Dynamical systems and Semi-algebraic geometry: interactions with Optimization and Deep Learning

University of Dalat, Vietnam

Supported by RCN grant 300814, VinIF and University of Dalat

July 17–21, 2023

Website: https://www.mn.uio.no/math/english/research/projects/granddrm/events/conferences/dynamical-systems-and-semi-algebraic-geometry-inte/

1 Lecture series

Applications of semialgebraic geometry in variational analysis

Aris Daniilidis (TU Wien)

Abstract:

Talk 1. Bounded orbits of a semi-algebraic sweeping process have finite length. (based on a recent work with D. Drusviatskiy)

Talk 2. Semialgebraic Lipschitz functions are semismooth. (based on an old work with J. Bolte and A. S. Lewis)

Transition Phenomena in Stochastic Dynamical Systems

Jinqiao Duan (Illinois Institute of Technology, Great Bay University)

Abstract:

Lecture 1: Introduction to Stochastic Dynamical Systems This lecture is about basics of stochastic dynamical systems, including Brownian motion, Levy motion, stochastic integration, and stochastic differential equations.

Lecture 2: Transition Phenomena and the Onsager-Machlup Theory This lecture covers transition phenomena in complex systems under random influences, as modeled by stochastic differential equations, and the Onsager-Machlup theory for defining the most probable transition pathways.

Lecture 3: Variational Methods for the Most Probable Transition Pathways This lecture presents variational methods for seeking the most probable transition pathways, together with some applications.

References:

1. Jinqiao Duan, An Introduction to Stochastic Dynamics. Cambridge University Press, 2015.

2. Detlef Dürr & Alexander Bach, The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process. Commun.Math. Phys. 1978. https://doi.org/10.1007/BF01609446

3. Ying Chao and Jinqiao Duan, The Onsager?Machlup function as Lagrangian for the most probable path of a jump-diffusion process. Nonlinearity, 2019. DOI 10.1088/1361-6544/ab248b

Deep Learning, Dynamics and Control

Qianxiao Li (National University of Singapore)

Abstract: In these lectures, we discuss some theoretical and algorithmic issues that arise on the intersection of deep learning, dynamics and control.

Part I: In the first part, we focus on a control theoretic viewpoint of deep learning, where approximation is connected to controllability and optimisation is connected to optimal control.

Part II: In the second part, we discuss a mathematical framework to analyse deep learning for sequence/time-series applications, revealing interesting connections between approximation, memory, sparsity and low rank phenomena that may guide the practical selection and design of these network architectures.

Lectures on Semialgebraic geometry

Loi Ta Le (University of Dalat)

Abstract: Semialgebraic geometry is the study of sets of real solutions of systems of polynomial equations and inequalities. These lectures present some basic results of semialgebraic geometry. An overview of o-minimal geometry is given.

Lecture 1: Semialgebraic sets - The Tarski-Seidenberg theorem.

Lecture 2: Cell decomposition - Stratification.

Lecture 3: Curve selection - Lojasiewicz inequalities.

Lecture 4: An introduction to o-minimal geometry

2 Research talks

Second-order inertial algorithms for smooth and non-smooth

large-scale optimization

Camille Castera (University of Tubingen)

Abstract: Non-convex non-smooth optimization has gained a lot of interest due to the efficiency of neural networks in many practical applications and the need to "train" them. Training amounts to solving very large-scale optimization problems. In this context, standard algorithms almost exclusively rely on inexact (sub-)gradients through automatic differentiation and mini-batch sub-sampling. As a result, first-order methods (SGD, ADAM, etc.) remain the most used ones to train neural networks. Driven by a dynamical system approach, we build INNA, an inertial and Newtonian algorithm, exploiting second-order information on the function only by means of first-order automatic differentiation and mini-batch sub-sampling. By analyzing together the dynamical system and INNA, we prove the almost-sure convergence of the algorithm to critical points of tame objective functions. We also show that despite its second-order nature, INNA is likely to avoid strict-saddle points. Practical considerations will be discussed and some deep learning experiments will be presented. Finally, we depart from non-smooth optimization and provide insights into a recent work that paves the way for designing faster second-order methods.

