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 p to distribution q, where $p, q \in S_{n-1}$. 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 $C(p, q)$ was computationally much less demanding than the original Wasserstein distance but does not define a distance because it is not minimized at $p = q$. 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.

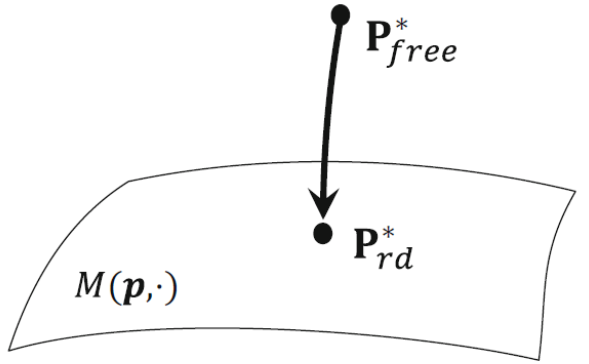


Fig. 3 *e*-projection in the rate-distortion problem

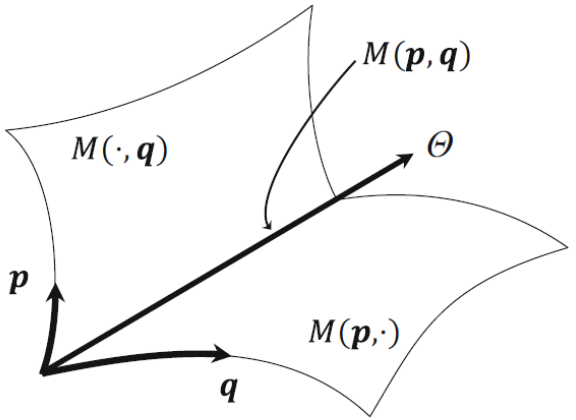


Fig. 4 *m*-flat submanifolds in the transportation problem

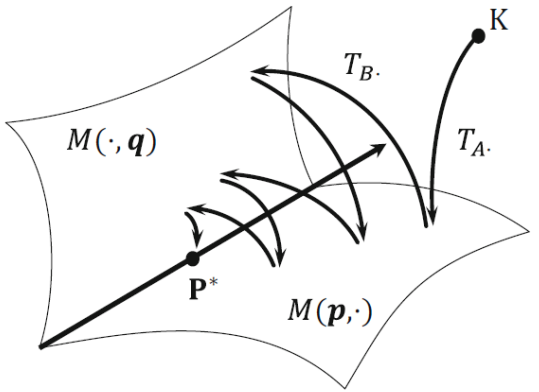
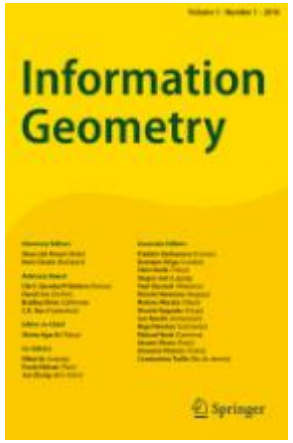


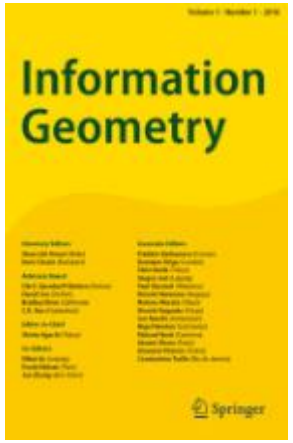
Fig. 6 Sinkhorn algorithm as iterative *e*-projections



Volume 1
Issue 1
Article 1

Logarithmic divergences from optimal transport and Rényi geometry

Ting-Kam Leonard Wong[1]



Volume 1
Issue 1
Article 2

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.

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

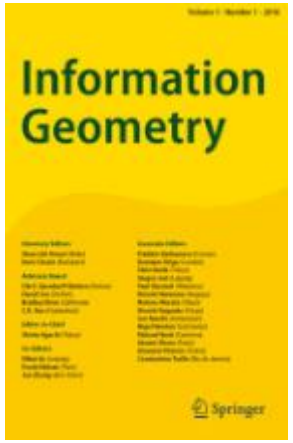
$$\mathbf{D}^{(\alpha)}\left[\xi : \xi'\right] = \frac{1}{\alpha}\log(1 + \alpha\nabla\varphi(\xi') \cdot (\xi - \xi')) - \left(\varphi(\xi) - \varphi(\xi')\right),$$

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

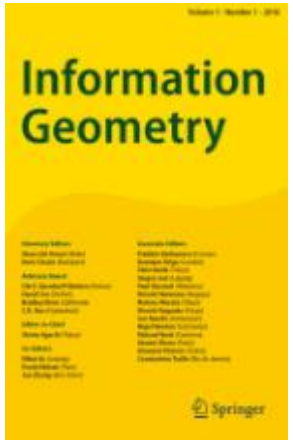
The standard model of information geometry, expressed as Fisher–Rao metric and Amari-Chensov tensor, reflects an embedding of probability density by log-transform. The present paper studies parametrized statistical models and the induced geometry using arbitrary embedding functions, comparing single-function 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

