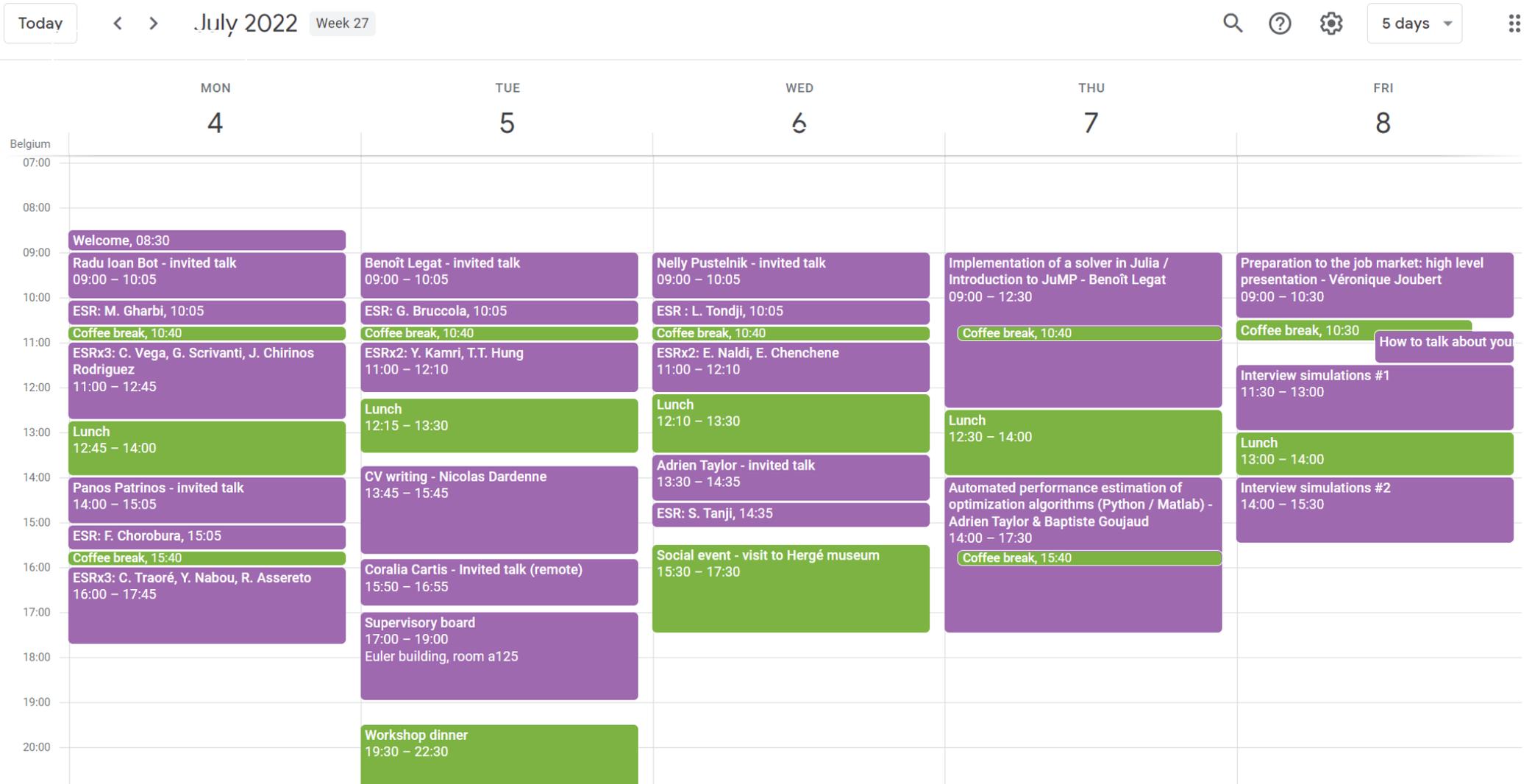


# TraDE-OPT workshop on Algorithmic and Continuous Optimization

July 4-8, 2022, UCLouvain, Belgium - <https://trade-opt-itn.eu/workshop.html>

All talks in the a002 room, Euler building

(av. G. Lemaître 4, <https://goo.gl/maps/1yXncxpTfm2zr3Qe8> )



All events will take place in the mathematical engineering department of UCLouvain (INMA), located in the **Euler building** (avenue Georges Lemaître, 4).

The main room for talks will be **Euler a002** (on the left after main entrance). See [this page for more details about the venue, maps, and accommodation](#).

