Symposium on Sparsity and Singular Structures 2024

Poster Abstracts

Robust Statistical Estimation for Multiscale SDE Systems

Jaroslav Borodavka

In many applications of the natural sciences, models are employed which are based on processes occurring across different length and time scales. While these multiscale systems may be constructed and modeled in several ways, an important class are those described by (random) dynamical systems, in particular stochastic differential equations. For such multiscale systems it is possible to rigorously derive a so-called surrogate (or effective) model, a simplified reduced order model that captures the system's most dominant features.

However, oftentimes it is not feasible to derive closed-form expressions for the drift and diffusion coefficient of the surrogate model analytically, because of the complexity of the underlying multiscale system or simply because the full model is not completely known. In this case one would resort to a data-driven modeling strategy in order to estimate the drift and diffusion coefficient of the surrogate model from observations of the multiscale system.

It has been shown that standard statistical learning techniques fail to provide consistent estimates when confronted with such observations. The maximum likelihood estimator is an example of a well-established statistical learning estimator that fails to converge for multiscale observations. In order to overcome these recent drawbacks, we are, in this project, mainly concerned with the development of novel unbiased and effective data-driven training strategies for the identification and analysis of surrogate models for multiscale stochastic differential equations.

To tackle this problem, we will opt for different approaches in different settings. One such approach is a minimum distance procedure in a parametric setting that is based on the difference between the characteristic function of the invariant density of a homogenized stochastic differential equation and its empirical counterpart in a suitable weighted L^2-space. Under certain assumptions, we can prove desirable statistical properties of the constructed minimum distance estimator, such as robustness and asymptotic normality under multiscale observations. Numerical (and further theoretical) case studies for a general first-order Langevin equation with a two-scale potential substantiate the theoretical findings.

Memorization With Neural Nets: Going Beyond the Worst Case

Patrick Finke

In practice, deep neural networks are often able to easily interpolate their training data. To understand this phenomenon, many works have aimed to quantify the memorization capacity of a neural network architecture: the largest number of points such that the architecture can interpolate any placement of these points with any assignment of labels. For real-world data, however, one intuitively expects the presence of a benign structure so that interpolation already occurs at a smaller network size than suggested by memorization capacity.

We investigate interpolation by adopting an instance-specific viewpoint. To this end, we introduce a simple randomized algorithm that, given a fixed finite data set with two classes, with high probability constructs an interpolating three-layer neural network in polynomial time. The required number of parameters is linked to geometric properties of the two classes and their mutual arrangement. As a result, we obtain guarantees that are independent of the number of samples and hence move beyond worst-case memorization capacity bounds. We illustrate the effectiveness of the algorithm in non-pathological situations with numerical experiments and link the insights back to the theoretical results.

This is joint work with Sjoerd Dirksen and Martin Genzel. The preprint can be found on arXiv: `https://arxiv.org/abs/2310.00327`.

Low-rank Tensor Recovery via Tractable Algorithms

Arinze Folarin

The project focuses on recovering low-rank tensors from given observations, employing tractable algorithms such as Iterative Hard Thresholding and Riemannian Gradient Iteration. The objective is to establish optimal estimates for the minimal number of measurements required to recover the low-rank tensor through the utilization of these tractable algorithms.

Scattering transforms of sparse signals

Max Getter

Mallat's windowed scattering transform is a convolutional neural network, typically with the modulus as nonlinearity and pre-specified filters. We study the impact of those filters on the response behavior of the scattering transform. An important criterion for assessing the strength of scattering networks is the speed of energy propagation across layers, which, as we show, can be arbitrarily slow in wavelet scattering networks.

Schwarz domain decomposition method applied to the conductor like screening model

Nibedita Ghosh[1], Arnold Reusken[1], and Benjamin Stamm[2]

[1] *Institute for Geometry and Practical Mathematics, RWTH Aachen University*
ghosh@igpm.rwth-aachen.de, reusken@igpm.rwth-aachen.de

[2] *Institute of Applied Analysis and Numerical Simulation, University of Stuttgart*
best@ians.uni-stuttgart.de

The conductor-like screening model or COSMO model has wide applications for computing the electrostatic interaction between a solvent and a particular solute molecule. However, mathematically this involves solving a Laplace boundary value problem on a domain with a number of intersecting balls. We propose to solve this problem by implementing a new numerical approach based on the Schwarz domain decomposition method.

