VGSCO

Vienna Graduate School on Computational Optimization

Second Vienna Workshop on Computational Optimization

March 15-17, 2023

Sky Lounge, 12th floor
Faculty of Mathematics, University of Vienna
(Oskar-Morgenstern-Platz 1, 1090 Vienna)

universität wien

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

ISTA Institute of Science and Technology Austria

WU WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Schedule VWCO23, March 15-17, Vienna

<u>Wednesday, March 15, 2023</u>

8.30-8.45 Registration
8.45-9.00 Opening

Section 1 – Chair Birgit Rudloff (WU)

9.00-9.45 Andrzej Ruszczyński (Rutgers University) *Stochastic optimization and learning with risk-averse objectives*
9.45-10.15 Niklas Hey (WU) *Computing an approximate set of Nash equilibria for convex Nash games via vector optimization*

10.15-10.45 Coffee break

10.45-11.30 René Henrion (Weierstrass Institute Berlin) *Turnpike phenomenon in discrete-time optimal control with probabilistic constraint*
11.30-12.00 Nurtai Meimanjan (WU) *Computation of systemic risk measures: a mixed-integer programming approach*
12.00-12.30 Aleksandr Shevchenko (ISTA) *Fundamental limits of two-layer autoencoders, and achieving them with gradient methods*

12.30-2.00 Lunch break

Section 2 – Chair Radu Boţ (University of Vienna)

2.00-2.45 Hedy Attouch (Université de Montpellier) *Fast convex optimization via damped inertial dynamics with Tikhonov regularization*
2.45-3.15 David Alexander Hulett (University of Vienna) *From maximally monotone operators to linearly constrained problems*

3.15-3.45 Coffee break

3.45-4.30 Andrea Walther (Humboldt Universität zu Berlin) *On nonsmooth optimization based on abs-linearization*
4.30-5.00 Axel Böhm (University of Vienna) *Beyond the golden ratio for variational inequality algorithms*
5.00-5.30 Mikhail Karapetyants (University of Vienna) *Fast continuous time approaches for convex nonsmooth optimization using Tikhonov regularization technique*

5.30 Welcome Reception (Sky Lounge)

<u>Thursday, March 16, 2023</u>

Section 1 – Chair Immanuel Bomze (University of Vienna)

9.00-9.45 Jean Bernard Lasserre (CNRS Toulouse) *The Christoffel function: Some of its applications and connections*
9.45-10.15 Jeferson Zapata *Hybrid methods in semidefinite programming*

10.15-10.45 Coffee break

10.45-11.30 Ivana Ljubic (ESSEC Business School of Paris) *Bilevel optimization under uncertainty*

11.30-12.00 Bo Peng (University of Vienna) *Lower bounds for structured nonconvex quadratic optimization problems*

12.00-12.30 Markus Gabl (Karlsruhe Institute of Technology) *Concave tents: A new approach to optimizing nonlinear convex functions over nonconvex set*

12.30-2.00 Lunch break

Section 2 – Chair Günther Raidl (TU Wien)

2.00-2.45 Sebastian Pokutta (TU Berlin) Conditional Gradients - an overview

2.45-3.15 Alexandra Peste (ISTA) *Effects of neural network compression beyond accuracy*

3.15-3.45 Coffee break

3.45-4.30 Pascal van Hentenryck (Georgia Tech) *Machine learning for optimization*

4.30-5.00 Marc Huber (TU Wien) *A policy-based learning beam search for combinatorial optimization*

5.00-5.30 Tommaso Mannelli Mazzoli (TU Wien) *Adaptive large neighborhood search for the bus driver scheduling problem*

5.30 Workshop Dinner (Restaurant Porzellan Lounge)

<u>Friday, March 17, 2023</u>

Section 1 – Chair Monika Henzinger (ISTA)

9.00-9.45 Michael Mitzenmacher (Harvard University) *Algorithms with predictions*

9.45-10.15 Lara Ost (University of Vienna) *Dynamic demand-aware link scheduling for reconfigurable datacenters*

10.15-10.45 Coffee break

10.45-11.30 Harald Räcke (TU München) *Hop-constrained expander decompositions, oblivious routing, and distributed universal optimality*

11.30-12.00 Sricharan Arunapuram (University of Vienna) *Electrical flows for polylogarithmic competitive oblivious routing*

12.00-12.30 Maximilian Vötsch (University of Vienna) *Online min-max paging*

12.30-2.00 Lunch break

Section 2 – Chair Nysret Musliu (TU Wien)

2.00-2.45 Guoyin Li (University of New South Wales Sydney) *Proximal methods for nonsmooth and nonconvex fractional programs: when sparse optimization meets fractional programs*

2.45-3.15 Ida Gjergji (TU Wien) *A large neighborhood search for the capacitated p-median problem*

3.15-3.45 Coffee break

3.45-4.15 Beryl Ramadhian Aribowo (University of Vienna) *Optimization in data science: 100 energies to learn molecular chemistry?*

4.15-4.45 Nikolaus Frohner (TU Wien) *Approaches to same-day delivery problems with soft deadlines*

Vienna Graduate School on Computational Optimization
Oskar-Morgenstern-Platz 1
1090 Vienna, Austria
T: +43-1-4277-50608
vgsco@univie.ac.at

Optimization in data science: 100 energies to learn molecular chemistry?