RESEARCH PAPER

Wasserstein Riemannian geometry of Gaussian densities

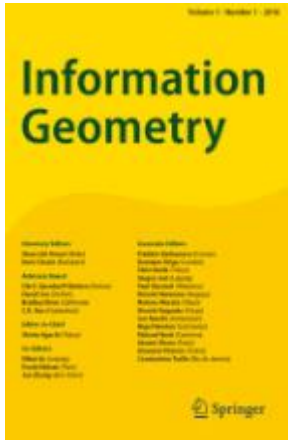
Luigi Malagò[1] · Luigi Montrucchio[2] · Giovanni Pistone[3]

Received: 29 May 2018 / Revised: 9 October 2018 / Published online: 20 November 2018

Abstract

The Wasserstein distance on multivariate non-degenerate Gaussian densities is a Riemannian distance. After reviewing the properties of the distance and the metric geodesic, we present an explicit form of the Riemannian metrics on positive-definite matrices and compute its tensor form with respect to the trace inner product. The tensor is a matrix which is the solution to a Lyapunov equation. We compute the explicit formula for the Riemannian exponential, the normal coordinates charts and the Riemannian gradient. Finally, the Levi-Civita covariant derivative is computed in matrix form together with the differential equation for the parallel transport. While all computations are given in matrix form, nonetheless we discuss also the use of a special moving frame.

Keywords Information geometry · Gaussian distribution · Wasserstein distance · Riemannian metrics · Natural gradient · Riemannian exponential · Normal coordinates · Levi-Civita covariant derivative · Optimization on positive-definite symmetric matrices



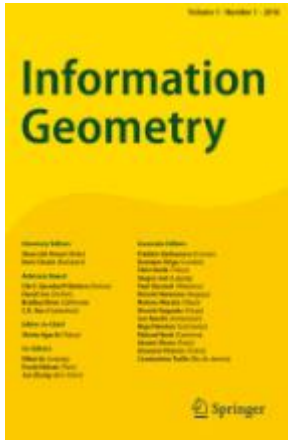
Volume 1
Issue 2
Article 5

RESEARCH PAPER

Natural gradient via optimal transport

Wuchen Li[1] · Guido Montúfar[2,3]

Received: 15 March 2018 / Revised: 27 August 2018 / Published online: 19 November 2018



Volume 1
Issue 2
Article 6

Abstract

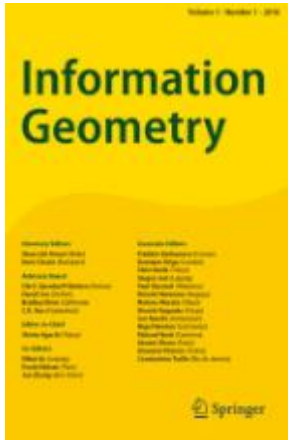
We study a natural Wasserstein gradient flow on manifolds of probability distributions with discrete sample spaces. We derive the Riemannian structure for the probability simplex from the dynamical formulation of the Wasserstein distance on a weighted graph. We pull back the geometric structure to the parameter space of any given probability model, which allows us to define a natural gradient flow there. In contrast to the natural Fisher–Rao gradient, the natural Wasserstein gradient incorporates a ground metric on sample space. We illustrate the analysis of elementary exponential family examples and demonstrate an application of the Wasserstein natural gradient to maximum likelihood estimation.

Keywords Optimal transport · Information geometry · Wasserstein statistical manifold · Displacement convexity · Machine learning

RESEARCH PAPER

Superharmonic priors for autoregressive models

Fuyuhiko Tanaka[1]



Volume 1
Issue 2
Article 7

Received: 24 July 2017 / Revised: 9 October 2017 / Published online: 1 December 2017

Abstract Tanaka and Komaki (Sankhya Ser A Indian Stat Inst 73-A:162–184, 2011) proposed superharmonic priors in Bayesian time series analysis as alternative to the famous Jeffreys prior. By definition the existence of superharmonic priors on a specific time series model with finite-dimensional parameter is equivalent to that of positive nonconstant superharmonic functions on the corresponding Riemannian manifold endowed with the Fisher metric. In the autoregressive models, whose Fisher metric and its inverse have quite messy forms, we obtain superharmonic priors in an explicit manner. To derive this result, we developed a systematic way of dealing with symmetric polynomials, which are related to Schur functions.

Keywords Jeffreys prior · Superharmonic priors · Autoregressive models · Noninformative priors · Kullback–Leibler divergence · Fisher metric

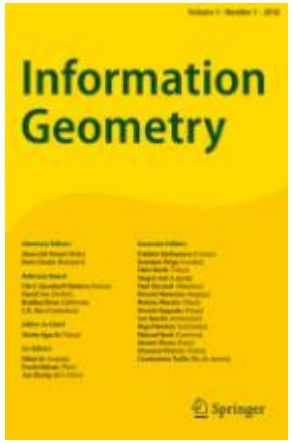
RESEARCH PAPER

Asymptotic dependency structure of multiple signals

Asymptotic equipartition property for diagrams of probability spaces

Rostislav Matveev[1] · Jacobus W. Portegies[2]

