Measure-theoretic Approaches and Optimal Transportation in Statistics

Geometry and Statistics in Data Science
21-25 November 2022 - IHP, Paris

Monday

Arthur Gretton
KALE Flow: A Relaxed KL Gradient Flow for Probabilities with Disjoint Support.

We study the gradient flow for a relaxed approximation to the Kullback-Leibler (KL) divergence between a moving source and a fixed target distribution. This approximation, termed the KALE (KL approximate lower-bound estimator), solves a regularized version of the Fenchel dual problem defining the KL over a restricted class of functions. When using a Reproducing Kernel Hilbert Space (RKHS) to define the function class, we show that the KALE continuously interpolates between the KL and the Maximum Mean Discrepancy (MMD). Like the MMD and other Integral Probability Metrics, the KALE remains well defined for mutually singular distributions. Nonetheless, the KALE inherits from the limiting KL a greater sensitivity to mismatch in the support of the distributions, compared with the MMD. These two properties make the KALE gradient flow particularly well suited when the target distribution is supported on a low-dimensional manifold. Under an assumption of sufficient smoothness of the trajectories, we show the global convergence of the KALE flow. We propose a particle implementation of the flow given initial samples from the source and the target distribution, which we use to empirically confirm the KALE's properties.

Blanche Buet
A varifold perspective on discrete surfaces.

Joint work with: Gian Paolo Leonardi (Trento), Simon Masnou (Lyon) and Martin Rumpf (Bonn).
We propose a natural framework for the study of surfaces and their different discretizations based on varifolds. Varifolds have been introduced by Almgren to carry out the study of minimal surfaces. Though mainly used in the context of rectifiable sets, they turn out to be

well suited to the study of discrete type objects as well. While the structure of varifold is flexible enough to adapt to both regular and discrete objects, it allows to define variational notions of mean curvature and second fundamental form based on the divergence theorem. Thanks to a regularization of these weak formulations, we propose a notion of discrete curvature (actually a family of discrete curvatures associated with a regularization scale) relying only on the varifold structure. We prove nice convergence properties involving a natural growth assumption: the scale of regularization must be large with respect to the accuracy of the discretization. We performed numerical computations of mean curvature and Gaussian curvature on point clouds in $\mathbb{R}^3$ to illustrate this approach. Building on the explicit expression of approximate mean curvature we propose, we perform mean curvature flow of point cloud varifolds beyond the formation of singularities and we recover well-known soap films.

Christophe Ley

Advances in statistics via tools from Stein's Method.

Stein's Method is becoming increasingly popular in statistics and machine learning. In this talk, I will describe how various components from the famous Stein Method, a well-known approach in probability theory for approximation problems, have been recently put to successful use in theoretical and computational statistics.

Tuesday

Théo Lacombe

On the existence of Monge maps for Gromov-Wasserstein problems.

The Gromov-Wasserstein (GW) problem provides a way to compare probability measures supported on (possibly) different spaces. It relies on a quadratic minimization problem over the transportation polytope using a cost function over each space. As for the optimal transportation (OT) problem, it is natural to study the structure of minimizers of GW, in particular to wonder whenever they are deterministic (i.e. induced by a map between the spaces). In this talk, we will characterize GW minimizers for two cost functions introduced in the literature, and prove that some of them are actual maps following some specific structure. In addition, we will provide numerical evidence for situations where the map structure does not hold, suggesting the sharpness of our assumptions.

Bodhisattva Sen

Distribution-free testing for Multivariate symmetry using Optimal Transport.

Break

Johan Segers

Graphical and uniform consistency of estimated optimal transport plans.

A general theory is provided delivering convergence of maximal cyclically monotone mappings containing the supports of coupling measures of sequences of pairs of possibly random probability measures on Euclidean space. The theory is based on the identification of such a mapping with a closed subset of a Cartesian product of Euclidean spaces and leveraging tools from random set theory. Weak convergence in the appropriate Fell space together with the maximal cyclical monotonicity then automatically yields local uniform convergence of the associated mappings. Viewing such mappings as optimal transport plans between probability measures with respect to the squared Euclidean distance as cost function yields consistency results for notions of multivariate ranks and quantiles based on optimal transport, notably the empirical center-outward distribution and quantile functions.

Gabriel Peyré

Unbalanced Optimal Transport across Metric Measured Spaces.