Beryl Ramadhian Aribowo
Faculty of Mathematics, University of Vienna
beryl.aribowo@univie.ac.at

joint work with Arnold Neumaier

Abstract

Ab initio quantum chemistry methods are powerful tools to accurately compute quantum properties of molecules; however, they require massive computational resources. In recent times, various machine learning models coupled with optimization methods have been proven to be flexible and efficient in approximating functions and are widely used in many fields, including molecular chemistry. Given an optimized machine learning model, it can predict quantum properties of molecules from a limited number of precomputed instances several orders of magnitude faster than the *ab initio* methods.

We present machine learning models which are simple to optimize and only have small computational cost. Machine learning models such as a combined supervised–unsupervised kernel and Gaussian kernel on molecular and atomic level are employed. Despite the simplicity, finding the best model configuration remains a challenge, hence we also employ an efficient hyperparameter optimization scheme using a derivative–free optimizer.

The QM9 dataset is chosen as a benchmark for the machine learning techniques. With only 100 sample molecules for training, a mean absolute error of 5 kcal/mol is the best currently achieved [1] across the whole 130k test molecules. Reducing the error to 1 kcal/mol (chemical accuracy) is the desired goal, which in the present needs around 1k sample molecules.

References

[1] von Lilienfeld, O.A., Müller, KR. & Tkatchenko, A., *Exploring chemical compound space with quantum-based machine learning*, Nat Rev Chem. **4** (2020), 347–358.

Electrical Flows for Polylogarithmic Competitive Oblivious Routing

A. R. Sricharan
University of Vienna
sricharana67@univie.ac.at

joint work with Gramoz Goranci, Monika Henzinger, Harald Räcke, Sushant Sachdeva

Abstract

Given routing requests between vertices of a graph, we study *oblivious routing schemes*, which use precomputed routing tables as opposed to dynamic routing schemes, which choose a congestion minimizing routing based on the currently observed traffic pattern in the graph. Congestion is the maximum traffic sent through any single edge of the graph when serving these requests. In addition to their simple structure, one can show that there are oblivious routings whose congestion is at most a $O(\log n)$ factor away from the optimal dynamic routing, and this factor is called the *competitive ratio* of the routing.

While there is an oblivious routing that takes a convex combination of m trees with competitive ratio $O(\log n)$, it takes m rounds of an iterative algorithm to find this routing, which is bad for parallel implementations. We show that with *electrical routings*, where the graph is viewed as an electrical network with resistances on the edges, one can bring down this round complexity to $O(\sqrt{m})$, while still maintaining a polylogarithmic competitive ratio of $O(\log^2 n)$.

References

[1] Ongoing work [GMHSS].

Fast convex optimization via damped inertial dynamics with Tikhonov regularization

Hedy Attouch
IMAG Université Montpellier, CNRS, France
hedy.attouch@umontpellier.fr

joint work with A. Balhag, Z. Chbani, H. Riahi

Abstract

In a Hilbert framework, for convex differentiable optimization, we analyze the convergence properties as time $t \to +\infty$ of trajectories generated by damped inertial dynamics with a Tikhonov regularization term whose coefficient vanishes asymptotically. An ad hoc adjustment of the damping terms and the Tikhonov regularization parameter yields trajectories that converge strongly to the minimum norm minimizer and with fast convergence properties. We propose two distinct approaches to these questions:

1. The Tikhonov regularization term induces a property of strong convexity which vanishes asymptotically. To take advantage of the fast convergence rates attached to the heavy ball method in the strongly convex case, we consider inertial dynamics where the viscous damping coefficient is proportional to the square root of the Tikhonov regularization parameter, and hence converges towards zero. The geometric damping, controlled by the Hessian of the function to be minimized, induces an attenuation of the oscillations. According to [1], based on Lyapunov analysis, we show that the trajectories enjoy the properties mentioned above.

2. In the second approach, we start from the Lyapunov analysis of the steepest descent with vanishing Tikhonov regularization. Then we apply to this first-order evolution equation the time scaling and averaging method recently introduced in [2]. We thus obtain an inertial dynamic which involves viscous damping associated with Nesterov's method, implicit Hessian damping and Tikhonov regularization. According to [3] and Jensen's inequality, we obtain the properties mentioned above.

Then we show that the corresponding proximal algorithms share similar properties. This study paves the way for extensions to hierarchical, multi-objective optimization, monotone inclusions, stochastic framework.

References

[1] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Damped inertial dynamics with vanishing Tikhonov regularization: Strong convergence towards the minimum norm solution*, J. Differ. Equ., **311** (2022), 29–58.

[2] H. Attouch, R.I. Boţ, D.-K. Nguyen, *Fast convex optimization via time scale and averaging of the steepest descent*, arXiv:2208.08260 [math.OC] Aug 2022.

[3] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization*, arXiv:2211.10140 [math.OC] Nov 2022.

Beyond the Golden Ratio for Variational Inequality Algorithms

Axel Böhm
University of Vienna
axel.boehm@univie.ac.at

joint work with Ahmet Alacaoglu, Yura Malitsky

Abstract

We improve the understanding of the *golden ratio algorithm*, which solves monotone variational inequalities (VI) and convex-concave min-max problems via the distinctive feature of adapting the step sizes to the local Lipschitz constants. Adaptive step sizes not only eliminate the need to pick hyperparameters, but they also remove the necessity of global Lipschitz continuity and can increase from one iteration to the next.

