Workshop on

Algorithmic Optimization and Data Science

July 20—July 22, 2022

Organizers:
Ralph Bergmann
Nicole Marheineke
Ralf Münnich
Achim Rettinger
Martin Schmidt
Volker Schulz

Trier University
Universitätsring 15
54296 Trier

www.alop.uni-trier.de

Workshop on Algorithmic Optimization and Data Science

20.7.-22.7.2022

Trier University, Lecture Hall HS 9 (E Building)

Program

Wednesday: 20. July 2022

8:00		*Registration*	*Foyer, E Building*
8:50 – 9:00	Welcome		
9:00 – 9:30	*Clemens Zeile*: A Gauss-Newton-Based Decomposition Algorithm for Nonlinear Mixed-Integer Optimal Control Problems		
9:30 – 10:00	*Jan Sokolowski*: Robust Parameter Estimation for Dynamical Systems with Convolutional Neural Networks		
10:00 – 10:30	*Lena Sembach*: Riemannian Optimization for Variance Estimation in Linear Mixed Models		
10:30 – 11:00		*Coffee Break*	*Room E 51*
11:00 – 11:30	*Lukas Mich*: Neural Networks Approximation for Stochastic Optimal Control Problems in the Context of Optimal Investment		
11:30 – 12:00	*Ivan Oseledets*: Geometry, Tensors and Optimization		
12:00 bis 14:00		*Lunch*	*Mensa*
14:00 – 15:00	**Invited Presentation: *Sebastian Pokutta*** **Structured ML Training via Conditional Gradients**		
15:00 – 15:30		*Coffee Break*	*Room E 51*
15:30 – 16:00	*Carina Moreira Costa*: Mixed-Integer Programming Techniques for the Minimum Sum-of-Squares Clustering Problem		
16:00 – 16:30	*Benjamin Schwendinger*: Holistic Generalized Linear Models		
16:30 – 17:00	*René Pinnau*: Optimal Control of Free Boundary Problems		
17:00 – 17:30	*Cristina Cipriani*: A Mean-Field Optimal Control Approach to the Training of NeurODEs		
17:30		*Welcome Reception/Meet and Greet*	*Room E 51*

Thursday: 21 July 2022

8:30		*Late Registration*	*Foyer, E Building*
9:00 – 9:30	*Nadja Vater*: A Preconditioner for Least Squares Problems with Application to Neural Network Training		
9:30 – 10:00	*Sebastian Hofmann*: An Optimal Control Based Supervised Learning Algorithm for Training Runge-Kutta Structured Neural Networks		
10:00 – 10:30		*Coffee Break*	*Room E 51*
10:30 – 11:30	**Invited Presentation: *Gitta Kutyniok*** **The World of Deep Learning: From the Mystery of Training to Foundational Limitations**		
11:30 bis 13:30		*Lunch*	*Mensa*
13:30 – 14:00	*Maximilian Würschmidt*: Convergence Rates for a Deep Learning Algorithm for Semilinear PDE's		
14:00 – 14:30	*Jens Püttschneider*: Turnpikes in Deep Learning		
14:30 – 15:00		*Coffee Break*	*Room E 51*
15:00 – 16:00	**Invited Presentation: *Enrique Zuazua*** **Control and Machine Learning**		
17:00		*Guided City Tour* *Meet at Tourist Info, Porta Nigra*	
19:00		*Wine tasting followed by Conference Dinner* *„Wirtshaus zur Glocke" in Trier*	