Sparse Recovery and Group Representations

Timm Gilles

It is known that sparse recovery by measurements from random convolutions provide good recovery bounds. We aim to generalize this to measurements that arise as a random orbit of a group representation.

Convergence rate of general regularization in functional data

Naveen Gupta

The functional linear regression (FLR) model is one of the main methods for analysing functional data. The model gained popularity due to its simplicity in dealing with high-dimensional functional data. For example, it is widely used in medicine, chemometrics, and economics. Mathematically, the FLR model is stated as

$$Y = \int_S X(t)\beta^*(t)dt + \epsilon$$

where $Y(\omega)$ is a real-valued random variable, $(X(\omega,t); t \in S)$ is a continuous time process, β^* is an unknown slope function and ϵ is a zero mean random noise, independent of X, with finite variance σ^2. We assume that X and β^* are in $L^2(S)$, and S is a compact subset of $\mathbb{R}^d$. In the context of the slope function, it is evident that

$$\beta^* := \arg \min_{\beta \in L^2(S)} \mathbb{E}\,[Y - \langle X, \beta\rangle]^2.$$

The goal is to construct an estimator $\hat{\beta}$ to approximate the slope function β^* using observed empirical data $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$, where X_i's are i.i.d. copies of random function X and Y_i's are scalar responses. One of the approach is to solve the above optimization problem in the reproducing kernel Hilbert space (RKHS) framework. In this framework, T. Cai and M. Yuan [2] used the Tikhonov regularization to get the minimax rates assuming that the slope function lies in the RKHS. The analysis in the RKHS framework involves two operators T and C, where T is the integral operator with reproducing kernel and C is the covariance operator, associated with the RKHS and the data, respectively. In [1], they observed that if the operators commute then the convergence analysis of FLR with Tikhonov regularization can be proved with weaker assumption on the smoothness of the slope function. It is well-known that Tikhonov regularization has a saturation over the smoothness of the slope function. As a consequence, a faster rate of convergence cannot be achieved even though the slope function possess higher order smoothness. In our analysis, we use general regularization to establish the convergence rates which includes the spectral cut-off and the Landweber iteration (gradient descent) methods. The use of general regularization techniques allows us to deal with the saturation problem of the Tikhonov regularization and gives us faster convergence rates by assuming over-smoothness on the slope function. Our rates are optimal under the given smoothness criteria, which we ensure by obtaining the lower rates for both commutative and non-commutative cases.

References

[1] Balasubramanian, K., Müller, H.-G., and Sriperumbudur, B. K. Unified RKHS methodology and analysis for functional linear and single-index models. *arXiv preprint arXiv:2206.03975* (2022).

[2] Cai, T. T., and Yuan, M. Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc. 107*, 499 (2012), 1201–1216.

Landau and Boltzmann Collision Operators in the Irreducible Burnett Ansatz

Andrea Hanke

Numerically solving the Boltzmann equation is computationally expensive in part due to the number of variables the distribution function depends upon. Another contributor to the complexity of the Boltzmann Equation is the the quadratic collision operator describing changes in the distribution function due to colliding particle pairs. Solving it as efficiently as possible has been a topic of recent research, e.g. [1, 2, 3]. Recently, we exploited results from representation theory to find a very efficient algorithm both in terms of memory and computational time for the evaluation of Boltzmann's quadratic collision operator [4]. With this novel approach we are also able to provide a meaningful interpretation of its structure, leading us to the separation between purely mathematical operations, represented in the coupling tensor, and physically relevant coefficients, collected in the impact tensor. In this talk we apply the Irreducible Burnett Ansatz not only to Boltzmann's collision operator, but also to the Landau operator, which describes grazing collisions, which occur e.g. in plasma. We will take a closer look at the structure of different impact tensors and explore how they impact the evolution of moments.