We first establish the equivalence of this algorithm with popular VI methods such as reflected gradient, Popov or optimistic gradient descent-ascent in the unconstrained case with constant step sizes. We then move on to the constrained setting and introduce a new analysis that allows to use larger step sizes, to complete the bridge between the golden ratio algorithm and the existing algorithms in the literature. Doing so, we actually eliminate the link between the golden ratio $\frac{1+\sqrt{5}}{2}$ and the algorithm. Moreover, we improve the adaptive version of the algorithm, first by removing the maximum step size hyperparameter (an artifact from the analysis) to improve the complexity bound, and second by adjusting it to nonmonotone problems with weak Minty solutions, with superior empirical performance.

Approaches to Same-Day Delivery Problems with Soft Deadlines

Nikolaus Frohner
Institute of Logic and Computation, TU Wien, Vienna, Austria
nfrohner@ac.tuwien.ac.at

joint work with Günther Raidl and Adrian Bracher

Abstract

Same-day delivery problems (SDDPs) [1] are a class of dynamic vehicle routing problems (DVRPs) [2] with the focus on near-term demand satisfaction of stochastic customers appearing in a delivery area throughout the day. In this presentation, we discuss anticipatory approaches to a real-world SDDP with tight soft deadlines [3, 4], based on scenario sampling [6], surrogate function [3] and value function learning [7], and compare these with myopic baseline policies. A key observation is that the potentially high computational demand of a scenario sampling approach can be replaced by an offline training of a machine learning model that is used effectively in the online decision making.

References

[1] S. A. Voccia, A. M. Campbell, and B. W. Thomas. The same-day delivery problem for online purchases. *Transportation Science*, 53(1):167–184, 2019.

[2] U. Ritzinger, J. Puchinger, and R. F. Hartl. A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research*, 54(1):215–231, 2016.

[3] N. Frohner and G. R. Raidl. A double-horizon approach to a purely dynamic and stochastic vehicle routing problem with delivery deadlines and shift flexibility. In *Proceedings of the 13th International Conference on the Practice and Theory of Automated Timetabling - PATAT 2022: Volume I*, Leuven, Belgium, 2022.

[4] Marlin W Ulmer. Delivery deadlines in same-day delivery. *Logistics Research*, 10(3):1–15, 2017.

[5] A. Bracher, N. Frohner, and G. R. Raidl. Learning surrogate functions for the short-horizon planning in same-day delivery problems. In *18th International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR'21)*, volume 12735 of *LNCS*, pages 283–298. Springer, 2021.

[6] R. W. Bent and P. Van Hentenryck. Scenario-based planning for partially dynamic vehicle routing with stochastic customers. *Operations Research*, 52(6):977–987, 2004.

[7] W. Joe and H. C. Lau. Deep reinforcement learning approach to solve dynamic vehicle routing problem with stochastic customers. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 30, pages 394–402, 2020.

Concave Tents: A new approach to optimizing nonlinear convex functions over nonconvex set

Markus Gabl
IOR, Karlsruhe Institute of Technology, Germany
markus.gabl@kit.edu

Abstract

Optimizing a nonlinear, convex function over nonconvex sets is challenging since solving convex relaxations where the nonconvex set is approximated by its convex hull my produce substantial relaxation gaps and infeasible solutions which have to be "rounded" to a feasible solution, often with uncontrollable losses in performance of the rounded solution. For this reason the aforementioned convex hull are especially useful if the objective function is at least concave. We propose the notion of concave tents which are concave overestimators of the convex objective function that agrees with the that objective in the extreme points of the convex hull of the nonconvex feasible set. We derive a general way to construct these concave tents under very mild assumptions as the optimal value function of conic optimization problem. Hence, evaluating our concave tents requires solving a conic problem. Interestingly, we can find superderivatives by solving the conic dual problem, so that differentiation is of the same complexity as evaluation. The construction of the concave tents exploits copositive optimization techniques at several stages so that our exposition establishes a connection between copositive optimization and nonlinear convex optimization. Preliminary numerical experiments show that these concave tents can be useful e.g. in global optimization where an infeasible solution is to be "rounded" to a feasible one.

Turnpike phenomenon in discrete-time optimal control with probabilistic constraint

René Henrion
Weierstrass Institute Berlin
henrion@wias-berlin.de

joint work with Martin Gugat (FAU, Erlangen) and Holger Heitsch (WIAS, Berlin)

Abstract

The turnpike phenomenon is well-known in mathematical economics or in optimal control with PDE constraints (see, e.g., [1]). It roughly states that for a large time interval, close to its middle, the dynamic optimal state/control is close to the static optimal state/control (obtained when setting time-derivatives equal to zero and leaving initial and terminal conditions free in the optimization problem). In this talk we are going to consider a simple time-discrete linear optimal control problem but with the peculiarity of adding a joint probabilistic (or chance) constraint to the system. The model may be understood in the context of reservoir managment (water, gas) with uncertain inflow and controlled release. More precisely, an initial filling level of the reservoir has to be steered in expectation to a given terminal level while staying all the time with high probability in a desired region. The objective function consists of a term of tracking type for the expected values and a control cost. We provide sufficient conditions that imply the optimal expected trajectories and the optimal control to remain close to corresponding trajectories that can be characterized as the solution of an optimal control problem without prescribed initial and terminal condition. The talk is based on the paper [2].

