Optimal Transport from Theory to Applications:

Interfacing Dynamical Systems, Optimization, and Machine Learning

Humboldt-Universität zu Berlin, March 11 – 15, 2024

Optimal transport (OT) is a theory connecting PDEs, geometry, and probability theory. Recent developments in numerical algorithms for OT problems have further opened new applications, e.g., in statistics, machine learning, imaging, and optimization. The OT-DOM 2024 workshop brings together international experts in OT theory as well as domain experts in applied areas, such as machine learning, optimization, and engineering, to gather in Berlin, Germany.

Weierstrass Institute for Applied Analysis and Stochastics

Berlin Mathematics Research Center MATH+

Deutsche Forschungsgemeinschaft

The organizers,

Pavel Dvurechensky,
Matthias Liero,
Gabriele Steidl, and
Jia-Jie Zhu

1 General Information

1.1 Venue

The workshop is hosted by the Weierstrass Institute and will be held at the Humboldt-Universität zu Berlin, lecture hall 1.101 in Dorotheenstraße 24, 10117 Berlin. The nearest underground stations are **Friedrichsstraße**, served by line U6 and **Unter den Linden**, served by lines U5 and U6. Please see www.bvg.de for more information on public transport in Berlin.

1.2 Workshop Dinner and Lunch

The workshop dinner will take place at the restaurant Tapas y Más, Neue Grünstraße 17-18, 10179 Berlin on Tuesday, March 12 at 6 pm. It is covered within the conference fee and includes the meal and one drink per person.

Here is a selection of eateries close to the venue:

- Via Nova II – Italian cuisine
- Da Vinci – pizza, pasta, and salads
- Deponie Nr. 3 – savoury German food
- Hua Rong – Vietnamese cuisine
- ama Café – coffee, tartes flambées, sweet and savoury snacks
- Chupenga – burritos, tacos, etc.
- Restaurant Jolly – Chinese cuisine
- beets&roots – bowls, wraps, salads, and soups
- Swing Kitchen – vegan fast food
- Buddha – Indian cuisine

2 Detailed Schedule

Monday, March 11

13:00–14:00	Arrival & Registration
14:00–14:15	Opening
14:15–15:00	Alexander Mielke (Weierstraß-Institut, Berlin, Germany) *Hellinger–Kantorovich (aka WFR) Spaces and Gradient Flows*
15:00–15:30	Coffee Break
15:30–16:15	Philippe Rigollet (MIT, Cambridge, USA) *The Emergence of Clusters in Self-attention Dynamics*
16:15–17:00	Olga Mula (Eindhoven University of Technology, Netherlands) *Reduced Models in Wasserstein Spaces for Forward and Inverse Problems*

Tuesday, March 12

09:00–09:45	Mark Peletier (Technical University of Eindhoven, Netherlands) *Singular-limit Analysis of Training with Noise Injection*
09:45–10:30	Bernhard Schmitzer (Georg-August-Universität Göttingen, Germany) *Entropic Transfer Operators for Data-driven Analysis of Dynamical Systems*
10:30–11:00	Coffee Break
11:00–11:45	Julie Delon (Université Paris-Cité, France) *Optimal Transport with Invariances between Gaussian Mixture Models*
11:45–12:30	Virginie Ehrlacher (ENPC & INRIA, Paris, France) *Sparse Approximation of Multi-marginal Quantum Optimal Transport Problems with Moment Constraints: Application to Quantum Chemistry*
12:30–14:00	Lunch Break
14:00–14:45	Daniel Kuhn (École Polytechnique Fédérale de Lausanne, Switzerland) *Distributionally Robust Linear Quadratic Control*
14:45–15:30	Hongseok Namkoong (Columbia University, New York, USA) *On the Need for a Language Describing Distribution Shifts: Illustrations on Tabular Datasets*
15:30–16:15	Chaoye Zhao (University of Washington, Seattle, USA) *Provably Convergent Policy Optimization via Metric-aware Trust Region Methods*
18:00	Conference Dinner at “Tapas Y Mas”

Wednesday, March 13

09:00–09:45	Gabriel Peyré (CNRS & ENS, Paris, France) *Sparsistency for Inverse Optimal Transport*
09:45–10:30	Jan Maas (IST Austria, Klosterneuburg) *Anisotropy and the Infinitesimal Model*
10:30–11:00	Coffee Break
11:00–11:45	Xue-Mei Li (École Polytechnique Fédérale de Lausanne, Switzerland) *Coarse Extrinsic Curvature*
11:45–12:30	Laetitia Chapel (Université de Bretagne-Sud & Institut Agro, France) *Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics*
12:30–14:00	Lunch Break
14:00–14:45	Massimiliano Pontil (IIT Genoa & UC London, Italy & UK) *Transfer Operator Learning*
14:45–15:30	Guillaume Carlier (Université Paris Dauphine - PSL, France) *Displacement Smoothness of Entropic Optimal Transport and Applications to some Evolution Equations and Systems*
15:30–16:00	Poster Pitches
16:00–17:00	Poster Session & Coffee

Thursday, March 14

09:00–09:45	EMTIYAZ KHAN (RIKEN, Tokyo, Japan) *The Bayesian Learning Rule*
09:45–10:30	ANNA KORBA (ENSAE & CREST, Palaiseau, France) *Sampling through Optimization of Discrepancies*
10:30–11:00	Coffee Break
11:00–11:45	JIAXIN SHI (Google DeepMind, London, UK) *Stein's Method for Modern Machine Learning: From Gradient Estimation to Generative Modeling*
11:45–12:30	OLIVER TSE (Eindhoven University of Technology, Netherlands) *Variational Acceleration Methods in the Space of Probability Measures*
12:30–14:00	Lunch Break
14:00–14:45	CLAUDIA TOTZECK (Bergische Universität Wuppertal, Germany) *Adjoint-based Optimal Control with Wasserstein-2 Metric*
14:45–15:30	TIM LAUX (Universität Regensburg, Germany) *Emergence of Mean Curvature Flow in Learning*
15:30–16:00	Poster Pitches
15:30–16:15	Poster Session & Coffee

Friday, March 15

09:00–09:45	Taiji Suzuki (University of Tokyo, Japan) *Convergence of Mean-field Langevin Dynamics and its Application to Neural Network Feature Learning*
09:45–10:30	Wuchen Li (University of South Carolina, Columbia, USA) *Information Gamma Calculus: Convexity Analysis for Stochastic Differential Equations*
10:30–11:00	Coffee Break
11:00–11:45	Mi Jung Park (Technical University of Denmark. Lyngby) *Privacy-preserving Data Generation in the Era of Foundation Models: Generative Transfer Learning with Differential Privacy*
11:45–12:30	Discussion & Closing

3 Poster Sessions

Wednesday, March 13

A1 Magnus Tronstad: *Modelling Trajectories of Cell Development in Single Cell Sequencing Data as an Optimal Transport Problem with Multiple Stages*

A2 Razvan-Andrei Lascu: *Entropic Mean-field Min-Max Problems via Fisher–Rao and Best Response Flows*

A3 Michael Quellmalz: *Sliced Optimal Transport on the Sphere*

A4 Fabius Krämer: *On Computation and Analysis of the Volume Constrained MBO Scheme*

A5 Fabian Altekrüger: *Posterior Sampling via Sliced MMD Flows with the Negative Distance Kernel*

A6 Linfeng Wang: *Measure Transport with Kernel Mean Embeddings*

A7 Giacomo Cristinelli: *Conditional Gradients for Total Variation Regularization with PDE Constraints: A Graph-cuts Approach*

A8 Ilja Klebanov: *QMC and Sparse Grids beyond Uniform Distributions on Cubes*

A9 Gergely Neu: *Computing Optimal Transport Distances for Markov Chains via Linear Programming*

A10 Leonardo Del Grande: *A General Theory for Exact Sparse Representation Recovery in Convex Optimization*

A11 Jonas Bresch: *Rényi Regularized Optimal Transport*

A12 Zhao Jiaxi: *Scaling Limits of the Wasserstein Information Matrix on Gaussian Mixture Models*

A13 Mengmeng Li: *Transition-robust MDPs with Wasserstein Uncertainty Sets*

A14 Filippo Quattrocchi: *Gradient Flow Interpretation of the Fokker–Planck equation with Dirichlet Boundary Conditions*

Thursday, March 14

B1 Viet Hoang Tran: *Revisiting Over-smoothing and Over-squashing Using Ollivier–Ricci Curvature*

B2 Stefan Schrott: *The Wasserstein Space of Stochastic Processes in Continuous Time*

B3 Chun Yin Lam: *Variational Convergence of Exchange-driven Stochastic Particle Systems*

B4 Vitalii Aksenov: *Learning Distributions with Regularized JKO Scheme and Low-Rank Tensor Decompositions*

B5 Viktor Stein: *Gradient Flows of Moreau Envelopes of f-Divergences*

B6 Hong Ye Tan: *Noise-free Sampling Algorithms via Regularized Wasserstein Proximals*

