Tentative Agenda for the

Symposium on Machine Learning and Dynamical Systems

February 11th to February 13th, 2019

Imperial College London

Monday morning

1. Mauro Maggioni (JHU)

    - Title: Geometric methods in statistical learning problems for (stochastic) dynamical systems
    - Abstract: We discuss methods that measure and exploit quantitative geometric properties of high-dimensional data to perform inference and statistical estimation tasks, avoiding the curse of dimensionality. Examples include stochastic systems in high-dimensions, which have a large number of fast degrees of freedom but a relatively small number of slower, ``effective" degrees of freedom, which are unknown. We introduce geometric-based techniques for learning such unknown degrees of freedom, and learn surrogate, lower-dimensional dynamical systems cheaply, but with theoretical guarantees on their performance in predicting large-time properties of the system. We will also discuss the problem of learning the influence function in interacting agent systems.

2. Tim Sauer
    - Title: Consistent manifold learning from data
    - Abstract: Characterizing attracting sets of dynamical systems is a key concern for data-driven algorithms.
      We take a geometric approach by assuming the data points lie on a Riemannian manifold, and reconstruct the Laplace-Beltrami operator. In particular, we show how build a graph Laplacian that converges, pointwise and spectrally, to the continuous operator in the large data limit. If this can be achieved, geometric and topological information about the manifold can be inferred from a single graph. Since real data is typically sampled irregularly, it is necessary to introduce a criterion called Continuous k-Nearest Neighbors (CkNN) for the graph construction that implies convergence for arbitrary sampling.

3. Michael Dellnitz (U. Paderborn), Andreas Bittracher (FU Berlin), Sebastian Peitz (Universität Paderborn)
    - Title: Data based computation of invariant objects for dynamical systems
    - Abstract:

4. Steven Brunton (UW)

    - Title: Discovering dynamics and enforcing known symmetries and constraints with machine learning
    - Abstract: Accurate and efficient reduced-order models are essential to understand, predict, estimate, and control high-dimensional nonlinear dynamical

systems. These models should ideally be generalizable, interpretable, and based on limited training data. This talk will explore the use of machine learning and sparse regression to uncover interpretable dynamical systems models from data. We will discuss how it is possible to enforce known constraints, such as energy conserving quadratic nonlinearities in fluid dynamics, to essentially "bake in" known physics. Next, we will demonstrate that higher-order nonlinearities can approximate the effect of truncated modes, resulting in more accurate models of lower order than Galerkin projection. Finally, we will discuss the use of intrinsic measurement coordinates to build nonlinear models, circumventing the well-known issue of continuous mode deformation associated with methods based on the proper orthogonal decomposition. This approach will be demonstrated on several relevant systems with low-dimensional dynamics.

5. Kevin Lin (U. Arizona)

- Title: Data-driven modeling of chaotic dynamics: a model reduction perspective
- Abstract: Nonlinear dynamic phenomena often require a large number of dynamical variables for their description, only a small fraction of which are of direct interest. Reduced models using only these relevant variables can be very useful in such situations, both for computational efficiency and insights into the underlying dynamics. Unfortunately, except in special cases, deriving reduced models from first principles can be quite challenging. This has motivated interest in both parametric and nonparametric data-driven modeling in the model reduction community. In this talk, I will review a discrete-time version of the Mori-Zwanzig (MZ) projection operator formalism from nonequilibrium statistical mechanics, which provides a simple and general framework for model reduction. I will then discuss data-driven modeling and model reduction for chaotic dynamical systems within the MZ framework, highlighting some of the theoretical and practical issues that arise.

6. AmirAli Ahmadi (Princeton)

- Title: Nonnegative polynomials, learning, and control
- Abstract: The problem of recognizing nonnegativity of a multivariate polynomial has a celebrated history, tracing back to Hilbert's 17th problem. In recent years, there has been much renewed interest in the topic because of a multitude of applications in applied and computational mathematics and the observation that one can optimize over an interesting subset of nonnegative polynomials using "sum of squares optimization".
  In this talk, we give a brief overview of the recent developments in this field and show how they can be applied to problems in learning, dynamics and control, as well as the intersection of the two. Examples include the problem of learning a Lyapunov function subject to shape constraints (e.g., convexity or monotonicity),

and that of learning a dynamical system subject to qualitative knowledge of the behavior of trajectories (e.g., stability, invariance, or collision avoidance).