References

[1] M. Gugat, *On the turnpike property with interior decay for optimal control problems*, Math. Control Signals Syst. **33** (2021), 237–258.

[2] M. Gugat, H. Heitsch, and R. Henrion, *A turnpike property for optimal control problems with probabilistic constraints*, J. Convex Analysis, to appear.

Computing an approximate set of Nash equilibria for convex Nash games via vector optimization

Niklas Hey
WU Wien
nihey@wu.ac.at

joint work with Zachary Feinstein, Birgit Rudloff

Abstract

It was shown in [1] that the set of all possible Nash equilibria for an N-player noncooperative Nash game equals the intersection of the set of Pareto optimal elements of finitely many vector optimization problems. This result does hold for general Nash games. However when it comes to practical computations it might not be possible to derive the exact set of Pareto optimal elements for the corresponding vector optimization problems. Therefore the concept of approximate Pareto optimal elements is used. It can be shown that, for all $\epsilon > 0$, the set of ϵ-approximate Nash equilibria can be derived via the intersection of ϵ- Pareto optimal elements for finitely many vector optimization problems. Taking this fact into account we propose an algorithm that outputs a set which contains the set of true Nash equilibria and is upper-bounded by the set of ϵ-approximate Nash equilibria for Nash games with convex cost functions and independent convex constraint sets for each player.

References

[1] Z. Feinstein, B. Rudloff, *Characterizing and Computing the Set of Nash Equilibria via Vector Optimization*, Preprint. (2022).

A Policy-Based Learning Beam Search for Combinatorial Optimization

Marc Huber
Algorithms and Complexity Group,
Institute of Logic and Computation, TU Wien, Austria
mhuber@ac.tuwien.ac.at

Rupert Ettrich, Günther R. Raidl

Abstract

Beam search (BS) is a popular incomplete breadth-first search widely used to find near-optimal solutions to hard combinatorial optimization problems in limited time. Its central component is an evaluation function that estimates the quality of nodes encountered on each level of the search tree. While this function is usually manually crafted for a problem at hand, we propose a Policy-Based Learning Beam Search (P-LBS) that learns a policy to select the most promising nodes at each level offline on representative random problem instances in a reinforcement learning manner. In contrast to an earlier learning beam search, the policy function is realized by a neural network (NN) that is applied to all the expanded nodes at a current level together and does not rely on the prediction of actual node values. Different loss functions suggested for beam-aware training in an earlier work, but there only theoretically analyzed, are considered and evaluated in practice on the well-studied Longest Common Subsequence (LCS) problem. To keep P-LBS scalable to larger problem instances, a bootstrapping approach is further proposed for training. Results on established sets of LCS benchmark instances show that P-LBS with loss functions "upper bound" and "cost-sensitive margin beam" is able to learn suitable policies for BS such that results highly competitive to the state-of-the-art can be obtained.

References

[1] R. Ettrich, M. Huber, and G. R. Raidl, *A Policy-Based Learning Beam Search for Combinatorial Optimization*, Evolutionary Computation in Combinatorial Optimization – 23nd European Conference, EvoCOP 2023 (Leslie Pérez Cáceres, Thomas Stützle, eds.), 2023, Springer, to appear.

From maximally monotone operators to linearly constrained problems

David Alexander Hulett
University of Vienna
david.alexander.hulett@univie.ac.at

joint work with R. I. Boţ & D.-K. Nguyen

Abstract

The continuous fast OGDA scheme [2] addresses the problem $V(z_*) = 0$, where V is a monotone and continuous operator, through a second order differential equation. For a trajectory $z(\cdot)$ generated by this system, the authors show a rate of $\mathcal{O}\left(\frac{1}{t\beta(t)}\right)$ for $\|V(z(t))\|$ and $\langle z(t)-z_*, V(z(t))\rangle$, where β is a nonnegative function which satisfies a growth condition. We give an overview of this dynamical system and show that for a suitable choice of V, we enter the setting of linearly constrained convex optimization problems: we obtain a primal-dual dynamical system very similar to the (PD-AVD) dynamics in [1], where the fast rates transferred from fast OGDA translate into fast rates for the objective function values and feasibility gap along the generated trajectories.

References

[1] R. I. Boţ & D.-K. Nguyen, *Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping*, J. Differential Equations **303** (2021), 369–406.

[2] R.I. Boţ, E.R. Csetnek & D.-K. Nguyen, *Fast OGDA in continuous and discrete time*, arXiv: 2203.10947 (2022).

