Book of Abstracts for the

Symposium on Machine Learning and Dynamical Systems

February 11th to February 13th, 2019

Imperial College London

Part I - Talks

1. Markus Abel (Ambrosys, markus.abel@ambrosys.de)

    - 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 synchronous behaviour of its parts, and an airfoil controlled to show high-lift.

2. AmirAli Ahmadi (Princeton, a_a_a@princeton.edu)

    - 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).

3. Rossella Arcucci and Yi-Ke Guo (Imperial College London, r.arcucci@imperial.ac.uk)
    - 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.

4. Peter Ashwin (Exeter, P.Ashwin@exeter.ac.uk)

    - Title: Network attractors and the functional dynamics of recurrent neural networks
    - Abstract: Attractors for dynamical systems come in a variety of types and structures, and understanding this can help to predict the behaviour of the system to inputs. Network attractors, which consist of heteroclinic or excitable networks of states in phase space give a useful paradigm to explain how a dynamical system may perform broadly accurate finite state computation, for example, as a Turing machine. Recurrent neural networks (RNNs), such as Echo State Networks, are themselves complex dynamical systems that are trained to respond dynamically to inputs in a task-dependent way. In this talk I will discuss some attempts to understand the computational abilities of RNNs using their nonautonomous dynamical behaviour and suggest ways to efficiently design RNNs suited to specific tasks. This is joint work with Andrea Ceni, Lorenzo Livi and Claire Postlethwaite.

5. Pascal Bianchi (Télécom ParisTech, pascal.bianchi@telecom-paristech.fr)

    - 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.

6. Erik Bollt (Clarkson, ebollt@clarkson.edu)

    - 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

7. Daan Crommelin (CWI Amsterdam, Daan.Crommelin@cwi.nl)
    - Title: Data-driven stochastic modeling for multiscale dynamical systems
    - Abstract: Modeling and simulation of multiscale dynamical systems (e.g. atmosphere and ocean) is challenging due to the wide range of spatio-temporal scales that need to be taken into account. To tackle this issue, one can employ stochastic models to represent the feedback from dynamical processes at the small/fast scales onto processes at larger scales. I will discuss ongoing work on extracting such stochastic subgrid-scale (i.e., small-scale) models from data, including results with resampling methods and discrete models. A key aspect is

the two-way coupling of the data-driven subgrid-scale model to a given (e.g. physics-based) large-scale model. Furthermore, the involved systems often display spatial and temporal correlations (or even memory) that should be accounted for.

[1] https://ir.cwi.nl/pub/23851
[2] http://journals.ametsoc.org/doi/abs/10.1175/2008JAS2566.1

8. Tiago Pereira DaSilva (Universidade de São Paulo, tiagophysics@gmail.com)

   - 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.

9. Michael Dellnitz (Paderborn, dellnitz@upb.de), Andreas Bittracher (FU Berlin, bittracher@mi.fu-berlin.de) and Sebastian Peitz (Paderborn, speitz@math.upb.de)

   - Title: Low-dimensional data-based surrogate models for analysis, simulation and control of high-dimensional dynamical systems
   - Abstract: In this presentation, we discuss the importance of low-dimensional coordinates for the efficient analysis, simulation and control of high- or infinite-dimensional systems, and we present data-based methods for their computation and use in control schemes.