Optimization Challenges in Robust Machine Learning

Volkan Cevher (EPFL)

Abstract: Thanks to neural networks (NNs), faster computation, and massive datasets, machine learning (ML) is under increasing pressure to provide automated solutions to even harder real-world tasks beyond human performance with ever faster response times due to potentially huge technological and societal benefits. Unsurprisingly, the NN learning formulations present a fundamental challenge to the back-end learning algorithms despite their scalability, in particular due to the existence of traps in the non-convex optimization landscape, such as saddle points, that can prevent algorithms from obtaining “good” solutions.

In this talk, we describe our recent research that has demonstrated that the non-convex optimization dogma is false by showing that scalable stochastic optimization algorithms can avoid traps and rapidly obtain locally optimal solutions. Coupled with the progress in representation learning, such as over-parameterized neural networks, such local solutions can be globally optimal.

Unfortunately, this talk will also demonstrate that the central min-max optimization problems in ML, such as generative adversarial networks (GANs), robust reinforcement learning (RL), and distributionally robust ML, contain spurious attractors that do not include any stationary points of the original learning formulation. Indeed, we will describe how algorithms are subject to a grander challenge, including unavoidable convergence failures, which could explain the stagnation in their progress despite the impressive earlier demonstrations. We will conclude

with promising new preliminary results from our recent progress on some of these difficult challenges.

Minimizing a separable sum coupled by a difference-of-convex function and linear constraints

Minh Dao (RMIT)

Abstract: In this work, we develop a splitting algorithm for solving a broad class of linearly constrained composite optimization problems whose objective function is the separable sum of possibly nonconvex nonsmooth functions and a smooth function, coupled by a difference-of-convex function. This structure encapsulates numerous significant nonconvex and nonsmooth optimization problems in the current literature including the linearly constrained difference-of-convex problems. Relying on successive linearization and alternating direction method of multipliers, the proposed algorithm exhibits the subsequential convergence to a stationary point of the underlying problem. We also establish the convergence of the full sequence generated by our algorithm under the Kurdyka–Łojasiewicz property and some mild assumptions. The efficiency of the proposed algorithm is tested on a robust principal component analysis problem and a nonconvex optimal power flow problem.

Detecting Cheapfakes - lLESSONS LEARNED FROM THREE YEARS OF ORGANIZING THE GRAND CHALLENGE

Minh-Son Dao (NICT Japan)

Abstract: The term "Cheapfake" has been recently coined to describe non-AI manipulations ("cheap" manipulations) of multimedia content. Cheapfakes are more widespread than deepfakes. Cheapfake media can be produced using image/video editing software or even without any software, by manipulating the context of an image/video through misleading claims. This contextual manipulation is known as out-of-context (OOC) misuse of media. OOC media is more difficult to identify compared to fake media, as the images and videos themselves remain unaltered. In this challenge, our focus is on detecting OOC images, specifically the misuse of authentic photographs accompanied by conflicting image captions in news items. This talk will discuss the lessons learned from organizing the grand challenge, which aims to develop and evaluate models capable of detecting re-contextualized cheapfakes.

A convex function satisfying the Lojasiewicz inequality but failing the gradient conjecture both at zero and infinity

Aris Daniilidis (TU Wien)

Abstract: We construct an example of a smooth convex function on the plane with a strict minimum at zero, which is real analytic except at zero and satisfies the Lojasiewicz gradient inequality, yet fails Thom's gradient conjecture both at zero and infinity. Based on a recent joint work with Olivier LEY and Mounir HADDOU.

Geometric properties of neural networks' risk functions

Vu Dinh (University of Delaware)

Abstract: In recent years, neural networks have become one of the most popular models for learning systems for their strong approximation properties and superior predictive performance. Despite their successes, their "black box" nature provides little insight into how predictions are being made. This lack of interpretability and explainability stems from the fact that there has been little work that investigates the statistical properties of neural networks due to their severe unidentifiability. In this talk, I outline a few main difficulties in studying the geometric properties of the set of risk minimizers of neural network models due to the model's unidentifiability. Using the context of neural network's feature selection in a regression setting as the primary example, I will highlight the relations between the optimality of the Lojasewizc exponent of the risk function around the set of its minimizers and the rate of convergence of the empirical risk minimizers. Built upon this example, I will introduce a few open questions about (semi-analytical) geometric properties of neural networks and their direct implications in explainable deep learning.

