OptML: Optimization and Machine Learning

Southampton, 18 October 2019

1. About the workshop

Optimization is a driving force behind the development of efficient tools for machine learning. Considering the ever-growing number of challenges related to existing and new machine learning tools and complicated structures appearing in data sets, there is a need for optimization and machine learning experts to come together to address these issues, while anticipating on future ones. At the same time, to make sure that new solutions from the research community meet real-world demand, practitioners need to be involved in this effort. Hence, the workshop will be a forum for optimization and machine learning experts, from both academia and industry, to discuss approaches to tackle optimization challenges in machine learning tasks and develop collaborations that will lead to new discoveries.

The event features invited talks from top machine learning and optimization experts, as well as practitioners from companies using and/or developing machine learning tools to solve real-world problems. A poster session dedicated to algorithms and applications of machine learning will also take place during the workshop. The programme allows for ample time for discussions and networking activity between participants from academia and industry.

The event is sponsored by The Alan Turing Institute, the national institute for data science and artificial intelligence. Find out more at turing.ac.uk

Support from the School of Mathematics, University of Southampton, is also acknowledged.

2. Venue

The Workshop will take place in the Cinema room located in Building 42 (Students' Union) of the main campus, Highfield, of the University of Southampton. The easiest way to access the Building is from the University Road, which is the main road that crosses the campus. For directions to the building, see the University map at the end of this document.

For accommodation, a list of hotels with preferential rates for university visitors can be found here. If you are interested in visiting Southampton during the workshop, some introductory information about the city is provided here.

For any queries related to the event, please email optml@southampton.ac.uk

3. Invited speakers

Coralia Cartis, Mathematical Institute, University of Oxford, UK

Coralia Cartis is Associate Professor in Numerical Optimisation in the Mathematical Institute, University of Oxford since 2013, and a Turing Fellow since 2016; previously, she held academic and research positions at University of Edinburgh and Rutherford Appleton Laboratory, respectively. She holds a PhD degree from Cambridge University and a BSc in Mathematics from Babesh-Bolyai University, Cluj, Romania. Her research interests are in nonlinear optimisation algorithm analysis and implementation, and in diverse applications of optimisation from climate modelling to signal processing and machine learning.

Alexander Culley, INTL FCStone Inc., London, UK

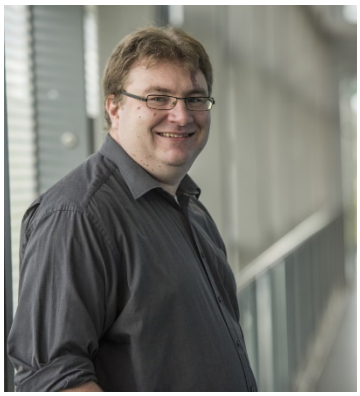
Alexander Culley is Chief Compliance Officer for the EMEA & Asia regions of INTL FCStone,

one of the world's leading non-bank brokerage firms. Called to the Bar of England and Wales in 2010, he is also fellow of the Chartered Institute for Securities and Investments and is currently studying at the University of Southampton for a Doctorate of Business Administration with a focus on how to design ethical conduct into trading firms' algorithmic business models.

Robert M. Gower, Télécom ParisTech, Paris, France

Robert M. Gower joined Telecom ParisTech as an Assistant Professor in 2017, he is interested in designing and analyzing new stochastic algorithms for solving big data problems in Machine Learning and scientific computing. A mathematician by training, his academic studies started with a Bachelors and a Masters degree in applied mathematics at the state University of Campinas (Brazil), where he designed the current state-of-art algorithms for automatically calculating high order derivatives using back-propagation. His PhD in stochastic numerics at the University of Edinburgh earned him the 2nd place of the 2017 Leslie Fox prize in numerical analysis. After which in 2016 he was granted the Fondation Sciences Mathématiques de Paris postdoctoral Laureate fund to continue his work as a postdoc in ENS.

Jonathon Hare, Electronics and Computer Science, University of Southampton, UK



Jonathon Hare is an Associate Professor in the School of Electronics & Computer Science. He holds a BEng degree in Aerospace Engineering and PhD in Computer Science, both from the University of Southampton. Jonathon's research interests lie at the convergence of machine learning and computer vision. He has a particular interest in deep-learning approaches for embedding multiple modalities of data in latent spaces that capture their semantics. He is also interested in models that take inspiration from biological systems in order to achieve efficient processing. The long-term goal of his research is to innovate techniques that can allow machines to efficiently understand the information conveyed by visual data and use that information for fulfil the information needs of humans. He has published over 90 articles in top peer-reviewed conferences and journals.

