TU Clausthal

International Workshop

Mathematics in the Age of Data

September 8–10, 2025
Institute for Mathematics
Clausthal University of Technology
Clausthal-Zellerfeld, Germany

Program

August 26, 2025

Contents

Workshop Mathematics in the Age of Data

In a time of rapidly growing amounts of data, data science plays a central role in modern research and technology. The increasing availability of data is revolutionising the way we do business, the way we develop technology, the way we do research, and the way we live. On the one hand, applied mathematics plays a driving role in this revolution, a fact that is often overlooked by non-experts. On the other hand, the focus on data also provides transformative research questions within the field of applied mathematics.

The goal of this workshop is an in-depth discussion of the current chances and challenges in applied mathematics in the face of the ever-growing amount of available data.

Website and registration:

`https://www.mathematik.tu-clausthal.de/workshop-mathdata`

Organizers: Dominic Breit, Olaf Ippisch, Philipp Öffner, Andreas Potschka, Benjamin Säfken, Andreas Tillmann, Aleksandra Zimmermann

Contact: `mitaod@tu-clausthal.de`

Program / Schedule

The workshop will consist of invited talks, contributed research spotlight talks and poster presentations, with ample time for intensive discussions after talks and throughout.

time	Monday Sep. 8	Tuesday Sep. 9	Wednesday Sep. 10
8:30 – 9:30	Registration & Opening	Andreas Prohl	Robert Scheichl
9:30 – 10:30	Emily King	Jan Hesthaven	Alexandre Ern
10:30 – 11:00	Coffee Break	Coffee Break	Coffee Break
11:00 – 12:00	Sebastian Sager	David Siska	Lars Schewe
12:00 – 13:00	Flore Nabet	Hergen Schultze	Maria Han Veiga
13:00 – 15:00	Lunch Break	Lunch Break	Closing & Lunch
15:00 – 16:30	Research Spotlights	Social Event	
17:00 – 19:00	Poster Session & Drinks		
19:00 –		Conference Dinner	

Invited Talks

Emily King, Colorado State University [Mon., 9:30 –10:30]

Interpretable, Explainable, and Adversarial AI: Data Science Buzzwords and You (Mathematicians)

Many state-of-the-art methods in machine learning are black boxes which do not allow humans to understand how decisions are made. In a number of applications, like medicine and atmospheric science, researchers do not trust such black boxes. Explainable AI can be thought of as attempts to open the black box of neural networks, while interpretable AI focuses on creating clear boxes. Adversarial attacks are small perturbations of data that cause a neural network to misclassify the data or act in other undesirable ways. Such attacks are potentially very dangerous when applied to technology like self-driving cars. The goal of this talk is to introduce mathematicians to problems they can attack using their favorite mathematical tools.

Sebastian Sager, OVGU Magdeburg [Mon., 11:00 –12:00]

Corleone.jl: COmplexity Reduction and Lifting Enhanced Optimization of Nonlinear processes and Experiments

We are interested in active learning for complex processes with expensive-to-obtain training data.

Complex dynamic systems are typically either modeled using expert knowledge in the form of differential equations or via data-driven universal approximation models such as artificial neural networks (ANN) or Markov decision processes. While the first approach has advantages with respect to interpretability, transparency, data-efficiency, and extrapolation, the second approach is able to learn completely unknown functional relations from data and may result in models that can be evaluated more efficiently. To combine the complementary advantages, universal differential equations (UDE) have been suggested. They replace unknown terms in the differential equations with ANN. Such hybrid models allow to both encode prior domain knowledge, such as first principles, and to learn unknown mechanisms from data. Often, data for the training of UDE can only be obtained via costly experiments. We consider active learning / optimal experimental design (OED) for planning of experiments and generating data needed to train UDE. The number of weights in the embedded ANN usually leads to an overfitting of the regression problem.

We present our interpretation of active learning as a specially structured optimal control problem, different dimension reduction methods that are based on lumping of weights and singular value decomposition of the Fisher information matrix (FIM), the open-source software package Corleone.jl, and applications from chemical engineering.

Flore Nabet, CMAP, École Polytechnique Paris-Saclay [Mon., 12:00 –13:00]

Numerical analysis of a finite-volume method for stochastic heat equation

In this talk we will focus on the numerical analysis of the stoichastic heat equation. To do this, we will use a finite-volume method for the space dicretization. First, we consider a multiplicative Ito noise for which we prove the convergence of the numerical scheme to the unique variational solution of the continuous problem, with particular emphasis on the stochastic compactness argument. We then turn to the problem of Stratonovich-type transport noise and detail the advantages of this choice, the difficulties it entails and the differences with the previous analysis.

This work was carried out in collaboration with C. Bauzet, K. Schmitz and A. Zimmermann and with A. De Bouard and L. Goudenège.

Andreas Prohl, University of Tübingen [Tue., 8:30 –9:30]

Convergent Discretization of the Fokker-Planck equation in Higher Dimensions

I present a new algorithm that combines the Euler scheme to solve the associated SDE with a data-dependent density estimator. I will discuss its convergence proof, which also employs tools from statistics.