Genericity of Lyapunov spectrum of bounded random compact operators on infinite-dimensional Hilbert spaces

Thai-Son Doan (Hanoi Institute of Mathematics)

Abstract: In this talk, we present our recent result on genericity of Lyapunov spectrum of bounded random compact operators on infinite-dimensional Hilbert spaces. Our main contribution is about stability of Lyapunov exponents and simplicity of Lyapunov spectrum for bounded random compact operators on a separable infinite-dimensional Hilbert space from a generic point of view generated by the essential supremum norm. Firstly, we show the density of both the set of bounded random compact operators having finite number Lyapunov exponents and the set of bounded random compact operators having countably infinite number Lyapunov exponents. Meanwhile, the set of bounded random compact operators having no Lyapunov exponent is nowhere dense. Finally, for any $k \in \mathbf{N}$ we show that the set of bounded random compact operators satisfying that the Lyapunov spectral corresponding to their first Lyapunov exponents are simple and continuous contains an open and dense set. The preprint of the work can be found in https://arxiv.org/abs/2303.14359.

Detecting Transition Pathway in Stochastic Dynamical Systems through Optimal Control and Machine Learning

Ting Gao (Huazhong University of Science and Technology)

Abstract: Many complex real world phenomena exhibit abrupt, intermittent, or jumping behaviors, which are more suitable to be described by stochastic differential equations under non-Gaussian Lévy noise. Among these complex phenomena, the most likely transition paths between metastable states are important since these rare events may have a high impact in certain scenarios.

One of the challenges to calculate the most likely transition path for stochastic dynamical systems under non-Gaussian Lévy noise is that the associated rate functional cannot be explicitly expressed by paths. For this reason, we formulate the original variational problem into an optimal control problem to obtain the optimal state as the most likely transition path.

In this talk, we will present three types of efficient ways to solve this issue and the corresponding numerical analysis on the convergence and computational efficiency, including supervised learning and reinforcement learning as well as FBSDE with Pontryagin?s Maximum Principle. Various stochastic dynamical systems in applications will be discussed.

References:

J. Guo, T. Gao*, P. Zhang, J. Duan, Deep Reinforcement Learning in Finite-Horizon to Explore the Most Probable Transition Pathway, arXiv:2304.12994, 2023.

J. Chen, T. Gao*, Y. Li and J. Duan, Detecting the Most Probable High Dimensional Transition Pathway Based on Optimal Control Theory, arXiv:2303.00385, 2023

W. Wei, T. Gao, X. Chen*, J. Duan, An Optimal Control Method to Compute the Most Likely Transition Path for Stochastic Dynamical Systems with Jumps, Chaos, 32, 051102, 2022.

Fixed points, saddle graphs, and numerical algebraic geometry

Jonathan Hauenstein (University of Notre Dame)

Abstract: A key first step in studying maps and dynamical systems is to compute and analyze fixed points. For polynomial maps and polynomial dynamical systems, computational tools from algebraic geometry, such as numerical algebraic geometry, can be used to compute the fixed points. By understanding the local behavior near the fixed points along with possible connections between fixed points, one is able to construct a graphical representation of the landscape which we call a saddle graph. This talk will summarize some computational approaches for computing fixed points using numerical algebraic geometry and describe saddle graphs for polynomial optimization problems. These techniques will be demonstrated using several examples. This talk will cover joint work with Cinzia Binzi and Tuyen Truong (fixed points of maps), and Aravind Baskar and Mark Plecnik (saddle graphs).

The computational abilities of biological neurons and the relation to their neural network counterparts

Martin Hornkjøl (University of Oslo)

Abstract: In a standard neural network the neurons have the ability to multiply its inputs with weights, adding a bias and summing the modified inputs together. The neurons will also pass the output through an activation function. But what kind of functions does our brain neurons have? In this talk I will explore the computational and mathematical abilities of biological neuron and how this compare to the artificial ones we use in deep learning. I will look at neurological theory and various prominent experiments in this field.

Global convergence of the gradient method for functions definable in o-minimal structures

Cedric Josz (Columbia University)

Abstract: We consider the gradient method with variable step size for minimizing functions that are definable in o-minimal structures on the real field and differentiable with locally Lipschitz gradients. We prove that global convergence holds if continuous gradient trajectories are bounded, with the minimum gradient norm vanishing at the rate o(1/k) if the step sizes are greater than a positive constant. If additionally the gradient is continuously differentiable, all saddle points are strict, and the step sizes are constant, then convergence to a local minimum holds almost surely over any bounded set of initial points.