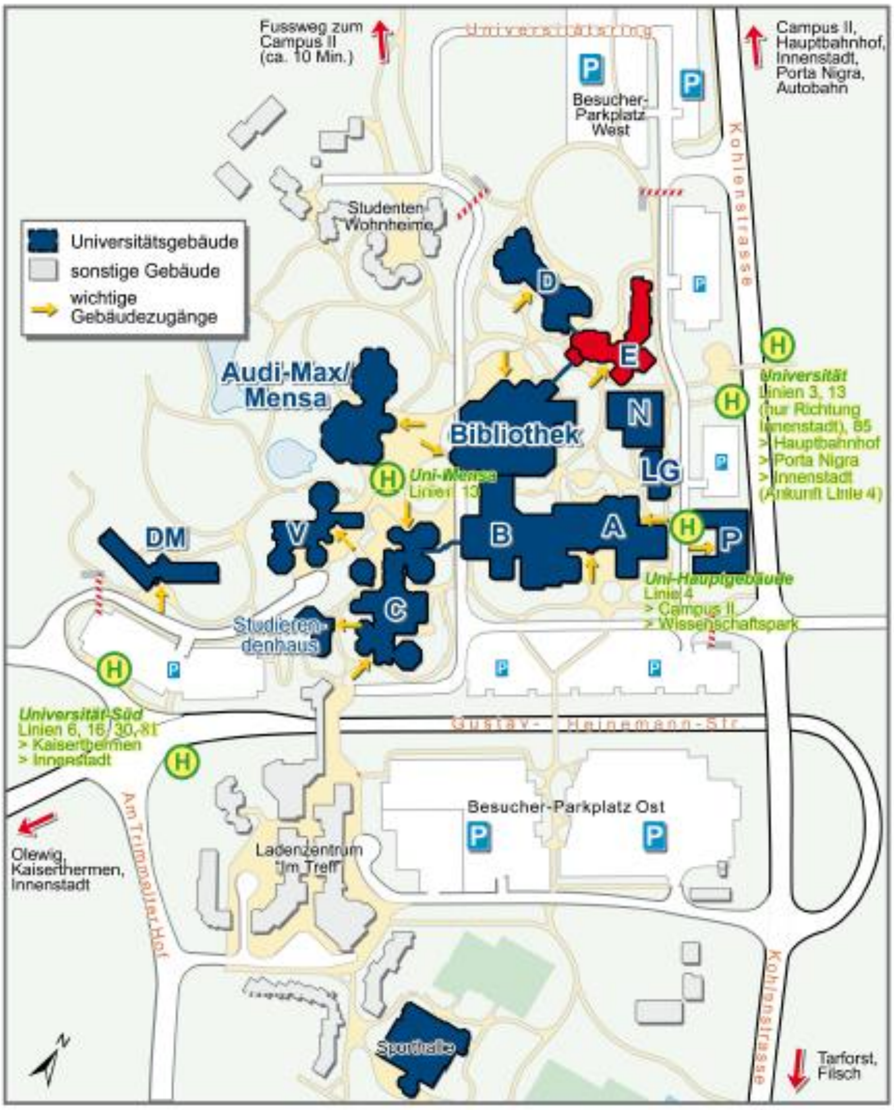
Friday: 22 July 2022

8:30 – 9:00	*Konstantin Riedl*: Consensus-Based Optimization: Theory and Application
9:00 – 9:30	*Ayush Tiwari*: Data Science and Optimization in Scale: A Study of Changing Landscape
9:30 – 10:00	*Moritz Link*: Multiobjective Optimization of Decentralized Energy Supply Networks
10:00 – 10:30	*Coffee Break* *Room E 51*
10:30 – 11:00	*Alfio Borzì*: The Sequential Quadratic Hamiltonian Algorithm in Action: PDE Optimal Control and Inverse Problems
11:00 – 11:30	*Jan Bartsch*: A Stochastic Gradient Approach for the Calibration of Stochastic Nonlinear Bath Models
11:30 - 13:30	*Lunch* *Mensa*
13.30 – 14:30	**Invited Presentation: *Suvrit Sra*** **Some Surprising Gaps between Optimization Theory and Practice in Machine Learning**
14:30 – 15:00	Closing Comments

The workshop will take place in Building E on the Trier University Campus I.

Please refer to the map below for a detailed Campus layout.

For directions to the University, please refer to the [website]()



We gratefully acknowledge the support of

Deutsche Forschungsgemeinschaft

A stochastic gradient approach for the calibration of stochastic nonlinear bath models

J. Bartsch

University Konstanz; jan.bartsch@uni-konstanz.de

Nonlinear bath models, as for example the Stochastic Prantl-Tomlison model, are key tools to study properties of complex fluids [1, 2]. While simple linear models have been widely investigated in the Markovian framework, less is known for the non-linear case in which the Markovian framework is no longer applicable. To study the properties of fluids, systems of coupled nonlinear stochastic differential equations (SDEs) representing the movement of particles are investigated. The SDEs typically possess the structure of nonlinear Langevin equations. The bath models contain several parameters that have to be chosen carefully to match the experimental data and validate the effectiveness of the model.
In this talk, a discretize-before-optimize approach similar to the one in [3] is presented in order to calibrate a model with two or more particles. For this purpose, an optimality system characterizing the solution of the calibration problem is derived within the Lagrange framework and used to formulate a stochastic gradient method.
Monte Carlo methods are applied to solve the arising equations and due to long computational times caused by a high equilibration time, parallelization strategies are exploited. The model and the optimization strategy are tested and validated using data that are generated in physical experiments.

References

[1] Rohit Jain, Félix Ginot, Johannes Berner, Clemens Bechinger, and Matthias Krüger. Two step micro-rheological behavior in a viscoelastic fluid. *The Journal of Chemical Physics*, 154(18):184904, 2021.

[2] Rohit Jain, Félix Ginot, and Matthias Krüger. Micro-rheology of a particle in a nonlinear bath: Stochastic Prandtl-Tomlinson model. *Physics of Fluids*, 33(10):103101, October 2021.

[3] C. Kaebe, J. H. Maruhn, and E. W. Sachs. Adjoint-based Monte Carlo calibration of financial market models. *Finance Stoch.*, 13(3):351–379, 2009.