The first part is concerned with the numerical computation of low-dimensional attractors and manifolds using embedding-techniques according to Takens, Sauer et al. and Robinson. If the system under consideration possesses a low-dimensional structure, then a generic low-dimensional observable of the system can be used to obtain a one-to-one image of the essential dynamics. In combination with set-oriented numerical algorithms, this is exploited in order to compute attractors and unstable manifolds of partial differential equations.
In the second part, we then focus on the task of obtaining dynamically meaningful coordinates for the system's essential dynamics from data. The information about the system's longtime behaviour is contained within the geometric structure of the so-called transition manifold, a low-dimensional object in the space of probability densities. By embedding this manifold into more tangible spaces, either again through generic low-dimensional observables or by kernel embedding techniques, its structure is made accessible to a wide variety of machine learning algorithms, which can then be used to compute optimal essential coordinates and other characteristic variables. This approach has proven especially useful in the conformational analysis of high-dimensional biomolecular systems.
Finally, in part three we will construct a low-dimensional surrogate model for a set of meaningful coordinates using the Koopman operator. The Koopman operator is a linear but infinite-dimensional operator describing the dynamics of observables. Using numerical methods such as Extended Dynamic Mode Decomposition (EDMD), we can construct a finite-dimensional approximation of the operator which yields a highly efficient linear reduced model. Using a recent convergence result for EDMD, optimality of the resulting reduced control problem can be guaranteed. The efficiency of this approach is shown for the Burgers equation as well as the flow around a cylinder governed by the Navier-Stokes equations.

10. Peter Dueben (ECMWF, Peter.Dueben@ecmwf.int)

- 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/

11. Vincent Fortuin (ETH, fortuin@inf.ethz.ch)

- 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

12. Marco Gallieri (NNAISENSE, marco@nnaisense.com)

- 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.

13. Hamza Ghadyali (SAS Institute, Cary, NC, Hamza@hmgxyz.com)

- Title: Brains, Waves, and Oribts: Topological Data Analysis with Machine Learning for Data-Driven Discovery of Dynamical Systems
- Abstract: For time-varying dynamical systems, the geometric properties of an orbit can indicate the state of the system. Multi-dimensional time-series data from real-world applications can be modelled as representing such a dynamical system through Takens-style delay-embeddings, however these embeddings are at best approximations to the underlying latent dynamical systems and can be very high dimensional. Analyzing non-linear or non-local features of high dimensional data is challenging due to the curse of dimensionality, which is where Topological Data Analysis (TDA) plays a useful role. TDA is a set of techniques for robustly extracting geometric information from data, even when the data is high-dimensional. For some problems, where a rule cannot be easily formulated for associating geometric configurations of the system to invariant sets of the dynamics, machine learning can be used to find that mapping from observed data. Deep learning, as a special case, has also proven useful in that regard and can find that mapping through hierarchical data-driven feature representations. We evaluate these ideas in the context of EEG-based epileptic-seizure detection and prediction, asking the question of whether we can use TDA and machine learning to discover characteristic geometric signatures of the anomalous electrical-activity in the underlying dynamical system which is the network of the brain's neurons together with its environment. We also evaluate it in the context of climate dynamics looking at weather data and obtaining new metrics for analyzing variation in periodic phenomenon.

14. Bernard Haasdonk (Stuttgart, haasdonk@mathematik.uni-stuttgart.de) , Gabriele Santin (Stuttgart) and T. Brunnette (Stuttgart)

- 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.

15. Boumediene Hamzi (Imperial College London, boumediene.hamzi@gmail.com)

- 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

16. Heather Harrington (University of Oxford, harrington@maths.ox.ac.uk)

- Title: Topological data analysis for investigation of dynamics and networks

- Abstract: Persistent homology (PH) is a technique in topological data analysis that allows one to examine features in data across multiple scales in a robust and mathematically principled manner, and it is being applied to an increasingly diverse set of applications. We investigate applications of PH to dynamics and networks, focusing on two settings: dynamics *on* a network and dynamics *of* a network.

  Dynamics on a network: a contagion spreading on a network is influenced by the spatial embeddedness of the network. In modern day, contagions can spread as a wave and through the appearance of new clusters via long-range edges, such as international flights. We study contagions by constructing 'contagion maps' that use multiple contagions on a network to map the nodes as a point cloud. By analyzing the topology, geometry, and dimensionality of manifold structure in such point clouds, we reveal insights to aid in the modelling, forecast, and control of spreading processes.

Dynamics of a network: one can construct static graph snapshots to represent a network that changes in time (e.g. edges are added/removed). We show that partitionings of a network of random-graph ensembles into snapshots using existing methods often lack meaningful temporal structure that corresponds to features of the underlying system. We apply persistent homology to track the topology of a network over time and distinguish important temporal features from trivial ones. We define two types of topological spaces derived from temporal networks and use persistent homology to generate a temporal profile for a network. We show that the methods we apply from computational topology can distinguish temporal distributions and provide a high-level summary of temporal structure.

