Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem

Shun-ichi Amari[1,3] · Ryo Karakida[2] · Masafumi Oizumi[1,3]

Abstract Two geometrical structures have been extensively studied for a manifold of probability distributions. One is based on the Fisher information metric, which is invariant under reversible transformations of random variables, while the other is based on the Wasserstein distance of optimal transportation, which reflects the structure of the distance between underlying random variables. Here, we propose a new information-geometrical theory that provides a unified framework connecting the Wasserstein distance and Kullback–Leibler (KL) divergence. We primarily considered a discrete case consisting of n elements and studied the geometry of the probability simplex S_{n-1}, which is the set of all probability distributions over n elements. The Wasserstein distance was introduced in S_{n-1} by the optimal transportation of commodities from distribution $\boldsymbol{p}$ to distribution $\boldsymbol{q}$, where $\boldsymbol{p}, \boldsymbol{q} \in S_{n-1}$. We relaxed the optimal transportation by using entropy, which was introduced by Cuturi. The optimal solution was called the entropy-relaxed stochastic transportation plan. The entropy-relaxed optimal cost $C(\boldsymbol{p}, \boldsymbol{q})$ was computationally much less demanding than the original Wasserstein distance but does not define a distance because it is not minimized at $\boldsymbol{p} = \boldsymbol{q}$. To define a proper divergence while retaining the computational advantage, we first introduced a divergence function in the manifold $S_{n-1} \times S_{n-1}$ composed of all optimal transportation plans. We fully explored the information geometry of the manifold of the optimal transportation plans and subsequently constructed a new one-parameter family of divergences in S_{n-1} that are related to both the Wasserstein distance and the KL-divergence.

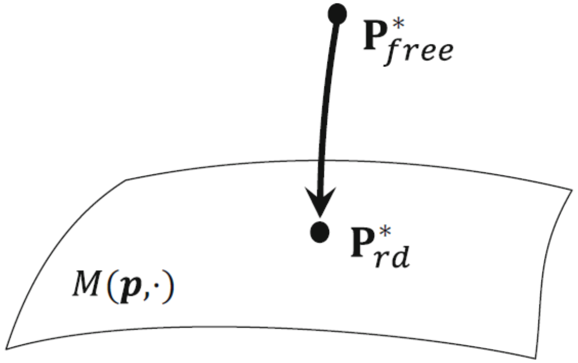


Fig. 3 e-projection in the rate-distortion problem

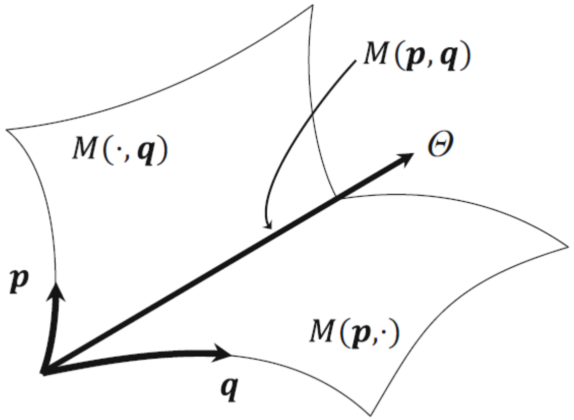


Fig. 4 m-flat submanifolds in the transportation problem

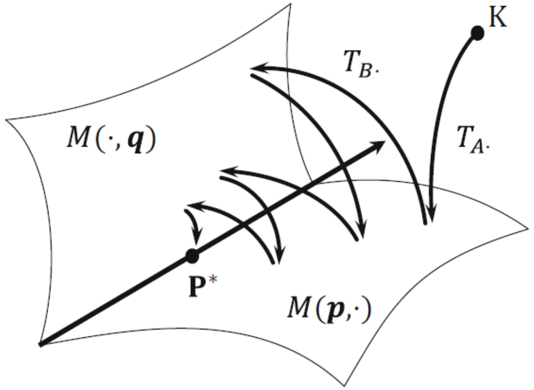
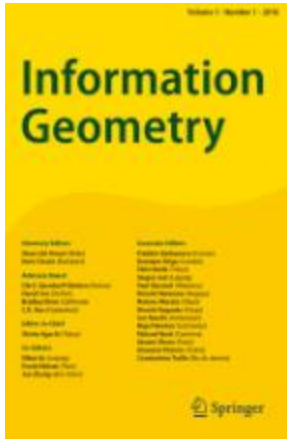


Fig. 6 Sinkhorn algorithm as iterative e-projections

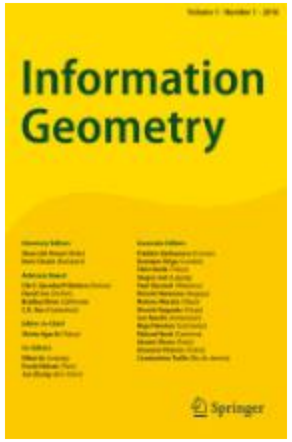


Volume 1
Issue 1
Article 1

Logarithmic divergences from optimal transport and Rényi geometry

Ting-Kam Leonard Wong[1]