Jan Hesthaven, Karlsruhe Institute of Technology [Tue., 9:30 –10:30]

Enabling digital twins through modeling, data, and learning

The ability to build efficient physics-based surrogate models with predictive accuracy is an old dream, in particular for complex engineering systems. However, realizing this remains a challenge due to the need to combine advanced computational modeling, reduced complexity models and external data for calibration, updating and uncertainty management. Nevertheless, the perspectives of using such models, often referred to as digital twins, for design, predictive maintenance, operational optimization, and risk analysis are substantial and the potential for impact significant, from safety, planning, and financial points of view.

In this talk we shall first discuss the importance of efficient and accurate reduced models as a key element in the development of such models for complex systems. This provides a path to generate data which can subsequently be combined with various modern learning techniques to arrive at physics-based models with predictive value and, ultimately, provide decision support. We shall illustrate this through a number of examples which combine computational first principle models with reduced order models and data driven technologies, e.g., classifiers, Gaussian regression and neural networks, to enable failure analysis, optimal sensor placement, closures for the accurate simulation of drone dynamics and, time permitting, multi-fidelity methods and risk analysis for rare events. These are all elements of the workflow that needs to be realized to address the challenge of building predictive surrogate models – Digital Twins – for complex applications.

David Siska, University of Edinburgh [Tue., 11:00 –12:00]

Convergence of policy gradient methods in discrete and continuous time reinforcement learning

Policy gradient methods have been used to develop extremely efficient reinforcement learning (RL) algorithms over the last decade. The most empirically efficient is the Proximal Policy Optimization (PPO) method, which has been inspired by the Trust Region Policy Optimization (TRPO) method and mirror descent, but contains intuition-based modifications that make its convergence analysis an open problem.

This talk will review recent results on the convergence of policy-gradient-type algorithms for RL and stochastic control, focusing on the mirror descent method with and without function approximation, its connection to the natural policy gradient (NPG), and the role of convexity (or lack thereof) in the objective function. Finally, we will highlight the gap between the mathematical understanding of convergence on one hand and PPO-based actor-critic algorithms with function approximation on the other.

Hergen Schultze, BASF SE Ludwigshafen [Tue., 12:00 –13:00]

Artificial Intelligence for Reaction Optimization

Optimization of reaction conditions has a long history in chemical industry. While progress was originally based on random discoveries and the intuition of chemists, in the last century we developed techniques and statistical methods for process optimization. Now, we are taking the next step, using artificial intelligence to automate and accelerate reaction optimization, not least to better address the major chemical industry challenges such as CO2 management and the transformation of energy and raw materials. We believe that all significant decisions in R&D should be driven by evidence, that means by well designed, executed, and analyzed experiments. To accelerate this process we create machines, that "close the loop" for autonomous research and development. Such a machine is basically a fully automized lab setup. It is controlled by a lab management and execution software, which is orchestrated by a higher-level optimization algorithm. The context and goals are set by the scientist, but the augmentation by generative large language models is perceivable and currently under development. Whereas academic groups focus on proving the principle for new technologies and applications, we in industry focus on establishing the technologies in a reliable way.

In the first part of the talk, we describe the concept: For the digital level of the autonomous research machine, we collaborate with external partners on open-source algorithm development. We provide the capabilities internally as digital products, that means the optimization functionality is readily available and maintained long-term by an internal team. For a certain application, these capabilities are coupled with a professional lab execution system. In the second part of the talk, we describe our technical set-up, scope and limitations of our present system and potential future improvements. Based on some case studies, we finally conclude with a discussion on how empirical approaches, highly automated and digitized, could effectively short-cut traditional R&D especially when combined with large multi-modal models.

Robert Scheichl, Heidelberg University [Wed., 8:30 –9:30]

Scalable Bayesian Inference and Optimal Experimental Design

In the last decade, alongside the rise of data science and machine learning there has also been a vast growth in interest and contributions from numerical analysis and scientific computing to high-dimensional Bayesian statistics, with the aim to efficiently combine data and physical models to better understand and control engineering problems with a quantitative measure of the remaining uncertainty. Simply opening the leading journals in the field or looking at recent job adverts will reveal this fact. But what are the problems and challenges that people are aiming to address, what are potential contributions and how can it benefit the field most effectively? In this talk, I will try to summarise some of the main areas of research where there are opportunities for numerical analysis to have an impact, but also the difficulties and barriers encountered. More specifically, I will present two exemplary approaches that use surrogates to significantly accelerate Bayesian computation and optimal experimental design in high-dimensional PDE-constrained applications: multilevel delayed acceptance MCMC, as well as a measure-transport approach based on low-rank tensor approximations.