B7 Sebastian Gutiérrez Hernández: *Control Problems in Wasserstein Space via Conditional Flow Matching*

B8 Julia Sanders: *Optimal Control of Underdamped Systems: Numerical Predictions*

B9 Maciej Buze: *Anisotropic Power Diagrams for Modelling of Polycrystalline Materials: Efficient Generation of Curved Grains via Optimal Transport*

B10 Robert Lasarzik: *Minimizing Movement Scheme for Non-gradient Flows*

B11 Florian Beier: *Multi-marginal Gromov–Wasserstein Transport and Barycenters*

B12 Giacomo Borghi: *Optimization in Wasserstein Space via Measure-valued Agents*

B13 Raphaël Barboni: *Training Infinitely Deep and Wide ResNets with Conditional Optimal Transport*

B14 Pratik Rai: *State Estimation with Wasserstein Reduced Models*

4 Abstracts (in Alphabetical Order)

4.1 Invited Talks

Guillaume Carlier (Université Paris Dauphine - PSL, France)

Displacement Smoothness of Entropic Optimal Transport and Applications to some Evolution Equations and Systems

The entropic approximation of optimal transport has received a lot of attention since the influential work of Cuturi who popularized Sinkhorn's algorithm in the OT and ML communities. In the first part of my talk, based on a joint work with L. Chizat and M. Laborde, I will show that the potentials from entropic optimal transport (EOT) depend in a smooth way on the marginals when the latter are interpolated by (non necessarily optimal) displacement. This implies well posedness of Wasserstein gradient flows of functionals involving EOT but also of other dynamical systems which are not gradient flows. I will describe the case of entropic semi-geostrophic equations, an Hamiltonian system motivated by frontogenesis in large-scale atmospheric flows (joint work with Hugo Malamut).

Laetitia Chapel (Université de Bretagne-Sud & Institut Agro, France)

Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics

I will discuss a new proxy of the squared WD, coined min-SWGG, that is based on the transport map induced by an optimal one-dimensional projection of the two input distributions. I will draw connections between min-SWGG and Wasserstein generalized geodesics in which the pivot measure is supported on a line. This gives a new closed form for the exact Wasserstein distance in the particular case of one of the distributions supported on a line allowing one to derive a fast computational scheme that is amenable to gradient descent optimization. I will show that min-SWGG is an upper bound of WD and that it has a complexity similar to as Sliced-Wasserstein, with the additional feature of providing an associated transport plan. I will also investigate some theoretical properties such as metricity, weak convergence, computational and topological properties.

Julie Delon (Université Paris-Cité, France)

Optimal Transport with Invariances between Gaussian Mixture Models

Gaussian Mixture Models (GMMs) are ubiquitous in statistics and machine learning and are especially useful in applied fields to represent probability distributions of real datasets. Optimal transport can be used to compute distances or geodesics between such mixture models, but the corresponding Wasserstein geodesics do not preserve the property of being a GMM. It has been shown in https://arxiv.org/abs/1907.05254 that restricting the set of possible coupling measures to GMMs transforms the original infinitely dimensional optimal transport problem into a finite dimensional problem with a simple discrete formulation, well suited to applications where a clustering structure is present in the data. In this talk, we present two possible extensions of this Wasserstein-type distance between GMMs that remain invariant to isometries. Inspired by the Gromov–Wasserstein distance, these extensions can also be used to compare GMMs of different dimensions.

Virginie Ehrlacher (École des Ponts ParisTech & Institut National de Recherche en Informatique et en Automatique, France)

Sparse Approximation of Multi-marginal Quantum Optimal Transport Problems with Moment Constraints: Application to Quantum Chemistry

The aim of this talk is to present new sparsity results about numerical approximations of multimarginal quantum optimal transport problems, using moment constraints. We apply this approach for the computation of the so-called Lieb functional, which is a particular instance of symmetric multi-marginal quantum optimal transport problem arising in Density Functional Theory for electronic structure calculations for molecules. More precisely, given an electronic density for a system of N electrons, which may be seen as a probability density on $\mathbb{R}3$, the value of the Lieb functional for this density is defined as the solution of a quantum multi-marginal optimal transport problem, which reads as a minimization problem defined on the set of trace-class operators acting on the space of electronic wave-functions that are anti-symmetric L2 functions of $\mathbb{R}3N$, with partial trace equal to the prescribed electronic density. We introduce a relaxation of this quantum optimal transport problem where the full partial trace constraint is replaced by a finite number of moment constraints on the partial trace of the set of operators. We show that, under mild assumptions on the electronic density, there exist sparse minimizers to the resulting moment constrained approximation of the Lieb (MCAL) functional that read as operators with rank at most equal to the number of moment constraints. We also prove under appropriate assumptions on the set of moment functions that the value of the MCAL functional converges to the value of the exact Lieb functional as the number of moments go to infinity. We also prove some rates of convergence on the associated approximation of the ground state energy. We finally study the mathematical properties of the associated dual problem. We also present an iterative numerical algorithm based on this sparsity structure for the resolution of the problem.

Emtiyaz Khan (Institute of Physical and Chemical Research (RIKEN), Japan)

The Bayesian Learning Rule

Humans and animals have a natural ability to autonomously learn and quickly adapt to their surroundings. How can we design machines that do the same? In this talk, I will present Bayesian principles to bridge such gaps between humans and machines. I will show that a wide-variety of machine-learning algorithms are instances of a single learning-rule derived from Bayesian principles. I will show our recent result on scaling up variational learning to large deep networks (e.g., GPT-2). Time permitting, I will also briefly discuss the dual perspective yielding new mechanisms for knowledge transfer in learning machines.

[1] M.E. Khan, H. Rue: *The Bayesian Learning Rule.* arXiv:2107.04562, 2021.

[2] P. Nickl, L. Xu, D. Tailor, T. Möllenhoff, M.E. Khan: *The Memory Perturbation Equation: Understanding Model's Sensitivity to Data.* arXiv:2310.19273, 2023.

[3] Y. Shen, N. Daheim, B. Cong, P. Nickl, G.M. Marconi, C. Bazan, R. Yokota, I. Gurevych, D. Cremers, M.E. Khan, T. Möllenhoff: *Variational Learning is Effective for Large Deep Networks.* arXiv:2402.17641, 2024.

Anna Korba (École nationale de la statistique et de l'administration économique & Centre de recherche en économie et statistique, France)

Sampling through Optimization of Discrepancies

Sampling from a target measure when only partial information is available (e.g. unnormalized density as in Bayesian inference, or true samples as in generative modeling) is a fundamental problem in computational statistics and machine learning. The sampling problem can be formulated as an optimization over the space of probability distributions of a well-chosen discrepancy (e.g. a divergence or distance). In this talk, we'll discuss several properties of sampling algorithms for some choices of discrepancies (well-known ones, or novel proxies), both regarding their optimization and quantization aspects.

Daniel Kuhn (École Polytechnique Fédérale de Lausanne, Switzerland)

Distributionally Robust Linear Quadratic Control

Linear-Quadratic-Gaussian (LQG) control is a fundamental control paradigm that is studied in various fields such as engineering, computer science, economics, and neuroscience. It involves controlling a system with linear dynamics and imperfect observations, subject to additive noise, with the goal of minimizing a quadratic cost function for the state and control variables. In this talk, we consider a generalization of the discrete-time, finite-horizon LQG problem, where the noise distributions are unknown and belong to Wasserstein ambiguity sets centered at nominal (Gaussian) distributions. The objective is to minimize a worst-case cost across all distributions in the ambiguity set, including non-Gaussian distributions. Despite the added complexity, we prove that a control policy that is linear in the observations is optimal for this problem, as in the classic LQG problem. We propose a numerical solution method that efficiently characterizes this optimal control policy. Our method uses the Frank-Wolfe algorithm to identify the least-favorable distributions within the Wasserstein ambiguity sets and computes the controller's optimal policy using Kalman filter estimation under these distributions.

Tim Laux (Universität Regensburg, Germany)

Emergence of Mean Curvature Flow in Learning

Mean curvature flow (MCF) is a classical geometric evolution equation. In this talk, I will give two examples of how this flow emerges in machine learning applications. The first example is a basic algorithm in data clustering, the MBO scheme. Together with Jona Lelmi, we show that in the large-data limit and a certain scaling regime, the boundary between clusters moves by MCF. The second example comes from adversarial learning. In ongoing work with Leon Bungert and Kerrek Stinson, we analyze certain adversarial training methods and prove that in the limit of vanishing adversarial budget, the algorithms regularize the decision boundary of the non-robust classifier by evolving it via MCF.