Optimal transport (OT) has recently gained a lot of interest in machine learning. It is a natural tool to compare in a geometrically faithful way probability distributions. It finds applications in both supervised learning (using geometric loss functions) and unsupervised learning (to perform generative model fitting). OT is however plagued by several issues, and in particular: (i) the curse of dimensionality, since it might require a number of samples which grows exponentially with the dimension, (ii) sensitivity to outliers, since it prevents mass creation and destruction during the transport (iii) impossibility to transport between two disjoint spaces. In this talk, I will review several recent proposals to address these issues, and showcase how they work hand-in-hand to provide a comprehensive machine learning pipeline. The three key ingredients are: (i) entropic regularization which defines computationally efficient loss functions in high dimensions (ii) unbalanced OT, which relaxes the mass conservation to make OT robust to missing data and outliers, (iii) the Gromov-Wasserstein formulation, introduced by Sturm and Memoli, which is a non-convex quadratic optimization problem defining transport between disjoint spaces. More information and references can be found on the website of our book "Computational Optimal Transport".

Quentin Berthet
Mirror Sinkhorn: Fast Online Optimization on Transport Polytopes.

Wednesday

Axel Munk

Transport Dependency: Optimal Transport Based Dependency Measures.

Finding meaningful ways to determine the dependency between two random variables ξ and ζ is a timeless statistical endeavor with vast practical relevance. In recent years, several concepts that aim to extend classical means (such as the Pearson correlation or rank-based coefficients like Spearman's ρ) to more general spaces have been introduced and popularized, a well-known example being the distance correlation. In this talk, we propose and study an alternative framework for measuring statistical dependency, the transport dependency $\tau \geq 0$ (TD), which relies on the notion of optimal transport and is applicable in general Polish spaces. It can be estimated via the corresponding empirical measure, is versatile and adaptable to various scenarios by proper choices of the cost function. It intrinsically respects metric properties of the ground spaces. Notably, statistical independence is characterized by $\tau = 0$, while large values of τ indicate highly regular relations between ξ and ζ . Based on sharp upper bounds, we exploit three distinct dependency coefficients with values in $[0, 1]$, each of which emphasizes different functional relations: These transport correlations attain the value 1 if and only if $\zeta = \phi(\xi)$, where ϕ is a) a Lipschitz function, b) a measurable function, c) a multiple of an isometry. Besides a conceptual discussion of transport dependency, we address numerical issues and its ability to adapt automatically to the potentially low intrinsic dimension of the ground space. Monte Carlo results suggest that TD is a robust quantity that efficiently discerns dependency structure from noise for data sets with complex internal metric geometry. The use of TD for inferential tasks is illustrated for independence testing on a data set of trees from cancer genetics. Furthermore, we illustrate other optimal transport based dependency concepts for protein colocalization in cell biology.
This is joint work with Giacomo Nies and Thomas Staudt. Based on:

- Transport dependency: Optimal transport based dependency measures, TG Nies, T Staudt, A Munk,
- Colocalization for super-resolution microscopy via optimal transport C Tameling, S Stoldt, T Stephan, J Naas, S Jakobs, A Munk.

Giovanni Peccati

Some bounds on probabilistic distances via integration by parts formulae.

I will review a recent stream of research, dealing with the control of Wasserstein-type (and more general) probabilistic distances, both in uni- and multi-variate settings, by using infinite-dimensional integration by parts formulae. I will illustrate my presentation with examples from stochastic geometry (random geometric graphs and the geometry of Gaussian fields), and evoke several connections with generalized logarithmic Sobolev and concentration estimates.

Break

Agnès Desolneux
A Wasserstein-type distance in the space of Gaussian mixture models.

In this talk, we introduce a Wasserstein-type distance on the set of Gaussian mixture models. This distance is defined by restricting the set of possible coupling measures in the optimal transport problem to Gaussian mixture models. We derive a very simple discrete formulation for this distance, which makes it suitable for high dimensional problems. We also study the corresponding multi-marginal and barycenter formulations. We show some properties and propose some possible extensions of this Wasserstein-type distance, and we illustrate its practical use with some examples in image processing. This is a joint work with Julie Delon (Université Paris Cité).

Nicolas Courty
Sliced Wasserstein on Manifolds: Spherical and Hyperbolical cases.

Many variants of the Wasserstein distance have been introduced to reduce its original computational burden. In particular the Sliced-Wasserstein distance (SW), which leverages one-dimensional projections for which a closed-form solution of the Wasserstein distance is available, has received a lot of interest. Yet, it is restricted to data living in Euclidean spaces, while the Wasserstein distance has been studied and used recently on manifolds. In this talk I will discuss novel methodologies to transpose SW to the Riemannian manifold case. By appropriately choosing a proper Radon transform, we show how fast and differentiable algorithms can be designed in two cases: Spherical and Hyperbolic manifolds. After discussing some of the theoretical properties of those novel discrepancies, I will showcase applications in machine learning problems, where data naturally live on those spaces.