The Alan Turing Institute

UNIVERSITY OF Southampton

Ivan Markovsky, ELEC Department, Vrije Universiteit Brussel, Brussels, Belgium

Ivan Markovsky obtained MS degree in Control and Systems Engineering from the Technical University of Sofia in July 1998 and Ph.D. degree in Electrical Engineering from the Katholieke Universiteit Leuven in February 2005. From January 2007 to September 2012 he was a lecturer at the School of Electronics and Computer Science of the University of Southampton. Since October 2012 he is with the department ELEC of the Vrije Universiteit Brussel. His current research interests are structured low-rank approximation, system identification, and data-driven control.

Mohammad Mesgarpour, Data Science and Research, Microlise, Nottingham, UK

Dr Mohammad Mesgarpour has led a team of PhD-level data scientists and researchers at Microlise since 2017. Mohammad also manages a number of high-profile R&D projects with universities and other business partners. Prior to this, Mohammad acted as a Research Associate at the University of Nottingham and as a Technical Research Analyst at Microlise. Mohammad received his Ph.D. in Mathematics-Operational Research from the University of Southampton in 2012. He also holds Master's degrees in Information Systems and Industrial Engineering.

Microlise telematics, real-time journey management and proof of delivery solutions support local and global organisations to unlock profitability, strengthen safety, efficiency and compliance and reduce environmental impact. Operating for over thirty years, Microlise is privately-owned and proudly holds two concurrent Queen’s Awards for Enterprise.Based in Nottingham in the UK, with offices in Europe, India and Australia, Microlise supports 14 of the UK’s 15 largest retailers as well as JCB, MAN Truck & Bus UK and Tata Motors.

Laura Palagi, Computer Science and Control, Sapienza University of Rome, Rome, Italy

Laura Palagi is an Associate Professor in Operations Research at Sapienza University of Rome (Department of Computer, Control, and Management Engineering Antonio Ruberti) since 2005. She graduated in Electronic Engineering summa cum laude at the Sapienza University of Rome in 1990 and she received a Ph.D. in "Operations Research" from the Sapienza University of Rome in 1995. Her research activity is focused on the analysis and development of algorithms for the solution of nonlinear continuous optimization problems, with a focus on those arising in training machine learning (Deep Network and Support Vector Machines), and of mixed-integer nonlinear problems (MINLP). She authored more than 40 papers in international journals and books. She is a member elected of the managing board of EUROPT (The Continuous Optimization Working Group of EURO). She participated in many workshops both as a speaker and as a member of the Program Committee. She was co-chair of the Organizing Committee of EUROPT 2013 Meeting of the organized in Florence.

Andrea Simonetto, Ireland Research Lab, IBM, Dublin, Ireland

Andrea Simonetto is a research staff member in the optimization and control group of IBM Research Ireland, in Dublin. He received his PhD in systems and control from Delft University of Technology, The Netherlands in 2012, and spent 3+1 years as postdoc, first in the signal processing group in the electrical engineering department in Delft, then in the applied mathematics department of the Université catholique de Louvain, in Belgium. He was a visiting researcher at Carnegie Mellon University, University of Pennsylvania, and KTH, Sweden. He joined IBM Research in February 2017. His interests span optimization, control, and signal processing, with applications in smart energy, smart transportation, and personalized health.

Dave Woods, School of Mathematical Sciences, University of Southampton, UK

Dave Woods has research interests in the statistical design and analysis of experiments, particularly the development of new methods and criteria for design selection and assessment under linear and nonlinear models. A particular emphasis of his work is finding efficient designs when there is uncertainty in one or more aspects of the model for the response. He develops methodology that combines theory and computation to solve problems motivated by real experiments in science and industry. Application areas include engineering, chemistry and the pharmaceutical, automotive and aeronautics industries.

In 2019, he was elected as a Fellow of the American Statistical Association. From 2012-2017, he held a 5-year EPSRC Fellowship to conduct research into design of experiments for the complex nonparametric and mechanistic models required for modern scientific and industrial problems.

He has been an investigator on research grants and contracts totalling more than £18 million and has extensive research links with industry and government, including GlaxoSmithKline and Dstl. He is an Associate Editor for Technometrics and the SIAM/ASA Journal of Uncertainty Quantification, and a member of the EPSRC Peer Review College and the NSERC Mathematics and Statistics Evaluation Group.

4. Organizers

Alain Zemkoho, School of Mathematical Sciences, University of Southampton
Mahesan Niranjan, School of Electronics and Computer Science, University of Southampton
Houduo Qi, School of Mathematical Sciences, University of Southampton

5. Program of the workshop

Each invited talk is scheduled for 35 minutes, including about 5 minutes for questions. The Poster session with networking lunch will take place from 11:45 to 13:15 after the first session of talks. The workshop, including welcome + registration and the afternoon coffee break, will take place in Building 42, while the networking lunch will be at the Hartley Suite located in Building 38. Buildings 38 and 42 are very close to each other; cf. map below.

