Data-driven methods for Multi-Scale Physics and Complex Systems
(A summer school and workshop for 2017)

Organizers

Prof. J. Nathan Kutz, Applied Mathematics, University of Washington
Prof. Steven L. Brunton, Mechanical Engineering, University of Washington
Prof. Claudio Conti, Physics, University of Roma "La Sapienza"
Prof. Eugenio del Rey, Physics, University of Roma "La Sapineza"

Summer School

Dates: July 24-28, 2017
Location: University of Rome "La Sapienza"

Workshop

Dates: July 31-August 4, 2017
Location: University of Washington Rome Center, Campo di Fiori, Roma

SUMMER SCHOOL SCHEDULE AT A GLANCE

Rome Summer School and Workshop 2017

Data-driven methods for Multi-Scale Physics and Complex Systems

Time	Mon	Tue	Wed	Thu	Fri	Sat	Sun
8:30 AM			Coffee	Coffee	Coffee		
9:00 AM			**Benjamin Erichson** Part 1: Randomized Linear Dimension Reduction - A Guided Tour	**Steven L. Brunton** Part 1: Machine Learning Control	**Krithika Manohar** Part 1: Sparse Measurements for Complex Systemsr		
9:30 AM							
10:00 AM	Coffee and Orientation						
10:30 AM	**Guido Caldarelli** Part 1: Introduction to Complex Networks		Coffee	Coffee	Coffee		
11:00 AM			**Benjamin Erichson** Part 2: Randomized Linear Dimension Reduction - A Guided Tour	**Steven L. Brunton** Part 2: Machine Learning Control	**Krithika Manohar** Part 2: Sparse Measurements for Complex Systemsr		
11:30 AM							
12:00 PM	Lunch Lunch						
12:30 PM			Lunch Lunch	Lunch Lunch	Lunch Lunch		
1:00 PM							
1:30 PM		Coffee					
2:00 PM	**Guido Caldarelli** Part 2: Introduction to Complex Networks	**J. Nathan Kutz** Part 1: Reduced Order Modeling	**Steven L. Brunton and J. Nathan Kutz** Part 1: Data-Driven Model Discovery	**Claudio Conti and Eugenio Del Re** Part 1: Complex Physics	**Claudio Castellano** Part 1: Dynamical processes on complex networks		
2:30 PM							
3:00 PM							
3:30 PM	Coffee	Coffee	Coffee	Coffee	Coffee		
4:00 PM	**Guido Caldarelli** Part 3: Introduction to Complex Networks	**J. Nathan Kutz** Part 2: Reduced Order Modeling	**Steven L. Brunton and J. Nathan Kutz** Part 2: Data-Driven Model Discovery	**Claudio Conti and Eugenio Del Re** Part 2: Complex Physics	**Claudio Castellano** Part 2: Dynamical processes on complex networks		
4:30 PM							
5:00 PM							
5:30 PM	Adjourn	Adjourn	Adjourn	Adjourn	Adjourn		
6:00 PM							
6:30 PM							

SUMMER SCHOOL SCHEDULE

Monday, July 24: **MARCONI BUILDING: Aula Conversi (first floor)**

Tutorial 1: **Guido Caldarelli** (IMT-Lucca): *Introduction to Complex Networks*

10:00 Coffee and Orientation
10:30 Part 1
12:00 Lunch
2:00 Part 2
3:30 Coffee Break
4:00 Part 3
5:30 Adjourn

Tuesday, July 25: **FERMI BUILDING: Aula Cabibbo (ground floor)**

Tutorial 2: **J. Nathan Kutz** (Washington): *Reduced Order Modeling*

Abstract: The theoretical study of nonlinear dynamical systems pervades the physical, biological and engineering sciences. Today, these studies are increasingly driven by computational simulations that are of growing complexity and dimension due to increasing computational power and resolution in numerical discretization schemes. Yet most dynamics of interest are known ultimately to be low-dimensional in nature, thus contrasting, and in antithesis to, the high-dimensional nature of scientific computing. Reduced-order models (ROMs) are of growing importance in scientific applications and computing as they help reduce the computational complexity and time needed to solve large-scale, engineering systems. Specifically, ROMs provide a principled approach to approximating high-dimensional spatio-temporal systems. This tutorial provides an overview of how dimensionality reduction techniques can be used to build efficient models for characterizing complex systems.