## Talks and abstracts

### Invited speakers

- [Radu Ioan Boț](#) (University of Vienna)

Monday July 4, 9:00



#### **Fast methods for linearly constrained convex optimization problems**

In this talk we address the minimization of a continuously differentiable convex function under linear equality constraints. We consider a second-order dynamical system with asymptotically vanishing damping term formulated in terms of the Augmented Lagrangian associated with the minimization problem. Time discretization leads to an inertial algorithm with a general rule for the inertial parameters that covers the classical ones by Nesterov, Chambolle-Dossal and Attouch-Cabot used in the context of fast gradient methods. In both settings we prove fast convergence of the primal-dual gap, the feasibility measure, and the objective function value along the generated trajectory/iterates, and also weak convergence of the primal-dual trajectory/iterates to a primal-dual optimal solution. For the unconstrained minimization of a convex differentiable function we rediscover all convergence statements obtained in the literature for Nesterov's accelerated gradient method. If time permits, an alternative approach relying on the fast solving of monotone equations will be also presented.

- [Panos Patrinos](#) (KU Leuven)

Monday July 4, 14:00



#### **Escaping limit cycles: Global convergence for nonmonotone inclusions and constrained nonconvex-nonconcave minimax problems**

In this work we present a general projective framework for solving nonmonotone inclusions. Specializing the framework to nonconvex-nonconcave minimax problems, we introduce a new extragradient-type algorithm for a class of nonconvex-nonconcave minimax problems. It is well-known that finding a local solution for general minimax problems is computationally intractable. This observation has recently motivated the study of structures sufficient for convergence of first order methods in the more general setting of variational inequalities when the so-called weak Minty variational inequality (MVI) holds. This problem class captures non-trivial structures as we demonstrate with examples, for which a large family of existing algorithms provably converge to limit cycles. Our results require a less restrictive parameter range in the weak MVI compared to what is

previously known, thus extending the applicability of our scheme. The proposed algorithm is applicable to constrained and regularized problems, and involves an adaptive stepsize allowing for potentially larger stepsizes. Our scheme also converges globally even in settings where the underlying operator exhibits limit cycles. Finally, we show how our framework can be employed to provide provably convergent versions of Douglas-Rachford splitting for finding a zero of the sum of two (possibly) nonmonotone set-valued operators. To this end, we introduce the concept of semimonotonicity and provide sufficient conditions for global convergence of DRS involving the sum of two semimonotone operators. Our analysis relies on establishing a connection between DRS and a preconditioned proximal point algorithm studied under a certain oblique weak Minty assumption on the underlying primal-dual operator.

- [Benoît Legat](#) (MIT and UCLouvain)

Tuesday July 5, 9:00



### **Low-Rank Univariate Sum-of-Squares Has No Spurious Local Minima**

This talk studies the problem of decomposing a polynomial into a sum of  $r$  squares by minimizing a quadratically-penalized objective. This objective is non-convex and is equivalent to the rank- $r$  Burer-Monteiro factorization of a semidefinite program (SDP) encoding the sum of squares decomposition. We show that for all univariate polynomials  $p$ , if  $r$  is at least 2 then the objective function has no spurious second-order critical points, showing that all local optima are also global optima. This is in contrast to previous work showing that for general SDPs, in addition to genericity conditions,  $r$  has to be roughly the square root of the number of constraints (the degree of the polynomial) for there to be no spurious second-order critical points. We also show that by choosing a norm based on sampling equally-spaced points on the circle, the gradient of the objective function can be computed in nearly linear time using fast Fourier transforms. Experimentally we demonstrate that this method has very fast convergence using first-order optimization algorithms such as L-BFGS, with near-linear scaling to million-degree polynomials.

- [Coralia Cartis](#) (University of Oxford)

Tuesday July 5, 15:50



### **Evaluation complexity of algorithms for nonconvex optimization**

This talk presents an overview, as well as recent developments, regarding global rates of convergence and the worst-case evaluation (also called oracle) complexity of methods for nonconvex smooth optimization problems. We show how the popular methods of steepest descent and Newton's enjoy similar (sharp) bounds on their performance, despite the expectation that the latter would be 'fast(er)'. We argue the benefits of second-order regularization methods, which yield superior complexity bounds to first-order algorithms, and even optimal ones within a wide class of methods containing Newton's. Then we consider the advantages of having and incorporating higher- (than second-) order derivative information inside regularization frameworks, generating higher-order regularization algorithms that have better complexity, universal properties and can certify higher-

order criticality of candidate solutions. Time permitting, we also discuss inexact settings where derivatives and even function evaluations may only be sufficiently accurate occasionally, but whose worst-case complexity can still be quantified. A story of robust methods emerges, with continuously-varying, often sharp and sometimes even optimal, complexity.