Wuchen Li (University of South Carolina, USA)

Information Gamma Calculus: Convexity Analysis for Stochastic Differential Equations

We study the Lyapunov convergence analysis for degenerate and non-reversible stochastic differential equations (SDEs). We apply the Lyapunov method to the Fokker–Planck equation, in which the Lyapunov functional is chosen as a weighted relative Fisher information functional. We derive a structure condition and formulate the Lyapunov constant explicitly. Given the positive Lyapunov constant, we prove the exponential convergence result for the probability density function towards its invariant distribution in the L1 norm. Several examples are presented: underdamped Langevin dynamics with variable diffusion matrices, quantum SDEs in Lie groups (Heisenberg group, displacement group, and Martinet sub-Riemannian structure), three oscillator chain models with nearest-neighbor couplings, and underdamped mean field Langevin dynamics (weakly self-consistent Vlasov–Fokker–Planck equations).

Xue-Mei Li (École Polytechnique Fédérale de Lausanne, Switzerland)

Coarse Extrinsic Curvature

Coarse Ricci curvature is a concept proposed by Ollivier. We have extended this to generalised coarse r Ricci curvature,and more recently to coarse extrinsic curvature. The idea is to replace the geometric differential concepts with Wasserstein distances between measures supported about points. They are interesting concepts, expected useful on studying graphs, and also for curvature retrievals with data.

Jan Maas (Institute of Science and Technology Austria)

Anisotropy and the Infinitesimal Model

We prove upper bounds on the L^∞-Wasserstein distance between strongly log-concave probability densities and log-Lipschitz perturbations. In the simplest setting, such a bound amounts to a transport-information inequality involving the L^∞-Wasserstein metric andthe relative L^∞-Fisher information. We show that this inequality can be sharpened significantly in situations where the involved densities are anisotropic. Our proof is based on probabilistic techniques using Langevin dynamics. As an application of these results, we obtain sharp exponential rates of convergence in Fisher's infinitesimal model from quantitative genetics, generalising recent results by Calvez, Poyato, and Santambrogio in dimension 1 to arbitrary dimensions. This is joint work with Ksenia Khudiakova (ISTA) and Francesco Pedrotti (ISTA).

Alexander Mielke (Weierstraß-Institut für Angewandte Analysis und Stochastik)

Hellinger–Kantorovich (aka WFR) Spaces and Gradient Flows

The gradient-flow theory for geodesic metric spaces has proved to be a flexible tool in many applications, most recently also in data science. Often, pure optimal transport needs to be replaced by unbalanced transport to account for reweighting or generation of new data points.

So far, the Hellinger-Kantorovich distance (which is also known as Wasserstein-Fisher-Rao distance) is the only geodesic distance that describes unbalanced transport and allows for a complete description of the geodesics by mixture of transport and positive and negative growth. This distance was introduced in the mid 2010s independently in Paris [CP15], Lisbon [KMV16], and Berlin/Pavia [LMS18]. The restriction of HK to probability measures leads to the geodesic spherical HK distance, see [LaM19] where the set of measures is considered as a cone over the probability measures.

We discuss the characterization of geodesic semi-convexity obtained in [LMS23]. Necessary and sufficient conditions on E are given, such that $\mathcal{E}(\mu) = \int_\Omega E(\rho)\,\mathrm{d}x$ for $\mu = \rho\,\mathrm{d}x$

is geodesically λ-convex. Using this, we discuss how to construct solutions to Evolutionary Variational Inequalities via the minimizing movement scheme, see [LaM23]. For this we rely on the unpublished part II of [MuS20].

[CP15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard: *An interpolating distance between optimal transport and Fisher—Rao metrics.* Found. Comput. Math. 18:1 (2015) 1–44.

[KMV16] Kondratyev, L. Monsaingeon, and D. Vorotnikov: *A new optimal transport distance on the space of finite Radon measures.* Adv. Differ. Eqns. 21:11/12 (2016) 1117–1164.

[LaM19] V. Laschos and A. Mielke: *Geometric properties of cones with applications on the Hellinger–Kantorovich space, and a new distance on the space of probability measures.* J. Funct. Analysis 276:11 (2019) 3529–3576.

[LaM23] V. Laschos and A. Mielke: *Evolutionary Variational Inequalites on the Hellinger–Kantorovich and the spherical Hellinger–Kantorovich spaces.* Submitted (2023), arXiv:2207.09815v3.

[LMS18] M. Liero, A. Mielke, and G. Savaré: *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures.* Invent. math. 211 (2018) 969–1117.

[LMS23] M. Liero, A. Mielke, and G. Savaré: *Fine properties of geodesics and geodesic λ-convexity for the Hellinger–Kantorovich distance.* Arch. Rational Mech. Anal. 247:112 (2023) 1–73.

[MuS20] M. Muratori and G. Savaré: *Gradient flows and evolution variational inequalities in metric spaces. I: structural properties.* J. Funct. Analysis 278:4 (2020) 108347/1–67.

Olga Mula (Eindhoven University of Technology, Netherlands)

Reduced Models in Wasserstein Spaces for Forward and Inverse Problems

Forward model order reduction of parametric Partial Differential Equations and inverse state and parameter estimation are two major problems that arise in countless applications. Although these are mature fields and many algorithms have been developed over the years, most works focus on problems posed on Hilbert and Banach spaces, and they are severely challenged by transport dominated PDEs. In this talk, I will present recent works aiming at addressing such challenges and limitations by working in the space of Wasserstein measures.

Hongseok Namkoong (Columbia University, USA)

On the Need for a Language Describing Distribution Shifts: Illustrations on Tabular Datasets

Different distribution shifts require different algorithmic and operational interventions. Methodological research must be grounded by the specific shifts they address. Although nascent benchmarks provide a promising empirical foundation, they implicitly focus on covariate shifts, and the validity of empirical findings depends on the type of shift, e.g., previous observations on algorithmic performance can fail to be valid when the $Y|X$ distribution changes. We conduct a thorough investigation of natural shifts in 5 tabular datasets over 86,000 model configurations, and find that $Y|X$-shifts are most prevalent. To encourage researchers to develop a refined language for distribution shifts, we build WhyShift, an empirical testbed of curated real-world shifts where we characterize the type of shift we benchmark performance over. Since $Y|X$-shifts are prevalent in tabular settings, we identify covariate regions that suffer the biggest $Y|X$-shifts and discuss implications for algorithmic and data-based interventions. Our testbed highlights the importance of future research that builds an understanding of how distributions differ.

Mi Jung Park (Technical University of Denmark)

Privacy-preserving Data Generation in the Era of Foundation Models: Generative Transfer Learning with Differential Privacy

Creating impactful machine learning solutions for real-world applications often requires access to personal data that may compromise privacy, raising ethical and legal concerns. To address this, differentially private data generation aims to generate synthetic data that safeguards individual privacy while retaining statistical properties. In the era of foundation models, where pre-trained large models are readily available, the focus shifts to utilizing pre-trained models for generating high-quality synthetic data with a high privacy guarantee. In this talk, I will introduce two recently developed methods by my group for generative transfer learning with differential privacy (DP). The first method involves building off a good, relevant representation from a public, pre-trained model, then learning to model the private data with that representation. In particular, we minimize the maximum mean discrepancy (MMD) between private target data and a generator's distribution, using a kernel based on perceptual features learned from a public dataset. With the MMD, we can simply privatize the data-dependent term once and for all, significantly improving privacy-utility trade-offs, compared to introducing noise at each step of optimization as in differentially private stochastic gradient descent (DP-SGD, Abadi et al, 2016). The second method involves fine-tuning diffusion models with differential privacy constraints. In particular, we adopt pre-trained Latent Diffusion Models (LDMs), equipped with powerful autoencoders that map the high-dimensional pixels into a lower-dimensional latent space, in which diffusion models are trained, yielding a more efficient and faster training of diffusion models. Rather than fine-tuning the entire LDMs, we fine-tune only the attention modules of LDMs with DP-SGD, reducing the number of trainable parameters by roughly 90%, achieving a better privacy-accuracy trade-off than existing methods on DP diffusion models. These two methods offer efficient and practical frameworks for DP generative transfer learning.

Mark Peletier (Technical University of Eindhoven, Netherlands)

Singular-limit Analysis of Training with Noise Injection

Many training algorithms inject some form of noise in the training. The classical example is the mini-batch noise in Stochastic Gradient Descent, but other examples are dropout, data augmentation, 'noise nodes', 'label noise', and input-data noise. While the additional noise is generally believed to improve generalisation performance, there is little mathematical understanding of how this is achieved. In this talk I will describe recent work, together with Anna Shalova (TU/e) and André Schlichting (Münster), in which we analyse a fairly general class of iterative training schemes with noise injection. In the limit of small noise, we prove convergence of the appropriately rescaled time courses to solutions of an auxiliary evolution equation. This auxiliary equation is a gradient flow driven by a functional for which we obtain an explicit expression, thus opening the door to understanding the different types of regularisation generated by different types of noise injection.

