. (6) The identification of hidden momentum variables in CNNs. (7) The generalization of CNNs/ResNets to arbitrary Hilbert spaces (possibly infinite-dimensional and functional).

2.5 Publications

The following book will be published by Springer

- Kernel mode decomposition and the programming of kernels. H. Owhadi, C. Scovel and G. R. Yoo. Springer, 2021. [arXiv:1907.08592]

We also report the following publications.

- Data-driven geophysical forecasting: Simple, low-cost, and accurate baselines with kernel methods. 2021 B. Hamzi, R. Maulik, H. Owhadi [arXiv:2103.10935]. To appear in Proc. Royal Society A.
- Do ideas have shape? Plato's theory of forms as the continuous limit of artificial neural networks. 2020 H. Owhadi [arXiv:2008.03920]
- Deep regularization and direct training of the inner layers of Neural Networks with Kernel Flows. 2020. G. R. Yoo and H. Owhadi. Physica D: Nonlinear Phenomena, 2021
- Consistency of Empirical Bayes And Kernel Flow For Hierarchical Parameter Estimation. Mathematics of Computation, 2021. Y. Chen, H. Owhadi, A. M. Stuart [arXiv:2005.11375]
- Learning dynamical systems from data: a simple cross-validation perspective, part I: Parametric kernel flows. B. Hamzi and H. Owhadi Physica D: Nonlinear Phenomena, Volume 421, 2021
- Kernel Flows: from learning kernels from data into the abyss. H. Owhadi, G. R. Yoo, Journal of Computational Physics, volume 389, Pages 22-47, 2019

- De-noising by thresholding operator adapted wavelets. G. R. Yoo and H. Owhadi, Statistics and Computing, volume 29, number 6, 1185–1201, 2019.
- Competitive gradient descent. F. Schäfer and A. Anandkumar. In Advances in Neural Information Processing Systems (pp. 7625-7635), 2019.
- Implicit competitive regularization in GANs Schäfer, Florian, Zheng, Hongkai, and Anandkumar, Anima In the 37th International Conference on Machine Learning (ICML 2020).
- Robust Reinforcement Learning: A Constrained Game-theoretic Approach Yu, Jing, Gehring, Clement, Schäfer, Florian, and Anandkumar, Animashree In Proceedings of the 3rd Conference on Learning for Dynamics and Control 2021.
- Competitive Mirror Descent Schäfer, Florian, Anandkumar, Anima, and Owhadi, Houman, 2021. [arXiv:2004.14455]

3 Accomplishments in Uncertainty Quantification

There are essentially three kinds of approaches to Uncertainty Quantification (UQ): (A) robust optimization (min and max) (B) Bayesian (conditional average) (C) decision theory (minimax). Although (A) is robust, it is unfavorable with respect to accuracy and data assimilation. (B) requires a prior, it is generally non-robust (brittle) with respect to the choice of that prior and posterior estimations can be slow. Although (C) leads to the identification of an optimal prior, its approximation suffers from the curse of dimensionality and the notion of loss/risk used to identify the prior is one that is averaged with respect to the distribution of the data. We have introduced a 4th kind which is a hybrid between (A), (B), (C) and and hypothesis testing. It can be summarized as, after observing a sample x, (1) defining a likelihood region through the relative likelihood and (2) playing a minmax game in that region to define optimal estimators and their risk. The resulting method has several desirable properties (a) an optimal prior is identified after measuring the data and the notion of loss/risk is a posterior one, (b) the determination of the optimal estimate and its risk can be reduced to computing the minimum enclosing ball of the image of the likelihood region under the quantity of interest map(such computations are fast and do not suffer from the curse of dimensionality). The method is characterized by a parameter in $[0, 1]$ acting as an assumed lower bound on the rarity of the observed data (the relative likelihood). When that parameter is near 1, the method produces a posterior distribution concentrated around a MLE with tight but low confidence UQ estimates. When that parameter is near 0, the method produces a maximal risk posterior distribution with high confidence UQ estimates. In addition to navigating the accuracy-uncertainty tradeoff, the proposed method addresses the brittleness of Bayesian inference by navigating the robustness-accuracy tradeoff associated with data assimilation.

4 Honors and awards

PI H. Owhadi has been awarded the 2019 SIAM Dahlquist prize. Florian Schäfer has been awarded the 2021 W.P. Carey prize in applied and computational mathematics.

5 Training

Florian Schäfer defended his Ph.D. in June 2021 and will be joining Georgia Tech as an assistant professor in computational science and engineering this August 2021. Gene Ryan Yoo defended his Ph.D. in May 2020 and joined Susquehanna International Group as a Quantitative Researcher.

References

[1] The trilinos project website, 2020.

[2] JIONG CHEN, FLORIAN SCHÄFER, JIN HUANG, and MATHIEU DESBRUN. Multiscale cholesky preconditioning for ill-conditioned problems. *Sci-Graph*.

[3] Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M Stuart. Solving and learning nonlinear pdes with gaussian processes. *arXiv preprint arXiv:2103.12959*, 2021.

[4] Yifan Chen, Houman Owhadi, and Andrew M Stuart. Consistency of empirical bayes and kernel flow for hierarchical parameter estimation. *arXiv preprint arXiv:2005.11375*, 2020.

[5] Denis Demidov. Amgcl: An efficient, flexible, and extensible algebraic multigrid implementation. *Lobachevskii Journal of Mathematics*, 40(5):535–546, 2019.

[6] Boumediene Hamzi, Romit Maulik, and Houman Owhadi. Data-driven geophysical forecasting: Simple, low-cost, and accurate baselines with kernel methods. *arXiv preprint arXiv:2103.10935*, 2021.

[7] Boumediene Hamzi and Houman Owhadi. Learning dynamical systems from data: A simple cross-validation perspective, part i: Parametric kernel flows. *Physica D: Nonlinear Phenomena*, 421:132817, 2021.

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

[9] Houman Owhadi. Do ideas have shape? Plato's theory of forms as the continuous limit of artificial neural networks. *arXiv preprint arXiv:2008.03920*, 2020.

[10] Houman Owhadi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*, volume 35. Cambridge University Press, 2019.

[11] Houman Owhadi, Clint Scovel, and Florian Schäfer. Statistical numerical approximation. *Notices of the AMS*, 2019.

[12] Houman Owhadi and Gene Ryan Yoo. Kernel flows: From learning kernels from data into the abyss. *Journal of Computational Physics*, 389:22–47, 2019.

[13] Houman Owhadi and Gene Ryan Yoo. Kernel flows: From learning kernels from data into the abyss. *Journal of Computational Physics*, 389:22–47, 2019.

[14] Florian Schäfer and Anima Anandkumar. Competitive gradient descent. *arXiv:1905.12103*, 2019.

[15] Florian Schäfer, Anima Anandkumar, and Houman Owhadi. Competitive mirror descent. *arXiv preprint arXiv:2006.10179*, 2020.

[16] Florian Schäfer, Matthias Katzfuss, and Houman Owhadi. Sparse Cholesky factorization by Kullback-Leibler minimization. *SIAM SISC*, 2020. arXiv preprint arXiv:2004.14455.

[17] Florian Schäfer, Timothy John Sullivan, and Houman Owhadi. Compression, inversion, and approximate pca of dense kernel matrices at near-linear computational complexity. *SIAM MMS*, 2017. arXiv preprint arXiv:1706.02205.

[18] Ting Wang, Kenneth W Leiter, Petr Plecháč, and Jaroslaw Knap. Accelerated scale bridging with sparsely approximated gaussian learning. *Journal of Computational Physics*, 403:109049, 2020.