The Sequential Quadratic Hamiltonian Algorithm in Action: PDE Optimal Control and Inverse Problems

A. Borzì *

* Institut für Mathematik, Universität Würzburg, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany;
alfio.borzi@mathematik.uni-wuerzburg.de [1]

A sequential quadratic hamiltonian (SQH) algorithm for solving different classes of non-smooth and non-convex optimization problems governed by partial differential equations (PDEs) is presented.
The novel SQH algorithm belongs to the class of iterative schemes known as the method of successive approximations that are based on the characterization of optimality in control problems by the Pontryagin maximum principle (PMP); see, e.g., [1, 2, 3, 4].
In this talk, theoretical and numerical results are presented that demonstrate the ability of the SQH scheme in solving different optimal control and inverse problems. In particular, the case of PDE control problems with nonlinear control mechanisms, discontinuous costs of the controls, and a mixed-integer setting are discussed. The illustration of the application of the SQH algorithm to solve a diffusion identification problem concludes this talk.

References

[1] T. Breitenbach and A. Borzì. *A sequential quadratic Hamiltonian method for solving parabolic optimal control problems with discontinuous cost functionals*. *Journal of Dynamical and Control Systems*, 25:403–435, 2019.

[2] T. Breitenbach and A. Borzì. *On the SQH scheme to solve nonsmooth PDE optimal control problems*. *Numerical Functional Analysis and Optimization*, 40:1489–1531, 2019.

[3] T. Breitenbach and A. Borzì. *A sequential quadratic Hamiltonian scheme for solving non-smooth quantum control problems with sparsity*. *Journal of Computational and Applied Mathematics*, 369:112583, 2020.

[4] T. Breitenbach and A. Borzì. *The Pontryagin maximum principle for solving Fokker–Planck optimal control problems*. *Computational Optimization and Applications*, 76:499–533, 2020.

[1] Partially supported by the BMBF-Project iDeLIVER : Intelligent MR Diagnosis of the Liver by Linking Model and Data-Driven Processes.

A Mean-Field Optimal Control Approach to the Training of NeurODEs

Benoît Bonnet[1], Cristina Cipriani*[2], Massimo Fornasier[3] and Hui Huang[4]

[1] LAAS - CNRS, Toulouse, France; benoit.bonnet@laas.fr
[2] Technical University Munich, Department of Mathematics, Munich, Germany; cristina.cipriani@ma.tum.de
[3] Technical University Munich, Department of Mathematics, Munich, Germany; massimo.fornasier@ma.tum.de
[4] University of Calgary, Department of Mathematics and Statistics, Calgary, Canada; hui.huang1@ucalgary.ca

Inspired by the pioneering works [1] and [2], we investigate particular types of neural networks called NeurODEs and present a reformulation of their training. These networks contain residual block regularisations, proposed in [7] for stability reasons, that have an important mathematical advantage: the training of NeurODEs can be interpreted as a stochastic optimal control problem. Starting from this interpretation and extending it to its mean-field version, in [4] we derive first order optimality conditions in the form of a mean-field version of the Pontryagin Maximum Principle. Our result is based on a novel and generalized version of the Lagrange multiplier theorem on convex sets of spaces of measures, combined with a more intrinsic viewpoint developed in [3, 5, 6] based on geometric considerations on Wasserstein spaces. Our approach provides a unique control solution, which is also Lipschitz continuous. Finally, some explanatory and easy-to-read numerical examples in low dimension give insights into the resulting algorithm which provides an alternative way of performing the training of these neural networks.

References

[1] W. E, J. Han and Q. Li. Mean-Field Optimal Control Formulation of Deep Learning. *Research in the Mathematical Sciences* **6** (10), 2019.

[2] E. Haber and L. Ruthotto. Stable architectures for deep neural networks. *Inverse Problems* **34** (2017dec), no. 1, 014004.

[3] B. Bonnet. A Pontryagin Maximum Principle in Wasserstein Spaces for Constrained Optimal Control Problems. *ESAIM COCV* **25** (52), 2019.

[4] B. Bonnet, C. Cipriani, M. Fornasier and H. Huang. A Measure Theoretical Approach to the Mean-field Maximum Principle for Training NeurODEs. *arxiv preprint arXiv:2107.08707*, 2022.

[5] B. Bonnet and H. Frankowska. Necessary Optimality Conditions for Optimal Control Problems in Wasserstein Spaces. *Applied Mathematics and Optimization* **84** 1281-1330, 2021.

[6] B. Bonnet and F. Rossi. The Pontryagin Maximum Principle in the Wasserstein Space. *Calculus of Variations and Partial Differential Equations***58** 11, 2019.

[7] He, X. Zhang, S. Ren and S. Jian. Deep Residual Learning for Image Recognition. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 770-778, 2016.

Mixed-Integer Programming Techniques for the Minimum Sum-of-Squares Clustering Problem

J.P. Burgard[1], C.M. Costa*[2], C. Hojny[3], T. Kleinert[4], and M. Schmidt[2]

[1] Trier University, Department of Economic and Social Statistics, Universitätsring 15, 54296 Trier, Germany; burgardj@uni-trier.de

[2] Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany; costa@uni-trier.de, martin.schmidt@uni-trier.de

[3] Eindhoven University of Technology, Department of Mathematics and Computer Science, Combinatorial Optimization Group, P.O. Box 513, 5600 MB Eindhoven, The Netherlands; c.hojny@tue.nl

[4] Horváth Steering Lab, Ganghoferstrae 39, 80339 München, Germany; TKleinert@horvath-partners.com

The minimum sum-of-squares clustering problem is a very important problem in data mining and machine learning with very many applications in, e.g., medicine or social sciences. However, it is known to be NP-hard in all relevant cases and to be notoriously hard to be solved to global optimality in practice. We develop and test different tailored mixed-integer programming techniques to improve the performance of state-of-the-art MINLP solvers when applied to the problem—among them are cutting planes, propagation techniques, branching rules, or primal heuristics. Our extensive numerical study shows that our techniques significantly improve the performance of the open-source MINLP solver SCIP. We now solve many instances that are not solvable without our techniques and we obtain much smaller gaps for those instances that can still not be solved to global optimality.