Received: 7 February 2018 / Revised: 16 August 2018 / Published online: 12 October 2018



Volume 1
Issue 2
Article 8

Abstract

We formalize the notion of the dependency structure of a collection of *multiple* signals, relevant from the perspective of information theory, artificial intelligence, neuroscience, complex systems and other related fields. We model multiple signals by commutative diagrams of probability spaces with measure-preserving maps between some of them. We introduce the asymptotic entropy (pseudo-)distance between diagrams, expressing how much two diagrams differ from an information-processing perspective. If the distance vanishes, we say that two diagrams are asymptotically equivalent. In this context, we prove an asymptotic equipartition property: any sequence of tensor powers of a diagram is asymptotically equivalent to a sequence of homogeneous diagrams. This sequence of homogeneous diagrams expresses the relevant dependency structure.

RESEARCH PAPER

Asymptotic dependency structure of multiple signals

Asymptotic equipartition property for diagrams of probability spaces

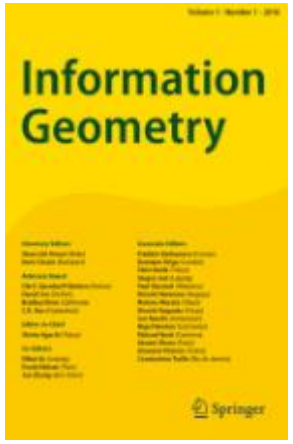
Rostislav Matveev[1] · Jacobus W. Portegies[2]

Received: 7 February 2018 / Revised: 16 August 2018 / Published online: 12 October 2018

Abstract

We formalize the notion of the dependency structure of a collection of *multiple* signals, relevant from the perspective of information theory, artificial intelligence, neuroscience, complex systems and other related fields. We model multiple signals by commutative diagrams of probability spaces with measure-preserving maps between some of them. We introduce the asymptotic entropy (pseudo-)distance between diagrams, expressing how much two diagrams differ from an information-processing perspective. If the distance vanishes, we say that two diagrams are asymptotically equivalent. In this context, we prove an asymptotic equipartition property: any sequence of tensor powers of a diagram is asymptotically equivalent to a sequence of homogeneous diagrams. This sequence of homogeneous diagrams expresses the relevant dependency structure.

Keywords Asymptotic equipartition property · Entropy distance · Diagrams of probability spaces · Multiple signals

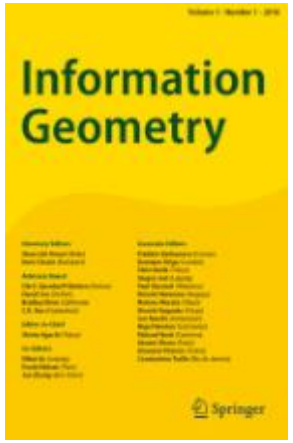


Volume 1
Issue 2
Article 9

RESEARCH PAPER

Ordering positive definite matrices

Cyrus Mostajeran[1] · Rodolphe Sepulchre[1]



Volume 1
Issue 2
Article 10

Received: 3 June 2017 / Revised: 2 May 2018 / Published online: 15 May 2018

Abstract We introduce new partial orders on the set S_n^+ of positive definite matrices of dimension n derived from the affine-invariant geometry of S_n^+. The orders are induced by affine-invariant cone fields, which arise naturally from a local analysis of the orders that are compatible with the homogeneous geometry of S_n^+ defined by the natural transitive action of the general linear group $GL(n)$. We then take a geometric approach to the study of monotone functions on S_n^+ and establish a number of relevant results, including an extension of the well-known Löwner-Heinz theorem derived using differential positivity with respect to affine-invariant cone fields.

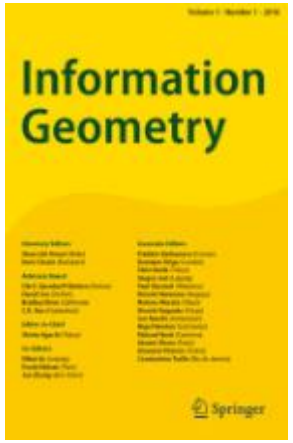
Keywords Positive definite matrices · Partial orders · Monotone functions · Monotone flows · Differential positivity · Matrix means

RESEARCH PAPER

Local equivalence problem in hidden Markov model

Masahito Hayashi[1,2,3,4]

Received: 30 March 2018 / Revised: 7 May 2019 / Published online: 10 June 2019



Volume 2
Issue 1
Article 11

Abstract

In the hidden Markov process, there is a possibility that two different transition matrices for hidden and observed variables yield the same stochastic behavior for the observed variables. Since such two transition matrices cannot be distinguished, we need to identify them and consider that they are equivalent, in practice. We address the equivalence problem of hidden Markov process in a local neighborhood by using the geometrical structure of hidden Markov process. For this aim, we introduce a mathematical concept to express Markov process, and formulate its exponential family by using generators. Then, the above equivalence problem is formulated as the equivalence problem of generators. Taking this equivalence problem into account, we derive several concrete parametrizations in several natural cases.