Monday afternoon

Kernel Methods and Dynamical Systems

7. Eniko Szekely (EPFL)
   - Title: Data-driven kernel methods for dynamical systems with application to atmosphere ocean science
   - Abstract: Datasets generated by dynamical systems are often high-dimensional, but they only display a small number of patterns of interest. The underlying low-dimensional structure governing such systems is generally modeled as a manifold, and its intrinsic geometry is well described by local measures that vary smoothly on the manifold, such as kernels, rather than by global measures, such as covariances. In this talk, a kernel-based nonlinear dimension reduction method, namely nonlinear Laplacian spectral analysis (NLSA), is used to extract a reduced set of basis functions that describe the large-scale behavior of the dynamical system. These basis functions are the leading Laplace-Beltrami eigenfunctions of a discrete Laplacian operator. They can be further employed as predictors to quantify the regime predictability of a signal of interest using clustering and information-theoretic measures. In this talk, NLSA will be employed to extract physically meaningful spatiotemporal patterns from organized tropical convection covering a wide range of timescales, from interannual to annual, semiannual, intraseasonal and diurnal scales.
8. Jehan AlSwaihli
   - Title: Kernel Reconstruction for Delayed Neural Field Equations
   - Abstract
9. B. Haasdonk, G. Santin, and T. Brunnette
   - Title: Accelerating Implicit Integrators for Parametric ODE Systems by Greedy Kernel Approximation
   - Abstract: In this presentation, we want to demonstrate how kernel-based approximation methods can contribute to a paradigm of "Data-based Numerical Mathematics". Parametric MOR has intensively focussed on approximately solving parametric high-dimensional PDE and ODE systems during the last decades. We want to widen this view to more general parametric problems in numerical mathematics, that might benefit from the same concept: 1) gathering data from solving some specific problem instantiations, 2) processing this data and obtain a surrogate that can be used for 3) rapidly solving or approximating the original parametric problem. We will first review some basic tools in kernel

approximation for the reconstruction of high-dimensional functions, both in input and output. These methods allow to construct approximants to general target functions defined on arbitrary domains by means of scattered samples, i.e., without requiring any structure on the sampling locations. We will then focus on greedy algorithms, in particular the VKOGA [3], which constructs approximants based on a small subset of the data sites, thus being faster to evaluate, while still providing a good accuracy, which can even be proven to be quasi-optimal in some cases [2]. The proof actually makes elegant use of known results for Reduced Basis Methods. These theoretical and computational features make greedy kernel-based algorithms particularly attractive for the construction of surrogate models. Then we will exemplify an application in data-driven numerical mathematics, namely acceleration of implicit ODE integrators by forecasting. A set of state trajectories precomputed with a high-accuracy ODE solver is used to train a kernel model which predicts the next point in the trajectory as a function of the previous one. This model is cheap to evaluate, and it is used in the online phase to provide a good initialization point for the nonlinear solver of the implicit integrator. The accuracy of the surrogate model results in a significant reduction of the number of required steps of the solver, thus providing an overall speedup of the full simulation. Despite the acceleration, the method maintains the accuracy of the original model. Although the method can be potentially applied to a large variety of solvers and different ODEs, we will present in detail its use with the implicit Euler method (VKOGA-IE) in the solution of e.g., the Burgers equation, which is an important test case to demonstrate the method's features [1].

References [1] T. Brunnette, G. Santin, and B. Haasdonk. Greedy kernel methods for accelerating implicit inte- grators for parametric ODEs. In Proc. ENUMATH 2017, 2018. Submitted. [2] G. Santin and B. Haasdonk. Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. Dolomites Research Notes on Approximation, 10:68–78, 2017. [3] D. Wirtz and B. Haasdonk. A vectorial kernel orthogonal greedy algorithm. Dolomites Research Notes on Approximation, 6:83–100, 2013.