Gabriel Peyré (Centre national de la recherche scientifique & École normale supérieure, France)

Sparsistency for Inverse Optimal Transport

Optimal Transport is a useful metric to compare probability distributions and to compute a pairing given a ground cost. Its entropic regularization variant (eOT) is crucial to have fast algorithms and reflect fuzzy/noisy matchings. This work focuses on Inverse Optimal Transport (iOT), the problem of inferring the ground cost from samples drawn from a coupling that solves an eOT problem. It is a relevant problem that can be used to infer unobserved/missing links, and to obtain meaningful information about the structure of the ground cost yielding the pairing. On one side, iOT benefits from convexity, but on the other side, being ill-posed, it requires regularization to handle the sampling noise. This work presents an in-depth theoretical study of the l_1 regularization to model for instance Euclidean costs with sparse interactions between features. Specifically, we derive a sufficient condition for the robust recovery of the sparsity of the ground cost that can be seen as a far reaching generalization of the Lasso's celebrated "Irrepresentability Condition". To provide additional insight into this condition, we work out in detail the Gaussian case. We show that as the entropic penalty varies, the iOT problem interpolates between a graphical Lasso and a classical Lasso, thereby establishing a connection between iOT and graph estimation, an important problem in ML.

This is a joint work with Francisco Andrade and Clarice Poon.

Massimiliano Pontil (Istituto Italiano di Tecnologia & University College London, Italy & UK)

Transfer Operator Learning

Non-linear dynamical systems are elegantly described by the associated transfer operator, whose action evolves every observable of the system forward in time. These operators are instrumental to forecasting and, leveraging the spectral decomposition, interpreting the system dynamics, with extensive applications in science and engineering. The talk will start with a short introduction to this topic, and then present recent ongoing work leading to improved operator regression estimators. Specifically, we will introduce a reduced-rank estimator, designed to learn more effectively the spectral decomposition of the transfer operator. Our statistical learning analysis is driven by the simultaneous control of the operator norm error and a novel metric distortion functional of the estimated eigenfunctions. This approach is crucial for avoiding spurious spectra and eigenvalues, a long-standing challenge that prevented reliable use of transfer operator estimators in practice. Our learning bounds further motivate a differentiable score function which can be used to learn a good representation of the dynamical system. We show that the score can be optimized over neural networks leading to improved interpretability and forecasting power of the estimators. Lastly, we discuss how our learning methods can be employed to evolve distributions and for long term forecasting.

Philippe Rigollet (Massachusetts Institute of Technology, USA)

The Emergence of Clusters in Self-attention Dynamics

Since their introduction in 2017, Transformers have revolutionized large language models and the broader field of deep learning. Central to this success is the groundbreaking self-attention mechanism. In this presentation, I'll introduce a mathematical framework that casts this mechanism as a mean-field interacting particle system, revealing a desirable long-time clustering behavior. This perspective leads to a trove of fascinating questions with unexpected connections to Kuramoto oscillators, sphere packing, and Wasserstein gradient flows. Primarily based on `https://arxiv.org/abs/2312.10794` as well as more recent results from our group.

Bernhard Schmitzer (Georg-August-Universität Göttingen, Germany)

Entropic Transfer Operators for Data-driven Analysis of Dynamical Systems

The transfer operator is an elegant way to capture the behaviour of a (stochastic) dynamical system as a linear operator. Spectral analysis can then in principle reveal (almost) invariant measures, cyclical behaviour, as well as separation of the dynamics into different time scales. In practice this analysis can rarely be done analytically, due to the complexity of the operator or since it may not be known in closed form. A central objective is therefore to numerically approximate this operator (or its ajoint: the Koopman operator) or to estimate it from data. In this talk we introduce a new estimation method based on entropic optimal transport and show convergence to a smoothed version of the original operator as more data becomes available. This involves an interplay between three different length scales: the discretization scale given by the data, the blur scale introduced by entropic transport, and the spatial scale of eigenfunctions of the operator.

Jiaxin Shi (Google DeepMind, UK)

Stein's Method for Modern Machine Learning: From Gradient Estimation to Generative Modeling

Stein's method, originating from Stein's seminal 1972 paper, provides a set of powerful mathematical tools for characterizing probability distributions and bounding distances between them. Despite predating the modern machine learning era, these ideas have recently gained traction in the ML community due to deep connections with diffusions, PDEs, and optimal transport, alongside rising attention on generative model applications that build on these concepts. In this talk, I will provide an overview of our latest developments of Stein's method for addressing modern machine learning challenges. I will begin by revisiting Stein's foundational identity and Barbour's generalization via stochastic processes. Next, I will introduce a general recipe for constructing practical Stein operators for characterizing constrained and discrete distributions. Then, I will show how we can apply this method to achieve a new state-of-the-art in gradient estimation—approximating gradients of an expectation with respect to distribution parameters—a problem central to training probabilistic models and policy gradient-based reinforcement learning. Finally, we will explore the implications of Stein's method to generative modeling, connecting score-based and diffusion models. We conclude by a few open questions motivated by applying Stein's method to solve generative machine learning problems.

Taiji Suzuki (University of Tokyo, Japan)

Convergence of Mean-field Langevin Dynamics and its Application to Neural Network Feature Learning

The mean-field Langevin dynamics (MFLD) is a nonlinear generalization of the gradient Langevin dynamics (GLD) that minimizes an entropy regularized convex function defined on the space of probability distributions, and it naturally arises from the optimization of two-layer neural networks via (noisy) gradient descent. In this talk, I will present the convergence result of MFLD and explain how the convergence of MFLD is characterized by the log-Sobolev inequality of the so-called proximal Gibbs measure corresponding to the current solution. Moreover, I will provide a general framework to prove a uniform-in-time propagation of chaos for MFLD that takes into account the errors due to finite-particle approximation, time-discretization, and stochastic gradient approximation. In the latter half, I will discuss the generalization error analysis of neural networks trained by MFLD. Addressing a binary classification problem, we have a general form of a test classification error bound that provides a fast learning rate based on a local Rademacher complexity analysis. By applying this general framework to the k-sparse parity problem, we demonstrate how the feature learning improves its sample complexity compared with the kernel methods.

Claudia Totzeck (Bergische Universität Wuppertal, Germany)

Adjoint-based Optimal Control with Wasserstein-2 Metric

In this talk we consider optimal control problems with state space given by the space of probability measures equipped with Wasserstein-2 metric. Due to its structure, the Wasserstein-2 metric poses several challenges when it comes to optimization and in particular the derivation of the first-order optimality system. We will consider the Wasserstein 2-metric which facilitates this derivation and the statement of the adjoint system, discuss the relation to optimization in L^2-sense and show two numerical examples that underline the feasibility of the approach.

Oliver Tse (Eindhoven University of Technology, Netherlands)

Variational Acceleration Methods in the Space of Probability Measures

Acceleration of gradient-based optimization methods is an issue of significant practical and theoretical interest, particularly in machine learning applications. Most research has focused on optimization over Euclidean spaces, but given the need to optimize over spaces of probability measures in many machine learning problems, it is of interest to investigate accelerated gradient methods in this context too.

In this talk, I will give a brief overview of variational acceleration methods in Euclidean space and describe one way of lifting these methods to the space of probability measures. I will then discuss their convergence rates under suitable assumptions on the functional to be minimized.

This talk is based on joint work with Shi Chen, Qin Li, and Stephen J. Wright.

Chaoyue Zhao (University of Washington, USA)

Provably Convergent Policy Optimization via Metric-aware Trust Region Methods

Trust-region methods based on Kullback–Leibler divergence are pervasively used to stabilize policy optimization in reinforcement learning. In this paper, we exploit more flexible metrics and examine two natural extensions of policy optimization with Wasserstein and Sinkhorn trust regions, namely Wasserstein policy optimization (WPO) and Sinkhorn policy optimization (SPO). Instead of restricting the policy to a parametric distribution class, we directly optimize the policy distribution and derive their closed-form policy updates based on the Lagrangian duality. Theoretically, we show that WPO guarantees a monotonic performance improvement, and SPO provably converges to WPO as the entropic regularizer diminishes. Moreover, we prove that with a decaying Lagrangian multiplier to the trust region constraint, both methods converge to global optimality. Experiments across tabular domains, robotic locomotion, and continuous control tasks further demonstrate the performance improvement of both approaches, more robustness of WPO to sample insufficiency, and faster convergence of SPO, over state-of-art policy gradient methods.

4.2 Posters

Vitalii Aksenov (Weierstraß-Institut für Angewandte Analysis und Stochastik, Germany)

Learning Distributions with Regularized JKO Scheme and Low-rank Tensor Decompositions

The theory of Wasserstein gradient flows gives a framework for sequentially constructing approximations to complicated distributions, which may arise from applications such as Bayesian inversion, importance sampling and generative modelling. For high dimensional problems, Eulearian methods are not usually used, because in general, they suffer from the curse of dimensionality. In the current talk, we attempt to "rehabilitate" the Eulerian methods in high dimension by using the developments in low-rank Tensor Train decomposition. The approach is based on the Fisher Information regularization of the dynamic JKO scheme. This essentially is a proximal step with respect to the Wasserstein distance and can be viewed as a generalization of the implicit gradient descent method. The nonlinear transformation of variables allows to represent the problem as a system of heat equations with nonlinear coupling in the initial and terminal condition. The system is formulated as a fixed-point equation, and we explore the possibility of solving it with accelerated methods. An ODE governing the evolution of particles can be defined with help of intermediate variables, giving a deterministic sampling algorithm for the approxinate distribution.