An optimal control based supervised learning algorithm for training Runge-Kutta structured neural networks [1]

Sebastian Hofmann *

* Institut für Mathematik, Universität Würzburg, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany;
sebastian.hofmann@mathematik.uni-wuerzburg.de

The dynamical systems approach to machine learning is applied to a supervised learning (SL) problem for residual networks with Runge-Kutta (RK) structure. In this context, the SL problem is formulated as an optimal control problem, where the cost functional is minimized subject to a nonlinear dynamical equation representing the neural network (NN) with its weights playing the role of the controls. Within this framework, the optimality of the NNs weights is characterized by the discrete Pontryagin maximum principle (PMP) that also provides the foundation of the to be presented, sequential quadratic Hamiltonian (SQH) training method for RK-NNs [1]. The SQH method represents a novel extension to the so-called successive approximation methods and has its strength in a dynamic hyperparameter-adaptation technique that guarantees the decrease of the objective functional's value in every iterations, independent of the initial hyperparameters choice. Moreover, a stopping condition serves as a control mechanism for the accuracy of the networks approximation and allows to stop training at an iteration that is optimal in the sense of the discrete PMP.
Convergence and stability results for the SQH scheme are provided in the framework of training residual neural networks with RK structure. Moreover, it is shown numerically that inter alia due to the integrated, dynamic hyperparameter-adaptation technique, the efficiency and robustness of the training process of the SQH scheme is superior to a comparable algorithm, called extended method of successive approximation [2].

References

[1] S.Hofmann and A.Borzì. *A Sequential Quadratic Hamiltonian Algorithm for Training Explicit RK Neural Networks*, Journal of Computational and Applied Mathematics 405 (2022) 113943.

[2] Q.Li, L.Chen, C.Tai and W.E . *Maximum Principle Based Algorithms for Deep Learning*, Journal of Machine Learning Research 18 (2018) 1-29.

[1]Partially supported by the BMBF-Project iDeLIVER : Intelligent MR Diagnosis of the Liver by Linking Model and Data-Driven Processes.

The World of Deep Learning: From the Mystery of Training to Foundational Limitations

Prof. Dr. Gitta Kutyniok

Bavarian AI Chair for Mathematical Foundations of Artificial Intelligence, LMU Munich
Adjunct Professor for Machine Learning, University of Tromsø

Ludwig-Maximilians-Universität München
Mathematisches Institut

Despite the outstanding success of deep neural networks in real-world applications, ranging from science to public life, most of the related research is empirically driven and a comprehensive mathematical foundation is still missing. At the same time, these methods have already shown their impressive potential in mathematical research areas such as imaging sciences, inverse problems, or numerical analysis of partial differential equations, sometimes by far outperforming classical mathematical approaches for particular problem classes. The goal of this lecture is to first provide an introduction into this new vibrant research area. We will then delve more deeply into the training process, discuss its mysteries and also showcase different approaches. Finally, we will discuss fundamental limitations of deep neural networks in terms of computability (of the optimization problem), which seriously affects their reliability.

Multiobjective Optimization of Decentralized Energy Supply Networks

Moritz Link and Stefan Volkwein

In view of the climate change, energy supply network planning can no longer focus only on the economic objective, but also on other criteria such as minimizing the carbon emissions. This leads to multiobjective optimization problems for decentralized energy supply networks. Moreover, taking into account stationary energy flow equations together with modeling practical decisions in the network yields a mixed-integer nonlinear structure of the resulting problems. In this talk, we present first results towards tackling real-world network instances.

Neural Network Approximation for Stochastic Optimal Control Problems in the Context of Optimal Investment

Lukas Mich*[1] and Frank Seifried[2]

[1] University of Trier, Department IV – Mathematics, Germany; mich@uni-trier.de

[2] University of Trier, Department IV – Mathematics, Germany; seifried@uni-trier.de

Continuous Stochastic Optimal Control Problems (SOCPs) are the key for modelling and analyzing prob-lems of optimal investment in finance. Solving these problems traditionally involves the solution of highly non-linear PDEs which – in most cases – are not available in closed form. On the other hand, numerical resolution methods of these PDEs are numerically expensive in higher dimensions (curse of dimensionality). We take the quite recent idea of using neural networks as approximation tool for the op-timal control policy and prove convergence of the time-discretized approximated SOCP (DA-SOCP) to the original continuous formulation. The DA-SOCP then is a finite-dimensional problem and is (in con-trast to the C-SOCP) itself numerically tractable. This eliminates the necessity of proving equivalence of the solutions of the PDE and the C-SOCP.