Fast continuous time approaches for convex nonsmooth optimization using Tikhonov regularization technique

Mikhail Karapetyants
University of Vienna
mikhail.karapetyants@univie.ac.at

joint work with Professor Radu I. Bot and Robert E. Csetnek

Abstract

In a Hilbert setting we study a second order in time differential equation, combining viscous and Hessian-driven damping, containing (optionally) a time scaling parameter function and a Tikhonov regularization term. The dynamical system is related to the problem of minimization of a nonsmooth convex function. In the formulation of the problem as well as in our analysis we use the Moreau envelope of the objective function and its gradient and heavily rely on their properties. We show that there is a setting where the system preserves and even improves the well-known fast convergence properties of the function and Moreau envelope along the trajectories and also of the gradient of Moreau envelope due to the presence of time scaling. Moreover, we prove strong convergence of the trajectories to the element of the minimal norm from the set of all minimizers of the objective. The talk will be a brief overview of the well-known results in this area concluding with the presentation of our contribution to this topic, including several numerical examples.

References

[1] Robert E. Csetnek, Mikhail Karapetyants, *A fast continuous time approach for non-smooth convex optimization with time scaling and Tikhonov regularization*, preprint.

The Christoffel function: Some of its applications and connections

Jean B. Lasserre
LAAS-CNRS
Toulouse, France
lasserre@laas.fr

Abstract

We introduce the Christoffel function (CF), a well-known tool in theory of approximation and orthogonal polynomials. We will briefly describe how the CF turns out to also provide a quite easy-to-use tool to help solve interesting problems in data analysis [1,2] (e.g., outlier detection, support inference, density approximation, classification). We will also discuss some of its links (in the author's opinion some surprising) with seemingly unrelated topics, like e.g. certificates of positivity in real algebraic geometry, equilibrium measure of compact sets, polynomial Pell's equation, and duality in polynomial optimization [3,4].

References

[1] J.B. Lasserre, E. Pauwels, M. Putinar. *The Christoffel-Darboux Kernel for Data Analysis*, Cambridge University Press, Cambridge, UK, 2022

[2] J.B. Lasserre, *On the Christoffel function and classification in data analysis*, Comptes Rendus Mathématique **360** (2022), 919–928. Lasserre J.B. (2022)

[3] J.B. Lasserre, *A disintegration of the Christoffel function*, Comptes Rendus Mathématique **360** (2022), 1071–1079.

[4] J.B. Lasserre, *Pell's equation, sum-of-squares and equilibrium measures on a compact set*, Comptes Rendus Mathématique **361** (2023). To appear.

Proximal methods for nonsmooth and nonconvex fractional programs: when sparse optimization meets fractional programs

Guoyin Li
The University of New South Wales (UNSW Sydney)
g.li@unsw.edu.au

Based on joint work with R.I. Boţ, M. Dao, T.K. Pong and P. Yu

Abstract

Nonsmooth and nonconvex fractional programs are ubiquitous and also highly challenging. It includes the composite optimization problems studied extensively lately, and encompasses many important modern optimization problems arising from diverse areas such as the recent proposed scale invariant sparse signal reconstruction problem in signal processing, the robust Sharpe ratio optimization problems in finance and the sparse generalized eigenvalue problem in discrimination analysis. In this talk, we will introduce extrapolated proximal methods for solving nonsmooth and nonconvex fractional programs and analyse their convergence behaviour. Interestingly, we will show that the proposed algorithm exhibits linear convergence for sparse generalized eigenvalue problem with either cardinality regularization or sparsity constraints. This is achieved by identifying the explicit desingularization function of the Kurdyka-Łojasiewicz inequality for the merit function of the fractional optimization models. Finally, if time permits, we will present some preliminary encouraging numerical results for the proposed methods for sparse signal reconstruction and sparse Fisher discriminant analysis

References

[1] R.I. Boţ, M. Dao and G. Li, Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs, Mathematics of Operations Research, **47** (2022) 2415–2443.

[2] R.I. Boţ, M. Dao and G. Li, Inertial proximal block coordinate method for a class of nonsmooth sum-of-ratios optimization problems, to appear in SIAM Journal on Optimization, 2023.

[3] P. Yu, G. Li and T.K. Pong, Kurdyka-Łojasiewicz exponent via inf-projection. Foundations of Computational Mathematics **22**, (2022), 1171–1217.

Bilevel Optimization Under Uncertainty

Ivana Ljubić
ESSEC Business School of Paris, Cergy-Pontoise, France
ljubic@essec.edu

joint work with Yasmine Beck, Martin Schmidt

Abstract

Significant algorithmic advances in the field of computational bilevel optimization allow us to solve much larger and also more complicated problems today compared to what was possible two decades ago. This results in more and more challenging bilevel problems that researchers try to solve today. In this talk, we will focus on one of these more challenging classes of bilevel problems: bilevel optimization under uncertainty.

We will discuss classical ways of addressing uncertainties in bilevel optimization using stochastic or robust techniques. The sources of uncertainty in bilevel optimization can be much richer than for usual, single-level problems, since not only the problem's data can be uncertain but also the (observation of the) decisions of the two players can be subject to uncertainty. The talk is based on the two articles by [1, 2]

References

[1] Yasmine Beck, Ivana Ljubić, Martin Schmidt, *A Survey on Bilevel Optimization Under Uncertainty*, EJOR, forthcoming (2023), `https://optimization-online.org/2022/06/8963/`

[2] Yasmine Beck, Ivana Ljubić, Martin Schmidt, *Exact Methods for Discrete* Γ*-Robust Interdiction Problems with an Application to the Bilevel Knapsack Problem* (2022), `https://optimization-online.org/2021/11/8678/`