2:00 Part 1
3:30 Coffee Break
4:00 Part 2
5:30 Adjourn

Wednesday, July 26: **MARCONI BUILDING: Aula Conversi (first floor)**

Tutorial 3: **Benjamin Erichson** (Washington) *Randomized Linear Dimension Reduction-A Guided Tour*

Abstract: Linear dimension reduction techniques are fundamental tools for analyzing high dimensional data. However, massive data pose a computational challenge for traditional techniques. Fortunately, the fascinating and powerful concept of randomness has recently emerged as a computational strategy to ease the computational load of deterministic matrix and data algorithms. We give a tutorial overview of randomized algorithms for dimension reduction. Specifically, we present randomized algorithms to compute the singular value decomposition (SVD), principal component analysis (PCA), the dynamic mode decomposition (DMD) as well as the CUR decomposition. Finally, we give an introduction to the generalized concept of randomized tensor factorizations.

9:00 Part 1
10:30 Coffee Break
11:00 Part 2
12:30 Lunch

Tutorial 4: **Steven L. Brunton and J. Nathan Kutz** (Washington): *Data-Driven Model Discovery*

Abstract: Data-driven discovery methods, which have been enabled in the past decade by the plummeting cost of sensors, data storage, and computational resources, have a transformative impact on the sciences, facilitating a variety of innovations for characterizing high-dimensional data generated from experiments. Less well understood is how to uncover underlying physical laws and/or governing equations from time series data that exhibit spatiotemporal activity. Traditional theoretical methods for deriving the underlying dynamics are rooted in conservation laws, physical principles, and/or phenomenological behaviors. These first-principles derivations lead to many of the canonical models ubiquitous in physics, engineering, and the biological sciences. However, there remain many complex systems that have eluded quantitative analytic descriptions or even characterization of a suitable choice of variables (for example, neuroscience, power grids, epidemiology, finance, and ecology). We propose an alternative method to derive governing equations based solely on time series data collected at a fixed number of spatial locations. Using innovations in sparse regression, we propose a methodology whereby we can discover the terms of the governing dynamics that most accurately represent the data from a large library of potential candidate functions. Measurements can be made in an Eulerian framework, where the sensors are fixed spatially, or in a Lagrangian framework, where the sensors move with the dynamics. We demonstrate the success of the method by rediscovering a broad range of physical laws solely from time series data.

2:00 Part 1
3:30 Coffee Break
4:00 Part 2
5:30 Adjourn

Thursday, July 27: **FERMI BUILDING: Aula Cabibbo (ground floor)**

Tutorial 5: **Steven L. Brunton** (Washington): *Machine Learning Control*

Background: The data-driven control of complex, nonlinear dynamical systems is currently undergoing a revolution. Advances in data science and machine learning are providing powerful new techniques to attack canonically challenging control systems, such as are found in turbulence control, epidemiology, and neuroscience. These systems are all characterized by high-dimensional, nonlinear, and multiscale phenomena, along with a lack of simple models suitable for control design. Increasingly, machine learning is able to solve large nonconvex optimization problems, such as discovering nonlinear controllers. This tutorial will provide an overview of the use of machine learning methods for feedback control. These methods include machine learning for system identification and model reduction, as well as for the direct design of control laws. An emphasis will be placed on emerging methods and practical concerns in real-world applications, including high-dimensionality, sparse sampling, nonlinearity, and robustness.

9:00 Part 1
10:30 Coffee Break
11:00 Part 2
12:30 Lunch

Tutorial 6: **Claudio Conti and Eugenio Del Re** (Roma): *Complex Physics*

2:00 Part 1
3:30 Coffee Break
4:00 Part 2
5:30 Adjourn

Friday, July 28: **FERMI BUILDING: Aula Cabibbo (ground floor)**

Tutorial 7: **Krithika Manohar** (Washington): *Sparse Measurements for Complex Systems*

Abstract: Sensor placement or selection is one of the most challenging unsolved problems in engineering. Optimizing sensor placement to characterize a high-dimensional system is critical for the downstream tasks of prediction, estimation, and control. The central focus of this tutorial is to highlight methods for selecting the best and fewest measurement locations for reconstruction in a given basis. This goal implicitly assume that the high-dimensional state has a compact representation in some known basis, and this inherent compressibility enables sparse sensing. We compare two leading perspectives: 1) compressed sensing with random measurements in a universal basis (Fourier, wavelets, etc.), and 2) optimized measurements in a tailored basis, leveraging the ability of machine learning to exploit patterns in data. We synthesize concepts from sensor selection and model reduction to find the fewest sensors for recovery in a data-driven proper orthogonal decomposition (POD) basis. Optimized sensors are efficiently chosen using QR column pivots of the POD basis. Sensors obtained through this method are shown to surpass the performance of random sensors in the reconstruction of high-dimensional states in several compelling real-world datasets. Reducing the number of sensors through principled selection may be critically enabling when sensors are costly, and may also enable faster state estimation for low latency, high bandwidth control. MATLAB code is demonstrated.