Keywords Hidden Markov · Equivalence problem · Information geometry · Exponential family

RESEARCH PAPER

Information geometry of modal linear regression

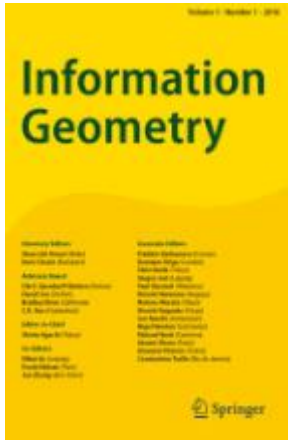
Keishi Sando[1] · Shotaro Akaho[2,3] · Noboru Murata[3,4] · Hideitsu Hino[3,5]

Received: 29 August 2018 / Revised: 12 June 2019 / Published online: 1 July 2019

Abstract

Modal linear regression (MLR) is used for modeling the conditional mode of a response as a linear predictor of explanatory variables. It is an effective approach to dealing with response variables having a multimodal distribution or those contaminated by outliers. Because of the semiparametric nature of MLR, constructing a statistical model manifold is difficult with the conventional approach. To overcome this difficulty, we first consider the information geometric perspective of the modal expectation–maximization (EM) algorithm. Based on this perspective, model manifolds for MLR are constructed according to observations, and a data manifold is constructed based on the empirical distribution. In this paper, the *em* algorithm, which is a geometric formulation of the EM algorithm, of MLR is shown to be equivalent to the conventional EM algorithm of MLR. The robustness of the MLR model is also discussed in terms of the influence function and information geometry.

Keywords Modal linear regression · Information geometry · Kernel density estimation



Volume 2
Issue 1
Article 12

RESEARCH PAPER

From Hessian to Weitzenböck: manifolds with torsion-carrying connections

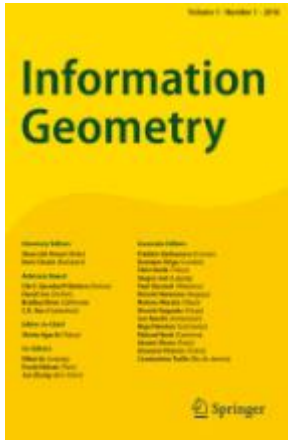
Jun Zhang[1] · Gabriel Khan[1]

Received: 22 December 2018 / Revised: 4 July 2019 / Published online: 18 July 2019

Abstract

We investigate affine connections that have zero curvature but not necessarily zero torsion. Slightly generalizing from what is known as Weitzenböck connections, such non-flat connections (we call "pseudo-Weitzenböck connections") can be constructed from any frame (a set of linearly independent vector fields), not just the orthonormal frames, with their torsions vanishing if and only if the frame is a coordinate frame. In such situations, the notion of biorthogonal frames generalizes the notion of biorthogonal coordinates of a Hessian manifold, with respect to any given Riemannian metric g (not necessarily Hessian). Our main theorem shows that the pair of pseudo-Weitzenböck connections, each adapted to one of the pair of g-biorthogonal frames, are g-conjugate to each other. As a result, the pseudo-Weitzenböck connection pair generalize dually flat connections characteristic of Hessian manifolds, by being both curvature-free yet admitting (generally unequal) torsions. These results allow us to construct a pseudo-Weitzenböck connection for the manifold of parametric statistical models and treat it as a "statistical manifold admitting torsion".

Keywords Codazzi coupling · Conjugate connection · Biorthogonal · Curvature · Torsion · Dually flat · Partially flat



Volume 2
Issue 1
Article 13

RESEARCH PAPER

Alpha-Beta Log-Determinant Divergences Between Positive Definite Trace Class Operators

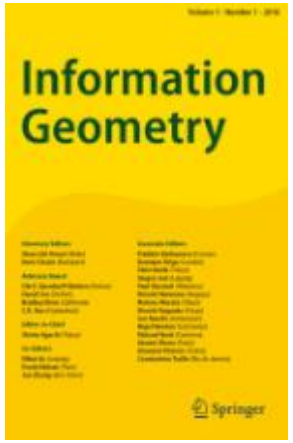
Hà Quang Minh[1]

Received: 2 May 2017 / Revised: 22 July 2019 / Published online: 26 August 2019

Abstract

This work presents a parametrized family of divergences, namely Alpha-Beta Log-Determinant (Log-Det) divergences, between positive definite unitized trace class operators on a Hilbert space. This is a generalization of the Alpha-Beta Log-Determinant divergences between symmetric, positive definite matrices to the infinite-dimensional setting. The family of Alpha-Beta Log-Det divergences is highly general and contains many divergences as special cases, including the recently formulated infinite-dimensional affine-invariant Riemannian distance and the infinite-dimensional Alpha Log-Det divergences between positive definite unitized trace class operators. In particular, it includes a parametrized family of metrics between positive definite trace class operators, with the affine-invariant Riemannian distance and the square root of the symmetric Stein divergence being special cases. For the Alpha-Beta Log-Det divergences between covariance operators on a Reproducing Kernel Hilbert Space (RKHS), we obtain closed form formulas via the corresponding Gram matrices.