Together, these two investigations illustrate that persistent homology can be very illuminating in the study of networks and their applications.

17. Axel Hutt (German Meteorological Service, digitalesbad@gmail.com)

- 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)

18. Isao Ishikawa (Riken, isao.ishikawa@riken.jp)
    - Title: Metric on nonlinear dynamical systems with Perron-Frobenius operators
    - Abstract: The development of a metric for structural data is a long-term problem in pattern recognition and machine learning. In my talk, we introduce a general framework to construct a metric for comparing nonlinear dynamical systems that is defined with Perron-Frobenius operators in reproducing kernel Hilbert spaces. Our metric includes the existing fundamental metrics for dynamical systems, which are basically defined with principal angles between some appropriately-chosen subspaces, as its special cases. We also describe the estimation of our metric from finite data. We empirically illustrate our metric with an example of rotation dynamics in a unit disk in a complex plane, and evaluate the performance with real-world time-series data.

19. Qianxiao Li (Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore, liqix@ihpc.a-star.edu.sg)

    - 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

20. Kevin Lin (U. Arizona, klin@math.arizona.edu)

- 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.

21. Robert MacKay (University of Warwick, R.S.MacKay@warwick.ac.uk)

- 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.

22. Mauro Maggioni (JHU, mauro@math.jhu.edu)

- 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.

23. Pierre-Yves Masse (ENS Cachan, PIERRE-YVES.MASSE@ens-cachan.fr)

- 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.

24. Kevin McGoff (UNCC, kmcgoff1@uncc.edu)

- Title: Empirical risk minimization over dynamical models

- Abstract: This talk concerns the fitting of a parametrized family of dynamical models to an observed real-valued stochastic process using empirical risk minimization. The limiting behavior of the minimum risk parameters is studied in a general setting. We establish a general convergence theorem for minimum risk estimators and ergodic observations. We then study conditions under which empirical risk minimization can effectively separate the signal from the noise in various observational noise models. The key, necessary condition in the latter results is that the family of dynamical models has limited complexity, which is quantified through a notion of entropy for families of infinite sequences.

25. Edward OTT (UMD, edott@umd.edu)
    - Title: Machine Learning for Prediction and Analysis of Chaotic Dynamics (Including that of Large Spatiotemporally Chaotic Systems)
    - ABSTRACT: We first review the basic idea of using machine learning to to construct from time series data a closed-loop, autonomous, dynamical system that mimics the dynamics of the unknown system that generated the data and predicts the future evolution of the measurements [1]. Using the reservoir computing type of machine learning, we then present examples of extensions and applications of this idea. These will include a parallel implementation enabling forecasting of the states very large spatiotemporally chaotic systems with local interactions [2], a hybrid scheme where a knowledge-based model component is combined with a limited-size machine learning component to achieve prediction performance much better than that of either of the components acting alone [3], Kalman filtering by a purely machine learning approach and by a hybrid machine-learning / knowledge-based approach, etc. Finally, we will use concepts from nonlinear dynamics to obtain understanding and conditions for effective operation of closed-loop machine learning prediction systems [4].

      [1] Jaeger, Haas, Science (2004).

      [2] Pathak, Hunt, Girvan, Lu, Ott, Phys Rev Lett (2018).

      [3] Pathak, Wikner, Fussell, Chandra, Hunt, Girvan. Ott, CHAOS (2018).

      [4] Lu, Hunt, Ott, CHAOS (2018).

26. Juan-Pablo ORTEGA (Universität Sankt Gallen, juan-pablo.ortega@unisg.ch)

    - 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.

27. Ioannis Panageas (SUTD, panageasj@gmail.com)

- 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.

28. Panos Parpas (Imperial College London, panos.parpas@imperial.ac.uk)

- 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.