We apply this approach to an optimal investment problem where the optimal control is both time- and state-dependent and numerically solve the DA-SOCP formulation with a variable sample size gradient descent (VSS-GD) method. We introduce appropriate error measures and compare the numerical results to the (existing) closed form solution of the problem.

Geometry, tensors and optimization

Ivan Oseledets

Humboldt Fellow, ITWM Kaiserslautern, Germany

In this talk, I will cover topics. First, I will discuss training of generative adversarial models (GANs) and analysis of local convergence of standard gradient ascent-descent method in the functional space. It can be shown that local dynamics is completely determined by the eigenvalues of the weighted Laplace operator.
In the second part, I will give an overview for optimization on low-rank tensor manifolds. Such problems appear as subproblems in many data analysis tasks. The Riemannian optimization is one of the most promising tools, and many algorithmic tasks need to be solved, such as automatic differentiation and computation of Riemmanian gradient, efficient preconditioning of the iterative process and many others.

Optimal Control of Free Boundary Problems

René Pinnau

Technische Universität Kaiserslautern, Department of Mathematics
Kaiserslautern, Germany
Email: pinnau@mathematik.uni-kl.de

In this talk we present FBPs in different areas of application, e.g. filter production, and the corresponding optimal control problems. These include the production of the fibre material as well as the control of the filtration process. The problems are analysed and we use the adjoint variables to derive the derivative information which is needed for the respective numerical solution. Further, we discuss possible relaxations and approximations.

Structured ML Training via Conditional Gradients

Prof. Dr. Sebastian Pokutta

Vice President and Division Head
Mathematical Algorithmic Intelligence
AI in Society, Science, and Technology (AIS²T)
Zuse Institute Berlin (ZIB) Professor for Optimization and Machine Learning

Institute of Mathematics
Electrical Engineering and Computer Science (courtesy)
Technische Universität Berlin

Conditional Gradient methods are an important class of methods to minimize (non-)smooth convex functions over (combinatorial) 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.

Turnpikes in Deep Learning

J. Püttschneider *[1] and T. Faulwasser[2]

[1] TU Dortmund University; jens.puettschneider@tu-dortmund.de

[2] TU Dortmund University; timm.faulwasser@tu-dortmund.de

This talk investigates the turnpike phenomena in optimal control approaches to deep learning, i.e. the training of neural network [1, 2]. To this end, we formulate the training of residual neural networks as an optimal control problem and investigate its dissipaitivity properties [3] when introducing a stage cost based on the loss function. This was already done for quadratic losses [2, 4] and in this talk we consider the cross-entropy, the classical loss used in classification t a sks. T he c onsidered t urnpikes a re verified for the MNIST benchmark problem and they are shown to generalize to test data. Based on this, those layer that constitute the turnpike can be pruned since they do not contribute to the transformation learned. This results in an network with the minimum number of layers required for learning without the need for extensive hyperparameter tuning. Moreover, the training with and without a stage cost is compared.

References

[1] Q. Li, L. Chen, C. Tai, and E. Weinan. *Maximum principle based algorithms for deep learning.* In *The Journal of Machine Learning Research*, vol. 18, 2017.

[2] C. Esteve Yagüe, B. Geshkovski, D. Pighin, and E. Zuazua. *Large-time asymptotics in deep learning. arXiv preprint:2008.02491*, 2020.

[3] D. Angeli, R. Amrit, and J. Rawlings. *On Average Performance and Stability of Economic Model Predictive Control.* In *EEE Transactions on Automatic Control*, vol. 57, 2012.

[4] T. Faulwasser, A.-J. Hempel, S. Streif. *On the Turnpike to Design of Deep Neural Nets: Explicit Depth Bounds. arXiv preprint:2101.03000*, 2021.

Consensus-Based Optimization: Theory and Applications

Massimo Fornasier[1], Timo Klock[2], and Konstantin Riedl[3]

[1] Technical University of Munich, Department of Mathematics, Munich, Germany; massimo.fornasier@ma.tum.de
[2] Simula Research Laboratory, Department of Numerical Analysis and Scientific Computing, Oslo, Norway; timo@simula.no
[3] Technical University of Munich, Department of Mathematics, Munich, Germany; konstantin.riedl@ma.tum.de

Consensus-based optimization (CBO) is a multi-agent metaheuristic derivative-free optimization method that can find global minimizers of nonconvex nonsmooth objective functions. In this talk we shed a light on the internal mechanisms of CBO, which are responsible for its success. Based on an experimentally supported intuition that, in the mean-field limit, CBO always performs a gradient descent of the squared Euclidean distance to the global minimizer, we derive a novel technique for proving global convergence in mean-field law for a rich class of objective functions. In particular, we prove that CBO performs a convexification of a very large class of optimization problems as the number of optimizing agents tends to infinity. From the result of CBO convergence in mean-field law it becomes apparent that the hardness of any global optimization problem is necessarily encoded in the mean-field approximation, or, more precisely, in the way how the empirical measure of the finite particle dynamics is used to approximate the mean-field limit. In consideration of the central significance of such approximation with regards to the overall computational complexity of the implemented numerical scheme we establish a novel probabilistic quantitative result about the convergence of the interacting particle system towards the corresponding mean-field dynamics. By combining both results we provide the first holistic convergence proof of CBO methods on the plane. To demonstrate the practicability of CBO we moreover provide experiments for both a well-understood high-dimensional machine learning problem and the phase retrieval problem in signal processing. Towards the end we discuss extensions of the theoretical analysis to other metaheuristics such as particle swarm optimization as well as the potential of CBO when it comes to more challenging machine learning tasks, like reinforcement learning.

