Information Geometry (2020) 3:1–32
https://doi.org/10.1007/s41884-020-00028-0

RESEARCH PAPER

Optimal transport natural gradient for statistical manifolds with continuous sample space

Yifan Chen[1] · Wuchen Li[2]

Received: 21 May 2018 / Revised: 22 March 2020 / Published online: 11 May 2020

Abstract

We study the Wasserstein natural gradient in parametric statistical models with continuous sample spaces. Our approach is to pull back the L^2-Wasserstein metric tensor in the probability density space to a parameter space, equipping the latter with a positive definite metric tensor, under which it becomes a Riemannian manifold, named the Wasserstein statistical manifold. In general, it is not a totally geodesic sub-manifold of the density space, and therefore its geodesics will differ from the Wasserstein geodesics, except for the well-known Gaussian distribution case, a fact which can also be validated under our framework. We use the sub-manifold geometry to derive a gradient flow and natural gradient descent method in the parameter space. When parametrized densities lie in $\mathbb{R}$, the induced metric tensor establishes an explicit formula. In optimization problems, we observe that the natural gradient descent outperforms the standard gradient descent when the Wasserstein distance is the objective function. In such a case, we prove that the resulting algorithm behaves similarly to the Newton method in the asymptotic regime. The proof calculates the exact Hessian formula for the Wasserstein distance, which further motivates another preconditioner for the optimization process. To the end, we present examples to illustrate the effectiveness of the natural gradient in several parametric statistical models, including the Gaussian measure, Gaussian mixture, Gamma distribution, and Laplace distribution.

Keywords Optimal transport · Information geometry · Wasserstein statistical manifold · Wasserstein natural gradient

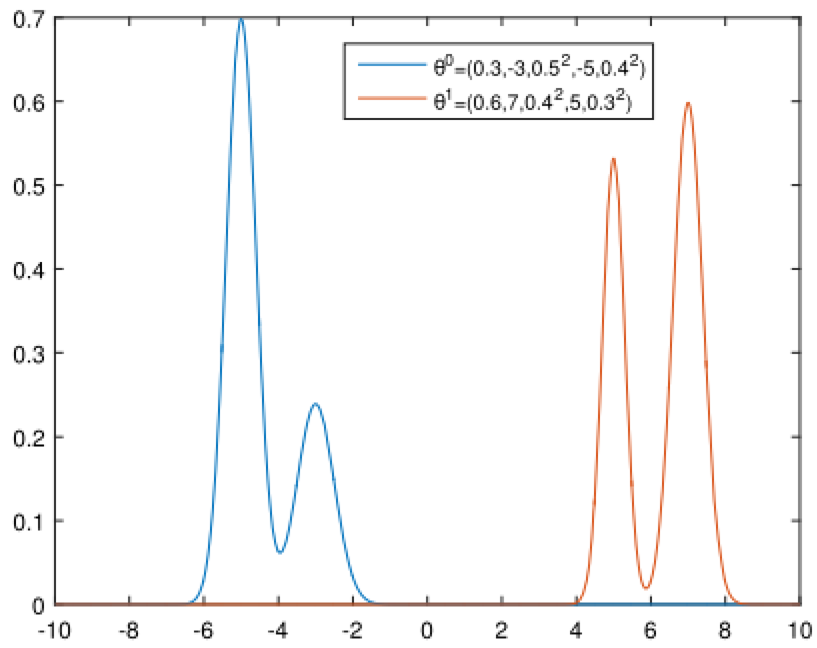


Fig. 1 Densities of Gaussian mixture distribution

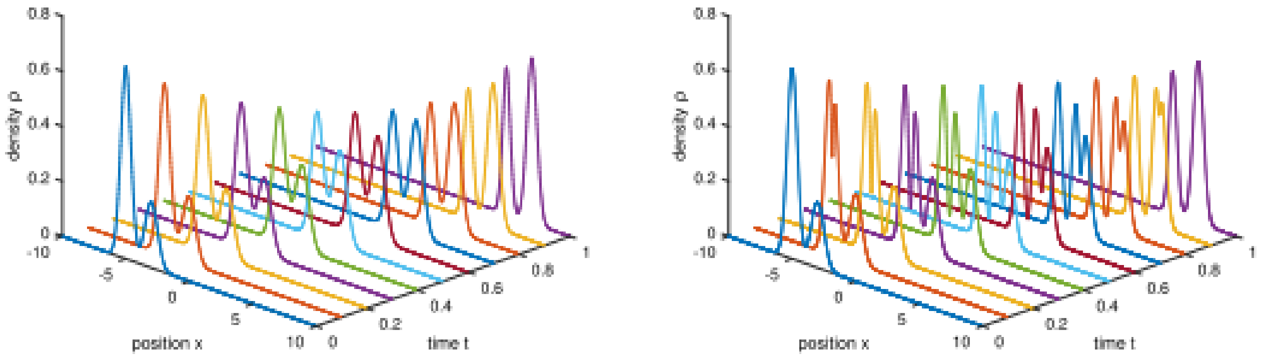
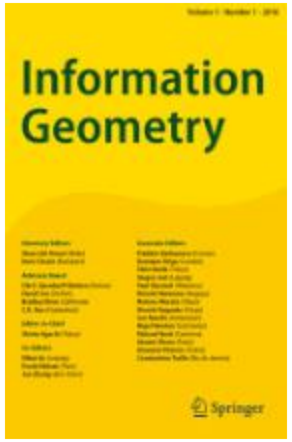


Fig. 2 Geodesic of Gaussian mixtures; left: in the Wasserstein statistical manifold; right: in the whole density space