Fabian Altekrüger (Humboldt-Universität zu Berlin, Germany)

Posterior Sampling via Sliced MMD Flows with the Negative Distance Kernel

Maximum mean discrepancy (MMD) flows suffer from high computational costs in large scale computations. We show that MMD with Riesz kernels $K(x,y) = -|x-y|^r$, $rin(0,2)$ coincides with the MMD of its sliced version. As a consequence, the computation of gradients of MMDs can be performed in the one-dimensional setting. For $r = 1$, a simple sorting algorithm can be applied to reduce the complexity from $O(MNN^2)$ to $O((MN)log(MN))$ for two measures with M and N support points. For the implementations, we approximate the gradient of the sliced MMD by using only a finite number P of slices and show that the resulting error has complexity $O(sqrtd/P)$, where d is the data dimension. These results enable us to train generative models by approximating MMD gradient flows by neural networks even for image applications. By approximating the joint distribution of ground truth and observations, we use them for posterior sampling in inverse problems and conditional generative modelling.

Raphaël Barboni (École normale supérieure - PSL, France)

Training Infinitely Deep and Wide ResNets with Conditional Optimal Transport

We study the convergence of gradient flow for the training of deep Neural Networks. If Residual Neural Networks are a popular example of very deep architectures, their training constitutes a challenging optimization problem due notably to the non-convexity and the non-coercivity of the objective. Yet, in applications, those tasks are successfully solved by simple optimization algorithms such as gradient descent. To better understand this phenomenon, we focus here on a mean-field model of infinitely deep and arbitrarily wide ResNet, parameterized by probability measures over the product set of layers and parameters and with constant marginal on the set of layers. Indeed, in the case of shallow Neural Networks, mean field models have proven to benefit from simplified loss-landscapes and good theoretical guarantees when trained with gradient flow for the Wasserstein metric on the set of probability measures. Motivated by this approach, we propose to train our model with gradient flow w.r.t. the Conditional Optimal Transport distance: a restriction of the classical Wasserstein distance which enforces our marginal condition. Relying on the theory of gradient flows in metric spaces we first show the well-posedness of the gradient flow equation and its consistency with the training of ResNets at finite width. Performing a local Polyak–Lojasiewicz analysis, we then show convergence of the gradient flow for well-chosen initializations: if the number of features is finite but sufficiently large and the loss is sufficiently small at initialization, the gradient flow converges towards a global minimizer of the risk. This is the first result of this type for infinitely deep ResNets.

In addition, this work is the occasion to study in more detail the Conditional Optimal Transport distance, particularly its dynamic formulation. Some of our results in this direction may be interesting on their own.

Florian Beier (Technische Universität Berlin, Germany)

Multi-marginal Gromov–Wasserstein Transport and Barycenters

Gromov–Wasserstein (GW) distances are combinations of Gromov–Hausdorff and Wasserstein distances that allow the comparison of two different metric measure spaces or, more generally, gauge measure spaces. Due to their invariance under measure- and gauge-preserving transformations, they are well suited for many applications in graph and shape analysis. In this contribution, we consider the concept of multi-marginal GW transport between more than two gauge measure spaces and show its relation to GW barycenter problems. The potential of the concept is illustrated by various numerical examples.

This is joint work with Robert Beinert and Gabriele Steidl (both Technische Universität Berlin).

Giacomo Borghi (Rheinisch-Westfälische Technische Hochschule Aachen, Germany)

Optimization in Wasserstein Space via Measure-valued Agents

Optimization over probability measures is a common tasks in Machine Learning as many objects of interest like images, data clouds, or uncertainties, can be modelled as measures. In this contribution, we present a novel approach based on an interacting multi-agent system for global optimization. The measure-valued agents self-organize through a consensus-type dynamics to concentrate around a global minimum of the objective functional. Numerically, the dynamics is implemented via efficient optimal transport solvers. We will also discuss an application to variational inference problems and future research directions.

Joint work with Michael Herty, Andrey Stavitskiy and José Carrillo.

Jonas Bresch (Technische Universität Berlin, Germany)

Rényi-regularized Optimal Transport

Regularizing the optimal transport problem for given marginals $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and a distance function c using the Shannon entropy (or, equivalently: the Kullback–Leibler divergence) is well known. We regularize using the family of α-Rényi-divergences for $\alpha \in (0, 1)$, which recovers the Kullback–Leibler divergence in the limit $\alpha \nearrow 1$. This approach is novel in part because the α-Rényi divergence is not even an f-divergence and not separable with respect to the entries of the matrix it is applied to. We use a mirror descent algorithm for the primal formulation and a subgradient method for solving the dual problem. We show that both on real and synthetic data sets Rényi-regularized optimal transport plans outperform their Kullback–Leibler regularized or Tsallis entropy regularized counterparts. Additionally, we show that in the dual problem the constraint that the sum of the dual potentials is smaller than the cost function is not penalized in a soft way like for Kullback-Leibler divergence-regularized OT, but instead in a hard way, which is why we prefer to solve the primal problem instead.

Maciej Buze (Heriot-Watt University, UK)

Anisotropic Power Diagrams for Modelling of Polycrystalline Materials: Efficient Generation of Curved Grains via Optimal Transport

The microstructure of metals and foams is often modelled using power diagrams, a general class of tessellations which includes the well-known Voronoi diagrams. While power diagram-based approaches can generate complex microstructures in a matter of seconds and require a relatively small number of parameters, the idealised grains they produce are inherently unrealistic - they have straight boundaries and any spatial anisotropy they possess is solely determined by the relative location of seed points of neighbouring grains and not by the preferred growth directions of each grain. Curved boundaries and control over the anisotropy of individual grains can be achieved by employing anisotropic power diagrams (APDs), with several promising APD-based approaches explored in recent years by various authors. One obstacle in the wider adoption of APDs as a practical tool for modelling the microstructure of metals is the computational cost of generating them. Known efficient methods for generating power diagrams do not translate to the anisotropic setup and known techniques for generating APDs are drastically slower - while the usual runtime to generate a power diagram with grains of given volumes is (tens of) seconds, for APDs it ranges from (tens of) minutes to (tens of) hours. In my presentation, I will begin by providing a brief overview of (anisotropic) power diagram methods in geometric modelling of polycrystalline materials and subsequently present a novel approach to generating APDs with prescribed statistical properties, in which we combine semi-discrete optimal transport techniques with modern GPU-oriented computational tools, originally developed for the Sinkhorn algorithm. Our method succeeds in bringing the runtime to generate optimal APDs down to (tens of) seconds, which is fast enough to be used, e.g. in computational homogenisation. I will finish by showcasing the speed and the versatility of our method with several examples, including ones based on Electron Backscatter Diffraction (EBSD) measurements provided by our industrial partner, Tata Steel.

This is joint work with David Bourne (Heriot-Watt), Steve Roper (Glasgow), Jean Feydy (Inria Paris) and Karo Sedighiani (Tata Steel).

Giacomo Cristinelli (University of Twente, Netherlands)

Conditional Gradients for Total Variation Regularization with PDE Constraints: A Graph-cuts Approach

Sparse optimization methods constitute a powerful paradigm in a variety of applications including inverse problems and machine learning, promoting both computational efficiency and interpretability in the outcomes. Within this framework, generalized conditional gradient methods are particularly relevant, exploiting the convex geometry of a regularization term to approximate minimizers as conical combinations of simple objects. In this paper, we present a conditional gradient method for a broad class of minimization problems with total variation regularization, comprising both continuous and discretized control spaces. We give a detailed account of the practical realization of such a geometry-exploiting algorithm, delving into the analysis of the extremal points of the respective total variation unit balls and employing an efficient max-flow min-cut method to address the resulting linear sub-problem.

Leonardo Del Grande (University of Twente, Netherlands)

A General Theory for Exact Sparse Representation Recovery in Convex Optimization

In this contribution we analyze the recovery of the sparse representation of data in general infinite-dimensional optimization problems regularized by convex functionals. By assuming a suitable non-degeneracy condition on the problem we establish that, for small regularization parameters and noise levels, the minimizer is unique and is uniquely represented as a linear combination of n extreme points of the ball of the regularizer. Such non-degeneracy condition extends the classical non-degeneracy source condition (NDSC) for total variation regularized inverse problems introduced by Duval and Peyré. More precisely, it is connected to the behaviour of the solution of the dual problem when evaluated on the set of extreme points of the ball of the regularizer, seen as a metric space. This justifies the name Metric Non-Degenerate Source Condition (MNDSC). Finally, we obtain explicit formulations of the MNDSC, which lead us to specific results of sparse recovery for three problems of interest:

- Total variation regularized deconvolution problems, where we show that the classical NDSC implies our MNDSC;
- 1-dimensional BV functions regularized with their BV-seminorm;
- Pairs of measures regularized with their mutual 1-Wasserstein distance.