29. Georgios Piliouras (SUTD, georgios.piliouras@gmail.com)

- 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.

30. Massimiliano Pontil (UCL and IIT, massimiliano.pontil@gmail.com)

- 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

31. Saverio Salzo (IIT, salzo.uni@gmail.com)

- 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.

32. Tim Sauer (GMU, tsauer@gmu.edu)

- 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.

33. Sebastian van Strien (Imperial College London, svanstrien@gmail.com)

- Title: Reinforcement learning in the multi-agent setting
- Abstract: Reinforcement learning is where a player keeps track of the success of actions in the past and chooses new actions accordingly. In the setting of a stationary stochastic environment such models are well understood. However, when other players also use some sort of learning then this leads to interesting and complicated dynamics. In this talk I will focus on a few specific models, such as Q learning. I will show that, depending on the precise implementation of the model, one obtains dynamics similar to that of a biological population model (replicator-like systems) but also chaotic Lorenz-like behaviour. This talk is based on joint work with Björn Winckler.

34. Grzegorz Swirszcz (DeepMind, grzegorz.swirszcz@gmail.com)

- 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.

35. Eniko Szekely (EPFL, eniko.szekely@epfl.ch)

- 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.

Part II, Posters

1. Jehan AlSwaihli (Reading, jehanalswaihli@gmail.com
2. Djalel Benbouzid (Volkswagen, djalel@argmax.ai)
3. Anthony Caterini (Oxford, anthony.caterini@st-hughs.ox.ac.uk)
4. Karim Cherifi (Max Planck Institute of Complex Dynamical Systems, Magdeburg/Germany and University of Boumerdes/Algeria, cherifikarimdz@gmail.com)
5. Giulia Denevi (IIT Genova, giulia.denevi@gmail.com)
6. Lynn Houthuys (KU Leuven, lynn.houthuys@esat.kuleuven.be)
7. Sanket Kamthe (Imperial College, s.kamthe15@imperial.ac.uk)
8. Maximilian Karl (Volkswagen, karlma@argmax.ai)
9. Martin Lellep (Marburg, lellepgu@staff.uni-marburg.de)
10. Lazaros Mitskopoulos (Crete, med1p1040135@med.uoc.gr)
11. Siyakha Mthunzi (Staffordshire, Siyakha.Mthunzi@staffs.ac.uk)
12. Jaideep Pathak (UMD, jpathak@umd.edu)
13. Samuel Rudy (UW, shrudy@uw.edu)
14. Maximilian Soelch (Volkswagen, m.soelch@argmax.ai)
15. Anastasios Tsourtis (Crete, tasoskrhs@gmail.com)
16. Miguel Xochicale (Birmingham, perez.xochicale@gmail.com)

1. Jehan AlSwaihli (Reading)

    - Title: Kernel Reconstruction for Delayed Neural Field Equations
    - Abstract: Modelling the dynamical systems mathematically is a challenge and an inspiration to researchers from different fields such as physics, mathematics, engineering, computer science, medical science and neuroscience. In recent years, contributions from those fields had convergent together to improve modelling and understanding of dynamical systems.
      In this contribution, we provide an integral equation approach to reconstruct the kernel of a delay neural field equation. We reformulate the inverse problem into a family of integral equations of the first kind. Then, due to the ill-posedness of this problem, we employ spectral regularization techniques for its stable solution. Numerical examples are provided to support our discussion.

2. Djalel Benbouzid and Maximilian Karl (Volkswagen)

    - Title: Unsupervised Real-Time Control through Variational Empowerment
    - Abstract:

3. Elhadj Benkhelifa and Siyakha Mthunzi (Staffordshire University)

    - Title: Prey-Inspired Dynamics for Survivability in Cloud Systems
    - Abstract: Cloud computing is increasingly relied upon at micro and macro-level critical systems, service assurance becomes a key component of secure, reliable and resilient service provision. Survivability is hence emphasised as the desirable capacity to meet the service provisioning mission. Currently, Service Level Objectives (SLA) are guaranteed by controlling resources that are already used by cloud providers (substrate networking). However, these are inefficient due their deterministic approach. The combination tenant concurrency, heterogeneity of resource and uncertain, unobserved or unobservable risks (UUUR) characterise the unpredictability and complexity of cloud. Effective survivability in dynamic environments requires optimal decision-making on survivability objectives on the fly. Decision-making requires monitoring and is particularly critical but majorly challenging where information is not known with certainty and future systems behaviours are unpredictable. For cloud-driven/cloud-based critical systems, it is optimally strategic that a monitoring system produce unambiguous and precise system information to facilitate optimal strategic responses under UUUR. In this work, a novel approach is presented that is robust for survivability of cloud systems under UUUR. Our proposition consists of a prey-inspired (Pi) survivability feedback-communication-control loop termed SDSSurv

(sense-decide-act) to dynamically actuate escalating countermeasures, akin to escalating anti-predation responses in prey animals. Partially Observable MDP achieve the optimal decision points by extending the observability element of MDP, i.e. inferred partial information to provide probabilistic information about future states (accurately premised on inference, observation and observability in survival prey animals). Cloud service provisioning survivability cloud benefit from dynamic decision-making and strategic escalating responses to UUURs.

4. Anthony Caterini and Dino Sejdinovic (Oxford)
   - Title: Hamiltonian Variational Auto-Encoder
   - Abstract: Variational Auto-Encoders (VAEs) have become very popular techniques to perform inference and learning in latent variable models as they allow us to leverage the rich representational power of neural networks to obtain flexible approximations of the posterior of latent variables as well as tight evidence lower bounds (ELBOs). Combined with stochastic variational inference, this provides a methodology scaling to large datasets. However, for this methodology to be practically efficient, it is necessary to obtain low-variance unbiased estimators of the ELBO and its gradients with respect to the parameters of interest. While the use of Markov chain Monte Carlo (MCMC) techniques such as Hamiltonian Monte Carlo (HMC) has been previously suggested to achieve this [23, 26], the proposed methods require specifying reverse kernels which have a large impact on performance. Additionally, the resulting unbiased estimator of the ELBO for most MCMC kernels is typically not amenable to the reparameterization trick. We show here how to optimally select reverse kernels in this setting and, by building upon Hamiltonian Importance Sampling (HIS) [17], we obtain a scheme that provides low-variance unbiased estimators of the ELBO and its gradients using the reparameterization trick. This allows us to develop a Hamiltonian Variational Auto-Encoder (HVAE). This method can be reinterpreted as a target-informed normalizing flow [20] which, within our context, only requires a few evaluations of the gradient of the sampled likelihood and trivial Jacobian calculations at each iteration.

5. Karim Cherifi (Boumerdes, Algeria)

   - Title: Introduction to Port Hamiltonian systems for realizations based on Machine learning techniques
   - Abstract: Realizations allow to model control systems accurately in a state space form. Moreover, we want the order of the system to be minimal. Models computed analytically are in general sensitive and cannot be computed in large scale systems. The modern approach is to use input/output data to interpolate data driven realizations. The state of the art in data driven methods are machine learning methods. However, its application is not easy as the models which we are dealing with have some constraints that have to be dealt with in an appropriate manner using appropriate methods. In this talk, we will focus on a special type of systems called Port Hamiltonian systems. These systems are energy based

systems and have interesting properties which made them subject of interest in recent years.
The objective of this talk to introduce the attendees to problems arising in realization theory for control systems in general and Port Hamiltonian systems in particular. Any ideas, suggestions or contribution to the solutions of these problems using machine learning is welcome.

6. Giulia Denevi (IIT)

   - Title: Incremental Learning-to-Learn with Statistical Guarantees
   - Abstract: In learning-to-learn (LTL) the goal is to infer a learning algorithm that works well on tasks sampled from an unknown meta distribution. We consider a scenario where tasks are presented sequentially and the underlying algorithm is Ridge Regression parametrized by a positive semidefinite matrix, implicitly defining a linear representation shared among the tasks. We propose to learn this matrix by minimizing the future empirical risk of the algorithm. We show that the objective function is convex and we apply a stochastic approach to minimize it. We give a non-asymptotic learning rate for the meta algorithm which is comparable to previous bounds for batch LTL.