References

[1] Cai & Torrilhon, On the Holway-Weiss debate: Convergence of the Grad-moment-expansion in kinetic gas theory (2019)

[2] Wang & Cai, Approximation of the Boltzmann collision operator based on Hermite spectral method (2019)

[3] Cai & Fan & Wang, Burnett spectral method for the spatially homogeneous Boltzmann equation (2020)

[4] Hanke & Torrilhon, Representation Theory Based Algorithm to Compute Boltzmann's Bilinear Collision Operator in the Irreducible Spectral Burnett Ansatz Efficiently (2023)

On the adversarial training of deep learning models

Vinh Hoang[1], Sebastian Krumscheid[2], Holger Rauhut[3], and Raúl Tempone[1]

[1] *RWTH Aachen University*
hoang@uq.rwth-aachen.de

[2] *Karlsruhe Institute of Technology*

[3] *Ludwig-Maximilians-Universität*

This study addresses the critical issue of adversarial vulnerability in deep learning classifiers, a pressing concern in the field of machine learning. Adversarial attacks, which involve crafting additional noise into the input data to manipulate the output label, pose significant challenges to the integrity and reliability of these models. The focus of our study is to explore and evaluate various training methodologies aimed at enhancing the adversarial robustness of deep learning systems. By reviewing different approaches to train these models, we aim to develop provably effective training methods to mitigate the risks posed by adversarial attacks.

Uncertainty quantification for learned ISTA

Frederik Hoppe

Machine learning plays a major role in our everyday life and its impact is vastly growing. However, in some safety-critical areas such as medical imaging the use is still limiting due to its lack of explainability. In medical magnetic resonance imaging, for example, sparse images are often generated from noisy undersampled data via the solution of a high-dimensional optimization problem. In order to decide if a reconstructed image is good or not, uncertainty quantification must be performed for such high-dimensional problems. One of the very few methods, that is able to rigorously perform this task, is the so-called debiased LASSO. Due to its asymptotic normality, it allows for the construction of confidence intervals for every pixel of a sparse MR image. The poster explores how this estimator can be applied to a learned version of the LASSO, LISTA, and provides confidence intervals for this data-driven solution.

This is joint work with Felix Krahmer, Hannah Laus, Claudio Mayrink Verdun, Marion I. Menzel, and Holger Rauhut.

Singularity formation in dissipative harmonic flows

Chunxi Jiao

We aim to study the formation of singularities in models of ferromagnetism, liquid crystals and skyrmions. An important model for magnetisation dynamics is the Landau-Lifshitz-Gilbert (LLG) equation, for which partial regularity and singularities (for example, bubbling in dimension 2) of solution have already been studied. In practice, noises are non-negligible for small magnetic devices, and singularity formation is often accompanied with topological changes. We prove the existence of Struwe solution (smooth except at most finitely many singular points) for a stochastic LLG equation, and propose large deviation and control problems as first steps in the investigation of singularity creation and transport.

Fisher-Rao gradient flows of linear programs and state-action natural policy gradients

Semih Çayci[1], Guido Montúfar[2,3], and Johannes Müller[1,3]

[1] *Department of Mathematics, RWTH Aachen University*

[2] *Departments of Mathematics and Statistics, University of California, Los Angeles*

[3] *Max Planck Institute for Mathematics in the Sciences, Leipzig*

cayci@mathc.rwth-aachen.de, montufar@math.ucla.edu, mueller@mathc.rwth-aachen.de

The convergence of Kakade's natural policy gradient method has been studied extensively in the last years showing linear convergence with and without regularization. We study a competing natural gradient method based on the Fisher information matrix of the state-action distributions which has received little attention from the theoretical side. Here, the state-action distributions follow the Fisher-Rao gradient flow inside the state-action polytope with respect to a linear potential. Therefore, we study Fisher-Rao gradient flows for general polytopes and show linear convergence with explicit control of the exponent and the coefficient. Equivalently, this yields estimates on the regularization error of entropy-regularized linear programs improving known results. Finally, this provides linear convergence rates for state-action natural policy gradients.

