Abstracts for CNA Ki-Net Workshop
Dynamics and Geometry from High Dimensional Data-driven

Day 1:

Nathan Kutz (University of Washington)

Title:
Data-driven discovery of dynamical systems in the engineering, physical and biological sciences

Abstract:
We demonstrate that we can use emerging, large-scale time-series data from modern sensors to directly construct, in an adaptive manner, governing equations (differential and partial differential equations), even nonlinear dynamics, that best model the system measured using sparsity-promoting techniques. Our sparse identification of nonlinear dynamics (SINDy) algorithm can be integrated with principles of model selection which allows for the consideration of a combinatorially large number of candidate models governing a dynamical system. The innovation circumvents a disadvantage of standard model selection which typically limits the number candidate models considered due to the intractability of computing information criteria. Using SINDy, the sub-selection of candidate models near the Pareto frontier allows for a tractable computation of AIC (Akaike information criteria) or BIC (Bayes information criteria) scores for the remaining candidate models. The information criteria hierarchically ranks the most informative models, enabling the automatic and principled selection of the model with the strongest support in relation to the time series data. Specifically, we show that AIC scores place each candidate model in the strong support, weak support or no support category, thus allowing us to select the best model for a given set of time series data. The methodology generalizes to the discovery of nonlinear PDEs as well.

Christof Schütte (Freie Universität Berlin)

Title:
Finding Reaction Coordinates in Molecular Dynamics

Abstract:
The talk will address the question of whether, for a given molecular system, there is a low-dimensional reaction coordinate manifold on which the slow dynamics is essentially happening and how to find the effective dynamics on this manifold. In particular, it will address how to reliably uncover the manifold and the effective dynamics from simulation data and why this question is deeply related to the geometry present in the simulation data.

Andrew Stuart (Caltech)

Title:
Uncertainty Quantification in the Classification of High Dimensional Data

Abstract:
Classification of high dimensional data finds wide-ranging applications. In many of these applications equipping the resulting classification with a measure of uncertainty may be as important as the classification itself. In this talk we introduce, develop algorithms for, and investigate the properties of, a variety of Bayesian models for the task of binary classification; via the posterior distribution on the classification labels, these methods automatically give measures of uncertainty. The methods are all based around the graph formulation of semi-supervised learning. We provide a unified framework which brings together a variety of methods which have been introduced in different communities within the mathematical sciences. We study probit classification, generalize the level-set method for Bayesian inverse problems to the classification setting, and generalize the Ginzburg-Landau optimization-based classifier to a Bayesian setting. We introduce efficient numerical methods, suited to large data-sets, for both MCMC-based sampling as well as gradient-based MAP estimation. Through numerical experiments we study both classification accuracy and uncertainty quantification for our models; these experiments showcase a suite of datasets commonly used to evaluate graph-based semi-supervised learning algorithms. Joint work with Andrea L. Bertozzi and Xiyang Luo (UCLA), and Konstantinos C. Zygalakis (Edinburgh).

Daniel Sanz-Alonso (Brown University)

Title:
The role of dimension, order, and regularity in the learning of PDE inputs

Abstract:
The aim of this talk is to explore some theoretical, computational, and methodological aspects of the Bayesian learning of PDE models by combining ideas from PDE theory, statistics, machine learning, and beyond. On the theoretical side I will show ---relying on recently developed regularity theory--- well-posedness of the Bayesian learning of fractional order elliptic PDEs. On the computational side I will argue that the key challenge facing sampling algorithms is large distance between prior and posterior, rather than large dimension of the unknown or the data. Finally, and time permitting, I will present a new methodology to construct priors based on random fields of spatially varying regularity.
This is joint work with Nicolás García Trillos.