Keywords Positive definite operators · Trace class operators · Infinite-dimensional Log-Determinant divergences · Alpha-Beta divergences · Affine-invariant Riemannian distance · Stein divergence · Extended trace · Extended Fredholm determinant · Reproducing kernel Hilbert spaces · Covariance operators



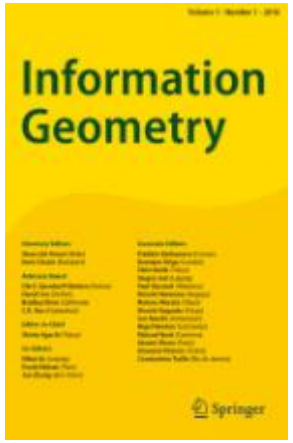
Volume 2
Issue 2
Article 14

RESEARCH PAPER

The geometry of recombination

Julian Hofrichter[1] · Jürgen Jost[1,2] · Tat Dat Tran[1]

Received: 18 March 2018 / Revised: 20 August 2019 / Published online: 29 August 2019



Volume 2
Issue 2
Article 15

Abstract

With the tools of information geometry, we can express relations between marginals of a joint distribution in geometric terms. We develop this framework in the context of population genetics and use this to interpret the famous Ohta–Kimura formula (cf. Ohta and Kimura in Genet Res 13(01):47–55, 1969) and discuss its generalizations for linkage equilibria in Wright–Fisher models with recombination with several loci. The state space associated with the Ohta–Kimura model is simply a Riemannian manifold of constant positive curvature. Furthermore, the equilibria states for recombination can be interpreted geometrically as a product of spheres. In the case of only 2 loci, we also derive the behavior of the mutual information between these two loci.

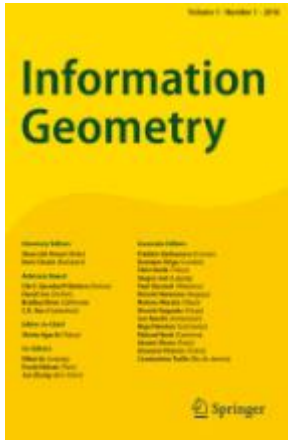
Keywords Wright–Fisher model · Random genetic drift · Recombination · Linkage · Fisher information metric · Equilibrium states · Compositionality

RESEARCH PAPER

Modelling election dynamics and the impact of disinformation

Dorje C. Brody[1]

Received: 1 April 2019 / Revised: 17 September 2019 / Published online: 7 October 2019



Volume 2
Issue 2
Article 16

Abstract

Complex dynamical systems driven by the unravelling of information can be modelled effectively by treating the underlying flow of information as the model input. Complicated dynamical behaviour of the system is then derived as an output. Such an information-based approach is in sharp contrast to the conventional mathematical modelling of information-driven systems whereby one attempts to come up with essentially *ad hoc* models for the outputs. Here, dynamics of electoral competition is modelled by the specification of the flow of information relevant to election. The seemingly random evolution of the election poll statistics are then derived as model outputs, which in turn are used to study election prediction, impact of disinformation, and the optimal strategy for information management in an election campaign.

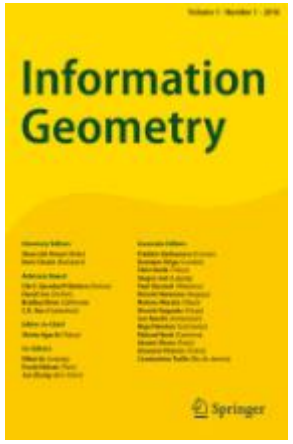
Keywords Information-based modelling · Electoral competition · Election prediction · Information control · Fake news

RESEARCH PAPER

Manifolds of classical probability distributions and quantum density operators in infinite dimensions

F. M. Ciaglia[1] · A. Ibort[2] · J. Jost[1] · G. Marmo[3]

Received: 12 June 2019 / Revised: 10 September 2019 / Published online: 24 October 2019



Volume 2
Issue 2
Article 17

Abstract

The manifold structure of subsets of classical probability distributions and quantum density operators in infinite dimensions is investigated in the context of C^*-algebras and actions of Banach-Lie groups. Specificaly, classical probability distributions and quantum density operators may be both described as states (in the functional analytic sense) on a given C^*-algebra $\mathscr{A}$ which is Abelian for Classical states, and non-Abelian for Quantum states. In this contribution, the space of states $\mathscr{S}$ of a possibly infinite-dimensional, unital C^*-algebra $\mathscr{A}$ is partitioned into the disjoint union of the orbits of an action of the group $\mathscr{G}$ of invertible elements of $\mathscr{A}$. Then, we prove that the orbits through density operators on an infinite-dimensional, separable Hilbert space $\mathcal{H}$ are smooth, homogeneous Banach manifolds of $\mathscr{G} = \mathcal{GL}(\mathcal{H})$, and, when $\mathscr{A}$ admits a faithful tracial state τ like it happens in the Classical case when we consider probability distributions with full support, we prove that the orbit through τ is a smooth, homogeneous Banach manifold for $\mathscr{G}$.