Adaptive Large Neighborhood Search for the Bus Driver Scheduling Problem

Tommaso Mannelli Mazzoli
TU Wien
tommaso.mazzoli@tuwien.ac.at

joint work with Lucas Kletzander, Nysret Musliu

Abstract

The Bus Driver Scheduling Problem (BDSP) is a complex combinatorial optimisation problem. The goal is to assign bus drivers to vehicles with predetermined routes to minimise a given objective function. The objective takes into account both operating costs and the satisfaction of employees. Since we must satisfy several rules from the collective agreement and European regulations, the BDSP is highly constrained, so in practice using exact methods to solve real-life-based instances is computationally too expensive. We propose an Adaptive Large Neighbourhood Search (ALNS) approach to solve the BDSP. This matheuristic method combines the strength and efficiency of exact methods with the flexibility of heuristics. As exact methods, we have used the Branch and Price implementation that is currently state-of-the-art for the small and medium instances, and we also proposed an entirely new Mixed Integer Programming MIP model. Different destroy operators are studied, analysed and compared. We compared ALNS with state-of-the-art approaches and showed that it provides good solutions.

Computation of Systemic Risk Measures: A Mixed-Integer Programming Approach

Nurtai MEIMANJAN
Vienna University of Economics and Business
nurtai.meimanjan@wu.ac.at

joint work with Çağın ARARAT

Abstract

Systemic risk is concerned with the instability of a financial system whose members are interdependent in the sense that the failure of a few institutions may trigger a chain of defaults throughout the system. Recently, several systemic risk measures have been proposed in the literature that are used to determine capital requirements for the members subject to joint risk considerations. We address the problem of computing systemic risk measures for systems with sophisticated clearing mechanisms. In particular, we consider an extension of Rogers-Veraart network model where the operating cash flows are unrestricted in sign. We propose a mixed-integer programming problem that can be used to compute clearing vectors in this model. Due to the binary variables in this problem, the corresponding (set-valued) systemic risk measure fails to have convex values in general. We associate nonconvex vector optimization problems to the systemic risk measure and provide theoretical results related to the weighted-sum and Pascoletti-Serafini scalarizations of this problem. Finally, we test the proposed formulations on computational examples and perform sensitivity analyses with respect to some model-specific and structural parameters.

References

[1] Ararat Ç., Rudloff B., *Dual representations for systemic risk measures*, Mathematics and Financial Economics. **14(1)** (2020), 139–174.

[2] Biagini F., Fouque J.P., Fritelli M., Meyer-Brandis T., *A unified approach to systemic risk measures via acceptance sets*, Mathematical Finance. **29(1)** (2019), 329–367.

[3] Chen C., Iyengar G., Moallemi C.C., *An axiomatic approach to systemic risk*, Management Science. **59(6)** (2013), 1373–1388.

[4] Eichhorn A., Römisch W., *Polyhedral risk measures in stochastic programming*, SIAM Journal on Optimization. **16(1)** (2005), 69–95.

[5] Eisenberg L., Noe T.H., *Systemic risk in financial systems*, Management Science. **47(2)** (2001), 236–249.

[6] Feinstein Z., Rudloff B., Weber S., *Measures of systemic risk*, SIAM Journal on Financial Mathematics. **8(1)** (2017), 672–708.

[7] Kabanov Yu., Mokbel R., El Bitar Kh., *Clearing in financial networks*, Theory of Probability and Its Applications. **62(2)** (2017), 311–344.

[8] Nobakhtian S., Shafiei N., *A Benson type algorithm for nonconvex multiobjective programming problems*, TOP. **25(2)** (2017), 271–287.

[9] Pascoletti A., Serafini P., *Scalarizing vector optimization problems*, Journal of Global Optimization. **42(4)** (1984), 499–524.

[10] Rogers L.C.G., Veraart L.A.M., *Failure and rescue in an interbank network*, Management Science. **59(4)** (2013), 882–898.

Algorithms with Predictions

Michael Mitzenmacher
Harvard University
michaelm@eecs.harvard.edu

joint work with several co-authors

Abstract

In this talk, I will survey results in the emerging area of Algorithms with Predictions (aka learning augmented algorithms, or data driven algorithms), with a focus on my own work on filters and scheduling. These methods incorporate predictors – which can correspond naturally to machine learning oracles – to adapt the algorithms' behavior to the properties of the input distribution. This can consequently improve their performance, such as the runtime, space, or quality of the solution. Ideally, the goal is to offer provable guarantees of improved performance when the predictions are good, while maintaining nearly identical worst-case or other performance guarantees when they are not. The field has recently boomed, with applications to classical streaming algorithms, online scheduling, clustering, and many other problems.

Dynamic Demand-Aware Link Scheduling for Reconfigurable Datacenters

Lara Ost
University of Vienna
lara.ost@univie.ac.at

joint work with Kathrin Hanauer, Monika Henzinger, Stefan Schmid

Abstract

Emerging reconfigurable datacenters allow to dynamically adjust the network topology in a demand-aware manner. These datacenters rely on optical switches which can be reconfigured to provide direct connectivity between racks, in the form of edge-disjoint matchings. While state-of-the-art optical switches in principle support microsecond reconfigurations, the demand-aware topology optimization constitutes a bottleneck.