Mauro Maggioni (Johns Hopkins University**)**

Title:
Geometric Methods for the Approximation of High-dimensional Dynamical Systems

Abstract:
We discuss a geometry-based statistical learning framework for performing model reduction and modeling of stochastic high-dimensional dynamical systems. We consider two complementary settings. In the first one, we are given long trajectories of a system, e.g. from molecular dynamics, and we discuss new techniques for estimating, in a robust fashion, an effective number of degrees of freedom of the system, which may vary in the state space of then system, and a local scale where the dynamics is well-approximated by a reduced dynamics with a small number of degrees of freedom. We then use these ideas to produce an approximation to the generator of the system and obtain, via eigenfunctions of an empirical Fokker-Planck question, reaction coordinates for the system that capture the large time behavior of the dynamics. We present various examples from molecular dynamics illustrating these ideas. In the second setting we only have access to a (large number of expensive) simulators that can return short simulations of high-dimensional stochastic system, and introduce a novel statistical learning framework for learning automatically a family of local approximations to the system, that can be (automatically) pieced together to form a fast global reduced model for the system, called ATLAS. ATLAS is guaranteed to be accurate (in the sense of producing stochastic paths whose distribution is close to that of paths generated by the original system) not only at small time scales, but also at large time scales, under suitable assumptions on the dynamics. We discuss applications to homogenization of rough diffusions in low and high dimensions, as well as relatively simple systems with separations of time scales, stochastic or chaotic, that are well-approximated by stochastic differential equations.

Frédéric Chazal (INRIA Saclay)

Title:
Subsampling Methods for Persistent Homology

Abstract:
Computational topology has recently seen an important development toward data analysis, giving birth to Topological Data Analysis. Persistent homology appears as a fundamental tool in this field. It is usually computed from filtrations built on top of data sets sampled from some unknown (metric) space, providing "topological signatures", the so-called persistence diagram, revealing the structure of the underlying space. When the size of the sample is large, direct computation of persistent homology often suffers two issues. First, it becomes prohibitive due to the combinatorial size of the considered filtrations and, second, it appears to be very sensitive to noise and outliers. The goal of this talk is to show that it is possible to overcome these issues computing persistent diagrams, and some related quantities, from several subsamples and combining them in order to efficiently infer robust and relevant topological information. This is a joint work with B. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman.

Jianfeng Lu (Duke University)

Title:
Path-integral molecular dynamics with surface hopping: High dimensional sampling with diffusion and jumps

Abstract:
In this talk, after reviewing the ideas of path-integral molecular dynamics, we will describe a novel ring polymer representation for multi-level quantum system for thermal average calculations. The proposed representation keeps the discreteness of the electronic states: besides position and momentum, each bead in the ring polymer is also characterized by a surface index indicating the electronic energy surface. A path integral molecular

dynamics with surface hopping (PIMDSH) dynamics is developed to sample the equilibrium distribution of ring polymer configurational space.

Day 2:

Rachel Ward (University of Texas, Austin)

Title:
Extracting governing equations in chaotic systems from highly corrupted data

Abstract:
Learning the governing equations for time-varying measurement data is of great interest across different scientific fields. When such data is moreover highly corrupted, for example, due to the recording mechanism failing over unknown intervals of time, recovering the governing equations becomes quite challenging. In this work, we show that if the data exhibits chaotic behavior, it is possible to recover the underlying governing nonlinear differential equations even if a large percentage of the data is corrupted by outliers, by solving an l1 minimization problem which assumes a polynomial representation of the system and exploits the joint sparsity in the variable representing the corrupted data. Theoretical reconstruction guarantees are obtained by combining recent results on central limit theorems for time-1 maps of chaotic flows with results from compressive sensing theory. This is joint work with Giang Tran, University of Texas, Austin.

Gilad Lerman (UMN)

Title:
A Well-Tempered Landscape for Non-convex Robust Subspace Recovery

Abstract:
We present a mathematical analysis of a gradient descent method for Robust Subspace Recovery. The optimization is cast as a minimization over the Grassmannian manifold, and gradient steps are taken along geodesics. We show that under a generic condition, the energy landscape is nice enough for the non-convex gradient method to exactly recover an underlying subspace. The condition is shown to hold with high probability for a certain model of data. This work is joint with Tyler Maunu and Teng Zhang.