A lot of other problems can be explored applying this result, particularly those considering optimal transport-type regularizers, as presented in our last example. Of particular interest for future works are dynamic problems regularized by the Benamou–Brenier energy, where extreme points have been already identified. This contribution is based on the preprint (https://arxiv.org/abs/2311.08072), joint work with Marcello Carioni.

K. Bredies, M. Carioni, S. Fanzon, and F. Romero. *On the extremal points of the ball of the Benamou–Brenier energy.* Bulletin of the London Mathematical Society, 53(5):1436–1452, 2021.

V. Duval and G. Peyré. *Exact support recovery for sparse spikes deconvolution.* Foundations of Computational Mathematics, 15(5):1315–1355, 2014.

M. Carioni and L. Del Grande. *A General Theory for Exact Sparse Representation Recovery in Convex Optimization.* arXiv:2311.08072, 2023.

Sebastian Gutiérrez Hernández (Georgia Institute of Technology, USA)

Control Problems in Wasserstein Space via Conditional Flow Matching

The space of squared integrable probability distributions in $\mathbb{R}^d$, equipped with the Wasserstein Metric, forms a manifold structure that becomes a rich domain for studying dynamical systems. Hamiltonian systems, a natural outcome in the solutions of mechanical systems and optimal control, find a natural extension in the form of Wasserstein Hamiltonian flows. However, this frontier remains underdeveloped, and computational techniques are scarce. In this poster, we propose a numerical framework for control problems within Wasserstein space. We combine two techniques, generalized Conditional Flow Matching and Parametric Wasserstein Hamiltonian Flows.

Samuel Howard (University of Oxford, UK)

Learning Transport Costs through Monge Map Parameterization

Within the field of optimal transport (OT), the choice of ground cost is crucial to ensuring that the optimality of a transport map corresponds to usefulness in real-world applications. Existing approaches in computational OT typically use the squared-Euclidean cost which, although convenient due to well-characterised transport maps, is often insufficiently tailored to the given problem. It is therefore desirable to learn more suitable cost functions from data. We introduce a novel approach for learning ground costs through optimising the resulting Monge map directly. We consider a class of ground costs consisting of a composition of convex and invertible neural networks, which enables direct parameterizations of the corresponding Monge map. By optimising such parameterizations, we can simultaneously learn both an OT map and a corresponding adapted cost function through a simple and amenable optimisation procedure.

Joint work with James Thornton, George Deligiannidis, Patrick Rebeschini.

Zhao Jiaxi (National University of Singapore)

Scaling Limits of the Wasserstein Information Matrix on Gaussian Mixture Models

We consider the Wasserstein metric on the Gaussian mixture models (GMMs), which is defined as the pullback of the full Wasserstein metric on the space of smooth probability distributions with finite second moment. It derives a class of Wasserstein metrics on probability simplices over one-dimensional bounded homogeneous lattices via a scaling limit of the Wasserstein metric on GMMs. Specifically, for a sequence of GMMs whose variances tend to zero, we prove that the limit of the Wasserstein metric exists after certain renormalization. Generalizations of this metric in general GMMs are established, including inhomogeneous lattice models whose lattice gaps are not the same, extended GMMs whose mean parameters of Gaussian components can also change, and the second-order metric containing high-order information of the scaling limit. We further study the Wasserstein gradient flows on GMMs for three typical functionals: potential, internal, and interaction energies. Numerical examples demonstrate the effectiveness of the proposed GMM models for approximating Wasserstein gradient flows.

Ilja Klebanov (Freie Universität Berlin, Germany)

QMC and Sparse Grids beyond Uniform Distributions on Cubes

While Monte Carlo and MCMC methods are generally applicable and have a dimension-independent convergence rate, this rate is rather slow and unfeasible for many applications. Sparse grids and Quasi Monte Carlo methods provide better convergence rates under certain assumptions, but have only been constructed for uniform distributions on cubes and several other very specific distributions such as Gaussians. In this talk, I will show how these methods can be generalized to mixtures of such specific distributions, e.g. Gaussian mixtures, by means of a properly constructed transport map, which is a crucial step towards combining QMC and sparse grids methods with state of the art importance sampling algorithms, that are often based on such mixtures. I will avoid technical details and present lots of illustrations and videos instead.

Fabius Krämer (Universität Bonn, Germany)

On Computation and Analysis of the Volume Constrained MBO Scheme

The MBO scheme has long been known as a robust scheme for simulating mean curvature flow. Nowadays, discrete versions on graphs are also becoming increasingly popular, for example to efficiently find clusters in networks. Using optimal transport theory, it is proven that the discrete, graph version converges back to the original, continuous version of the MBO scheme. We focus on the volume-constrained MBO scheme, where the number of nodes in each cluster is fixed in each iteration. The volume constraints naturally raise a discrete optimization problem that was previously solved approximately with a black-box algorithm. We give the problem a new geometric interpretation and describe an efficient algorithm to solve the problem exactly. Furthermore, we establish a connection to an L^2-estimate for the Lagrange multipliers corresponding to the optimization problem. This result improves the runtime analysis by a factor $O(\sqrt{h})$, where h is the diffusion time in each step of the MBO scheme.

The presentation is based on joint work with Tim Laux.

Chun Yin Lam (Universität Münster, Germany)

Variational Convergence of Exchange-driven Stochastic Particle Systems

We consider the thermodynamic limit of mean-field stochastic particle systems on a complete graph. The evolution of occupation number at each vertex is driven by particle exchange with its rate depending on the population of the starting vertex and the destination vertex, including zero-range and misanthrope process. We show that under a detailed balance condition and suitable growth assumptions on the exchange rate, the evolution equation of the law of the particle density can be seen as a generalised gradient flow equation related to the large deviation rate functional.

We show the variational convergence of the gradient structures based on the energy dissipation principle, which coincides with the large deviation rate function of the finite system. The convergence of the system in this variational sense is established based on compactness of the density and flux and Γ-lower-semicontinuity of the energy dissipation functional along solutions to the continuity equation. The driving free energy Γ-converges in the thermodynamic limit, after taking possible condensation phenomena into account.

Robert Lasarzik (Weierstraß-Institut für Angewandte Analysis und Stochastik, Germany)

Minimizing Movement Scheme for Non-gradient Flows

On this poster, we show how an adapted version of the minimizing movement scheme can be applied to get an approximation scheme for general thermodynamically consistent systems. This helps to prove existence of so-called energy-variational solutions for a class of highly nonlinear PDEs without a gradient-flow structure. Some preliminary results suggest that this scheme can also provide reasonable novel numerical schemes.

Razvan-Andrei Lascu (Heriot-Watt University & Maxwell Institute for Mathematical Sciences, UK)

Entropic Mean-field Min-Max Problems via Fisher–Rao and Best Response Flows

We investigate convergence properties of two continuous-time optimization methods, the Fisher–Rao (Mean-Field Birth-Death) and Mean-Field Best Response flows, for solving convex-concave min-max games with entropy regularization. We introduce suitable Lyapunov functions to establish exponential convergence to the unique mixed Nash equilibrium for both methods, albeit under slightly different conditions. Additionally, we derive time-discretizations for these flows, and prove convergence rates for the discrete-time schemes.

Gergely Neu (Universitat Pompeu Fabra, Spain)

Computing Optimal Transport Distances for Markov Chains via Linear Programming

We propose a new framework for formulating optimal transport distances between Markov chains. Previously known formulations studied couplings between the entire joint distribution induced by the chains, and derived solutions via a reduction to dynamic programming (DP) in an appropriately defined Markov decision process. This formulation has, however, not led to particularly efficient algorithms so far, since computing the associated DP operators requires fully solving a static optimal transport problem, and these operators need to be applied numerous times during the overall optimization process. In this work, we develop an alternative perspective by considering couplings between a "flattened" version of the joint distributions that we call discounted occupancy measures, and show that calculating optimal transport distances in the full space of joint distributions can be equivalently formulated as solving a linear program (LP) in this reduced space. This LP formulation formulation allows us to port several algorithmic ideas from other areas of optimal transport theory. In particular, our formulation makes it possible to introduce an appropriate notion of entropy regularization into the optimization problem, which in turn enables us to directly calculate optimal transport distances via a Sinkhorn-like method we call Sinkhorn Value Iteration (SVI). The LP formulation also allows us to develop a dual optimization problem that can be solved via stochastic optimization. Besides providing theoretical analysis to these methods, we experimentally compare them with some common heuristics for computing approximate "bisimulation metrics" (a peculiar notion of OT distances that is popular within the reinforcement-learning literature).

This is joint work with Sergio Calo, Anders Jonsson, and Javier Segovia.