- [Nelly Pustelnik](#) (ENS de Lyon)

Wednesday July 6, 9:00



### **Deep unfolded proximal algorithms**

In this talk we will detail how the proximal algorithms allows us to design a large class of new deep neural network architectures by exploiting the relation between proximity operators and activation functions. Additionally, we explore the interest of accelerated proximal algorithms to design more efficient architectures. Applications of these unfolded deep neural network to several inverse problems in image analysis will be presented and compared to state-of-the-art methods.

- [Adrien Taylor](#) (INRIA and ENS, Paris)

Wednesday July 6, 13:30



### **Constructive approaches to optimal first-order methods for convex optimization**

In this presentation, I want to provide a high-level overview of a recent approach for analyzing and designing first-order methods using symbolic computations and/or semidefinite programming (SDP). This approach is commonly referred to as “performance estimation”. The presentation will be example-based, as the main ingredients necessary for understanding the methodologies are already present in the analysis of base optimization schemes. Relying on those examples, I want to provide an overview on a few principled approaches to the construction of (optimal) first-order methods for convex optimization. The methodology is implemented within two open source packages, allowing to use the framework without the SDP modelling steps.

## **Progress presentations by doctoral students**

All 15 doctoral students (ESRs, see [this page](#)) from our network have confirmed their participation and will give a progress presentation at the workshop.

- [Mouna Gharbi](#) (CS)

Monday July 4, 10:05

### **GPU-based Implementations of MM algorithms. Application to Spectroscopy Signal Restoration**

Restoration of analytical chemistry data from degraded physical acquisitions is an important task for chemists to obtain accurate component analysis and sound interpretation. The high-dimensional nature of these signals and

the large amount of data to be processed call for fast and efficient reconstruction methods. Existing works have primarily relied on optimization algorithms to solve a penalized formulation. Although very powerful, such methods can be computationally heavy, and hyperparameter tuning can be a tedious task for non-experts. Another family of approaches explored recently consists in adopting deep learning to perform the signal recovery task in a supervised fashion. Although fast, thanks to their formulations amenable to GPU implementations, these methods usually need large annotated databases and are not explainable. In this work, we propose to combine the best of both worlds, by proposing unfolded Majorization-Minimization (MM) algorithms with the aim to reach fast and accurate methods for sparse spectroscopy signal restoration. Two state-of-the-art iterative MM algorithms are unfolded onto deep network architectures. This allows both the deployment of GPU-friendly tools for accelerated implementation, as well as the introduction of a supervised learning strategy for tuning automatically the regularization parameter. The effectiveness of our approach is demonstrated on the restoration of a large dataset of realistic mass spectrometry data.

- Cristian Vega (UniGe)

Monday July 4, 11:00

**Fast iterative regularization by reusing data** (with Cesare Molinari, Lorenzo Rosasco, and Silvia Villa)

Discrete inverse problems correspond to solving a system of equations in a stable way with respect to noise in the data. A typical approach to enforce uniqueness and select a meaningful solution is to introduce a regularizer. In this talk, we present two new iterative regularization methods, based on a primal-dual algorithm. The main contribution is the exploitation of some a priori knowledge about the solution set, reusing the data constraint.

- Gabriele Scrivanti (CS)

Monday July 4, 11:35

**A Convex-Non-Convex approach for Directional Total Variation** (with Emilie Chouzenoux and Jean-Christophe Pesquet)

A wide class of problems arising in image and signal processing is represented by sparsity-regularised variational models. The most natural sparsity inducing penalty is the  $l_0$ -pseudo-norm, but it forces the related problem to be NP hard and nonconvex. A popular convex variational surrogate is the  $l_1$ -norm, though it has the drawback of under-estimating the high amplitude components of the considered signal. Nonconvex variational regularisers manage to overcome this issue, at the cost of introducing suboptimal local minima in the objective function. An efficient solution to keep only the best traits of these regularisers is represented by Convex Non Convex strategies: they consist of building convex objective functionals that include nonconvex regularisation terms. Suitable CNC strategies have been designed for more and more general classes of problems. A CNC form of Total Variation has shown to get around the well known problems of boundary reduction and staircasing effect related to the classic TV regularisation for image Restoration and Segmentation. We propose a new variational formulation extending the CNC strategies to the recent Directional Total Variation model. We identify a Primal Dual procedure to efficiently address the resulting optimisation problems and provide experimental results that support the use of the presented regularisation method.

- Jonathan Chirinos Rodriguez (CAMELOT)

Monday July 4, 12:10

**A supervised learning approach to regularization of inverse problems** (with De Vito, E., Molinari, C., Rosasco, L., Villa, S.)