Massimo Fornasier (Technical University of Munich**)**

Title:
Learning and Sparse Control of Multiagent Systems

Abstract:
In the past decade there has been a large scope of studies on mathematical models of social dynamics. Self-organization, i.e., the autonomous formation of patterns, has been so far the main driving concept. Usually first or second order models are considered with given predetermined nonlocal interaction potentials, tuned to reproduce, at least qualitatively, certain global patterns (such as flocks of birds, milling school of fish or line formations in pedestrian flows, etc.). However, often in practice we do not dispose of a precise knowledge of the governing dynamics. In the first part of this talk we present a variational and optimal transport framework leading to an algorithmic solution to the problem of learning the interaction potentials from the observation of the dynamics of a multiagent system. Moreover, it is common experience that self-organization of a society does not always spontaneously occur. In the second part of the talk we address the question of whether it is possible to externally and parsimoniously influence the dynamics, to promote the formation of certain desired patterns. In particular, we address the issue of finding the sparsest control strategy for finite agent models in order to lead the dynamics optimally towards a given outcome. We eventually mention the rigorous limit process connecting finite dimensional sparse optimal control problems with ODE constraints to an infinite dimensional sparse mean-field optimal control problem with a constraint given by a PDE of Vlasov-type, governing the dynamics of the probability distribution of the agent population.

Giang Tran (University of Texas, Austin)

Title:
Learning governing equations from multiple samples using sparsity

Abstract:
TBD

Larry Wasserman (Carnegie Mellon University)

Title:
Statistical Estimation of Manifolds and Ridges

Abstract:
I will first discuss approaches to estimating manifolds based on noisy samples. In general, the best possible rates of convergence are very slow. As an alternative, I'll consider estimating the ridges of the density function. These ridges serve as approximations to the underlying manifold but they can be estimated at a much faster rate. I'll also discuss limiting distribution theory for ridge estimates. This is joint work with Chris Genovese, Marco Perone-Pacifico and Isabella Verdinelli.

Aaditya Ramdas (Berkeley)

Title:
Universality of Mallows' and degeneracy of Kendall's kernels for rankings

Abstract:
Kernel methods provide an attractive framework for aggregating and learning from ranking data, and so understanding the fundamental properties of kernels over permutations (the symmetric group) is a question of broad interest. We provide a detailed analysis of the Fourier spectra of the standard Kendall and Mallows kernels, and a new class of polynomial-type kernels. We prove that the Kendall kernel has exactly two irreducible representations at which the Fourier transform is non-zero, and moreover, the associated matrices are rank one. This implies that the Kendall kernel is nearly degenerate, with limited expressive and discriminative power. In sharp contrast, we prove that the Fourier transform of the Mallows kernel is a strictly positive definite matrix at all irreducible representations. This property guarantees that the Mallows kernel is both characteristic and universal. We introduce a family of normalized polynomial kernels of degree p that interpolates between the Kendall (degree one) and Mallows (infinite degree) kernels, and show that for d-dimensional permutations, the pth-degree kernel is characteristic when p is larger than d − 1, unlike the Euclidean case in which no finite-degree polynomial kernel is characteristic. We also derive an explicit finite-dimensional feature map for the Mallows kernel, which allow computations in primal form. We demonstrate applications to testing, regression and classification using a Eurobarometer survey dataset where we have access to respondents' features (age/gender/…) as well as their rankings over sources of information (internet/tv/radio/newspaper/...).

Sebastien Motsch (Arizona State University)

Title:
Tumor growth: from agent-based model to free-boundary problem

Abstract:
In this talk, we investigate the large time behavior of an agent based model modeling tumor growth. This microscopic model combines short-range repulsion and cell division. We derive the associated macroscopic dynamics leading to a porous media type equation. In order to capture the long-time behavior of the microscopic model, we have to modify the porous media in order to include a density threshold for the repulsion. The main difficulty is then to investigate the limit as the repulsion between cells becomes singular (modeling non-overlapping constraint). We show formally that such asymptotic limit leads to a free-boundary problem (Hele-Shaw type). Numerical results confirm the relevance of
such limit.