The Alan Turing Institute

UNIVERSITY OF Southampton

OptML workshop timetable

09:00–09:55 Welcome and registration
09:55–10:00 Opening remarks

10:00–11:45 – Session I: Machine learning

[Chair: Mahesan Niranjan]

10:00–10:35	**Jonathon Hare:** Learning to communicate: Challenges in optimising deep stochastic networks with categorical sampling
10:35–11:10	**Ivan Markovsky:** Sparsity in system identification and data-driven control
11:10–11:45	**Dave Woods:** Statistical learning through designed experiments

11:45–13:15 Poster session with networking lunch

13:15–15:00 – Session II: Optimization and machine learning in practice

[Chair: Alain Zemkoho]

13:15–13:50	**Alexander Culley:** Robocop: How financial regulators police machine learning
13:50–14:25	**Mohammad Mesgarpour:** Interaction of data science and optimisation across freight and logistics
14:25–15:00	**Andrea Simonetto:** Personalized optimization in a time-varying world

15:00–15:30 Coffee break

15:30–17:15 – Session III: Optimization methods for machine learning

[Chair: Houduo Qi]

15:30–16:05	**Coralia Cartis:** Stochastic variants of classical optimization methods, with complexity guarantees
16:05–16:40	**Robert M. Gower:** Expected smoothness is the key to understanding mini-batching for stochastic gradient methods
16:40–17:15	**Laura Palagi:** An overview on optimization issues in Deep network regression

17:15–18:00 Closing and networking

6. Abstracts

Stochastic variants of classical optimization methods, with complexity guarantees

Coralia Cartis, Mathematical Institute, University of Oxford, UK

Abstract. Optimization is a key component of machine learning application, as it helps with training of (neural net, nonconvex) models and parameter tuning. Classical optimization methods are challenged by the scale of machine learning applications and the lack of /cost of full derivatives, as well as the stochastic nature of the problem. On the other hand, the simple approaches that the machine learning community uses need improvement. Here we try to merge the two perspectives and adapt the strength of classical optimization techniques to meet the challenges of data science applications: from deterministic to stochastic problems, from small to large scale, from first- to second-order methods.

Robocop: How financial regulators police machine learning

Alexander Culley, INTL FCStone Inc., London, UK

Abstract. The proliferation of machine learning in financial services is prompting regulators around the world to rethink whether their human-oriented approach to supervision is still fit for purpose. In this talk, Alex will provide an overview of latest regulatory initiatives around the world, drawing upon both his research and extensive professional experience. What do developers need to be conscious of if they are developing machine learning tools for financial services businesses? Where could these tools be best deployed? What would the optimal business model of the financial services firm of the future look like? Alex will address these questions and more in this practical session.

Expected smoothness is the key to understanding mini-batching for stochastic gradient methods

Robert M. Gower, Télécom ParisTech, Paris, France

Abstract. Stochastic gradient methods are efficient when compared to full batch methods (gradient descent/ Newton) because for each iteration they need only a mini-batch of data. But how big should this mini-batch be? There is currently little to no theory that suggests what the size of the mini-batch should be, and so in practice a rule of thumb is used. Here I will show how to choose the mini-batch size in a way that results in the fastest execution (in other words that optimizes the total complexity). The key to choosing this optimal mini-batch size is in quantifying the expected smoothness constant, a definition I will introduce in my talk and that depends on the data and the sampling used. In particular, by quantifying the expected smoothness constant for sampling without replacement, I will show how we can choose the optimal mini-batch size and larger stepsizes for both the standard SGD (Stochastic gradient descent) methods and stochastic variance reduced methods such as SAGA and SVRG.

The Alan Turing Institute

UNIVERSITY OF Southampton

Learning to communicate: Challenges in optimising deep stochastic networks with categorical sampling

Jonathon Hare, Electronics and Computer Science, University of Southampton, UK

Abstract. This talk will give a whistle-stop tour of modern deep-learning, focussing on a problem setting in which two agents powered by deep networks learn a discrete communication protocol by playing collaborative games with images. Such a model is challenging to optimise using gradient-based methods. The talk will highlight some of the key techniques that make training the model possible, including: the use of automatic differentiation to compute the gradients; hardware acceleration to make training possible in a reasonable amount of time; and, differentiable relaxations of sampling, coupled with the straight-through operator, which allow for discrete categorical variables inside the model. The talk will end with some thoughts on some of the key future optimisation challenges in this line of work.