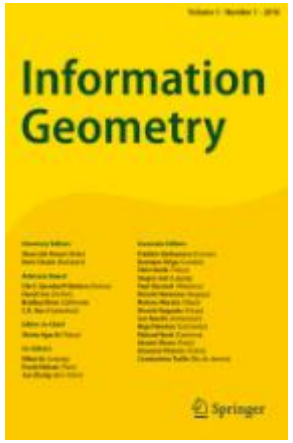
Keywords Probability distributions · Quantum states · C^*-algebras · Banach manifolds · Homogeneous spaces

RESEARCH PAPER

Spectral dimensionality reduction for Bregman information

Atsuya Kumagai[1]

Received: 17 March 2018 / Revised: 1 September 2019 / Published online: 7 October 2019



Volume 2
Issue 2
Article 18

Abstract

A spectral dimensionality reduction in dually flat spaces is formulated. On the basis of the expansion with respect to the affine coordinate, it is shown that some inner products compose Bregman divergence. Since the centroids are determined in two ways, depending on coordinate used for centering, Bregman information is also evaluated in two ways. In both cases, Bregman information is shown to be represented as a weighted sum of the eigenvalues of two inner product matrices. For given dissimilarities, the problem which finds the values of the inner product matrices is described as a semidefinite programming. As an example, numerical calculations are also presented.

Keywords Dimensionality reduction · Information geometry · Bregman divergence · Dually flat space

RESEARCH PAPER

A class of non-parametric statistical manifolds modelled on Sobolev space

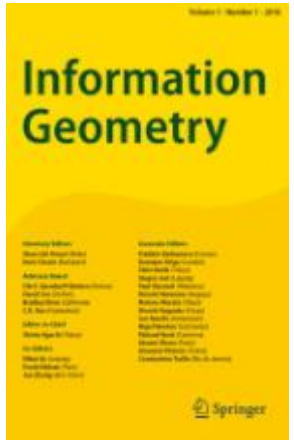
Nigel J. Newton[1]

Received: 13 December 2018 / Revised: 31 October 2019 / Published online: 25 November 2019

Abstract

We construct a family of non-parametric (infinite-dimensional) manifolds of finite measures on $\mathbb{R}^d$, each containing a smoothly embedded submanifold of *probability* measures. The manifolds are modelled on a variety of weighted Sobolev spaces, including Hilbert–Sobolev spaces and mixed-norm spaces, and support the Fisher–Rao metric as a weak Riemannian metric. Densities are expressed in terms of a *deformed exponential* function having linear growth. Unusually for the Sobolev context, and as a consequence of its linear growth, this "lifts" to a nonlinear superposition (Nemytskii) operator that acts continuously on a particular class of mixed-norm model spaces, and on the fixed norm space $W^{2,1}$; i.e. it maps each of these spaces continuously into itself. In contrast with non-parametric exponential manifolds, the density itself belongs to the model space, and the range of the chart is the whole of this space. Some of the results make essential use of a log-Sobolev embedding theorem, which also sharpens existing results concerning the regularity of statistical divergences on the manifolds. Applications to the stochastic partial differential equations of nonlinear filtering (and hence to the Fokker–Planck equation) are outlined.

Keywords Banach manifold · Fisher–Rao metric · Fokker–Planck equation · Log-Sobolev Embedding · Non-parametric statistics · Sobolev space



Volume 2
Issue 2
Article 19

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

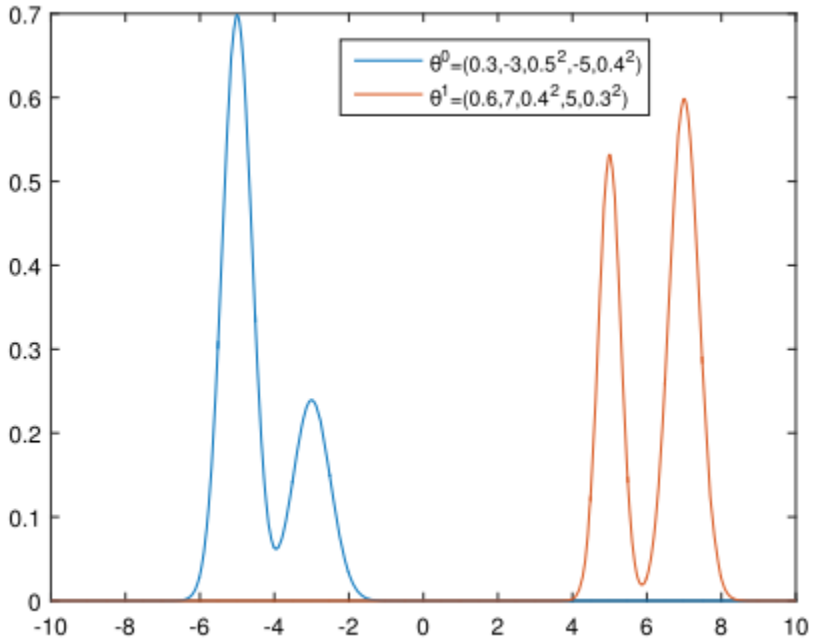
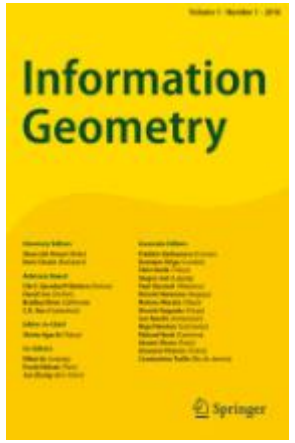