9:00 Part 1
10:30 Coffee Break
11:00 Part 2
12:30 Lunch

Tutorial 8: **Claudio Castellano** (ISC-CNR): *Dynamical processes on complex networks*

Abstract: Interaction patterns in many social, technological and natural systems exhibit complex

topological properties, that have a strong influence on the behavior of dynamical processes, giving rise to nontrivial phenomena. Traditional methods used to understand, predict and control processes on regular lattices are often unable to capture the relevant features of dynamics taking place on networks. New approaches must be developed to deal with the disordered topology. I will present the phenomenology of some fundamental dynamical processes on complex networks, the nontrivial questions raised and I will discuss the related theoretical methods, highlighting the strengths and limits of the different approaches.

2:00 Part 1
3:30 Coffee Break
4:00 Part 2
5:30 Adjourn

WORKSHOP SCHEDULE

Figure 1: *The Palazzo Pio is on the southern end of the campo. Go through the door marked welcome (left) and proceed to the first floor*

Monday, July 31: **Palazzo Pio, Campo de Fiori, Piazza del Biscione 95**

9:00 Plenary 1: **Simone Melchionna** (ISC-CNR, Rome)
Multiscale biofluidics: from kinetic theory to medicine

Abstract: Biofluidics involves the study of biological entities, from macromolecules to cells, in motion and confined in a physiological environment, where several forces of hydrodynamic, electrostatic and fluctuating origin act in multiple and unpredictable ways. To tackle such formidable problems, our multiscale approach is based on representing macromolecules and cells in suspension via the Lagrangian framework and the Eulerian/kinetic representation for the aqueous solvent. In general, this still involves a huge number of particles but one is mainly interested in the dynamics of a relatively small set of degrees of freedom. Therefore, a specific coarse-graining approach is needed to achieve biological realism and close the equations. During the talk, we will look into well known examples of protein diffusion in crowding environments, that is, in conditions similar to the cell interior, and the aggregation process of peptides at the onset of Alzheimer disease, a phenomenon greatly accelerated by solvent-mediated interactions, where diffusion and aggregation are either reduced or accelerated as compared to dilute conditions. We will then look into the flow of blood cells in phsyiological vessels, and its application in cardiovascular medicine. Finally, learning how to construct and manage a multi-purpose computational tools requires utilizing high-end hardware, based on the latest CPU-GPU platforms, and understanding its full potential particularly for multiscale computing.

10:00 Coffee Break

10:30 Stefania Melillo (Roma)
3D tracking of large groups of featureless objects

11:00 Jan Totz (Berlin)
Experimental Observation of Spiral Wave Chimera in a Very Large Network of Chemical Oscillators

11:30 Massimiliano Viale (Roma)
Information transfer and behavioural inertia in starling flocks

2:00 Plenary 2: **Bingni Brunton** (Washington)
Data-intensive approaches to understanding neural computations underlying naturalistic behaviors

Understanding the human brain in action requires exploration of naturalistic neural data acquired over many brain areas and monitored over long periods of time. I will talk about our recent work using computationally scalable approaches to analyze large-scale human intracranial brain recordings augmented with video and audio, all simultaneously and continuously monitoring subjects for many days. The size and scope of this dataset greatly exceeds all previously available human neural recordings, making it possible for us to take a deep learning approach to neural decoding and prediction. We believe data-driven models of natural neural dynamics are critical in advancing meaningful direct human-machine interactions and human neuroscience.