Thursday

Bharath Sriperumbudur
Spectral regularized kernel two-sample test.

Jérôme Dedecker
Some bounds for the Wasserstein distance between the empirical measure and the marginal distibution of a sequence of i.i.d. random variables.

Jean-Michel Loubes
Fair Learning with Optimal Transport.

Break

Jean Feydy
Computational optimal transport: mature tools and open problems.

Optimal transport is a fundamental tool to deal with discrete and continuous distributions of points. We can understand it either as a generalization of sorting to spaces of dimension $D > 1$, or as a nearest neighbor projection under an incompressibility constraint. Over the last decade, a sustained research effort on numerical foundations has led to a x1,000 speed-up for most transport-related computations. Computing Earth Mover's Distances or Wasserstein barycenters between 3D volumes and surfaces is now a matter of milliseconds. This has opened up a wide range of research directions in geometric data analysis, machine learning and computer graphics.
This lecture will discuss the consequences of these game-changing numerical advances from a user's perspective. We will focus on:

1. Mature libraries and software tools that can be used as of 2022, with a clear picture of the current state-of-the-art.

2. New ranges of applications in 3D shape analysis, with a focus on population analysis and shape registration.

3. Open problems that remain to be solved by experts in the field.

Thibaut Le Gouic

Sampler for the Wasserstein barycenter.

Wasserstein barycenters have become a central object in applied optimal transport as a tool to summarize complex objects that can be represented as distributions. Such objects include posterior distributions in Bayesian statistics, functions in functional data analysis and images in graphics. In a nutshell a Wasserstein barycenter is a probability distribution that provides a compelling summary of a finite set of input distributions. While the question of computing Wasserstein barycenters has received significant attention, this talk focuses on a new and important question: sampling from a barycenter given a natural query access to the input distribution. We describe a new methodology built on the theory of Gradient flows over Wasserstein space. This is joint work with Chiheb Daaloul, Magali Tournus and Jacques Liandrat.

Quentin Paris

Online learning with exponential weights in metric spaces with the measure contraction property.

This paper addresses the problem of online learning in metric spaces using exponential weights. We extend the analysis of the exponentially weighted average forecaster, traditionally studied in a Euclidean settings, to a more abstract framework. Our results rely on the notion of barycenters, a suitable version of Jensen's inequality and a synthetic notion of lower curvature bound in metric spaces known as the measure contraction property. We also adapt the online-to-batch conversion principle to apply our results to a statistical learning framework.

Friday

François-Xavier Vialard
Statistical estimation of optimal transport potentials.

We show how to break the curse of dimension for the estimation of optimal transport distance between two smooth distributions for the Euclidean squared distance. The approach relies on essentially one tool: represent inequality constraints in the dual formulation of OT by equality constraints with a sum of squares in reproducing kernel Hilbert space. By showing this representation is tight in the variational formulation, one can then leverage smoothness to break the curse. However, the constants associated with the algorithm a priori scale exponentially with the dimension.

Olga Mula
Structured Prediction with sparse Wasserstein barycenters.

Probability measures are fundamental objects in numerous applications from physics and machine learning. This talk discusses different notions of sparsity that are relevant in the context of approximating and interpolating parametrized families of measures. To this end, we introduce the notion of best n-term and sparse interpolation with Wasserstein barycenters from a dictionary of available measures. We discuss optimality properties, and numerical strategies to compute the sparse barycenters in practice.

Elsa Cazelles
Barycenters for probability distributions based on optimal weak mass transport.

We introduce weak barycenters of a family of probability distributions, based on the recently developed notion of optimal weak transport of mass [26], [9]. We provide a theoretical analysis of this object and discuss its interpretation in the light of convex ordering between probability measures. In particular, we show that, rather than averaging the input distributions in a geometric way (as the Wasser- stein barycenter based on classic optimal transport does) weak barycenters extract common geometric information shared by all the input distributions, encoded as a latent random variable that underlies all of them. We also provide an iterative algorithm to compute a weak barycenter for a finite family of input distributions, and a stochastic algorithm that computes them for arbitrary populations of laws. The latter approach is particularly well suited for the streaming setting, i.e., when distributions are observed sequentially. The notion of weak barycenter and our approaches to compute it are illustrated on synthetic examples, validated on 2D real-world data and compared to standard Wasserstein barycenters.