10. Alberto Muñoz García (Universidad Carlos III de Madrid)

- Title: Entropy Measures for Stochastic Processes with Applications in Machine Learning
- Abstract: We propose a definition of entropy for stochastic processes. We provide a reproducing kernel Hilbert space model to estimate entropy from a random sample of realizations of a stochastic process, namely functional data, and introduce approaches to estimate minimum entropy sets. Next we address the problem of combination of different sources of information in a Machine Learning context. Often, each source of information is given as a similarity, distance, or a kernel matrix. We propose a new class of methods which consists of producing, for anomaly detection purposes, a single Mercer kernel (that acts as a similarity measure) from a set of local entropy kernels and, at the same time, avoids the task of model selection.

This kernel is used to build an embedding of data in a variety that will allow the use of a (modified) one-class Support Vector Machine to detect outliers. In addition, we study several information combination schemes and their limiting behaviour when the data sample size increases within an Information Geometry context.

11. Boumediene Hamzi (Imperial College London)

- Title: Kernel Methods for Dynamical Systems
- Abstract: We introduce a data-based approach to estimating key quantities which arise in the study of nonlinear autonomous, control and random dynamical systems. Our approach hinges on the observation that much of the existing linear theory may be readily extended to nonlinear systems - with a reasonable expectation of success- once the nonlinear system has been mapped into a high or infinite dimensional Reproducing Kernel Hilbert Space. In particular, we develop computable, non-parametric estimators approximating controllability and observability energies and Lyapunov functions for nonlinear systems. We apply this approach to the problem of model reduction of nonlinear control systems. It is also shown that the controllability energy estimator provides a key means for approximating the invariant measure of an ergodic, stochastically forced nonlinear system. We also show how kernel methods can be used to detect critical transitions for some multi scale dynamical systems. Finally, we show how kernel methods can be used to approximate center manifolds for nonlinear ODEs.
  This is joint work with Jake Bouvrie (MIT, USA), Peter Giesl (University of Sussex, UK), Christian Kuehn (TUM, Munich/Germany), Sameh Mohamed (SUTD, Singapore), Martin Rasmussen (Imperial College London), Kevin Webster (Imperial College London), Bernard Hasasdonk, Gabriele Santin and Dominik Wittwar (University of Stuttgard).

  References: [1] https://arxiv.org/abs/1011.2952 [2] https://arxiv.org/abs/1108.2903 (http://aimsciences.org/article/doi/10.3934/jcd.2017001) [3] https://arxiv.org/abs/1204.0563 (https://epubs.siam.org/doi/abs/10.1137/14096815X) [4] https://arxiv.org/abs/1601.01568 [5]https://arxiv.org/abs/1804.09415 [6] https://arxiv.org/abs/1810.11329

Monday, Poster Session, 400pm to 530pm

- Arash Mehrjou
- Robert Polzin
- Kamthe, Sanke
- Karim Cherifi (Max Planck Institute of Complex Dynamical Systems)
- Francesco Carravetta (IASI-CNR, Rome)
- Lazaros Mitskopoulos
- Ioannis Tsamardinos
- Dino Sejdinovic and Anthony Caterini (Oxford)
- Friedrich Solowjow
- Pantelis R. Vlachas, Petros Koumoutsakos (ETH Zurich)
- Samuel Rudy (UW)
- Jaideep Pathak (UMD)
- Lynn Houthuys, Zahra Karevan and Johan Suykens (KU Leuven)
- Peter Coveney (UCL)