Volume 3
Issue 1
Article 20

RESEARCH PAPER

Cramér–Rao lower bounds arising from generalized Csiszár divergences

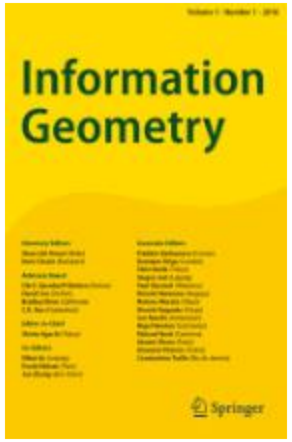
M. Ashok Kumar[1] · Kumar Vijay Mishra[2]

Received: 14 January 2020 / Revised: 23 May 2020 / Published online: 16 June 2020

Abstract

We study the geometry of probability distributions with respect to a generalized family of Csiszár f-divergences. A member of this family is the relative α-entropy which is also a Rényi analog of relative entropy in information theory and known as logarithmic or projective power divergence in statistics. We apply Eguchi's theory to derive the Fisher information metric and the dual affine connections arising from these generalized divergence functions. This enables us to arrive at a more widely applicable version of the Cramér–Rao inequality, which provides a lower bound for the variance of an estimator for an escort of the underlying parametric probability distribution. We then extend the Amari–Nagaoka's dually flat structure of the exponential and mixer models to other distributions with respect to the aforementioned generalized metric. We show that these formulations lead us to find unbiased and efficient estimators for the escort model. Finally, we compare our work with prior results on generalized Cramér–Rao inequalities that were derived from non-information-geometric frameworks.

Keywords Cramér–Rao lower bound · Csiszár f-divergence · Fisher information metric · escort distribution · relative entropy



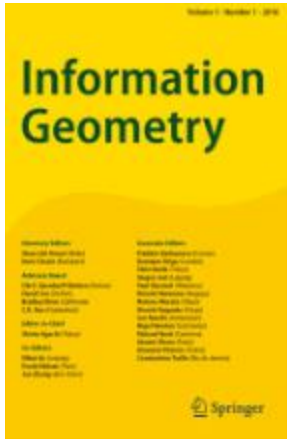
Volume 3
Issue 1
Article 21

RESEARCH PAPER

Tropical diagrams of probability spaces

R. Matveev[1] · J. W. Portegies[2]

Received: 5 June 2019 / Revised: 29 January 2020 / Published online: 7 April 2020



Volume 3
Issue 1
Article 22

Abstract

After endowing the space of diagrams of probability spaces with an entropy distance, we study its large-scale geometry by identifying the asymptotic cone as a closed convex cone in a Banach space. We call this cone the *tropical cone*, and its elements *tropical diagrams of probability spaces*. Given that the tropical cone has a rich structure, while tropical diagrams are rather flexible objects, we expect the theory of tropical diagrams to be useful for information optimization problems in information theory and artificial intelligence. In a companion article, we give a first application to derive a statement about the entropic cone.

Keywords Tropical probability · Entropy distance · Diagrams of probability spaces · Tropical cone

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X \arrow[dr] & & Y \arrow[dl] \\ & U & \end{tikzcd}$$

$$\begin{tikzcd} & T \arrow[dl] \arrow[d] \arrow[dr] & \\ U \arrow[d] \arrow[dr] & V \arrow[dl] \arrow[dr] & W \arrow[dl] \arrow[d] \\ X & Y & Z \end{tikzcd}$$

RESEARCH PAPER

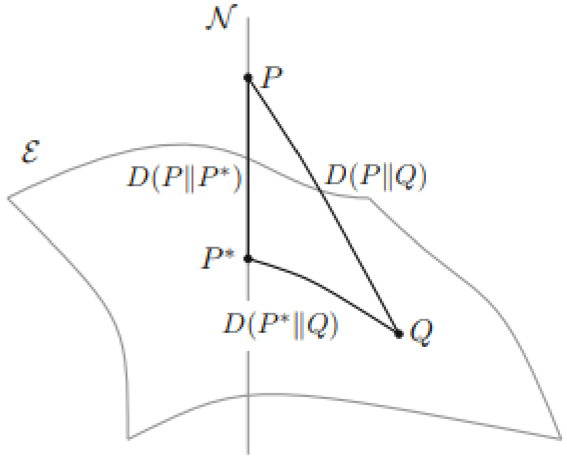
Ricci curvature for parametric statistics via optimal transport

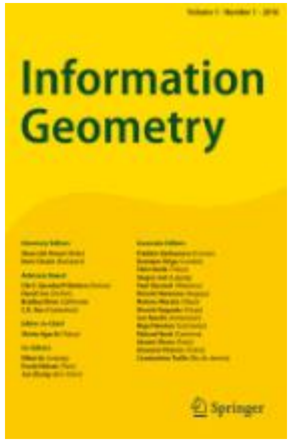
Wuchen Li[1] · Guido Montúfar[1]

Received: 18 July 2018 / Revised: 15 October 2019 / Published online: 31 January 2020

Abstract
We define the notion of a Ricci curvature lower bound for parametrized statistical models. Following the seminal ideas of Lott–Sturm–Villani, we define this notion based on the geodesic convexity of the Kullback–Leibler divergence in a Wasserstein statistical manifold, that is, a manifold of probability distributions endowed with a Wasserstein metric tensor structure. Within these definitions, which are based on Fisher information matrix and Wasserstein Christoffel symbols, the Ricci curvature is related to both, information geometry and Wasserstein geometry. These definitions allow us to formulate bounds on the convergence rate of Wasserstein gradient flows and information functional inequalities in parameter space. We discuss examples of Ricci curvature lower bounds and convergence rates in exponential family models.

Keywords Ricci curvature · Information projection · Wasserstein statistical manifold · Fokker–Planck equation on parameter space





Volume 3
Issue 1
Article 22