3:00 Coffee Break

3:30 Nancy Wang (Washington)
Deep Learning with Long-Term, Naturalistic ECoG Using Automated Video and Audio Feature Detection
4:00 David Caldwell (Washington)
Application of data-driven methods to intracranial human neural recordings during direct electrical cortical stimulation
4:30 Giulia Marcucci (Roma)
Simulations of Time Asymmetric Quantum Mechanics by Nonlocal Nonlinear Optics

Tuesday, August 1

9:00 Plenary 3: **Themis Sapsis** (MIT)
A variational method for probing extreme events in turbulent dynamical systems

Abstract: Prediction of extreme events for chaotic systems with intrinsically high-dimensional attractors is a formidable problem throughout science and engineering. These are especially challenging issues when real-time prediction is needed, such as, for example, in weather forecasting or prediction of extreme nonlinear waves. Thus, a major challenge in contemporary data-driven modeling of dynamical systems is the computation of low-energy patterns or signals, which systematically precede the occurrence of these extreme transient responses. Here, we propose a variational framework for probing conditions that trigger intermittent extreme events in high-dimensional nonlinear dynamical systems. These algorithms exploit in a combined manner some basic physical properties of the chaotic attractor, as well as, stability properties of the governing equations. Specifically, we seek the triggers as the probabilistically feasible solutions of an appropriately constrained optimization problem, where the function to be maximized is a system observable exhibiting intermittent extreme bursts. We apply the method to two challenging problems: i) the prediction of extreme intermittent bursts of energy dissipation in a prototype turbulent system, the body-forced incompressible Navier?Stokes equation, known as the Kolmogorov flow, and ii) the prediction of extreme events in a dispersive wave model for weak turbulence, the Majda-McLaughlin Tabak (MMT) model. We find that in both cases the intermittent bursts caused by the spontaneous transfer of energy from large scales to the mean flow via nonlinear interactions. The global maximizer of the corresponding variational problem identifies the responsible modes, providing in this way with precursors for the occurrence of extreme dissipation events. We assess the performance of the derived predictors through direct numerical simulations.

10:00 Coffee Break

10:30 Davide Pierangeli (Roma)
Rogue waves and turbulence in nonlinear optical propagation

11:00 Benjamin Strom (Washington)
Machine Learning Control of Cross-Flow Turbines

11:30 Sonja Molnos (Potsdam)
Winter precipitation forecast in the European region using cluster analysis

12:00 lunch

2:00 Plenary 4: **Peter Benner** (Max-Planck Institute, Magdeburg)
Computing Parametric Reduced-Order Models using Projection and Data

Abstract: In this talk, we discuss the construction of reduced-order parametric dynamical systems using projection and data. We present two approaches, both of which use the Loewner approach to find a realization of a linear dynamical system from frequency response data so that the realized system interpolates the data as suggested in a series of papers by A.C. Antoulas and his students. The first approach considers the situation when we already have a non-parametric model at one or several instances of the parameters obtained using H_2-optimal projection methods, plus some evaluations of the model at certain frequencies. The frequency-response data is then used to build a surrogate (using the Loewner approach) that can be employed for fast evaluation of the model at new parameter instances. The second approach is based on data only and does not assume the availability of a model. It constructs a parametric reduced-order model directly from frequency-response data using the Loewner approach and uses the fact that interpolation of the frequency-response data at several parameter values and interpolation of the Loewner-realized models at these parameter values commutes.

3:00 Coffee Break

3:30 Alessandro Alla (Florida State)
DMD-Galerkin approximation for nonlinear dynamical system

4:00 Krithika Manohar (Washington)
Data-Driven Sensor Placement Methods

4:30 Eurika Kaiser (Washington)
Sparsity enabled cluster-based reduced-order models for control

Wednesday, August 2

9:00 Plenary 5: **Bernard Haasdonk** (Stuttgart)
Sparse Surrogate Modelling by Greedy Kernel Approximation

Abstract: Many applications in CSE have as essential modular component one or several mappings from some input space to some output space which needs to be evaluated repeatedly and rapidly. If this mapping is computationally costly, e.g. requiring the solution of a PDE, this may be prohibitive for real-time situations and data-based surrogate models may be a promising solution. In this presentation, we focus on greedy kernel approximation schemes for construction of such surrogate models. Based on a large sample set of input-output data pairs, we incrementally construct approximating models by a greedy selection procedure. For a symmetric selection approach, i.e. centers being equal to the data-sites, we can observe and prove convergence of the approximation technique. A nonsymmetric approach is more versatile than the symmetric greedy approximation, since it comprises it as a special case and, moreover, we also allow the kernel centers to be different from the data-sites. We discuss different selection criteria for the test and trial spaces, and relate to some existing methods. Experimentally, we demonstrate the potential on artificial examples as well as on real world biomechanical applications.