This talk presents a dynamic algorithms approach to improve the performance of reconfigurable datacenter networks, by supporting faster reactions to changes in the traffic demand. This approach leverages the temporal locality of traffic patterns in order to update the interconnecting matchings incrementally, rather than recomputing them from scratch. In particular, we present six (batch-)dynamic algorithms and compare them to static ones. We conduct an extensive empirical evaluation on 176 synthetic and 39 real-world traces, and find that dynamic algorithms can both significantly improve the running time and reduce the number of changes to the configuration, especially in networks with high temporal locality, while retaining matching weight.

Lower bounds for structured nonconvex quadratic optimization problems

Bo Peng
VGSCO, University of Vienna
bo.peng@univie.ac.at

Abstract

The Nonconvex quadratic optimization model plays a significant role in many real-world applications, such as classification trees and sparse portfolio selection. In this talk, we consider the global optimization of nonconvex(mixed-binary) quadratic optimization problems and present several convex relaxations. We compare their bounds in terms of quality and computational cost. In some special cases, we also show some relaxations are equivalent to the original problems.

Effects of neural network compression beyond accuracy

Alexandra Peste
Institute of Science and Technology Austria
alexandra.peste@ist.ac.at

joint work with Eugenia Iofinova, Dan Alistarh

Abstract

Progress over the last few decades enabled us to build increasingly large neural networks for various tasks, and to efficiently optimize them. We have also learned that it is frequently possible to remove, by setting to zero, a large proportion (80-99%) of the weights of the network without incurring a penalty in the network's in-distribution generalization performance. However, these sparse solutions do not completely resemble the dense ones in aspects beyond accuracy. In this work, we analyze two key aspects of sparse neural network performance. The first is the generalizability of sparse neural networks for visual recognition, as measured by the ability to transfer the learned weights to a new task. The second is the impact of sparsity on systematic and categorical bias of the resulting model. Our results further suggest best practices for addressing potential issues arising from both aspects, which we hope will help improve the use of sparse neural networks in real-world applications.

Conditional Gradients – an overview

Sebastian Pokutta
Zuse Institute Berlin and TU Berlin
pokutta@zib.de

Abstract

Conditional Gradient methods are an important class of methods to minimize smooth convex functions over polytopes. Recently these methods received a lot of attention as they allow for structured optimization and hence learning, incorporating the underlying polyhedral structure into solutions. In this talk I will give a broad overview of these methods, their applications, as well as present some recent results both in traditional optimization and learning as well as in deep learning.

Hop-Constrained Expander Decompositions, Oblivious Routing, and Distributed Universal Optimality

Harald Raecke
Technische Universität München
raecke@in.tum.de

This is joint work with Bernhard Häupler and Mohsen Ghaffari.

Abstract

In a recent result Häupler, Ghaffari, and Zuzic [STOC 2021] showed that so-called h-hop oblivious routing schemes with a polylogarithmic competitive ratio exist. These can be used to solve packet routing problems almost optimally, i.e., with only a polylogarithmic loss in routing time compared to a globally optimal solution. In this talk we present a different way of constructing h-hop oblivious routing schemes that have a weaker (sub-polynomial) competitive ratio but which can be constructed very efficiently by a distributed algorithm. This result has important consequences in the area of distributed computing: it gives novel CONGEST algorithms for a large class of important optimization problems, including minimum-spanning tree, $(1 + \epsilon)$-min-cut, $(1 + \epsilon)$-shortest paths. Our algorithms solve these problems in sub-polynomial rounds on any network, as long as a sub-polynomial-round distributed algorithm exists for this network.

Stochastic Optimization and Learning
with Risk-Averse Objectives

Andrzej Ruszczyński
Rutgers University
rusz@rutgers.edu

joint work with Darinka Dentcheva, Mert Gürbüzbalaban, and Landi Zhu

Abstract

We shall discuss optimization and learning problems where the objective is to optimize a risk measure of a random performance (loss) function. Problems of this type involve adversarial learning, distributionally robust learning, and risk-averse reinforcement learning. A major difficulty in such settings is the problem of obtaining statistical estimates of subgradients of such a composition. We shall present specialized algorithms for selected measures of risk and study their convergence and rate of convergence for a broad class of nonsmooth and non-convex loss functions. Next, we shall introduce a new class of risk measures, called mini-batch risk forms, where the reference probability measure is an empirical measure and post-composition with expectation is employed. We shall derive their dual representation, study their pre-compositions with nonsmooth and nonconvex functions, and derive unbiased stochastic subgradients of such compositions. Finally, we study the dependence of mini-batch risk forms on perturbation of the probability measure and establish quantitative stability in terms of optimal transport metrics. We obtain finite-sample expected error estimates for mini-batch risk forms involving functions on a finite-dimensional space.

References

[1] D. Dentchava and A. Ruszczyński, *Mini-batch risk forms*, SIAM J. on Optimization. (2023), in press.

[2] M. Gürbüzbalaban, A. Ruszczyński, and L. Zhu, *A stochastic subgradient method for distributionally robust non-convex and non-smooth learning*, Journal of Optimization Theory and Applications. (19) (2022) 1014–1041.

