Workshop - Statistics in Metric Spaces

Date: Oct. 11, 12, 13

Location: ENSAE, Room 2033

Schedule

Wednesday, Oct. 11

10h-10h30: Registration (ENSAE entrance hall)

10h30-11h15 Quentin Paris, *Online learning with exponential weights in metric spaces with the measure contraction property*

11h15-12h Aleksei Kroshnin, *Robust k-means clustering in metric spaces*

12h-14h Lunch

14h-15h Stephan Huckemann, *The wald space for phylogenetic trees*

15h-15h30 Coffee break

15h30-16h15 Pierre Pansu, *Computing homology robustly: The geometry of normed chain complexes*

16h15-17h Gabriel Romon, *Generalized Fréchet means in metric trees*

18h30 Dinner

Thursday, Oct. 12

10h-11h Xavier Pennec, *Taylor expansion of geodesic triangles in Riemanian manifolds: A central tool to study the effect of curvature in geometric statistics*

11h-11h15 Coffee break

11h15-12h Jordan Serres, *Concentration of empirical barycenters in Non Positively Curved metric spaces*

12h-14h Lunch

14h-15h Kazuhiro Kuwae, *Hess-Schrader-Uhlenbrock inequality for the heat semigroup on differential forms over Dirichlet spaces tamed by distributional curvature lower bounds*

15h-15h30 Coffee break

15h30-16h15 Shin-ichi Ohta, *Discrete-time gradient flows in Gromov hyperbolic spaces*

16h15-17h Austin Stromme, *Global, dimension-free convergence of first-order methods for barycenters in the Bures-Wasserstein space*

18h30 Dinner

Friday, Oct. 13

10h-11h Miklós Pálfia, *Gradient flows and calculus of variations in CAT(1)-spaces*

11h-11h15 Coffee break

11h15-12h Christopher Criscitiello, *Curvature and Complexity: Lower bounds for geodesically convex optimization*

12h-14h Lunch

14h-15h Tom Nye, *1. Modelling manifold-valued time series data from EEG studies of patients with epilepsy, 2. Phylogenetic information geometry and the origin of wald space*

15h-15h30 Coffee break

15h30-16h15 Shin-ichi Ohta, *Barycenters and a law of large numbers in Gromov hyperbolic spaces*

16h15-17h Yunbum Kook, *Condition-number-independent convergence rate of Riemannian Hamiltonia Monte Carlo with numerical integrators*

19h30 Dinner

Abstracts

Christopher Criscitiello (Ecole Polytechnique Fédérale de Lausanne): *Curvature and Complexity: Lower bounds for geodesically convex optimization*

Abstract. We study the query complexity of geodesically convex (g-convex) optimization on a manifold. In a variety of settings (smooth or not; strongly g-convex or not; high- or low-dimensional), known upper bounds worsen with curvature. It is natural to ask whether this is warranted, or an artifact. We provide an initial set of lower bounds showing that the negative effect of curvature is indeed unavoidable in all settings. We then refine this initial set of lower bounds to further reduce the gap between best known upper and lower bounds, using a number of new techniques. We suspect these resulting bounds are not optimal. We conjecture optimal ones, and support them with a matching lower bound for a class of algorithms which includes subgradient descent, and a lower bound for a related game.

Stephan Huckemann (University of Göttingen): *The wald space for phylogenetic trees*