DSI's Distinguished Lecture, 530pm to 630pm, Edward Ott

Tuesday morning

Deep Learning

12. Rossella Arcucci and Yi-Ke Guo (Imperial College London)

- Title: Deep Data Assimilation: Integrating Deep Learning with Data Assimilation
- Abstract: Data assimilation is a methodology to incorporate observed data into a prediction model in order to improve numerical forecasting. Conventional methods for data assimilation include Kalman filters and variational approaches. In the past 20 years these methods has become a main component in the development and validation of mathematical models in meteorology, climatology, geophysics, geology and hydrology. Recently, data assimilation is also applied to numerical simulations of geophysical applications, medicine and biological science. Data assimilation methods have strongly increased in sophistication to better fit their application requirements and circumvent their implementation issues. Nevertheless, DA approaches are incapable of overcoming fully their unrealistic assumptions, particularly linearity, normality and zero error covariances. With the rapid developments in recent years, deep learning shows great capability in approximating nonlinear systems, and extracting high-dimensional features. Machine learning algorithms are capable of assisting or replacing the traditional methods in making forecasts, without the assumptions of the conventional methods. On the other side, the training data provided to deep learning technologies, include several numerical, approximation and round off errors which are trained in the deep learning forecasting model. This means that, in some safety and security-sensitive scenarios, deep learning are still not qualified to avoid unpredictable risks. Data assimilation can increase the reliability of the deep learning models reducing those errors by including information on physical meanings from observed data. The resulting cohesion of deep learning and data assimilation is then blended in the future generation of predictive models.

13. Peter Dueben (ECMWF)

- Title: Challenges and design choices for global weather and climate models based on machine learning

- Abstract: Can models that are based on deep learning and trained on atmospheric data compete with weather and climate models that are based on physical principles and the basic equations of motion? This question has been asked often recently due to the boom of deep learning techniques. The question is valid given the huge amount of data that is available, the computational efficiency of deep learning techniques and the limitations of today's weather and climate models in particular with respect to resolution and complexity. In this talk, the question will be discussed in the context of global weather forecasts. A toy-model for global weather predictions will be presented and used to identify challenges and fundamental design choices for a forecast system based on Neural Networks.

  References: [1] https://www.geosci-model-dev-discuss.net/gmd-2018-148/

14. Marco Gallieri (NNAISENSE)

- Title: NAIS-NET: Stable Deep Networks from Non-Autonomous Differential Equations
- Abstract: This talk introduces Non-Autonomous InputOutput Stable Network (NAIS-Net), a very deep architecture where each stacked processing block is derived from a time-invariant non-autonomous dynamical system. Non-autonomy is implemented by skip connections from the block input to each of the unrolled processing stages and allows stability to be enforced so that blocks can be unrolled adaptively to a pattern-dependent processing depth. NAIS-Net induces non-trivial, Lipschitz input-output maps, even for an infinite unroll length. We prove that the network is globally asymptotically stable so that for every initial condition there is exactly one input-dependent equilibrium assuming tanh units, and multiple stable equilibria for ReL units. An efficient implementation that enforces the stability under derived conditions for both fully-connected and convolutional layers is also presented. Experimental results show how NAIS-Net exhibits stability in practice, yielding a significant reduction in generalization gap compared to ResNets.

15. Grzegorz Swirszcz

- Title: Local minima in training of neural networks
- Abstract: There has been a lot of recent interest in trying to characterize the error surface of deep models. This stems from a long standing question. Given that deep networks are highly nonlinear systems optimized by local gradient methods, why do they not seem to be affected by bad local minima? It is widely believed that training of deep models using gradient methods works so well because the error surface either has no local minima, or if they exist they need to be close in value to the global minimum. It is known that such results hold under very strong

assumptions which are not satisfied by real models. In this paper we present examples showing that for such theorem to be true additional assumptions on the data, initialization schemes and/or the model classes have to be made. We look at the particular case of finite size datasets. We demonstrate that in this scenario one can construct counter-examples (datasets or initialization schemes) when the network does become susceptible to bad local minima over the weight space.

16. Peter Ashwin (Exeter)

- Title: Network attractors and the functional dynamics of RNN
- Abstract:

Tuesday afternoon

17. Heather Harrington (University of Oxford)

- Title: Analysing dynamics and networks using topological data analysis
- Abstract:

18. Hamza Ghadyali (Duke U.)

- Title: TDA for time-series analysis and dynamical systems
- Abstract:

Learning Theory for Dynamical Systems

19. Massimiliano Pontil

- Title: Incremental Learning-to-Learn with Statistical Guarantees (joint work with Giulia Denevi, Carlo Ciliberto and Dimitris Stamos)
- Abstract: In learning-to-learn the goal is to infer a learning algorithm that works well on a class of tasks sampled from an unknown meta distribution. In contrast to previous work on batch learning-to-learn, we consider a scenario where tasks are presented sequentially and the algorithm needs to adapt incrementally to improve its performance on future tasks. Key to this setting is for the algorithm to rapidly incorporate new observations into the model as they arrive, without keeping them in memory. We focus on the case where the underlying algorithm is ridge regression parameterized by a positive semidefinite matrix. We propose to learn this matrix by applying a stochastic strategy to minimize the empirical error incurred by ridge regression on future tasks sampled from the meta distribution. We study the statistical properties of the proposed algorithm and prove non-asymptotic bounds on its excess transfer risk, that is, the generalization performance on new tasks from the same meta distribution. We compare our online learning-to-learn approach with a state of the art batch method, both theoretically and empirically. https://arxiv.org/abs/1803.08089