7. Lynn Houthuys, Zahra Karevan and Johan Suykens (KU Leuven)

   - Title: Multi-view learning for black-box weather forecasting
   - Abstract: In multi-view regression, we have a regression problem where the input data can be represented in multiple ways. These different representations are called views. The aim of multi-view regression is to increase the performance of using only one view by taking into account the information available from all views. In this paper, we introduce a novel multi-view regression model called Multi-View Least Squares Support Vector Machines (MV LS-SVM) regression. This model is formulated in the primal-dual setting typical to Least Squares Support Vector Machines (LS-SVM) where a coupling term is introduced in the primal objective. This form of coupling allows for some degree of freedom to model the different representations while being able to incorporate the information from all views in the training phase. This work was motivated by the challenge of predicting temperature in weather forecasting. Black-box weather forecasting deals with a large number of observations and features and is one of the most challenging learning task around. In order to predict the temperature in a city, the historical data from that city as well as from the neighboring cities are taking into account. In the past, the data for different cities were usually simply concatenated. In this work, we use MV LS-SVM to do temperature prediction by regarding each city as a different view. Experimental results on the minimum and maximum temperature prediction in Brussels, show the improvement of the multi-view method with regard to previous work and that this technique is competitive to the existing state-of-the-art methods in weather prediction.

[1] https://ieeexplore.ieee.org/document/7965975/

8. Sanket Kamthe and Marc Deisenroth (ICL)

   - Title: Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control
   - Abstract: Trial-and-error based reinforcement learning (RL) has seen rapid advancements in recent times, especially with the advent of deep neural networks. However, the majority of autonomous RL algorithms require a large number of interactions with the environment. A large number of interactions may be impractical in many real-world applications, such as robotics, and many practical systems have to obey limitations in the form of state space or control constraints. To reduce the number of system interactions while simultaneously handling constraints, we propose a model-based RL framework based on probabilistic Model Predictive Control (MPC). In particular, we propose to learn a probabilistic transition model using Gaussian Processes (GPs) to incorporate model uncertainty into long-term predictions, thereby, reducing the impact of model errors. We then use MPC to find a control sequence that minimises the expected long-term cost. We provide theoretical guarantees for first-order optimality in the GP-based transition models with deterministic approximate inference for long-term planning. We demonstrate that our approach does not only achieve state-of-the-art data efficiency, but also is a principled way for RL in constrained environments.

     http://proceedings.mlr.press/v84/kamthe18a.html

9. Maximilian Karl and Maximilian Soelch (Volkswagen)

   - Title: Deep Variational Bayes Filters: Unsupervised Learning of State Space Models from Raw Data
   - Abstract:

10. Martin Lellep, Jonathan Prexl, Bruno Eckhardt (Philipps University Marburg, Germany)

    - Title: Predicting escape from a chaotic saddle using Machine Learning
    - Abstract: Chaotic saddles appear in chemical reactions, vortex interactions, transitional shear flows and many other systems. If the saddles are hyperbolic, the ensemble averaged lifetimes are exponentially distributed. Predicting whether a trajectory decays or remains on the saddle for some time to come is a challenging task on account of the typically intricate intermin- gling between escaping and remaining trajectories. We here use a neural network to predict whether a trajectory will escape from the chaotic sad- dle. The training input is a trajectory segment that includes not only the current position but also some interval into the past. The network is trained with N such samples, half of which remain on the

saddle and half of which escape. We study the Henon map as 2 dimensional chaotic saddle and investigate how N and the number of steps predicted in the future affect the prediction performance. The analysis can be extended to the Lorenz system and a low dimensional model of fluid turbulence as a step towards full direct numerical simulations of transitional flows.