Filippo Quattrocchi (Institute for Science and Technology Austria)

Gradient Flow Interpretation of the Fokker–Planck equation with Dirichlet Boundary Conditions

A classical finding by Jordan, Kinderlehrer, and Otto (1998) is that certain evolutionary PDEs with homogeneous Neumann boundary conditions can be seen as the gradient flow of an entropy functional in the space of probability measures endowed with the Wasserstein transportation metric. It has been found by Figalli and Gigli (2010) that also the heat equation with constant Dirichlet boundary conditions can be similarly interpreted as a gradient flow by taking a suitable modification of the Wasserstein distance. Later, Morales (2018) proved an analog theorem for the case of nonconstant Dirichlet boundary conditions and more general equations. I present these results along with some developments aimed at relaxing hypotheses (on the domain, on the initial datum, on the drift) and obtaining, in dimension 1, the EDI (energy dissipation inequality) gradient flow formulation for the Fokker–Planck equation with Dirichlet boundary conditions.

Michael Quellmalz (Technische Universität Berlin, Germany)

Sliced Optimal Transport on the Sphere

Sliced optimal transport is a powerful technique that simplifies optimal transport on high-dimensional domains by reducing it to transport on the real line. This is achieved through the concatenation of the Radon transform and the cumulative density transform, which analytically yields the solutions of the reduced transport problems and can be computed efficiently. Inspired by this concept, we analyze and compare two adaptions to measures on the sphere: firstly, the parallel slice transform, which can be interpreted as restriction of the Radon transform, integrates along all hyperplane sections, and, secondly, the semicircle transform integrates along all half great circles. We show that both notations yield sliced Wasserstein distances that are actually metrics. In case of parallel slicing, this metric is topologically equivalent to the classical Wasserstein distance, but can be computed much faster. To demonstrate the applicability of both approaches, we provide numerical examples dealing with barycenters and classifications of spherical probability measures. Our implementations rely on the singular value decompositions of both transforms and fast Fourier techniques.

This is joint work with Robert Beinert, Léo Buecher, Gabriele Steidl (Technische Universität Berlin).

Pratik Rai (Technical University of Eindhoven, Netherlands)

State Estimation with Wasserstein Reduced Models

This is a joint work with Olga Mula. In this work, we are concerned with the problem of optimal recovery of solution states to convection-dominated parametrized PDEs posed on the 2-Wasserstein space, through finitely many observations. To this end, we, first, reduce the complexity of the problem by approximating the parametric PDE solution manifold through a n-dimensional nonlinear space generated by a set of Wasserstein barycenters. This allows for a collective estimation of the elements of the solution manifold. We, then, design a recovery algorithm, based on barycenters, that recovers the solution states from the data collected through a set of sensors, in the form of linear functional evaluations. The performance of the recovery algorithm relies on a *stability* constant, that implicitly depends on the sensor locations, and the worst case error, which is attained when approximating the solution manifold. We, therefore, optimize over the sensor locations and provide bounds on the stability constant to ensure a reliable recovery of the solution state. Finally, we validate our algorithm for robustness against numerous parametric PDEs.

Julia Sanders (University of Helsinki, Finland)

Optimal Control of Underdamped Systems: Numerical Predictions

Optimal control theory finds protocols that steer a system from an assigned initial condition to an assigned final condition, while minimizing a trajectory-dependent cost function. Motivated by the design of nano-scale electronic components, where random noise and inertial effects cannot be ignored, we are interested in optimal control of underdamped systems. We introduce a physics-inspired, multi-scale perturbation theory around the overdamped limit. We use an iterative procedure to compute solutions in the overdamped dynamics, and, with the analytic results, we can then make predictions in the underdamped dynamics. In particular, we compute predictions for the momentum cumulants, which are observable experimentally. Furthermore, we develop a numerical algorithm directly minimizing the cost function by a gradient descent, and include some preliminary results in the overdamped dynamics.

Joint work with Marco Baldovin (CNR, Rome) and Paolo Muratore-Ginanneschi (University of Helsinki).

Stefan Schrott (Universität Wien, Austria)

The Wasserstein Space of Stochastic Processes in Continuous Time

Researchers from different areas have independently defined extensions of the usual weak topology between laws of stochastic processes. This includes Aldous' extended weak convergence, Hellwig's information topology and convergence in adapted distribution in the sense of Hoover–Keisler. We show that on the set of continuous processes with canonical filtration these topologies coincide and are metrized by a suitable *adapted Wasserstein distance* $\mathcal{AW}$. Moreover, we show that the resulting topology is the weakest topology that guarantees continuity of optimal stopping. While the set of processes with canonical filtration is not complete, we establish that its completion consists precisely in the space FP of stochastic processes with a general filtration. We also observe that $(\mathrm{FP}, \mathcal{AW})$ exhibits several desirable properties. Specifically, it is Polish, martingales form a closed subset and approximation results like Donsker's theorem extend to $\mathcal{AW}$.

Viktor Stein (Technische Universität Berlin, Germany)

Wasserstein Gradient Flows for Moreau Envelopes of f-divergences in Reproducing Kernel Hilbert Spaces

Most commonly used f-divergences of measures, e.g., the Kullback–Leibler divergence, are subject to limitations regarding the support of the involved measures. A remedy consists of regularizing the f-divergence by a squared maximum mean discrepancy (MMD) associated with a characteristic kernel K. In this paper, we use the so-called kernel mean embedding to show that the corresponding regularization can be rewritten as the Moreau envelope of some function in the reproducing kernel Hilbert space associated with K. Then, we exploit well-known results on Moreau envelopes in Hilbert spaces to prove properties of the MMD-regularized f-divergences and, in particular, their gradients. Subsequently, we use our findings to analyze Wasserstein gradient flows of MMD-regularized f-divergences. Finally, we consider Wasserstein gradient flows starting from empirical measures and provide proof-of-the-concept numerical examples with Tsallis-$alpha$ divergences. Joint work with Sebastian Neumayer (TU Chemnitz) and Gabriele Steidl (TU Berlin) available at https://arxiv.org/abs/2402.04613.

Hong Ye Tan (University of Cambridge, UK)

Noise-free Sampling Algorithms via Regularized Wasserstein Proximals

We consider the problem of sampling from a distribution governed by a potential function. This work proposes an explicit score-based MCMC method that is deterministic, resulting in a deterministic evolution for particles rather than a stochastic differential equation evolution. The score term is given in closed form by a regularized Wasserstein proximal, using a kernel convolution that is approximated by sampling. We demonstrate fast convergence on various problems and show improved dimensional dependence of mixing time bounds for the case of Gaussian distributions compared to the unadjusted Langevin algorithm (ULA) and the Metropolis-adjusted Langevin algorithm (MALA). We additionally derive closed form expressions for the distributions at each iterate for quadratic potential functions, characterizing the variance reduction. Empirical results demonstrate that the particles behave in an organized manner, lying on level set contours of the potential. Moreover, the posterior mean estimator of the proposed method is shown to be closer to the maximum a-posteriori estimator compared to ULA and MALA, in the context of Bayesian logistic regression.

Viet Hoang Tran (National Universterty of Singapore)

Revisiting Over-smoothing and Over-squashing Using Ollivier–Ricci Curvature

Graph Neural Networks (GNNs) had been demonstrated to be inherently susceptible to the problems of over-smoothing and over-squashing. These issues prohibit the ability of GNNs to model complex graph interactions by limiting their effectiveness in taking into account distant information. Our study reveals the key connection between the local graph geometry and the occurrence of both of these issues, thereby providing a unified framework for studying them at a local scale using the Ollivier-Ricci curvature. Specifically, we demonstrate that oversmoothing is linked to positive graph curvature, while over-squashing is linked to negative graph curvature. Based on our theory, we propose the Batch Ollivier–Ricci Flow, a novel rewiring algorithm capable of simultaneously addressing both over-smoothing and over-squashing.

Magnus Tronstad (Karolinska Institute & Karolinska University Hospital, Sweden)

Modelling Trajectories of Cell Development in Single Cell Sequencing Data as an Optimal Transport Problem with Multiple Stages

A key challenge in biological research is to develop more complete models of cell development. These models often assume that a cell's "state" is governed by some (unknown) dynamical system, so that its development can be represented mathematically as movement in state-space. Since the advent of single cell sequencing technologies, it is possible to obtain "snapshots" of the different cell states present in a biological sample, represented by gene expression levels. In certain systems, such as haematopoiesis (blood cell formation), cell development is a continuously ongoing process marked by the presence of well-defined initial and terminal cell states. Thus, in such systems, single cell sequencing captures cell states from a range of different developmental stages. A key problem when analyzing such data is to determine the most likely transitions from initial to terminal states. Previously, an extended optimal mass transport problem formulation that optimizes over multiple constrained marginals has successfully been applied in state estimation and tracking of ensembles of agents. In our work, we develop and extend this framework for modelling cell development as a discrete optimal mass transport problem, with marginals representing developmental stages. The optimal transport couplings can be interpreted in terms of cell-cell transition probabilities in a stationary absorbing Markov chain, which in turn can be used to compute the expected terminal cell state for a given intermediate cell.