20. Saverio Salzo

- Title: Bilevel Learning of the Group Lasso Structure
- Abstract: Regression with group-sparsity penalty plays a central role in high-dimensional prediction problems. However, most of existing methods require the group structure to be known a priori. In practice, this may be a too strong assumption, potentially hampering the effectiveness of the regularization method.

To circumvent this issue, we present a method to estimate the group structure by means of a continuous bilevel optimization problem where the data is split into training and validation sets. Our approach relies on an approximation scheme where the lower level problem is replaced by a suitable discrete dynamics which is smooth with respect to the hyperparameters of the group structure. We show the convergence of the approximate procedure to theexact problem and demonstrate the well behavior of the proposed method on synthetic and real experiments.

21. Erik Bollt (Clarkson)

- Title: How Entropic Regression Beats the Outliers Problem in Nonlinear System Identification
- Abstract: System identification (SID) is central in science and engineering applications whereby a general model form is assumed, but active terms and parameters must be inferred from observations. Virtually all methods for SI rely on optimizing some metric-based cost function that describes how a model fits observational data. The most common cost function employs a Euclidean metric and leads to a least squares estimate, while recently it becomes popular to also account for model sparsity such as in compressed sensing and Lasso methods. While the effectiveness of these methods has been demonstrated in various model systems, it remains unclear whether SID can be accomplished under more realistic scenarios where there may be large noise and outliers. We show that sparsity-focused methods such as compressive sensing, when used in the presence of noise, may result in "over sparse" solutions that are brittle to outliers. In fact, metric-based methods are prone to outliers because outliers by nature have an unproportionally large influence. To mitigate such issues of large noise and outliers encountered in practice, we develop an entropic regression approach for nonlinear SID, whereby true model structures are identified based on relevance in reducing information flow uncertainty, not necessarily sparsity. The use of information- theoretic measures as opposed to a metric-based cost function has the unique advantage, due to the asymptotic equipartition property of probability distributions, that outliers and other low-occurrence events are naturally and intrinsically de-emphasized.

References: [1] https://webspace.clarkson.edu/~ebollt/Papers/poCSEmain-180819-final.pdf

Wednesday morning

Dynamical Systems for Machine Learning

22. Qianxiao Li (Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore)

    - Title: A Dynamical Systems and Optimal Control Approach to Deep Learning
    - Abstract: Depsite its empirical success, deep learning lacks a concrete mathematical framework to study its algorithmic and theoretical properties. In this talk, we present a theoretical framework to study deep learning that draws on connections with dynamical systems and optimal control. In particular, supervised deep learning can be interpreted as an optimal control problem in a suitable "mean-field" sense [1]. This allows us to characterize precisely the optimality conditions of deep neural networks using maximum principles, as well as partial differential equations. Besides theoretical analysis, we will also discuss how the optimal control approach gives rise to novel training algorithms that are amenable to theoretical analysis and are effective in practice [2,3].

    References: [1] "A Mean-Field Optimal Control Formulation of Deep Learning". Weinan E, Jiequn Han, Qianxiao Li. arXiv preprint arXiv:1807.01083. 2018 [2] "Maximum Principle Based Algorithms for Deep Learning". Qianxiao Li, Long Chen, Cheng Tai, Weinan E. Journal of Machine Learning Research 18 165:1–165:29. 2018 [3] "An Optimal Control Approach to Deep Learning and Applications to Discrete-Weight Neural Networks". Qianxiao Li, Shuji Hao. Proceedings of the 35th International Conference on Machine Learning, ICML, 2018