11. Lazaros Mitskopoulos, J. Hizanidis and G. Tsironis (Crete)

- Title: Learning chaotic dynamical systems with recurrent neural networks
- Abstract: Data-driven modeling of chaotic nonlinear dynamical systems has recently entered a revolutionary phase where tools and techniques from machine learning have offered novel approaches to hard problems. In particular, learning algorithms based on recurrent neural networks such as Reservoir-Computing neural networks (RC) (Jaeger and Haas, 2004) and Long Short-Term Memory units (LSTM) (Hochreiter and Schmidhuber, 1997) have attracted considerable interest. The former have been demonstrated to perform remarkably well in the task of inferring the state of unmeasured state variables, after having been driven by one or a small set of the system variables during the training period (Lu et al, 2017). Furthermore, RC were shown to possess the capacity for accurate short-term prediction of future system states and phase space reconstruction, such that the lyapunov exponents that are calculated from the RC-generated output closely approximate the ones from the actual system of equations (Pathak et al, 2017; Pathak et al, 2018). Comparable success has been achieved using LSTM networks in high-dimensional dynamical systems, for short-term prediction (Vlachas et al, 2018) and for long-term prediction with observer nodes in the system (Neofotistos et al, 2018).
  In this study we aim to extend this line of work by investigating the performance of the RC as well as that of the LSTM networks in inference and future state prediction for the Hindmarsh-Rose neuron model (Hindmarsh and Rose, 1984). We perform a more conservative training procedure deploying roll-forward cross validation. Training is carried out with time series which are produced by integrating the Hindmarsh-Rose equations in various parameter regimes. Then, we conduct a detailed analysis of the state inference error in the hyperparameter spaces of the RC and the LSTMs. Lastly, we test the performance on short-term prediction and compare the Lyapunov exponents from the Hindmarsh-Rose system to the ones from the predicted output generated by the RC and the LSTMs.

  References

  -Hindmarsh, J. L., & Rose, R. M. (1984). A model of neuronal bursting using three coupled first order differential equations. Proc. R. Soc. Lond. B, 221(1222), 87-102.
  -Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

-Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. science, 304(5667), 78-80.
-Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brockett, R., & Ott, E. (2017). Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(4),041102.
-Neofotistos, G., Mattheakis, M., Barmparis, G. D., Hizanidis, J., Tsironis, G. P., & Kaxiras, E. (2018). Machin learning with observers predicts complex spatiotemporal behavior. arXiv preprint arXiv:1807.10758.
-Pathak, J., Hunt, B., Girvan, M., Lu, Z., & Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. Physical review letters, 120(2), 024102.
-Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., & Ott, E. (2017). Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(12), 121102.
-Vlachas, P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P., & Koumoutsakos, P. (2018). Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. Proc. R. Soc. A, 474(2213), 20170844.

12. Jaideep Pathak (UMD)

- Title: Using Machine Learning for Data-Driven Analysis of Ergodic Properties of Dynamical Systems
- Abstract: We demonstrate the effectiveness of machine learning for data-driven analysis of chaotic systems. Using a computationally efficient recurrent neural network called an echo state network or reservoir computer [1] we show that we can reconstruct the attractor of high dimensional chaotic dynamical systems with un- precedented fidelity. This reconstruction allows us to determine the ergodic properties (e.g., the spectrum of Lyapunov exponents) of a dynamical system purely from data [2]. We obtain excellent results using machine learning for these difficult tasks where traditional methods have had limited success.

[1] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 304(5667):78–80, 2004.
[2] Jaideep Pathak, Zhixin Lu, Brian R Hunt, Michelle Girvan, and Edward Ott. Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(12):121102, 2017.

13. Samuel Rudy (UW)

- Title: Deep Learning of Dynamics and Signal-Noise Decomposition
- Abstract: A critical challenge in the data-driven modeling of dynamical systems is producing methods robust to measurement error, particularly when data is limited.