This is joint work with Johan Karlsson (Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital) and Joakim Dahlin (Department of Mathematics, KTH Royal Institute of Technology), who contributed equally.

Linfeng Wang (King's College London, UK)

Measure Transport with Kernel Mean Embeddings

Kalman filters constitute a scalable and robust methodology for approximate Bayesian inference, matching first and second order moments of the target posterior. To improve the accuracy in nonlinear and non-Gaussian settings, we extend this principle to include more or different characteristics, based on kernel mean embeddings (KMEs) of probability measures into their corresponding Hilbert spaces. Focusing on the continuous-time setting, we develop a family of interacting particle systems (termed KME-dynamics) that bridge between the prior and the posterior, and that include the Kalman–Bucy filter as a special case. A variant of KME-dynamics has recently been derived from an optimal transport perspective by Maurais and Marzouk, and we expose further connections to (kernelised) diffusion maps, leading to a variational formulation of regression type. Finally, we conduct numerical experiments on toy examples and the Lorenz-63 model, the latter of which show particular promise for a hybrid modification (called Kalman-adjusted KME-dynamics).

Jia-Jie Zhu (Weierstraß-Institut für Angewandte Analysis und Stochastik, Germany)

Kernelization, Wasserstein, and Entropy-dissipation of Gradient Flows: From Wasserstein to Fisher–Rao

Motivated by various machine learning applications, we present a principled investigation of gradient flow dissipation geometry, emphasizing the Fisher-Rao type gradient flows and the connections with Wasserstein space. Using the dynamic Benamou–Brenier formulation, we reveal a few precise connections between those flow dissipation geometries and commonly used machine learning tools such as Stein flows, kernel discrepancies, and nonparametric regression. In addition, we present analysis results in terms of Łojasiewicz type functional inequalities, with an explicit threshold condition for a family of entropy dissipation along the Fisher–Rao flows. Finally, we establish rigorous evolutionary Γ-convergence for the Fisher-Rao type gradient flows obtained via regression, justifying the approximation beyond static point-wise convergence.

Joint work with Alexander Mielke.

5 List of Participants

Aksenov, Vitalii
Weierstraß-Institut
Berlin, Germany

Altekrüger, Fabian
Humboldt-Universität zu Berlin
Germany

Barboni, Raphaël
École normale supérieure - PSL
Paris, France

Bayer, Christian
Weierstraß-Institut
Berlin, Germany

Beier, Florian
Technische Universität Berlin
Germany

Beinert, Robert
Technische Universität Berlin
Germany

Bi, Hangcheng
Georg-August-Universität Göttingen
Germany

Bocchi, Gabriele
Università degli Studi di Roma Tor Vergata
Italy

Borghi, Giacomo
RWTH Aachen
Germany

Borkowski, Alexandra
King's College London
UK

Bresch, Jonas
Technische Universität Berlin
Germany

Brigati, Giovanni
Institute of Science and Technology Austria
Klosterneuburg

Butkovsky, Oleg
Weierstraß-Institut
Berlin, Germany

Buze, Maciej
Harriot-Watt University
Edinburgh, UK

Caiazzo, Alfonso
Weierstraß-Institut
Berlin, Germany

Calo Oliveira, Sergio
Universitat Pompeu Fabra
Barcelona, Spain

Carioni, Marcello
University of Twente
Enschede, Netherlands

Carlier, Guillaume
Université Paris Dauphine
France

Chapel, Laetitia
Université Bretagne Sud
Vannes, France

Chemseddine, Jannis
Technische Universität Berlin
Germany

Cristinelli, Giacomo
University of Twente
Enschede, Netherlands

David, Adam
Technische Universität Berlin
Germany

Deb, Ahana
Universitat Pompeu Fabra
Barcelona, Spain

Del Grande, Leonardo
University of Twente
Enschede, Netherlands

Delon, Julie
Université Paris-Cité
France

Dvurechensky, Pavel
Weierstraß-Institut
Berlin, Germany

Ehrlacher, Virginie
École des Ponts ParisTech & INRIA
France

Eigel, Martin
Weierstraß-Institut Berlin
Germany

Gonnella, Isabella Carla
SISSA
Trieste, Italy

Gladin, Egor
Humboldt-Universität zu Berlin
Germany

Gutiérrez Hernández, Sebastian
Georgia Institute of Technology
Atlanta, USA

Heinze, Georg
Weierstraß-Institut Berlin
Germany

Howard, Samuel
University of Oxford
UK

Jiaxi, Zhao
National University of Singapore

Kachaiev, Oleksii
Prague, Czech Republic

Khan, Emtiyaz
RIKEN Tokyo
Japan

Klebanov, Ilja
Freie Universität Berlin
Germany

Korba, Anna
ENSAE & CREST
Paris, France

Krämer, Fabius
Universität Bonn
Germany

Kroshnin, Aleksei
Weierstraß-Institut Berlin
Germany

Kuhn, Daniel
École Polytechnique Fédérale de Lausanne
Switzerland

Lam, Chun Yin
Universität Münster
Germany

Lasarzik, Robert
Weierstraß-Institut Berlin
Germany

Laschos, Vaios
Weierstraß-Institut Berlin
Germany

Lascu, Razvan-Andrei
Harriot-Watt University & Maxwell Institute
Edinburgh, UK

Laux, Tim
Universität Regensburg
Germany

Li, Mengmeng
École Polytechnique Fédérale de Lausanne
Switzerland

Li, Wuchen
University of South Carolina
Columbia, USA

Li, Xue-Mei
École Polytechnique Fédérale de Lausanne
Switzerland

Liero, Matthias
Weierstraß-Institut Berlin
Germany

Maas, Jan
Institute of Science and Technology Austria
Klosterneuburg

Manggala, Putra
Amazon.com, Inc.
Berlin, Germany

Martin, Ségolène
Technische Universität Berlin
Germany

Melnyk, Oleh
Technische Universität Berlin
Germany

Mielke, Alexander
Weierstraß-Institut
Berlin, Germany

Monina, Nataliia
University of Ottawa
Canada

Mula, Olga
Eindhoven University of Technology
Netherlands

Namkoong, Hongseok
Columbia University
New York, USA

Neu, Gergely
Universitat Pompeu Fabra
Barcelona, Spain

Park, Mi Jung
Technical University of Denmark
Lyngby

Peletier, Mark
Eindhoven University of Technology
Netherlands

Penka, Maximilian
Technische Universität München
Germany

Peyré, Gabriel
CNRS & ENS
Paris, France

Pontil, Massimiliano
University College London
UK

Portinale, Lorenzo
Universität Bonn
Germany

Quattrocchi, Filippo
Institute for Science and Technology Austria
Klosterneuburg

Quellmalz, Michael
Technische Universität Berlin
Germany

Rai, Pratik
Eindhoven University of Technology
Netherlands

Rigollet, Philippe
Massachusetts Institute of Technology
Cambridge, USA

Romor, Francesco
Weierstraß-Institut
Berlin, Germany

Ruiz-Balet, Domènec
Imperial College London
UK

Sanders, Julia
University of Helsinki
Finland

Schmitz, Leonard
Universität Greifswald
Germany

Schmitzer, Bernhard
Georg-August-Universität Göttingen
Germany

Schrott, Stefan
Universität Wien
Austria

Schütte, Janina
Weierstraß-Institut
Berlin, Germany

Shalova, Anna
Eindhoven University of Technology
Netherlands

Sharma, Akash
Chalmers University of Technology
Sweden

Shi, Jiaxin
Google DeepMind
London, UK

Steidl, Gabriele
Technische Universität Berlin
Germany

Stein, Viktor
Technische Universität Berlin
Germany

Sunkara, Vikram
Zuse-Institut Berlin
Germany

Suvorikova, Alexandra
Weierstraß-Institut
Berlin, Germany

Suzuki, Taiji
University of Tokyo
Japan

Tan, Hong Ye
University of Cambridge
UK

Tanguy, Eloi
Université Paris-Cité & Laboratoire MAP5
France

Totzeck, Claudia
Bergische Universität Wuppertal
Germany

Tran, Viet Hoang
National University of Singapore

Tronstad, Magnus
Karolinska Institute & Karolinska U Hospital
Stockholm, Sweden

Tse, Oliver
Eindhoven University of Technology
Netherlands

Vermeulen, Joop
Eindhoven University of Technology
Netherlands

Wagner, Alfred
RWTH Aachen
Germany

Wald, Christian
Technische Universität Berlin
Germany

Walter, Daniel
Humboldt-Universität zu Berlin
Germany

Wang, Linfeng
King's College London
UK

Zhao, Chaoyue
University of Washington
Seattle, USA

Zhu, Jia-Jie
Weierstraß-Institut
Berlin, Germany