Rujie Yin (Duke)

Title:
Convolution framelets: coupling local and nonlocal representations

Abstract:
We propose an image representation scheme combining the local and nonlocal characterization of patches in an image. Our representation scheme is shown to be equivalent to a tight frame constructed from convolving local bases (e.g. wavelet frames, discrete cosine transforms, etc.) with nonlocal bases (e.g. spectral basis from nonlinear embedding of patches), and we call the resulting frame elements convolution framelets. Insight gained from analyzing the proposed representation leads to a novel interpretation of a recent high-performance patch-based image inpainting algorithm, Low Dimension Manifold Model (LDMM). In particular, we show that

LDMM is weighted L2-regularization on the coefficients obtained by decomposing images into linear combinations of convolution framelets; we extend the original LDMM to a reweighted version that yields further improved inpainting results. Our framework can be potentially generalized to interpret more complex image processing algorithms.

Day 3:

Facundo Mémoli (Ohio State University)

Title:
Persistent Homology of Asymmetric Networks

Abstract:
We'll discuss recent work on trying to adapt PH methods to datasets that exhibit asymmetry. Natural candidates are the Rips and Cech filtrations. Whereas the Rips filtration can be generalized directly, generalizing the Cech filtration gives rise to two different versions: the sink and the source filtrations. It turns out that whereas the Rips filtration imposes a max-symmetrization on the data, the Cech filtrations does not, thus making it more suitable for the analysis of intrinsically asymmetric data. By generalizing a theorem of Dowker we can prove that the persistent homologies of these two Cech filtrations are isomorphic. Stability of these constructions takes place under a metric between networks that generalizes the Gromov-Hausdorff distance. I'll describe some results we have that characterize the persistence diagrams of some likely "motifs" in real (e.g. biological) networks: cycle-networks, which are directed analogues of the standard (discrete) circle. Finally, as an application, we'll show computational results about classifying simulated networks arising from ensembles of hippocampal cells.

Jerome Darbon (Brown University)

Title:
On convex finite-dimensional variational methods in imaging sciences, and Hamilton-Jacobi equations

Abstract:
We consider standard finite-dimensional variational models used in signal/image processing that consist in minimizing an energy involving a data fidelity term and a regularization term. We propose
new remarks from a theoretical perspective which give a precise description on how the solutions of the optimization problem depend on the amount of smoothing effects and the data itself. The dependence of the minimal values of the energy is shown to be ruled by Hamilton-Jacobi equations, while the minimizers u(x,t) for the observed images x and smoothing parameters t are given by

u(x,t)=x -\nabla H(\nabla E(x,t))

where E(x,t) is the minimal value of the energy and H is a Hamiltonian related to the data fidelity term. Various vanishing smoothing parameter results are derived illustrating the role played by the prior in such limits. Finally, we briefly present an efficient numerical numerical method for solving certain
Hamilton-Jacobi equations in high dimension and some applications in optimal control.

Eric Vanden-Eijnden (Courant Institute, NYU)

Title:
Markov State Models for data assimilation and interpretation

Abstract:
Markov State Models (MSMs) have emerged has a popular way to interpret and analyze time-series data generated e.g. in the context of molecular dynamics simulations. The basic idea of these models is to represent the dynamics of the system as memoryless jumps between predefined sets in the systems state space, i.e. as a Markov jump process (MJP). Performing this mapping typically involves two nontrivial questions: first how to define these sets and second how to learn the rate matrix of the MJP once the sets have been identified? Both these questions will be discussed in the talk, with emphasis put on the problem of model specification error. Some of the outputs of MSMs, in particular in terms of free energy, reaction rate, etc. will also be discussed.

Antonin Chambolle (Ecole Polytechnique)

Title:
A convex relaxation for the elastica functional

Abstract:
In this talk we will present a convex representation for curvature-dependent line energies, based on a lifting of the curves in the so-called roto-translational space, and show how it can be practically minimized. This is a joint work with Thomas Pock (TU Graz).