Abstract. Most existing metrics between phylogenetic trees directly measure differences in topology and edge weights, and are unrelated to the models of evolution used to infer trees. We describe metrics which instead are based on distances between the probability models of discrete or continuous characters induced by trees. We describe how construction of information-based geodesics leads to the recently [2] proposed wald space of phylogenetic trees. As a point set, it sits between the BHV space [1] and the edge-product space [4]. It has a natural embedding into the space of positive definite matrices, equipped with the information geometry. Thus, singularities such as overlapping leaves are infinitely far away, proper forests, however, comprising the "BHV-boundary at infinity", are part of the wald space, adding boundary correspondences to groves (corresponding to orthants in the BHV space). In fact the wald space contracts to the complete disconnected forest. Further, it is a geodesic space, exhibiting the structure of a Whitney stratified space of type (A) where strata carry compatible Riemannian metrics. We explore some more geometric properties, but the full picture remains open. We conclude by identifying open problems, we deem interesting [3].
[1] Billera, L., S. Holmes and K. Vogtmann (2001). Geometry of the space of phylogenetic trees. Advances in Applied Mathematics 27 (4), 733–767.
[2] Garba, M. K., T. M. Nye, J. Lueg and S. F. Huckemann (2021). Information geometry for phylogenetic trees. Journal of Mathematical Biology 82(3), 1–39.
[3] Lueg, J., M. K. Garba, T. M. Nye and S. F. Huckemann (2022). Foundations of wald space for statistics of phylogenetic trees. arXiv 2209.05332.
[4] Moulton, V. and M. Steel (2004). Peeling phylogenetic 'oranges'. Advances in Applied Mathematics 33(4), 710–727.

Yunbum Kook (Georgia Institute of Technology): *Condition-number-independent convergence rate of Riemannian Hamiltonian Monte Carlo with numerical integrators*

Abstract. We study the convergence rate of discretized Riemannian Hamiltonian Monte Carlo on sampling from distributions in the form of $e^{-f(x)}$ on a convex body $M \subset \mathbb{R}^n$. We show that for distributions in the form of $e^{-\alpha^\top x}$ on a polytope with m constraints, the convergence rate of a family of commonly-used integrators is independent of $\|\alpha\|_2$ and the geometry of the polytope. In particular, the implicit midpoint method (IMM) and the generalized Leapfrog method (LM) have a mixing time of $\widetilde{O}(mn^3)$ to achieve ϵ total variation distance to the target distribution. These guarantees are based on a general bound on the convergence rate for densities of the form $e^{-f(x)}$ in terms of parameters of the manifold and the integrator. Our theoretical guarantee complements the empirical results of [KLSV22], which shows that RHMC with IMM can sample ill-conditioned, non-smooth and constrained distributions in very high dimension efficiently in practice. This is joint work with Yin Tat Lee, Ruoqi Shen, and Santosh Vempala.

Aleksei Kroshnin (Weierstrass Institute): *Robust k-means clustering in metric spaces*

Abstract. In this talk, we consider robust algorithms for the k-means clustering (quantization) problem where a quantizer is constructed based on an i.i.d. sample. While the well-known asymptotic result by Pollard shows that the existence of two moments is sufficient for strong consistency of an empirically optimal quantizer in a Euclidean space, non-asymptotic bounds are usually obtained under the assumption of bounded support. We discuss a robust k-means in metric spaces based on trimming (similar to the one proposed by Cuesta-Albertos et al. in 1997), and prove novel non-asymptotic bounds on the excess distortion depending on the optimal distortion rather then the second moment of the distribution. In particular, this bound decreases the gap to the lower bound by Bartlett et al. (1998). The talk is based on the joint work with Y. Klochkov and N. Zhivotovskiy and on an ongoing project with A. Suvorikova and N. Zhivotovskiy.

Kazuhiro Kuwae (Fukuoka University): *Hess-Schrader-Uhlenbrock inequality for the heat semigroup on differential forms over Dirichlet spaces tamed by distributional curvature lower bounds*

Abstract. The notion of tamed Dirichlet space was proposed by Erbar, Rigoni, Sturm and Tamanini ('22) as a Dirichlet space having a weak form of Bakry-Émery curvature lower bounds in distribution sense. After their work, Braun ('21+) established a vector calculus for it, in particular, the space of L^2-normed L^∞-module describing vector fields, 1-forms, Hessian in L^2-sense. In this framework, we establish the Hess-Schrader-Uhlenbrock inequality for 1-forms as an element of L^2-cotangent bundles, (an L^2-normed L^∞-module), which extends the result on the Hess-Schrader-Uhlenbrock inequality under an additional condition by Braun ('21+).