Optimality Conditions for Approximate Solutions of Nonsmooth Semi-infinite Vector Optimization Problems

Do Sang Kim (Pukyong National University)

Abstract: In this talk, we establish fuzzy necessary optimality conditions for (weak) ϵ-Pareto solutions, and exact necessary optimality conditions for (weak) quasi ϵ-Pareto solutions to a nonsmooth semi-infinite vector optimization problems by employing some advanced tools of variational analysis and generalized differentiation. We also design some examples to analyze and illustrate the obtained results.

Existence theorems for optimal solutions in semi-algebraic optimization

Gue Myung Lee (Pukyong National University)

Abstract: The existence of optimal solutions for optimization problems has been an essential research topic in optimization theory. It is well-known that a linear function attains its infimum on a nonempty polyhedral set if it is bounded from below on the set. In 1956, Frank and Wolfe proved that a quadratic function attains its infimum on a nonempty polyhedral set if it is bounded from below on the set. In 2002, Belousov and Klatte established the existence of optimal solutions for convex polynomial optimization problems. Very recently, for a general polynomial optimization problem, Phạm provided necessary and sufficient conditions for the existence of optimal solutions of the problem as well as the boundedness from below and coercivity of the objective function on the constraint set, where the results are presented in terms of the tangency variety of the polynomials defining the problem. Since polynomials form a subclass of semi-algebraic functions, it is natural to extend these results for semi-algebraic optimization problems. In this talk, we intend to talk about such extension for semi-algebraic optimization problems.

In this talk, we consider the problem of minimizing a lower semi-continuous semi-algebraic function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ on an unbounded closed semi-algebraic set $S \subset \mathbb{R}^n$. Employing adequate tools of semi-algebraic geometry, we first present some properties of the tangency variety of the restriction of f on S. Then we give verifiable necessary and sufficient conditions for the existence of optimal solutions of the problem as well as the boundedness from below and coercivity of the restriction of f on S. We also present a computable formula for the optimal value of the problem.

This talk is mainly from the paper jointly worked with Dr. Jae Hyoung Lee and Prof. Tien Son Pham.

Stochastic neuron models and the inference of mesoscopic stochastic neural dynamics.

Andre Longtin (University of Ottawa)

Abstract: This talk will have two components, both with the goal of highlighting computational approaches to real biological neural nets that will endow future deep learning nets with greater temporal precision and processing capabilities. The first is a short review of single neuron spiking models used in biological neural network research. For each, I will emphasize the deterministic properties followed by the additional features caused by the driving with noise, usually in the form of Gaussian white noise. These include different forms of thresholding nonlinearities: the integrate-and-fire with additive or conductance-based inputs, quadratic integrate-and-fire, exponential integrate-and-fire, leaky integrate and fire with threshold or current adaptation or both (named generalized integrate-and-fire in this latter case). Adaptation, or neural fatigue, can lead to a broader set of patterns accessible to a network and reduce neural variability, thereby enhancing temporal precision.

The second component considers low-dimensional mesoscopic dynamical systems descrip-

tions of high-dimensional "microscopic" stochastic dynamics of all neurons in a network, grouped into one or more populations. The simpler mesoscopic equations can provide insight into how the high-dimensional system processes inputs, produces rhythms or learns. This approach relies on prior work to formulate mesoscopic approximations of population activity to the full dynamics, e.g. by averaging techniques. We derive the likelihood of both mesoscopic single-neuron and connectivity (weight) parameters given the microscopic activity. This can be used to optimally estimate parameters using gradient ascent, or perform Bayesian inference with the help of Markov chain Monte Carlo sampling. The issue of model selection will be discussed which arises especially in the case of latent (i.e. unmeasurable) parts of the network activity, as well as implications for explainable AI.