References

[1] M. Fornasier, T. Klock, and K. Riedl. Consensus-Based Optimization Methods Converge Globally. *arXiv:2103.15130* (2022).

[2] M. Fornasier, T. Klock, and K. Riedl. Convergence of Anisotropic Consensus-Based Optimization in Mean-Field Law. In J. L. Jiménez Laredo, J. I. Hidalgo, and K. O. Babaagba, editors, *Applications of Evolutionary Computation*, pages 738–754, Cham, 2022. Springer International Publishing.

Holistic Generalized Linear Models

Benjamin Schwendinger*[1] and Florian Schwendinger[2] and Laura Vana[1]

[1] TU Wien, Austria; benjamin.schwendinger@tuwien.ac.at
[2] University of Klagenfurt, Austria

Selecting a sensible model from the set of all possible models is an important but typically time-consuming task in the data analytic process. To simplify this process, Bertsimas et al. [1, 2] introduce the holistic linear model (HLM). The HLM is a constrained linear regression model in which the constraints aim to automate the model selection process. In particular, this can be achieved by utilizing quadratic mixed-integer optimization, where the integer constraints are used to place cardinality constraints on the linear regression model. Placing a cardinality constraint on the total number of variables allowed in the final model leads to the classical best subset selection problem [3].

$$\underset{\beta}{\text{minimize}} \quad \frac{1}{2}||y - X\beta||_2^2 \ \text{ subject to } \ \sum_{i=1}^{p} \mathbb{I}_{\{\beta_i \neq 0\}} \leq k \tag{1}$$

Further cardinality constraints placed on user-defined groups of predictors can be used to, e.g., limit the pairwise multicollinearity, select the best (non-linear) transformation for the predictors, include expert knowledge in the model estimation (such as forcing specific predictors to stay in the model) or ensure that certain sets of predictors are jointly included (excluded) in (from) the model.

In this talk, we present the **holiglm** package, a package for formulating and fitting holistic generalized linear models (HGLMs). To our knowledge, we are the first to suggest using conic optimization to extend the results presented for linear regression by [1, 2] to the class of generalized linear models. Moreover, we extend this framework by several additional constraints.
The **holiglm** package provides a flexible infrastructure for automatically translating constrained generalized linear models into conic optimization problems. The optimization problems are solved by utilizing the R optimization infrastructure package **ROI** [4]. Using **ROI** makes it possible that the user can choose between a wide range of commercial and open-source optimization solvers. Additionally, a high-level interface is provided, which can be used as a drop-in replacement for the `stats::glm()` function.
Using conic optimization instead of iteratively reweighted least squares (IRLS) has the advantages that no starting values are needed, the results are more reliable (optimality certificates) and the solvers are designed to handle constraints. These advantages come at the cost of a longer runtime. However, as shown by [5] for some GLMs, the speed of the conic formulation is similar to the IRLS implementation.

References

[1] Bertsimas, D. & King, A. *An Algorithmic Approach to Linear Regression*. Operations Research. 64, 2-16 (2015)

[2] Bertsimas, D. & Li, M. *Scalable Holistic Linear Regression*. Operations Research Letters. 48, 203-208 (2020)

[3] Miller, A. *Subset Selection in Regression*. (CRC Press, 2002, 4, 15)

[4] Theußl, S., Schwendinger, F. & Hornik, K. ***ROI**: An Extensible R Optimization Infrastructure*. Journal Of Statistical Software. 94, (2020)

[5] Schwendinger, F., Grün, B. & Hornik, K. *A Comparison of Optimization Solvers for Log Binomial Regression Including Conic Programming*. Computational Statistics. 36, 1721-1754 (2021)

Riemannian Optimization for Variance Estimation in Linear Mixed Models

Lena Sembach*[1], Jan Pablo Burgard[1] and Volker H.Schulz[1]

[1] Universität Trier, Universitaetsring 15, 54296 Trier; {sembach, burgardj, volker.schulz}@uni-trier.de

Linear Mixed Models are widely spread in statistics for modeling clustered data [1]. The main assumption is that the observed data is dependent of latent information which is not observable but captures an unneglectable part of the overall variance of the model. From an optimization perspective, the task lies in finding optimal estimates for the covariance matrices of the hidden data and the overall data. A common approach is maximizing the likelihood function (ML) or the restricted likelihood function (REML). A challenge for standard optimizers is the positive-definiteness constraint of the covariance matrices. In this talk, we discuss a Riemannian approach to variance estimation in linear mixed models on the manifold of positive definite matrices. We compare our method with state of the art optimizers and show numerical results.

