Seminar on Statistical Optimal Transport:

Foundations, Computation, and Statistics

Winter term 2023/24

Prof. Dr. Axel Munk and Shayan Hundrieser

Key information

Time:	October 27th, 2023 – February 09th, 2024, on Fridays, 10.15 – 11.45
Format:	in person, room 5.101 (Institute for Mathematical Stochastics)
Possible Modules:	B.Mat.3441: Seminar on applied and mathematical stochastics
	M.Mat.4841: Seminar on applied and mathematical stochastics
	B.Mat.3444: Seminar on mathematical statistics
	M.Mat.4844: Seminar on mathematical statistics
	B.Mat.3447: Seminar on statistical foundations of data science
	M.Mat.4847: Seminar on statistical foundations of data science
Instructors:	Prof. Dr. Axel Munk and Shayan Hundrieser
Intended Audience:	**Advanced Bachelor and beginning Master students**
Language:	English

Prerequisites

Participants must have successfully attended: Analysis I (B.Mat.0011), Analysis II (B.Mat.0021), Analytische Geometrie und Lineare Algebra I (B.Mat.0012), Maß- und Wahrscheinlichkeitstheorie (B.Mat.1400). Additionally, it is recommended that students have successfully attended a lecture on mathematical statistics or statistical data science.

Description

In recent years, optimal transport based data analysis has garnered immense interest for its unparalleled ability in quantifying the dissimilarity between probability measures in a way that aligns with the geometry of their ground space. The mathematical foundations of optimal transport are deep and rooted in the fields of probability, geometry, analysis, and optimization. Its profound significance has been recognized by several prestigious awards, including multiple Fields medals (2014, 2018) and the Nobel prize in economy (1975). Applications include econometrics, image retrieval, and more recently also statistics, machine learning and biology.

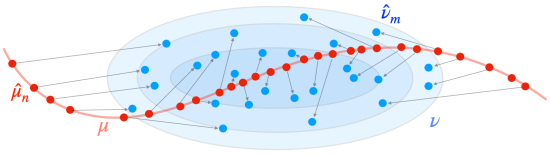

Figure: Two probability measures μ and ν, empirical measures $\hat{\mu}_n$ and $\hat{\nu}_n$, and a matching induced by optimal transport.

The focus of this seminar will be on carefully understanding basic mathematical principles, computational aspects, and statistical theory. Initial contents cover several basic topics on optimal transport, including the Monge problem, the Kantorovich relaxation along with its dual formulation, geometric facets of optimal transport in Euclidean spaces, the stability of transport plans, the concept of complementary slackness, the Wasserstein space along with its geodesics, and the notion of Wasserstein-Barycenters. On the computational front, we will delve into methods such as the Auction algorithm, the minimum-cost flow problem along with the network simplex algorithm,

as well as fast computational schemes for entropy-penalized optimal transport. This will lay the foundations for the statistical theory of optimal transport. We will explore crucial aspects such as minimax optimal rates of estimation for both the optimal transport cost and map, addressing challenges posed by the curse of dimensionality for high-dimensional data analysis, and investigate distributional limits. Certain real world applications will also be discussed. The seminar sets the foundation for possible Master thesis topics on statistical optimal transport.

Application and admission

To provide participants with the material to be presented at an early stage, we ask you to pre-register for this seminar until **Sunday, October 1st**. To this end, please email Anja Rentzsch (anja.rentzsch@uni-goettingen.de) and indicate your interest to give a seminar talk. Please include information about relevant courses you have taken in your email. In particular, recall the prerequisites mentioned above.

A preparatory virtual meeting, during which topics will be assigned to participating students, is scheduled for **Monday, October 2nd (2:00pm–3:30pm)**. The seminar is limited to 13 participants. Should preregistrations exceed 13, then participants will be chosen based on the information provided in their preregistration email. The link for the meeting is given by:

`https://uni-goettingen.zoom-x.de/j/63722842228`

Recommended literature

Main References

Presentations will be based on a selection of chapters of multiple references, all of which are fully accessible from within the university network, references will include:

- *"An Invitation to Statistics in Wasserstein Space"*, V.M. Panaretos and Y. Zemel (2020), Springer, `https://link.springer.com/book/10.1007/978-3-030-38438-8`.
- *"Computational Optimal Transport"*, G. Peyré and M. Cuturi (2019), Foundations and Trends® in Machine Learning, `https://arxiv.org/abs/1803.00567`.
- *"Optimal Transport: Fast Probabilistic Approximation with Exact Solvers"*, M. Sommerfeld, J. Schrieber, Y. Zemel, and A. Munk (2019), Journal of Machine Learning Research, `https://www.jmlr.org/papers/v20/18-079.html`.
- *"Optimal Estimation of Wasserstein Distance on a Tree With an Application to Microbiome Studies"* S. Wang, T. Cai, and H. Li (2020), Journal of the American Statistical Association, `https://www.tandfonline.com/doi/full/10.1080/01621459.2019.1699422`.
- *"Distribution-on-distribution regression via optimal transport maps"* L. Ghodrati and V.M. Panaretos (2022), Biometrika, `https://academic.oup.com/biomet/article-abstract/109/4/957/6515608`.

Background material from probability and statistics

- *"Probability and Measure"*, P. Billingsley (1999), Wiley.
- *"Empirical Processes with Applications to Statistics"*, G.A. Shorack and J.A. Wellner (1986), SIAM.
- *"Asymptotic Statistics"*, A.W. van der Vaart (1998), Cambridge University Press.
- *"Statistical Foundations of Data Science"*, A. Munk (2022/23), Lecture Notes IMS.