Tom Nye (Newcastle University): *Modelling manifold-valued time series data from EEG studies of patients with epilepsy*

Abstract. Electroencephalography (EEG) is a method for recording the brain's electrical activity and it is an important tool in the diagnosis and treatment of epilepsy. Researchers often study time series of covariance or correlation matrices between different locations in the brain as opposed to the raw time series of signals at each location, since these can be more informative. We propose a model for time series taking values on a Riemannian manifold and fit it to time series of covariance matrices in the space of symmetric positive definite (SPD) matrices. The aim of the study is two-fold: to develop a model with interpretable parameters for different possible modes of EEG dynamics, and to explore the extent to which modelling results are affected by the choice of geometry on SPD matrix space. The model specifies a distribution for the tangent direction vector at any time point, combining an autoregressive term, a mean reverting term and a form of Gaussian noise. Results distinguish between epileptic seizures and periods between seizures in patients: between seizures the dynamics have a strong mean reverting component and the auto regressive component is weaker, while during seizures the noise term has greater variance and the mean reverting effect is smaller. The affine invariant geometry is advantageous and it provides a better fit to the data.

Tom Nye (Newcastle University): *Phylogenetic information geometry and the origin of wald space*

Abstract. Phylogenetic trees, which represent evolutionary relationships between present-day species, are usually inferred from genetic sequence data. As such, each tree represents a Markovian model of sequence evolution. Collections or samples of alternative trees arise as a result of uncertainty when inferring trees from data. In recent years a number of powerful geometric methods have been developed for analysing samples of phylogenetic trees, for tasks such as computing sample means and performing principal component analysis. Typically, these methods adopt certain underlying geometric assumptions which regard trees as purely geometric objects in terms of branching shape and edge lengths. We develop an alternative approach which constructs the geometry on tree space by regarding trees directly as probabilistic models of sequence evolution. First we describe certain metrics on tree space which are

induced by metrics between distributions, such as the Hellinger distance. Secondly, we compute geodesics in tree space using the Riemannian metric defined by the Fisher information matrix. These geodesics are computationally expensive to construct, but we show via examples how they are closely approximated by geodesics in a more tractable geometry obtained by embedding tree space in the space of covariance matrices. This is the recently proposed wald space.

Shin-ichi Ohta (Osaka University): *Discrete-time gradient flows in Gromov hyperbolic spaces*

Abstract. The theory of gradient flows for convex functions on ”Riemannian” spaces (such as CAT(0)-spaces) has been making impressive progress. Nonetheless, much less is known for ”Finsler” spaces, especially the lack of contraction property is a central problem. As a class including some non-Riemannian Finsler manifolds, we employ Gromov hyperbolic spaces and study the contraction property of discrete-time gradient flows for convex functions.

Shin-ichi Ohta (Osaka University): *Barycenters and a law of large numbers in Gromov hyperbolic spaces*

Abstract. We investigate barycenters of probability measures on Gromov hyperbolic spaces, toward development of convex optimization in this class of metric spaces. We establish a contraction property in terms of the Kantorovich-Wasserstein distance, a deterministic approximation of barycenters of uniform distributions on finite points, and a kind of law of large numbers. These generalize the corresponding results on CAT(0)-spaces, up to additional terms depending on the hyperbolicity constant.

Miklós Pálfia (Corvinus University of Budapest): *Gradient flows and calculus of variations in CAT(1)-spaces*

Abstract. We generalize the theory of gradient flows of semi-convex functions established for CAT(0)-spaces to CAT(1)-spaces. We show that the so called commutativity property and semi-convexity of the squared distance function is enough to establish the uniqueness, EVI and contractivity of the gradient flow similarly to the CAT(0) setting using the Moreau-Yoshida resolvent. We consider discrete and continuous time flows, leading to law of large numbers in CAT(1)-spaces. Then generalize the results of Bačák on Mosco convergence established for CAT(0) spaces to the CAT(1) setting, so that Mosco convergence implies convergence of resolvents. The talk is based on earlier joint works with Shin-ichi Ohta and ongoing work with Hedvig Gál.