Many leading methods either rely on denoising prior to learning or on access to large volumes of data to average over the effect of noise. We propose a novel paradigm for data-driven modeling that simultaneously learns the dynamics and estimates the measurement noise at each observation. Our method explicitly accounts for measurement error in the map between observations, treating both the measurement error and the dynamics as unknowns to be identified, rather than assuming idealized noiseless trajectories. We model the unknown vector field using a neural network, imposing a Runge-Kutta integrator structure to isolate this vector field, even when the data has a non-uniform time-step, thus constraining and focusing the modeling effort. We demonstrate the ability of this framework to form predictive models on a variety of canonical test problems of increasing complexity and show that it is robust to substantial amounts of measurement error.

[1] https://arxiv.org/pdf/1808.02578.pdf

14. Anastasios Tsourtis (Crete)

- Title: Inference of Dynamical Systems from Population Data using the Fokker-Planck equation
- Abstract: Inferring the driving equations of a dynamical system from population or time-course data is important in several scientific fields such as biochemistry, epidemiology, financial mathematics and many others. Despite the existence of algorithms that infer the dynamics from trajectorial measurements there are no attempts to infer the dynamical system from population data. In this work, we employ and then computationally infer the Fokker-Planck equation which describes the evolution of the population's probability density. Then, following the USDL approach \cite{https://arxiv.org/abs/1710.00718}, we project the Fokker-Planck equation to a proper set of data-driven functions, transforming it into a linear system of equations. Finally, we apply sparse inference methods to induce the driving forces of the dynamical system. Our approach is illustrated in both synthetic and real data including non-linear, multimodal stochastic differential equations, biochemical reaction networks as well as mass cytometry biological measurements.

15. Miguel Xochicale (Birmingham)

- Title: Quantification of Dynamic Facial Expressions with Shannon Entropy in Human-Humanoid Interaction

- Abstract: Research on understanding and quantifying movement variability with nonlinear analyses has been well established in the last three decades in areas such as biome- chanics, sport science, psychology, cognitive science, and neuroscience (Davids et al., 2003). This work is hypothesising that nonlinear analyses can be used to quantify subtle variations of facial expressions that can be related to different mental states (i.e. anxiety, disinterest, relief, etc) (Back and Jordan, 2014). This hypothesis has then led the author to ask two research questions:
  (i) how the quantification of facial expressions can be related to the complexity of facial expressions?, and
  (ii) does the quantification of the complexity for facial expressions can tell us some- thing about the state of mind of a person?,
  In order to give insights into the raised questions, this work is proposing the use of Recurrence Quantification Analysis (RQA) to quantify the complexity of facial expressions which is based on previous investigations of the author with nonlinear dynamics to quantify movement variability in human-humanoid interaction (Xochi- cale, 2018). RQA computes measurements based on the density of recurrence points of diagonal line structures in the Recurrence Plots. For which, RQA provide under- standing of the dynamics of a system i.e. the determinism (predictability of a system) or Shannon entropy (complexity of a system) (Marwan et al., 2007). With that in mind, a pilot experiment is designed to show the complexity of facial expressions vari- ability. In the experiment one participant (the author) were asked to perform three levels of variability of face expressions: (i) neutral variations, (ii) slow variations, and (iii) faster variations. Then, using time-series data of 67 face landmarks collected with OpenFace (Baltrusaitis et al., 2018), 3D plots of RQA Entr (Shannon entropy) were computed in order to quantify the complexity of face expressions and therefore relate 3D plots of RQA Entr to both (i) the subtle variations of facial expressions and (ii) the state of mind of a person. Additionally, this work will present potential ap- plications in the context of human-humanoid interaction for automatic quantification of face expressions that can be related to person's state of mind.

  References

Back, E. and Jordan, T. R. (2014). Revealing variations in perception of mental states from dynamic facial expressions: A cautionary note. PLOS ONE, 9(1):1–5.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. (2018). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)(FG), volume 00, pages 59–66.

Davids, K., Glazier, P., Araujo, D., and Bartlett, R. (2003). Movement systems as dynamical systems. Sports Medicine, 33(4):245–260.

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. Physics Reports, 438(5):237 – 329.

Xochicale, M. (2018). Nonlinear Analyses to Quantify Movement Variability in Human-Humanoid Interaction. PhD dissertation as submitted, University of Birmingham, United Kingdom.