Alexandre Ern, École des Ponts / INRIA Paris [Wed., 9:30 –10:30]

Elasticity-based mesh morphing and application to reduced-order modeling with variable shapes

In the first part, we present a methodology for constructing morphings for a collection of target shapes that have identical topology. The morphings are obtained by deforming a reference shape onto every target shape, through the resolution of a sequence of linear elasticity equations. The approach does not assume any knowledge of a boundary parametrization, and the computation of the boundary deformation is not required beforehand. In the second part, we show how the proposed methodology can be integrated in the offline and online paradigm of reduced-order modeling involving variable shapes. The robustness and computational efficiency of the methodology is illustrated on two-dimensional test cases, including the regression problem of the drag and lift coefficients of airfoils of non-parameterized variable shapes. This is joint work with A. Kabalan (Safran Tech & Cermics), V. Ehrlacher (Cermics), and F. Casenave and F. Bordeu (Safran Tech).

Lars Schewe, University of Edinburgh [Wed., 11:00 –12:00]

Discrete Optimization for better energy networks: Outage planning, topology optimization, and future operation of hydrogen networks

In this talk, I will show how discrete optimization, specifically mixed-integer nonlinear optimization, is able to tackle real-world problem in energy systems. I will show examples from industry projects on outage planning, topology optimization of electricity networks, and give an outlook how mathematics can contribute to answering question about future hydrogen networks.

Maria Han Veiga, Ohio State University [Wed., 12:00 –13:00]

Mathematical aspects of generative machine learning models

Diffusion probabilistic models have become the state-of-the-art tool in generative methods, used to generate high-resolution samples from very high-dimension distributions (e.g. images). It relies on a forward-time stochastic differential equation (SDE) that slowly injects noise on the data distribution, transforming it into a known prior distribution, and a reverse-time SDE that transforms the prior distribution back into the data distribution by removing the noise. The reverse-time SDE follows the time-dependent gradient field (score) of the perturbed data distribution.

Although very effective, there are some drawback to this method: 1) as opposed to variational encoders, the dimension of the problem remains high during the generation process and 2) they can be prone to memorisation of the dataset. In this talk, we consider the second point: the learned score can overfit the finite dataset, making the reverse-time SDE sample mostly training points. In our work, we interpret the learned empirical score as a noisy approximation of the true score and study its covariance matrix, showing that in high-dimension for small time, the noise variance blows up. To reduce this variance, we introduce a smoothing operator on the empirical score and analyse its bias-variance properties. Consequences of this approach are better generalisation of the dataset and thus, a reduction on the amount of samples required to approximate a distribution.

Research Spotlight Talks

- Dmitry Gromov: *Structure- and data-driven methods for analysis and control of complex dynamical systems*
- Armin Iske: *Current topics of kernel-based approximation*
- Maha Youssef Ismail: *Diverse Applications of ML: Oil Industry & Food Service*
- Dustin Mixon: *Bilipschitz Invariants*
- Tsiry Avisoa Randrianasolo: t.b.a.
- Uwe Saint-Mont: *"Every time the amount of data increases by a factor of ten, we should totally rethink how we analyse it."*
- Niklas Sapountzoglou (TU Clausthal): *Convergence rates for a finite volume scheme of the stochastic heat equation*
- Jochen Schütz: *Multiderivative SDC schemes*
- Cédric Sultan: *Finite volume scheme for stochastic Allen-Cahn equation with constraint*

Poster Presentations

- De Ananyapam: t.b.a.
- Kurt Böhm: *Lineal: An Efficient, Hybrid-Parallel Linear Algebra Library*
- Anatole Gaudin: *Optimal regularity for the Stokes–Dirichlet problem: Elliptic Theory & Maximal Regularity in Rough Domains*
- Vinay Kumar: t.b.a.
- Nikolas Lykourinos: *Spanning Trees with Shortest Maximal Fundamental Cycle*
- Romeo Mensah: t.b.a.
- Dustin Mixon: *Bilipschitz Invariants*
- Tim Niemann: *Delay-Resistant Robust Vehicle Routing with Heterogeneous Time Windows*
- Philipp Öffner: *A structure-preserving scheme for the Navier-Stokes-Korteweg equations*
- Kavin Rajasekaran: *Large deviation principle for stochastic evolutionary p-Laplace type equation*

- Alexander Ritz: *Prediction Error for Quantile Regression*
- Cédric Sultan: *Seismic wave attenuation in a horizontally-stratified random soil layer*

Venue

All talks will take place in the Seminar Room A (B7-209) at the Institute of Mathematics (IfM), Clausthal University of Technology, Erzstr. 1, 38678 Clausthal-Zellerfeld, Germany.

The poster session will take place in the Aula of TU Clausthal (Aulastr. 8). There will be drinks, including delicious beers from our research brewery, at the session.

Lunch, Social Event & Conference Dinner

Lunch is included in the conference fee and will be available on all three workshop days at the Mensa (Leibnizstr. 3).

The social event will take us to the nearby medieval city of Goslar, a UNESCO world heritage site, for a guided tour (in English) followed by some time to explore further. We will remain in Goslar for the conference dinner at the restaurant "Platon" (Bäringerstr. 6, 38640 Goslar) at 7 pm. We have arranged for a charter bus transportation to and from Goslar; details t.b.a.

Dinner on Monday night is not included; however, we have reserved a large table at the Ju Bin Lou restaurant (Schulstr. 47) for workshop participants and welcome everyone who wants to join in.

Acknowledgments

The organizers gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG), project number 563143858.