Sparsity in system identification and data-driven control

Ivan Markovsky, ELEC Department, Vrije Universiteit Brussel, Brussels, Belgium

Abstract. Sparsity is a key underlying assumption in many applications. A challenge in using sparse approximation methods in new applications is finding bases that allow sparse representation of the data. This talk shows how sparsity appears naturally in problems in systems identification. The underlying assumption then is that the data is generated by a low complexity model. The system identification problem is then a structured low-rank approximation problem. The link between sparsity and model complexity allows us to use effective methods from the field of compressive sensing in systems and control. A specific example considered in this talk is data-driven control, i.e., obtaining a control signal directly from observed data without explicit model identification.

Interaction of data science and optimisation across freight and logistics

Mohammad Mesgarpour, Data Science and Research, Microlise, Nottingham, UK

Abstract. Traditionally optimisation solutions and transport management technologies deliver significant benefits to the freight and logistics sector. We have begun to recognise the impact of big data analytics and data science on the day-to-day management of commercial fleets. Some customer case studies will highlight current advances. However, the potential of two distinct platforms will never be realised without fully integrating the power of AI and ML across traditional optimisation and transport management solutions. The importance of the interaction between these domains will be examined.

The Alan Turing Institute

UNIVERSITY OF Southampton

An overview on optimization issues in deep network regression

Laura Palagi, Department of Computer, Control, and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

Abstract. Mathematical optimization plays a pillar role in Machine Learning (ML). Indeed, learning from available data means that parameters of a chosen system must be computed by solving to optimality a nonlinear optimization problem called the learning problem. We focus on methods for Deep Networks (DN) in a regression setting. We first recall the learning optimization paradigm for DN and we briefly discuss schemes for the joint choice of the network topologies and of the network parameters. The main part of the talk focuses on the core subproblem which is the continuous unconstrained (regularized) weights optimization problem with the aim of reviewing theoretical and practical aspects of widely used batch and online gradient methods.

Personalized optimization in a time-varying world

Andrea Simonetto, Ireland Research Lab, IBM, Dublin, Ireland

Abstract. Optimization is the cornerstone of many engineering systems and cyber-physical systems including smart homes, energy grids, and intelligent transportation systems. In many situations however, state-of-the-art optimization algorithms may fail to provide acceptable (and feasible) solutions e.g. because of the scale of the problem, because the problem is continuously changing in time, or because the problem is ill-posed (i.e., depends on a cost function that is unknown). In this talk, I will focus on how to build an online algorithm to solve a time-varying optimization problem with an objective that comprises a known time-varying cost and an unknown function. This problem structure arises in systems where the known function captures time-varying engineering costs, and the unknown function models user's satisfaction; in this context, the objective is to strike a balance between given performance metrics and user's satisfaction that has to be learned online and concurrently with the execution of the optimization algorithm. I will then touch upon applications in this area stemming from smart energy grids and vehicle control.

Statistical learning through designed experiments

Dave Woods, Southampton Statistical Sciences Research Institute, Southampton, UK

Abstract. Statistically designed experiments are the “gold standard” for learning about products, processes and systems through the collection of data. By deliberately introducing controlled variability, whilst working to minimise uncontrolled variation, we can establish causative relationships, screen for important variables and build predictive models. I will describe how design of experiments and statistical modelling can go beyond the usual factorial design methodology familiar to many physical and applied scientists. In particular, I will present methodology for (i) experiments with dynamic input variables; (ii) design to learn unknown parameters in empirical and first-principle nonlinear models; and (iii) Bayesian nonparametric learning through sequential experimentation. Where possible, methods will be illustrated on relevant examples.

UNIVERSITY OF Southampton

The Alan Turing Institute

University Buildings: Highfield Campus

Building 1	1	Energy Centre CHP	11	Faraday	21	EEE	32	Early Years Centre	41	Mountbatten	53
Arts	2	Tizard	13	Building 23	23	E&F Maintenance	33	Students' Union	42	Mathematics	54
Building 2 Annexe	2a	Wolfson	15a	EScience	25	Education	34	Shackleton	44	Human Performance Lab.	56
Law	4	Rayleigh	15b	Building 26	26	David Kiddle	35	Building 44a	44a	Students Union Shop	57
Eustice	5	Building 16	16	Chemistry	27	Library	36	Health Sciences	45	Murray	58
NST Campus	6	R J Mitchell Wind Tunnel	17	Froude	28	George Thomas	37	Physics	46	Zepler	59
Lanchester	7	Jubilee Sports Centre	18	Graham Hills	29	Catering Conf. & Hospitality	38	University Health Centre	48	Gower	60
A B Woods Laboratory	8	ISVR	19	Synthetic Chemistry	30	Social Statistics	39	Building 50	50	Nightingale	67
Engineering Workshop	9	Tony Davies High Voltage Lab.	20	Hartley Store	31	Garden Court	40	Turner Sims	52	Life Sciences	85