References

[1] Gumedze, FN and Dunne, TT. *Parameter estimation and inference in the linear mixed model.* Linear algebra and its applications 435.8 (2011): 1920-1944.

Robust Parameter Estimation for Dynamical Systems with Convolutional Neural Networks

Jan Sokolowski

1, rue du Campus
L-7795 Bissen, Luxembourg
jan.sokolowski@iee.lu

Parameter estimation for safety-critical systems in automotive industry is of immense interest. Common applications, as airbag functions, need to be automatically enabled to prevent children from being seriously injured in case of a crash. Therefore, smart systems can support mass-estimation of vehicle occupants to make applications more robust against environmental disturbances. We want to analyse parameter estimation problems for vehicle systems to compute appropriate values of mass-dependent spring - and damping - coefficients of a Quarter-Car-Model, which serves as an approximation of a more complex car model. Therefore, we numerically solve a randomly perturbed system of second order ordinary differential equations to generate a large amount of parameter-dependent realizations, serving as data for a Convolutional Neural Network. We investigate the estimation quality of pure data-driven models, as well as Hybrid Models, with respect to robustness against measurement noises and incompleteness of the data. For the latter, we develop a Hybrid Gauss-Newton method, using a set of pre-trained Convolutional Neural Networks, to estimate identifiable parameters of the complete Quarter-Car-Model. It is further shown, that such Hybrid Models, that make use of prior knowledge about the underlying data structure, give accurate estimates and are less volatile to corrupted data for a non-linear least-squares problem.

Some Surprising Gaps between Optimization Theory and Practice in Machine Learning

Prof. Dr. Suvrit Sra[1]
Haochuan Li[1], Hongzhou Lin[2], Ali Jadbabaie[1], Stefanie Jegelka[1], Jingzhao Zhang[3]

[1]**MIT Institute for Data, Systems and Society**, [2]Amazon, USA, [3]Tsinghua University, Beijing

It is well-known that there are large gaps between optimization theory and machine learning practice. However, there are two even more surprising gaps that have persisted at the fundamental level. The first one arises from ignoring the elephant in the room: non-differentiable non-convex optimization, e.g., when training a deep ReLU network. The second surprise is more disturbing: it uncovers a non-convergence phenomenon in the training of deep networks, and as a result it challenges existing convergence theory and training algorithms. Both these fundamental surprises open new directions of research, and I will talk about some of our theoretical progress on these, as well as potential research questions.

Data Science and Optimization in Scale: A Study of Changing Landscape

Ayush Tiwari

Data Scientist
IIT Kanpur

Data Science is one of the fastest-changing fields of study in the age of today's research. Advances in high-performance computer architectures and the way algorithms can take advantage of them, have been transformative for a variety of data science tasks. Researchers from across the world are working collaboratively under the shared goal to shape the future of computation for data science. With the advent of Machine Learning, Artificial Intelligence, and Cloud technologies it has taken an even giant leap. Optimization being at the core of any research in Data Science involves complex algorithms. As data science continues to grow as an industry and research sector, data-driven algorithms such as those required by deep learning multi-level networks that gradually identify things at higher levels of detail take up an increasing amount of valuable time and energy in data centers. This provokes a need to rethink how the technical challenges caused by this emerging new science are managed. The significant industry, government, and academic demand for data science skills create a supply problem, with the world facing a major skills gap that could inhibit the anticipated potential of data science and AI for our economy and society.

A Preconditioner for Least Squares Problems with Application to Neural Network Training

N. Vater

Institut für Mathematik, Universität Würzburg, Emil-Fischer-Strasse 30, 97074 Würzburg, Germany;
nadja.vater@mathematik.uni-wuerzburg.de

A preconditioned gradient descent scheme for a regularized nonlinear least squares problem arising from the aim to approximate given data by a neural network is presented.
Specifically, regularized over-parameterized nonlinear problems are considered, and the construction of a novel left preconditioner is illustrated. This preconditioner is based on randomized linear algebra techniques [1].
In this talk, theoretical and computational aspects of the proposed preconditioning scheme for a gradient method are discussed. Further, results of numerical experiments with the resulting preconditioned gradient applied to nonlinear least squares problems arising in neural network training are presented. These results demonstrate the effectiveness of preconditioning compared to a standard gradient descent method.

References

[1] N. Vater and A. Borzì. *Preconditioned Gradient Descent for Data Approximation with Shallow Neural Networks*. *International Conference on Machine Learning, Optimization and Data Science*, 2022 (to appear).

[1]Partially supported by the BMBF-Project iDeLIVER : Intelligent MR Diagnosis of the Liver by Linking Model and Data-Driven Processes.