Pierre Pansu (Université Paris-Saclay): *Computing homology robustly: The geometry of normed chain complexes*

Abstract. The conditioning number when computing the cohomology of a graph (or a simplicial complex) has geometric significance: it is known as Cheeger's constant or spectral gap. This indicates that (co-)chain complexes contain more information than their mere (co-)homology. We turn the set of normed chain complexes into a metric space.

Quentin Paris (Higher School of Economics): *Online learning with exponential weights in metric spaces with the measure contraction property*

Abstract. The talk 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.

Xavier Pennec (INRIA Sophia Antipolis): *Taylor expansion of geodesic triangles in Riemannian* *manifolds: A central tool to study the effect of curvature in geometric statistics.*

Abstract. The impact of manifold curvature on the statistical estimation or on the accuracy of algorithms to compute on manifolds is not evident to establish. We show in this talk that Gavrilov's Taylor expansions of the double exponential can be completed by a companion neighboring log expansion which measures how the Riemannian logarithm changes when its foot-point is moving. These two tensorial and coordinate free Taylor expansions constitute a complete toolbox for the polynomial approximations of problems related to infinitesimal geodesic triangles. Moreover they are valid in general affine connection manifolds and computations can surprisingly easily be pushed to higher orders if needed.

We exemplify the use of this toolbox for two of fundamental tools for geometric statistics: the estimation of the Fréchet mean and parallel transport. We first present a new non-asymptotic (small sample) expansion in high concentration conditions which quantifies the concentration of the empirical Fréchet mean towards the population mean. This shows a statistical bias on the empirical mean in the direction of the average gradient of the curvature and a modulation of the convergence speed which could partly explain smeary means in positive curvature spaces and sticky means in stratified spaces of negative curvature.

Parallel transport is a second major tool to compare local tangent information at different points of the manifold. We previously proposed a modification of Schild's ladder called pole ladder that is surprisingly exact in only one step on symmetric spaces. Iterating geodesic parallelograms was thought to be of first order but a real convergence analysis was lacking. We show that pole and Schild's ladder naturally converges with quadratic speed even when geodesics are approximated by numerical schemes. This contrasts with Jacobi fields approximations that are bound to linear convergence. The extra computational cost of ladder methods is thus easily compensated by a drastic reduction of the number of steps needed to achieve the requested accuracy.

Gabriel Romon (ENSAE): *Generalized Fréchet means in metric trees*

Abstract. We are interested in measures of central tendency for a population on a network, which is modeled by a metric tree. The location parameters that we study are generalized Fréchet means obtained by replacing the usual objective $\alpha \mapsto E[d(\alpha, X)^2]$ with $\alpha \mapsto E[\ell(d(\alpha, X))]$ where ℓ is a generic convex nondecreasing loss. We develop a notion of directional derivative in the tree, which helps up locate and characterize the minimizers.

Estimation is performed using a sample analog. We extend to a metric tree the notion of stickiness defined by Hotz et al. (2013). For generalized Fréchet means we obtain a sticky law of large numbers. For the Fréchet median we develop nonasymptotic concentration bounds and sticky central limit theorems.

Jordan Serres (ENSAE): *Concentration of empirical barycenters in Non Positively Curved metric spaces*

Abstract. Barycenters in non-Euclidean geometries are the most natural extension of linear averaging. They are widely used in shape statistics, optimal transport and matrix analysis. In this talk, I will present their asymptotic properties, i.e. law of large numbers. I will also present some more recent results about their non-asymptotic concentration rate.

Austin Stromme (ENSAE): *Global, dimension-free convergence of first-order methods for barycenters in the Bures-Wasserstein space*

Abstract. We consider the problem of computing the barycenter of a distribution supported on the space of multivariate Gaussians with the quadratic optimal transport metric, the Bures-Wasserstein space. First-order methods are known to work well for this problem, but a theoretical account of their performance is obstructed by the positive curvature of Bures-Wasserstein space and the ensuing non-convexity of the objective. We overcome this difficulty with a Polyak-Lojasiewicz (PL) inequality and guarantee it holds throughout the optimization trajectory by using new results on the geodesic convexity of

the minimum and maximum eigenvalue functionals, leading to global, dimension-free rates of convergence for gradient descent and stochastic gradient descent.