23. Ioannis Panageas (SUTD)

    - Title: The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization
    - Abstract: Motivated by applications in Optimization, Game Theory, and the training of Generative Adversarial Networks, the convergence properties of first order methods in min-max problems have received extensive study. It has been recognized that they may cycle, and there is no good understanding of their limit points when they do not. When they converge, do they converge to local min-max solutions? We characterize the limit points of two basic first order methods, namely Gradient Descent/Ascent (GDA) and Optimistic Gradient Descent Ascent (OGDA). We show that both dynamics avoid unstable critical points for almost all initializations. Moreover, for small step sizes and under mild assumptions, the set of \{OGDA\}-stable critical points is a superset of \{GDA\}-stable critical points, which is a superset of local min-max solutions (strict in some cases). The

connecting thread is that the behavior of these dynamics can be studied from a dynamical systems perspective. Joint work with Costis Daskalakis.

24. Georgios Piliouras (SUTD)

- Title: Online Optimization in Zero-Sum Games: A Dynamical Systems Approach
- Abstract: Zero-Sum games are basic staples of game theory. Their study is also closely connected with machine learning challenges such as training Generative Adversarial Networks (GANs). Online gradient descent and multiplicative weights in such settings can be interpreted as Hamiltonian dynamics. In zero-sum games Poincaré recurrence and novel no-regret bounds can be established as a result both in discrete and continuous-time.

25. Panos Parpas

- Title: Predict Globally, Correct Locally: A dynamical systems view of distributed multilevel learning
- Abstract: Recently several authors proposed a dynamical systems view of modern neural network architectures. This point of view enables the use of a rigorous mathematical framework to study essential properties of neural networks such as stability, and convergence. In this talk, we adopt the same point of view, but our aim is different. In particular, we show how to take advantage of the dynamical system point of view to develop a new class of distributed optimization algorithms based on a predictor-corrector framework. In the prediction phase, the algorithm propagates the dynamics of a coarse neural network. The corrector phase uses the results from the prediction phase to correct the weights of the full network in a distributed fashion. We study the worst case convergence rate of the proposed algorithm and report numerical results from benchmark test problems.

26. Pascal Bianchi

- Title: Convergence of the ADAM algorithm from a Dynamical System Viewpoint
- Abstract: Adam is a popular variant of the stochastic gradient descent for finding a local minimizer of a function. The objective function is unknown but a random estimate of the current gradient vector is observed at each round of the algorithm. This paper investigates the dynamical behavior of Adam when the objective function is non-convex and differentiable. We introduce a continuous-time version of Adam, under the form of a non-autonomous ordinary differential equation (ODE). The existence and the uniqueness of the solution are established, as well as the convergence of the solution towards the stationary points of the

objective function. It is also proved that the continuous-time system is a relevant approximation of the Adam iterates, in the sense that the interpolated Adam process converges weakly to the solution to the ODE.

27. Pierre-Yves Masse

- Title: Convergence of Online Training Algorithms for Recurrent Systems
- Abstract: Neural networks may be represented as parameterised dynamical systems: the set of activities of the neurons is the state of the system, the weights of the network are the parameter of the system, and the forward pass is the transition operator of the system. This representation holds in particular for the recurrent architectures, which aim at modelling time dependent data (like for instance climatic time series). Mainstream algorithms used to train these systems proceed by gradient descent on the parameter, as Truncated Backpropagation Through Time (TBTT), or the Real-Time Recurrent Learning algorithm (RTRL), which runs online. However, though they have been known for some decades, no proof of convergence was available as of today, up to our best knowledge. Our work thus organises as follows. We provide a mathematical representation of a parameterised dynamical system, gradient-descent optimisation algorithm, in a nonlinear setting. Under natural hypotheses on the components of this representation, we prove the local convergence of the training procedure. We then establish that TBTT and RTRL fit into this framework, and therefore prove their convergence. In particular, we devise a criterium of optimality for families of time dependent losses, which extends the classical hypotheses of Robbins and Monro. The RTRL algorithm works online, at the expense of great memory costs needed to maintain a huge tensor. The UORO algorithm circumvents this issue by approximating this tensor with a random, unbiased, rank-one tensor. We prove that, with probability arbitrarily close to one provided the initial step size of the descent is small enough, UORO converges to the same optimum as RTRL.