Convergence Rates for a Deep Learning Algorithm for Semilinear PDEs

Christoph Belak[1], Oliver Hager[2], Charlotte Reimers[3],
Lotte Schnell[4], Frank T. Seifried[5] and Maximilian Würschmidt*[6]

[1−4]Technische Universitaet Berlin, Institute of Mathematics, Straße des 17. Juni 136, 10623 Berlin
[5] University of Trier, Department IV – Mathematics, Universitaetsring 15, 54296 Trier, seifried@uni-trier.de
[6] University of Trier, Department IV – Mathematics, Universitaetsring 15, 54296 Trier, wuerschmidt@uni-trier.de

We derive convergence rates for a deep solver for semilinear partial differential equations which is based on a Feynman-Kac representation in terms of a forward-backward stochastic differential equation and a discretization in time. We show that the error of the deep solver is bounded in terms of its loss functional, hence yielding a direct measure to judge the quality in numerical applications, and that the loss functional converges sufficiently fast to zero to guarantee that the approximation error vanishes in the limit. As a consequence of these results, we show that the deep solver has a strong convergence rate of order 1/2.

References

[1] Belak, Christoph and Hager, Oliver and Reimers, Charlotte and Schnell, Lotte and Seifried, Frank Thomas and Würschmidt, Maximilian, *Convergence Rates for a Deep Learning Algorithm for Semilinear PDEs,* Available at SSRN: https://ssrn.com/abstract=3981933 or http://dx.doi.org/10.2139/ssrn.3981933

A Gauss-Newton-based Decomposition Algorithm for Nonlinear Mixed-Integer Optimal Control Problems

A. Bürger[1,2], **C. Zeile**[3], A. Altmann-Diese[1], S. Sager[3], and M. Diehl[2]

[1] Institute of Refrigeration, Air-Conditioning, and Environmental Engineering (IKKU), Karlsruhe University of Applied Sciences, Moltkestraße 30, 76133 Karlsruhe, Germany; adrian.buerger@h-ka.de

[2] Systems Control and Optimization Laboratory, Department of Microsystems Engineering (IMTEK), University of Freiburg, Georges-Koehler-Allee 102, 79110 Freiburg im Breisgau, Germany; moritz.diehl@imtek.uni-freiburg.de

[3] Institute for Mathematical Optimization, Faculty of Mathematics, Otto von Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany; clemens.zeile@ovgu.de

For the fast approximate solution of Mixed-Integer Non-Linear Programs (MINLPs) arising in the context of Mixed-Integer Optimal Control Problems (MIOCPs) a decomposition algorithm exists that solves a sequence of three comparatively less hard subproblems to determine an approximate MINLP solution. In this work, we propose a problem formulation for the second algorithm stage that is a convex approximation of the original MINLP and relies on the Gauss-Newton approximation. We analyze the algorithm in terms of approximation properties and establish a first-order consistency result. Then, we investigate the proposed approach considering an illustrative numerical example of Mixed-Integer Optimal Control (MIOC) of a simple nonlinear and unstable system and considering a more complex application that is a numerical case study of MIOC of a renewable energy system. The investigation shows that the proposed formulation can yield an improved integer solution regarding the objective of the original MINLP compared with the Combinatorial Integral Approximation (CIA) algorithm [1]. We discuss possible directions to use learning methods for improving the existing approach. This presentation is based on [2].

References

[1] Sager, S., Jung, M., & Kirches, C. (2011). *Combinatorial integral approximation*. Mathematical Methods of Operations Research, 73(3), 363-380.

[2] Bürger, A., Zeile, C., Altmann-Dieses, A., Sager, S., & Diehl, M., (2022). *A Gauss-Newton-based Decomposition Algorithm for Nonlinear Mixed-Integer Optimal Control Problems*, Optimization-Online preprint: *http://www.optimization-online.org/DB_HTML/2022/04/8890.html*

Control and Machine Learning

Prof. Dr. Enrique Zuazua

Friedrich Alexander Universität Erlangen Nürnberg - Alexander von Humboldt Professorship, Germany

Fundación Deusto, Bilbao

Universidad Autónoma de Madrid

In this lecture we shall present some recent results on the interplay between control and Machine Learning, and more precisely, Supervised Learning and Universal Approximation.
We adopt the perspective of the simultaneous or ensemble control of systems of Residual Neural Networks (ResNets). Roughly, each item to be classified corresponds to a different initial datum for the Cauchy problem of the ResNets, leading to an ensemble of solutions to be driven to the corresponding targets, associated to the labels, by means of the same control.
We present a genuinely nonlinear and constructive method, allowing to show that such an ambitious goal can be achieved, estimating the complexity of the control strategies.
This property is rarely fulfilled by the classical dynamical systems in Mechanics and the very nonlinear nature of the activation function governing the ResNet dynamics plays a determinant role. It allows deforming half of the phase space while the other half remains invariant, a property that classical models in mechanics do not fulfill.
The turnpike property is also analyzed in this context, showing that a suitable choice of the cost functional used to train the ResNet leads to more stable and robust dynamics.
This lecture is inspired in joint work, among others, with Borjan Geshkovski (MIT), Carlos Esteve (Cambridge), Domènec Ruiz-Balet (IC, London) and Dario Pighin (Sherpa.ai)