A classical and stable way to solve ill-posed inverse problems is to consider regularization functions that, in general, depend on a positive scalar that is known as regularization parameter, and choosing it remains a challenging problem nowadays. In the past years, there has been a huge attempt to consider data driven models for such issue with promising practical results. However, the lack of theoretical guarantees remains an open question. In this talk, we model such problem from a supervised learning point of view and provide generalization bounds for a broad class of regularization functions: those ones that can be written as filters.

- Flavia Chorobura (UPB)

Monday July 4, 15:05

**Random coordinate descent methods for nonseparable** (with Ion Necoara)

In this work we consider large-scale composite optimization problems having the objective function formed as a sum of two terms (possibly nonconvex), one has block coordinate-wise Lipschitz continuous gradient and the other is differentiable but nonseparable. Under these settings we derive and analyze two new random coordinate descent methods. The first algorithm, referred to as random coordinate proximal gradient method, considers the composite form of the objective function, while the other algorithm disregards the composite form of the objective and uses the partial gradient of the full objective, yielding a random coordinate gradient descent scheme with novel adaptive stepsize rules. We prove that these new stepsize rules make the random coordinate gradient scheme a descent method, provided that additional assumptions hold for the second term in the objective function. We also present a worst-case complexity analysis for these two new methods in both, convex and nonconvex settings. Preliminary numerical results also confirm the efficiency of the algorithms.

- Cheik Traoré (UniGe)

Monday July 4, 16:10

**Asynchronous parallel block-coordinate forward-backward algorithm**

We are going to present our work on the convergence properties of a randomized block-coordinate descent algorithm for the minimization of a composite convex objective function, where the block-coordinates are updated in parallel, asynchronously and randomly according to an arbitrary probability distribution. In that work, we prove that the iterates generated by the algorithm form a stochastic quasi-Fejér sequence and thus converge almost surely to a minimizer of the objective function. Moreover, we prove a general sublinear rate of convergence in expectation for the function values and a linear rate of convergence in expectation under an error bound condition of Tseng type. Under the same condition strong convergence of the iterates is provided as well as their linear convergence rate.

- Yassine Nabou (UPB)

Monday July 4, 16:35

**Efficiency of higher-order algorithms for minimizing general composite optimization** (with Ion Necoara)

In this paper, we design a higher-order majorization algorithmic framework for fully composite problems (possibly nonconvex). Our framework replaces each component with a higher-order surrogate such that the corresponding error function has a higher-order Lipschitz continuous derivative. We present convergence guarantees for our method for composite optimization problems with (non)convex and (non)smooth objective function. In particular, we prove stationary point convergence guarantees for general nonconvex (possibly nonsmooth) problems and under Kurdyka-Lojasiewicz (KL) property of the objective function we derive improved rates depending on the KL parameter. For convex (possibly nonsmooth) problems we also provide sublinear convergence rates.

- Rodolfo Assereto (U-GRAZ)

Monday July 4, 17:10

### **A multigrid approach to TV-Denoising**

In this talk we tackle the problem of accelerating the Total Variation regularization of a noisy image with Gaussian noise with a multigrid preconditioning approach. We show the equivalence of such preconditioning with the standard denoising problem, and we propose a computationally feasible splitting in the setting of Douglas-Rachford algorithms.

- Giovanni Brucola (SRI PAS)

Tuesday July 5, 10:05

### **Finding global solutions for a class of possibly nonconvex QCQP problems through the S-lemma** (with Ewa M. Bednarczuk)

In this talk we provide necessary and sufficient (KKT) conditions for global optimality for a new class of possibly nonconvex quadratically constrained quadratic programming (QCQP) problems, denoted by S-QCQP. The proof relies on a generalized version of the S-Lemma, stated in the context of general QCQP problems. Moreover, we prove the exactness of the SDP and the SOCP relaxations for S-QCQP.

- Yassine Kamri (UCLouvain)

Tuesday July 5, 11:00

### **On the worst-case analysis of coordinate gradient descent** (with F. Glineur and J. Hendrickx)

Coordinate gradient descent is a widely used first-order method in applications where computations on full vectors can be expensive such as machine learning or signal processing. In this talk, we extend the Performance estimation problem framework to cast the worst-case analysis of coordinate descent as a convex semidefinite problem (SDP) itself and we provide improved numerical worst-case bounds for coordinate descent in a variety of settings.