Wednesday afternoon

28. Markus Abel

- Title: Machine Learning for the control of complex systems.
- Abstract: The control of complex systems may be as complex as the system considered. We have developed a software prototype to learn a control law in situations where no analytical control can be found. It is developed for evolutionary optimization and symbolic regression. In this contribution we present two systems where we apply our methodology: a network with possibly chaotic dynamics controlled to synchronuous behaviour of its parts, and an airfoil controlled to show high-lift.

29. Axel Hutt

- Title: Recurrence structure analysis: revealing underlying metastable attractor dynamics from time series
- Abstract: Complex systems may exhibit intermittent temporal dynamics, such as transitions between slow and fast dynamics. Temporal sequences of such metastable attractors occur in biology, e.g. in bird songs, in neuroscience, e.g. as event-related components, or in the atmosphere as chaotic attractors. In order to understand and derive models of complex systems, it is insightful to compute the recurrence structure of observed time series. To this end, recurrence structure analysis (RSA) provides a set of tools to identify recurrent states in time series reflecting metastable attractors in the underlying complex system. RSA utilises recurrence plots originally invented by Hernri Poincare about 100 years ago and derives symbolic sequences from recurrence plots and corresponding transition probability models. Such models allow to compute the optimal recurrence box width where the recurrence is detected. To our best knowledge, this optimality criterion is one of the first to determine an optimal recurrence box width. Applications to well-known models, such as the Lorentz-attractor, and brain signals observed during cognitive experiments illustrate underlying metastable attractors. Moreover, application to intracranial animal brain signals observed under anaesthesia reveals a changing degree of recurrence complexity in the brain with the anaesthetic depth.

References: Hutt and P. beim Graben, *Sequences by metastable attractors: interweaving dynamical systems and experimental data*, Frontiers in Applied Mathematics and Statistics 3:11 (2017)

P. beim Graben, K. K. Sellers, F. Fröhlich and A. Hutt, *Optimal estimation of recurrence structures from time series*, Europhysics Letters 114(3): 38003 (2016)

P. beim Graben and A. Hutt, *Detecting event-related recurrences by symbolic analysis: Applications to human language processing,* Philosphical Transactions of the Royal Society A 373:20140089 (2015)
P. beim Graben and A. Hutt, *Detecting metastable states of dynamical systems by recurrence-based symbolic dynamics*,Physical Review Letters 110, 154101 (2013)

30. Robert MacKay (University of Warwick)

- Title: A Gaussian process to detect underdamped modes of oscillation
- Abstract: In many domains of data science, it is desired to detect modes of oscillation of a system, including estimating their frequency, damping rate, mode shape and amplitude. Here a Gaussian process solution is presented.

31. Vincent Fortuin (ETH)

- Title: Deep Self-Organization: Interpretable Discrete Representation Learning on Time Series
- Abstract: Human professionals are often required to make decisions based on complex multivariate time series measurements in an online setting, e.g. in health care. Since human cognition is not optimized to work well in high-dimensional spaces, these decisions benefit from interpretable low-dimensional representations. However, many representation learning algorithms for time series data are difficult to interpret. This is due to non-intuitive mappings from data features to salient properties of the representation and non-smoothness over time. To address this problem, we propose to couple a variational autoencoder to a discrete latent space and introduce a topological structure through the use of self-organizing maps. This allows us to learn discrete representations of time series, which give rise to smooth and interpretable embeddings with superior clustering performance. Furthermore, to allow for a probabilistic interpretation of our method, we integrate a Markov model in the latent space. This model uncovers the temporal transition structure, improves clustering performance even further and provides additional explanatory insights as well as a natural representation of uncertainty. We evaluate our model on static (Fashion-)MNIST data, a time series of linearly interpolated (Fashion-)MNIST images, a chaotic Lorenz attractor system with two macro states, as well as on a challenging real world medical time series application. In the latter experiment, our representation uncovers meaningful structure in the acute physiological state of a patient.
  References: [1] https://arxiv.org/abs/1806.02199