Fig. 1 Densities of Gaussian mixture distribution



Volume 3
Issue 1
Article 20

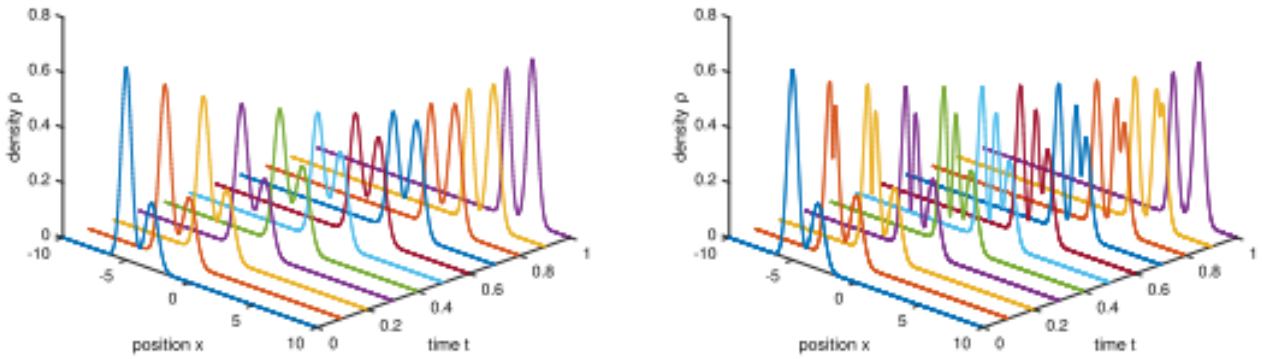


Fig. 2 Geodesic of Gaussian mixtures; left: in the Wasserstein statistical manifold; right: in the whole density space

RESEARCH PAPER

Cramér–Rao lower bounds arising from generalized Csiszár divergences

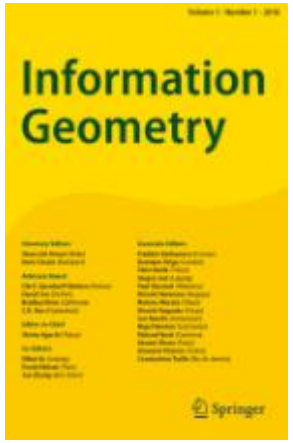
M. Ashok Kumar[1] · Kumar Vijay Mishra[2]

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



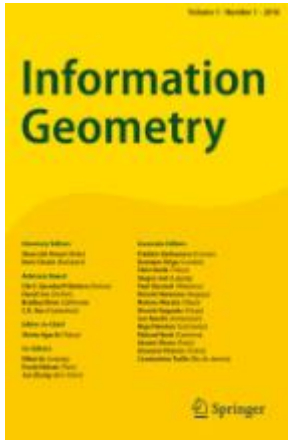
Volume 3
Issue 1
Article 21

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

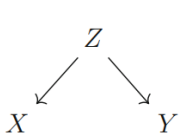


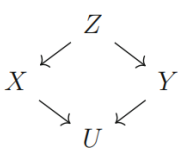
Volume 3
Issue 1
Article 22

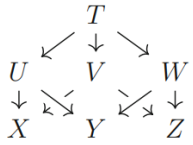
Abstract

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







RESEARCH PAPER

Ricci curvature for parametric statistics via optimal transport

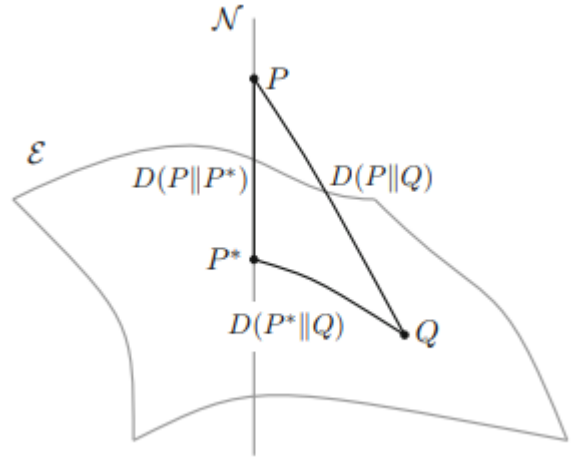
Wuchen Li[1] · **Guido Montúfar**[1]

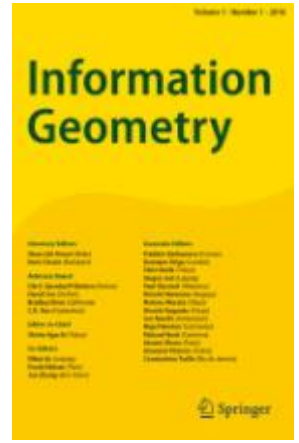
Received: 18 July 2018 / Revised: 15 October 2019 / Published online: 31 January 2020