Fundamental Limits of Two-layer Autoencoders, and Achieving Them with Gradient Methods

Alexander Shevchenko
Institute of Science and Technology Austria
alex.shevchenko@ist.ac.at

joint work with Kevin Kögler (ISTA), Hamed Hassani (University of Pennsylvania) and Marco Mondelli (ISTA)

Abstract

Autoencoders are a popular model in many branches of machine learning and lossy data compression. However, their fundamental limits, the performance of gradient methods and the features learnt during optimization remain poorly understood, even in the two-layer setting. In fact, earlier work has considered either linear autoencoders or specific training regimes (leading to vanishing or diverging compression rates). Our paper addresses this gap by focusing on non-linear two-layer autoencoders trained in the challenging proportional regime in which the input dimension scales linearly with the size of the representation. Our results characterize the minimizers of the population risk, and show that such minimizers are achieved by gradient methods; their structure is also unveiled, thus leading to a concise description of the features obtained via training. For the special case of a sign activation function, our analysis establishes the fundamental limits for the lossy compression of Gaussian sources via (shallow) autoencoders. Finally, while the results are proved for Gaussian data, numerical simulations on standard datasets display the universality of the theoretical predictions.

Machine Learning for Optimization

Pascal Van Hentenryck
NSF AI Institute for Advances in Optimization,
Georgia Institute of Technology, Atlanta, Georgia, USA
Email: pvh@gatech.edu

Abstract

This presentation reviews a number of advances at the intersection of machine learning and optimization, demonstrating that the fusion of these two methodologies open new avenues for decision making. It reviews the concepts of optimization proxies and learning to optimize, as well as the new methodologies that include just-in-time learning, self-supervised optimization learning, and compact learning. The methodologies are highlighted on applications in power systems, logistics, and supply chains. They highlight how both machine learning and optimization benefit from each other to accomplish results that they could achieve independently.

Online Min-Max Paging

Maximilian Vötsch
University of Vienna
maximilian.voetsch@univie.ac.at

joint work with Ashish Chiplunkar, Monika Henzinger, and Sagar Sudhir Kale

Abstract

Motivated by fairness requirements in communication networks, we introduce a natural variant of the online paging problem, called *min-max* paging, where the objective is to minimize the maximum number of faults on any page. While the classical paging problem, whose objective is to minimize the total number of faults, admits k-competitive deterministic and $O(\log k)$-competitive randomized algorithms, we show that min-max paging does not admit a $c(k)$-competitive algorithm for any function c. Specifically, we prove that the randomized competitive ratio of min-max paging is $\Omega(\log(n))$ and its deterministic competitive ratio is $\Omega(k \log(n)/\log(k))$, where n is the total number of pages ever requested. We design a fractional algorithm for paging with a more general objective — minimize the value of an n-variate differentiable convex function applied to the vector of the number of faults on each page. This gives an $O(\log(n) \log(k))$-competitive fractional algorithm for min-max paging. We show how to round such a fractional algorithm with at most a k factor loss in the competitive ratio, resulting in a deterministic $O(k \log(n) \log(k))$-competitive algorithm for min-max paging. This matches our lower bound modulo a $poly(\log(k))$ factor. We also give a randomized rounding algorithm that results in a $O(\log^2 n \log(k))$-competitive algorithm.

On Nonsmooth Optimization Based on Abs-Linearization

Andrea Walther
Humboldt-Universität zu Berlin
andrea.walther@math.hu-berlin.de

joint work with Sabrina Fiege, Andreas Griewank, and Timo Kreimeier

Abstract

For a so-called abs-smooth function, the concept of abs-linearization allows the generation of a piecewise linear local model that is provable of second order. Similar to the quadratic model generated by a truncated Taylor series in the smooth situation, this piecewise linear model can be used as a building block for optimization algorithms targeting nonsmooth problems of different kinds. In this talk, first we define abs-smooth functions covering a wide range of applications like clustering, image restoration, and robust gas network optimization. Also various mathematical models like complementarity problems or bilevel optimization tasks can be formulated as abs-smooth functions. Subsequently, the abs-linearization approach and the properties of the resulting local model will be illustrated. Then, we discuss an approach to solve piecewise linear optimization problems. Exploiting this methodology, we finally present an algorithm for the solution of abs-smooth functions including convergence results and numerical examples.

Hybrid methods in Semidefinite programming

Jeferson Zapata
ISTA
jzapata@ista.ac.at

Vladimir Kolmogorov, Simone Naldi

Abstract

Semidefinite programming (SDP) is a powerful generalization of Linear Programming (LP). Interior Point Methods (IPM) are the main methods to solve SDPs: Take a point in the interior of set of solutions and track more "accurate solutions" along a *central path.* But problems may arise when the interior of the feasible region is empty (weakly feasible problems [2]).

In [1], the authors describe a *symbolic algorithm* for solving SDPs in Linear Matrix inequality representation. Using this approach, it is possible to solve degenerate instances of SDPs under certain genericity conditions in the problem data. Our goal is to study and exploit the structures of (weakly feasible) SDP problems to find representations of exact solutions to such problems by combining numerical algorithms and symbolic computation.

References

[1] Henrion, D., Naldi, S. and El Din, M.S., 2016. Exact algorithms for linear matrix inequalities. *SIAM Journal on Optimization*, 26(4), pp.2512-2539.

[2] Pataki, G., 2017. Bad semidefinite programs: they all look the same. *SIAM Journal on Optimization*, 27(1), pp.146-172.