32. L. Magri, N.A.K. Doan, O. Schmidt, W. Polifke, P. Schmid

- Title: Data-driven prediction of rare and extreme events in turbulent reacting flows
- Abstract: Energy supply is one of the main challenges facing our society with an ever-increasing demand. Around 80% of the world's power still comes from burning fossil fuels and the use of fossil fuels is likely to continue for the foreseeable future. The novel MILD combustion technology enables environmentally-friendly combustion: By utilizing exhaust gas recirculation, MILD technology increases combustion efficiency by up to 30%, while reducing NOx emissions by two orders of magnitude and sound emissions by 10dB. Therefore, MILD combustion is attracting large interest in gas turbine applications for lean and clean power generation, but fundamental questions are still open. The intricate multi-physical, multi-scale interactions between turbulence and chemistry reactions is yet to be fully understood. In fact, direct numerical simulations of the turbulent reacting fluid dynamics of MILD combustion have shown the occurrence of sudden and violent phenomena, such as ignition kernels, which (seemingly) randomly appear in the mixture, and flame propagating in unpredictable directions. It is paramount to understand the physical precursors of auto-ignition kernels and the directions of flame propagation because, by exploiting them, engineers can maximise homogeneous reactions and temperature fields, which further improve efficiency, emissions and combustion stability. Whereas the calculation of the statistics of turbulent reacting flows may be accurate, no matter how accurate the simulation code is, the time and space prediction of rare and extreme events is very difficult to achieve because of the chaotic nature of turbulence. We propose to time- and space-accurately predict rare and extreme events in turbulent reacting flows by leveraging on data-driven algorithms in machine learning and dynamical systems' theory. The applicability of these methods hinges on the availability of sufficient data that captures enough rare events. First, we perform direct numerical simulations of two turbulent reacting cases as relevant to MILD combustion ("flame in a box" cases), which provide the database to apply data driven methods. Secondly, we post-process the data by using community clustering combined with phase- space embedding to separate the slow dynamics from the fast-intermittent manifold. Spectral proper orthogonal decomposition is used to efficiently extract a low-rank approximation of the slow dynamics from the data. This enables the development of a reduced order model of the turbulent reacting flow, whose parameters are trained by an ensemble Kalman filter. The algorithms are trained by several direct numerical simulations to capture a sufficient number of rare events. Finally, the physical precursors and flow structures that precede rare events are identified and classified. Data-driven techniques open up new possibilities for the prediction of rare and extreme events in turbulent reacting flows.

33. Juan-Pablo ORTEGA

- Title: The universality problem in dynamic machine learning.

- Abstract: The universal approximation properties with respect to $L ^\infty$ and $L ^p$-type criteria of three important families of reservoir computers with stochastic discrete-time semi-infinite inputs are shown. First, it is proved that linear reservoir systems with either polynomial or neural network readout maps are universal. More importantly, it is proved that the same property holds for two families with linear readouts, namely, state-affine systems and echo state networks, which are the most widely used reservoir systems in applications. The linearity in the readouts is a key feature in supervised machine learning applications. It guarantees that these systems can be used in high-dimensional situations and in the presence of large datasets. These results are illustrated with applications to the forecasting of high-dimensional financial realized covariance matrices.

34. Tiago Pereira Da Silva (Universidade de São Paulo (USP))

    - Title: Effective networks: a model to predict network structure and critical transitions from datasets
    - Abstract: Real-world complex systems, such as ecological communities, neuron networks, and power grids, are essential components of our everyday lives. These complex systems are composed of units, or nodes, which interact through intricate networks. By observing the dynamical behaviour of complex systems, statistical and machine-learning techniques can predict their future behaviour without knowing how the nodes interact. The ability to predict sudden changes in network behaviour, also known as critical transitions, is important to be able to avert potentially disastrous consequences of major disruptions in the complex systems. However, predicting such new behaviours is a major challenge. In this talk, we address this by building a model network, termed an effective network, consisting of the underlying local dynamics at each node and an accurate statistical description of their interactions. To illustrate the power of effective networks to predict critical transitions, we reconstruct the dynamics and structure of real networks using neuronal interactions in the cat cerebral cortex, and demonstrate the effective network's ability to predict critical transitions for parameters outside the observed range. This novel methodology raises the possibility that networks can be controlled to anticipate malfunctions.