Keywords: Natural policy gradient, Fisher-Rao geometry, Entropy regularized linear program

Finite Element Methods for the Harmonic Map Heat Flow From the Unit Disk to the Unit Sphere

Nam Anh Nguyen

We numerically confirm the optimal convergence orders of some numerical methods for the harmonic map heat flow present in the literature, that are based on finite element discretization in space combined with a time stepping scheme, in the case of smooth solutions. We also present a possible adaptive algorithm using a-posteriori error indicators in space and in time that can capture the blow-up of the harmonic map heat flow for the special case of spherically symmetric solutions.

Group actions and t-designs in sparse and low rank matrix recovery

Leonie Scheeren

The overarching question considered is “Which (unions of) group orbits lend themselves to obtaining suitable measurement schemes for low rank matrix recovery, and how do we construct and sample from them?"

Kinetic theory meets algebraic systems theory

Melanie Harms and Chiara Segala

We explore the structural properties of high-dimensional nonlinear systems involving interacting agents of varying sizes, along with their associated mean-field limit. We linearize the dynamics of each agent around an arbitrary equilibrium point, that we classify based on the system's spectrum analysis. We achieve stability implementing a sparse control closed-loop action, and we discuss our approach in a mean-field limit framework.

Hybrid hierarchical approach for efficient solution of multi-parametric PDEs

Tim Werthmann

Multi-parametric PDEs play a crucial role in scientific models with uncertain parameters. We focus on a well-known example PDE known as the "cookie problem", modeled by a diffusion equation with a parameter-dependent diffusion. The challenge arises when dealing with a large number of parameters, rendering classical methods impractical due to exponential scaling.

Our research goal is to develop a hybrid format combining the strengths of hierarchical matrices and the hierarchical Tucker format to efficiently solve parameter-dependent linear systems with high contrast in the diffusion. We aim to investigate and exploit additional structures inherent in such problems and address questions related to construction, update, computation of arithmetic operations, and the stability and accuracy of the proposed hybrid format.

To demonstrate the potential of our approach, we present initial numerical results on a cookie problem with 4 parameters.

Robust sparse low-rank approximation of multi-parametric partial differential equations

Markus Bachmayr[1] and Huqing Yang[1]

[1]Institut für Geometrie und Praktische Mathematik, RWTH Aachen

Parametric and stochastic partial differential equations with infinitely many parameters play a role in many applications. In uncertainty quantification, for instance, expanding input random fields for uncertain coefficients into random series expansions leads to sequences of countably many scalar parameters. With increasing index, these scalar parameters have successively decreasing influence on the corresponding solutions. This anisotropy in the problem can be used to construct numerical schemes with computational costs that are independent of the number of considered parameters. We consider modified low-rank tensor approximations of parameter dependent solution that combine representations in hierarchical tensor format with sparse expansions, aiming at the construction and analysis of approximations that improve the pre-asymptotic behavior in problematic cases, such as diffusion coefficients with high contrast or small correlation length, while at the same time preserving optimal asymptotic convergence rates.

References

[1] M. Bachmayr, A. Cohen, and W. Dahmen. Parametric PDEs: Sparse or low-rank approximations? (2018)

[2] M. Bachmayr and W. Dahmen. Adaptive near-optimal rank tensor approximation for high dimensional operator equations. (2015)

[3] M. Bachmayr and R. Schneider. Iterative methods based on soft thresholding of hierarchical tensors. (2017)

Theta Norms and Low Rank Tensor Recovery

Yuhuai Zhou

joint work with Arinze Folarin and Felix Röhrich
supervised by Prof. Dr. Ghislain Fourier and Prof. Dr. Holger Rauhut

The recovery problem aims to reconstruct signals via incomplete linear information. Once it comes to tensor structure data, the problem becomes more challenging as most constructions in tensor are NP-hard, for instance the nuclear norm. Therefore, we dedicate to find a tractable algorithm for low rank tensor recovering. Our method is based on the theta norm, which is a relaxation of a certain norm by replacing non-negativity with sum of squares. In particular, by interpreting unit balls as intersections of half spaces and substituting non-negativity with sum of squares, we obtain a sequence of relaxed norms, which converges to the original norm in general. Moreover, we apply the tool of Gaussian width to provide estimation of the sufficient number of measurements for recovery. Finally, we are hoping this method can achieve a better performance than matricization.