The Moment-SOS approach for machine learning: classification and regression

Hoang-Anh Ngoc Mai (University of Konstanz)

Abstract: Drawing inspiration from the Moment-Sums Of Squares (Moment-SOS) hierarchy for polynomial optimization, we provide the following practical methods that offer theoretical guarantees for machine learning: We first revisit the Christoffel–Darboux kernel-based method for supervised learning, introduced by the second author. We perform some experiments to illustrate the practical implementations of this method. Secondly, we present a parametric family of polynomials for maximum likelihood estimation, with applications to supervised learning as well. Based on Weierstrass' theorem and Putinar's Positivstellensatz, we guarantee the convergence of our polynomial estimations for exact probability density functions under mild conditions. Moreover, we show that our black-box optimization problem is a convex program with semidefinite constraints. Next, we apply Boyd's primal-dual subgradient method to solve this program numerically. Thirdly, we rely on the volume computation developed by Dabbene and Henrion to build up a probabilistic Moment-SOS hierarchy for classification. More precisely, with probability near one and a sufficiently large number of uniformly random samples in each class $A_r \subset \Omega$, for almost all points $\mathbf{a}$ in Ω, we can determine which class A_r the point $\mathbf{a}$ belongs to under mild conditions. Fourthly, we prove the theoretical guarantee of the kernel trick for bi-class support vector machine in cases where the feature maps are vectors of monomials. To establish this guarantee, we present a hierarchy of linear systems of increasing sizes that return a sequence of polynomials. Moreover, with probability near one and sufficiently large sizes of linear systems, the resulting polynomials determine algebraic hypersurfaces which separate two disjoint sets with given uniformly random samples. Fifthly, we aim to generalize the result for bi-class support vector machine to handle s classes. This can be achieved by considering a hierarchy of linear systems that grow in size, where the k-th system produces a sequence of s polynomials $(p_{k,r})_{r=1}^{s}$. We also demonstrate that when uniformly random samples are taken from each class, for a sufficiently large k, the set of indices $argmax_{r=1,\dots,s} p_{k,r}(\mathbf{a})$ can

accurately determine the class to which a given point $\mathbf{a}$ belongs, with a probability close to one. Sixthly, we establish the theoretical assurance of polynomial regression given moderate conditions. To do so, we provide a hierarchy of quadratic programs that returns a sequence of polynomials of increasing degrees. This sequence converges in a weak sense to the function that fits a nonlinear relationship between the input and output. In addition, we do a similar process for a classification method based on polynomial regression. Finally, we provide the numerical comparison of the above methods in some instances of classification and regression.

This is joint work with Jean-Bernard Lasserre, Victor Magron, and Srecko Durasinovic.

The Boosted Difference of Convex functions Algorithm for DC Programming

Vuong Phan (University of Southampton)

Abstract: We present a new algorithm, called Boosted Difference of Convex functions Algorithm (BDCA), for minimizing difference of convex (DC) functions. BDCA accelerates the convergence of the classical difference of convex functions algorithm (DCA) thanks to an additional line search step. We show that there is complete freedom in the choice of the trial step size for the line search, which is something that can further improve the performance of BDCA. We prove that any limit point of the BDCA iterative sequence is a critical point of the problem under consideration and that the corresponding objective value is monotonically decreasing and convergent. The global convergence and convergence rate of the iterations are obtained under the Kurdyka–Łojasiewicz property. Applications and numerical experiments for two problems in data science are presented, demonstrating that BDCA outperforms DCA. Specifically, for the minimum sum-of-squares clustering problem, BDCA was on average 16 times faster than DCA, and for the multidimensional scaling problem, BDCA was 3 times faster than DCA. Finally, we will discuss the extension of BDCA to DC programming with linear constraints and apply the result to test the copositivity of a given matrix, and to solve trust-region subproblems.

Convergence analysis of optimization and root-finding algorithms via Lyapunov functions

Quoc Tran-Dinh (University of North Carolina at Chapel Hill)

Abstract: In this talk, we will discuss some recent progress of Lyapunov analysis techniques to analyze convergence rates of optimization and root-finding algorithms. We start with two well-known algorithms in convex optimization: gradient descent and Nesterov's accelerated gradient methods. Then, we expand our presentation to recent root-finding algorithms such as extragradient schemes, which covers fixed-point and minimax methods as special cases. Finally, if time permits, we will also discuss possible applications of these methods in machine learning and robust learning.

On structural stability and first-order optimization with time-dependent adaptive step policy

Xiao Wang (Shanghai University of Finance and Economics)

Abstract: First-order methods in optimization have widespread ML applications due to their ability to adapt to non-convex landscapes.

Their convergence guarantees are typically stated in terms of vanishing gradient norms. In recent years, non-convergence of

first-order methods to spurious critical points has been studied extensively by mahcine learning and optimization community.

In this talk, we summerize some recent progress on this topic. It is clear that classic methods from dynamical system plays a

prominent role in understanding behaviors of first-order algorithms.