Abstract
Divergences, also known as contrast functions, are distance-like quantities defined on manifolds of non-negative or probability measures. Using the duality in optimal transport, we introduce and study the one-parameter family of $L^{(\pm\alpha)}$-divergences. They extrapolate between the Bregman divergence corresponding to the Euclidean quadratic cost, and the L-divergence introduced by Pal and the author in connection with portfolio theory and a logarithmic cost function. They admit natural generalizations of exponential family that are closely related to the α-family and q-exponential family. In particular, the $L^{(\pm\alpha)}$-divergences of the corresponding potential functions are Rényi divergences. Using this unified framework we prove that the induced geometries are dually projectively flat with constant sectional curvatures, and a generalized Pythagorean theorem holds true. Conversely, we show that if a statistical manifold is dually projectively flat with constant curvature $\pm\alpha$ with $\alpha > 0$, then it is locally induced by an $L^{(\mp\alpha)}$-divergence. We define in this context a canonical divergence which extends the one for dually flat manifolds.



Volume 1
Issue 1
Article 2

The key objects of study in this paper are the $L^{(\pm\alpha)}$-divergences defined for $\alpha > 0$. The $L^{(\alpha)}$-divergence is defined by

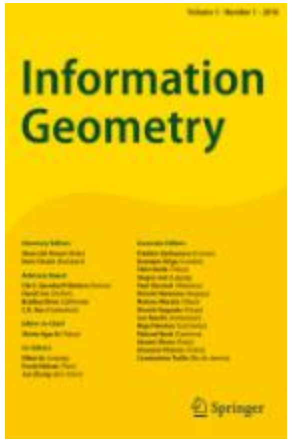
$$\mathbf{D}^{(\alpha)}\left[\xi : \xi'\right] = \frac{1}{\alpha}\log(1 + \alpha\nabla\varphi(\xi')\cdot(\xi - \xi')) - \left(\varphi(\xi) - \varphi(\xi')\right),$$

where the function φ is α-exponentially concave, i.e., $e^{\alpha\varphi}$ is concave.

Rho–tau embedding and gauge freedom in information geometry

Jan Naudts[1] · Jun Zhang[2]

Received: 1 November 2017 / Revised: 27 June 2018 / Published online: 20 August 2018



Volume 1
Issue 1
Article 3

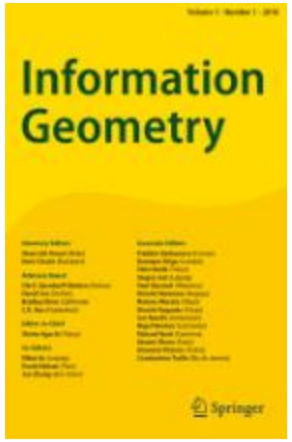
Abstract
The standard model of information geometry, expressed as Fisher–Rao metric and Amari-Chensov tensor, reflects an embedding of probability density by log-transform. The present paper studies parametrized statistical models and the induced geometry using arbitrary embedding functions, comparing single-function approaches (Eguchi's U-embedding and Naudts' deformed-log or phi-embedding) and a two-function embedding approach (Zhang's conjugate rho-tau embedding). In terms of geometry, the rho-tau embedding of a parametric statistical model defines both a Riemannian metric, called "rho-tau metric", and an alpha-family of rho-tau connections, with the former controlled by a single function and the latter by both embedding functions ρ and τ in general. We identify conditions under which the rho-tau metric becomes Hessian and hence the ± 1 rho-tau connections are dually flat. For any choice of rho and tau there exist models belonging to the phi-deformed exponential family for which the rho-tau metric is Hessian. In other cases the rho–tau metric may be only conformally equivalent with a Hessian metric. Finally, we show a formulation of the maximum entropy framework which yields the phi-exponential family as the solution.

Keywords Phi-embedding · U-embedding · Rho–tau embedding · Rho–tau metric · Rho–tau divergence · Rho–tau cross-entropy · U cross-entropy · Phi-exponential model · Escort distribution · Hessian metric · Gauge freedom

Chentsov's theorem for exponential families

James G. Dowty[1]

Received: 5 December 2017 / Revised: 20 June 2018 / Published online: 8 August 2018



Volume 1
Issue 1
Article 4

Abstract

Chentsov's theorem characterizes the Fisher information metric on statistical models as the only Riemannian metric (up to rescaling) that is invariant under sufficient statistics. This implies that each statistical model is equipped with a natural geometry, so Chentsov's theorem explains why many statistical properties can be described in geometric terms. However, despite being one of the foundational theorems of statistics, Chentsov's theorem has only been proved previously in very restricted settings or under relatively strong invariance assumptions. We therefore prove a version of this theorem for the important case of exponential families. In particular, we characterise the Fisher information metric as the only Riemannian metric (up to rescaling) on an exponential family and its derived families that is invariant under independent and identically distributed extensions and canonical sufficient statistics. We then extend this result to curved exponential families. Our approach is based on the central limit theorem, so it gives a unified proof for discrete and continuous exponential families, and it is less technical than previous approaches.

Keywords Chentsov's theorem · Fisher information metric · Information geometry · Curved exponential families