10:00 Coffee Break

10:30 Benjamin Erichson (Washington)
On Randomized Nonnegative Matrix Factorizations

11:00 Bethany Lusch (Washington)
Shape constrained tensor decompositions

11:30 Kicic Ivica (ETH Zurich)
Fast Motion of Heaving Airfoils

2:00 Plenary 6: **Petro Koumoutsakos** (ETH Zurich)
Data + Computers + Thinking = Computing

Abstract: Computing, the study of information processing, is a domain of knowledge and one of the most potent instruments for advancing Science and Engineering. Today, fueled by the advents of Exascale Computers and Big Data it shows unprecedented potential for discovery. We ride the wave made by a hundred years of advances in numerical methods and algorithms. In the rumble of this wave we may also hear that machines that learn may soon replace humankind. In this talk, I wish to take a critical stand on these premises and discuss the power and limitation of computing in Understanding and Prediction. Computers and Big Data require the allocation of important societal resources such as energy and human capital and, even more, may endanger the personality, liberty and property of its members. This talk makes the thesis that Computing can assist Science but should not be considered as replacement for thinking and imagination.

3:00 Coffee Break

3:30 Zhe Bai (Washington)
Compressive system identification using DMD

4:00 Yang Yang
Flame and flow field analysis using POD and DMD

4:30 Ion Victor Gosea (Magdeburg)
Identification and reduction of linear and nonlinear systems from frequency and time domain data

5:00 Stefano Trillo (Ferrara)
Dispersive hydrodynamics: recent advances

Thursday, Aug 3

9:00 Plenary 7: **Christof Schuette** (FU Berlin)
Transition manifolds of complex metastable systems: Theory and data-driven computation of effective dynamics

Abstract: We consider complex dynamical systems showing metastable behavior but no local separation of fast and slow time scales. The talk raises the question of whether such systems exhibit a low-dimensional manifold supporting its effective dynamics. For answering this question, we aim at finding nonlinear coordinates, called reaction coordinates, such that the projection of the dynamics onto these coordinates preserves the dominant time scales of the dynamics. We show that, based on a specific reducibility property, the existence of good low-dimensional reaction coordinates preserving the dominant time scales is guaranteed. Based on this theoretical framework, we develop and test a novel numerical approach for computing good reaction coordinates. The proposed algorithmic approach is fully local and thus not prone to the curse of dimension with respect to the state space of the dynamics. Hence, it is a promising method for data-based model reduction of complex dynamical systems such as molecular dynamics.

10:00 Coffee Break

10:30 Stefan Klus (Berlin)
Dynamic mode decomposition, time-lagged independent component analysis, and applications in molecular dynamics
11:00 Peter Schmidt (ICFO-Barcelona)
Intersubband transitions in two-dimensional materials
11:30 David Carvalho (Heriot-Watt)
Nonlinear Optics of 2D Relativistic Media

2:00 Plenary 8: **Claudio Castellano** (ISC-CNR, Rome)
Epidemic spreading on complex networks

Abstract: Spreading phenomena are ubiquitous in our increasingly interconnected world. Understanding how topologically complex interaction patterns affect spreading processes is a goal of huge interest, both theoretically and for applications. In the investigation of these phenom-

ena, simple models for disease transmission play a paradigmatic role. In this talk, I will give an overview of the most relevant questions in this area and the main results obtained in the last years. I will in particular focus on the most basic models for the spreading of infectious diseases (the SIR and SIS dynamics). The interplay between the simple dynamics and the topological properties of the substrate gives rise to nontrivial emergent behavior that requires refined theoretical approaches to be understood.

3:00 Coffee Break

3:30 Dumitru Calugaru (Cambridge)
Explosive Synchronization in Complex Networks of Relaxation Oscillators
4:00 Swaraj Nanda (TU Delft)
High-speed tomographic particle image velocimetry of the flow field around cavitating tip-vortices off marine propellers
4:30 Sven Giorno (Michigan)
The Use of Data-Driven Methods for Subscale Modeling
5:00 Nora Lüthen (ETH Zurich)
Vorticity field prediction with LSTM networks

Friday, August 4

9:00 Panel Discussion: The future of data + modeling + computation

10:30 Coffee Break

11:00 Steven L. Brunton (Washington)
Machine learning to model and control fluids
11:30 J. Nathan Kutz (Washington)
Parametrized model discovery and ROMs