- The Hung Tran (SRI PAS)

Tuesday July 5, 11:35

### **Duality for Composite Optimization Problem within the Framework of Abstract Convexity**

We study conjugate and Lagrange dualities for composite optimization problems within the framework of abstract convexity. We provide conditions for zero duality gap in conjugate duality. For Lagrange duality, intersection property is applied to obtain zero duality gap. Connection between Lagrange dual and conjugate dual is also established. Examples related to convex and weakly convex functions are given.

- Lionel Tondji (TUBS)

Wednesday July 6, 10:05

### **Faster Randomized Block Sparse Kaczmarz by Averaging**

The standard randomized sparse Kaczmarz (RSK) method is an algorithm to compute sparse solutions of linear systems of equations and uses sequential updates, and thus, does not take advantage of parallel computations. In this work, we introduce a mini batch version of RSK based on averaging several Kaczmarz steps. Naturally, this method allows for parallelization and we show that it can also leverage large over-relaxation. We prove linear expected convergence and show that, given that parallel computations can be exploited, the method provably provides faster convergence than the standard method.

- Emanuele Naldi (TUBS)

Wednesday July 6, 11:00

### **Graph extensions of the Douglas-Rachford method**

We propose several generalizations of the Douglas-Rachford method to solve monotone inclusion problems involving  $N$  maximal monotone operators. In particular, our objective is to show that, starting from a directed connected graph with a topological ordering, it is possible to construct a generalization of the Douglas-Rachford algorithm with a communication structure that respects the connections of such a graph. We describe our methods in the (degenerate) Preconditioned Proximal Point framework and, in this way, we are able to prove convergence of every derived algorithm in a unifying way. Specifically, we show that the proposed methods are instances of the proximal point algorithm. The companion talk will describe how the graph extension of the DRS method can be leveraged to design new fully distributed protocols.

- Enis Chenchene (U-GRAZ)

Wednesday July 6, 11:35

### **Distributed extension of the Douglas-Rachford method with applications**

In this talk, we describe how the graph-based extensions of the Douglas-Rachford Splitting (DRS) method introduced in our companion's talk can be leveraged to design new fully distributed protocols for solving non-smooth convex optimization problems composed of many terms. The proposed method is based on a two layers architecture, consisting of what we call base and state graphs. We discuss the role of the two layers in detail, observing with an application to a congested optimal transport problem an interesting influence of the algebraic connectivity of the base graph on the convergence speed of the method. However, while a distributed protocol with in some sense minimal communication cost can be designed without effort when the base graph is a tree, the design of an efficient distributed scheme for general (i.e., more connected) base-graphs intertwines with purely graph-theoretical challenges. In this setting, we focus in particular on the class of the so-called chordal base graphs, where an efficient distributed optimization protocol can be designed via sparse factorization techniques. We conclude with an application to distributed Support Vector Machines, showing that the devised distributed schemes reach highly competitive performances compared to state-of-art methods such as P-EXTRA and a distributed Primal-Dual method while requiring significantly lower storage capabilities.

- Sofiane Tanji (UCLouvain)

Wednesday July 6, 14:35

### **Snacks: a fast large-scale kernel SVM solver**

Kernel methods provide a powerful framework for non parametric learning. They are based on kernel functions and allow learning in a rich functional space while applying linear statistical learning tools, such as Ridge Regression or Support Vector Machines. However, standard kernel methods require a  $O(n^2)$  time and memory complexity where  $n$  is the number of data points and thus, have limited applications in large scale learning. In this paper, we propose Snacks, a new large-scale solver for Kernel Support Vector Machines. Specifically, Snacks relies on a Nyström approximation of the kernel matrix and a carefully tuned variant of the stochastic subgradient method. Experimentally, we demonstrate that this method competes with other SVM solvers on a variety of benchmark datasets.

## **Algorithmic bootcamp**

The algorithmic bootcamp will be organized by [Benoît Legat](#), [Adrien Taylor](#) and [Baptiste Goujaud](#). Two topics will be covered:

- *Implementation of a solver in Julia / Introduction to JuMP* by Benoît Legat  
Thursday July 7, 9:00-12:30



- *Automated performance estimation of optimization algorithms (Python / Matlab)* by Adrien Taylor and Baptiste Goujaud (Ecole Polytechnique)  
Thursday July 7, 14:00-17:30



## Job market preparation

- Boost your CV with your doctoral skills, by [Nicolas Dardenne](#) Friday July 8, 9:00-10:30



- Job market: high level presentation, by [Véronique Joubert](#) Friday July 8, 9:00-10:30



- How to talk about yourself, animated by Véronique Joubert Friday July 8, 10:45-11:30
- Interview simulations #1, animated by Véronique Joubert Friday July 8, 11:30-13:00
- Interview simulations #1, animated by Véronique Joubert Friday July 8, 14:00-15:30