Abstract

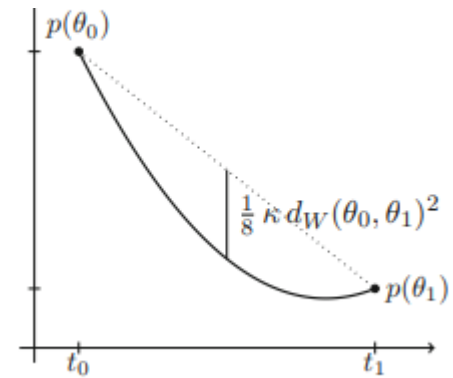
We define the notion of a Ricci curvature lower bound for parametrized statistical models. Following the seminal ideas of Lott–Sturm–Villani, we define this notion based on the geodesic convexity of the Kullback–Leibler divergence in a Wasserstein statistical manifold, that is, a manifold of probability 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.

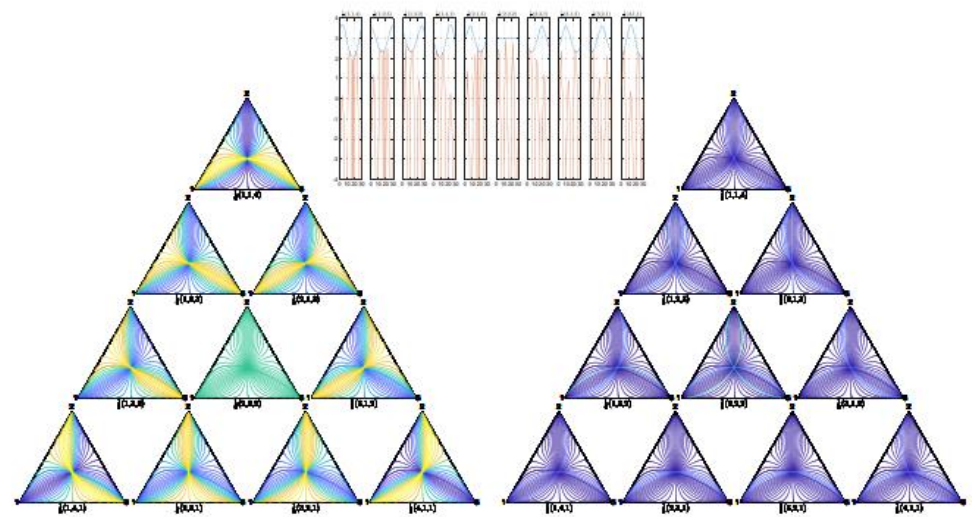
Keywords Ricci curvature · Information projection · Wasserstein statistical manifold · Fokker–Planck equation on parameter space





Volume 3
Issue 1
Article 22





RESEARCH PAPER

High dimensional nuisance parameters: an example from parametric survival analysis

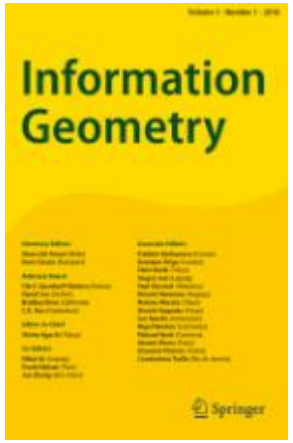
H. S. Battey[1] · D. R. Cox[2]

Received: 20 April 2020 / Revised: 12 July 2020 / Published online: 9 August 2020

Abstract

Parametric statistical problems involving both large amounts of data and models with many parameters raise issues that are explicitly or implicitly differential geometric. When the number of nuisance parameters is comparable to the sample size, alternative approaches to inference on interest parameters treat the nuisance parameters either as random variables or as arbitrary constants. The two approaches are compared in the context of parametric survival analysis, with emphasis on the effects of misspecification of the random effects distribution. Notably, we derive a detailed expression for the precision of the maximum likelihood estimator of an interest parameter when the assumed random effects model is erroneous, recovering simply derived results based on the Fisher information in the correctly specified situation but otherwise illustrating complex dependence on other aspects. Methods of assessing model adequacy are given. The results are both directly applicable and illustrate general principles of inference when there is a high-dimensional nuisance parameter. Open problems with an information geometrical bearing are outlined.

Keywords Conditional likelihood · Exponential distribution · Marginal likelihood · Matched pairs · Model comparison · Poisson process · Random effects · Model misspecification



Volume 3
Issue 2
Article 23

RESEARCH PAPER

Holonomic extended least angle regression

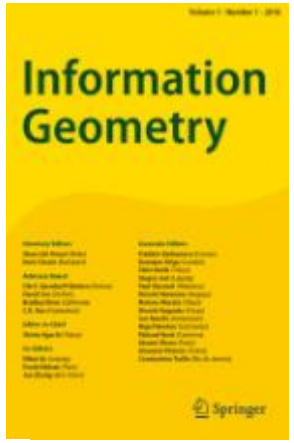
Marc Härkönen[1] · Tomonari Sei[2] · Yoshihiro Hirose[3]

Received: 30 August 2018 / Revised: 19 March 2020 / Accepted: 15 September 2020 /
Published online: 8 October 2020

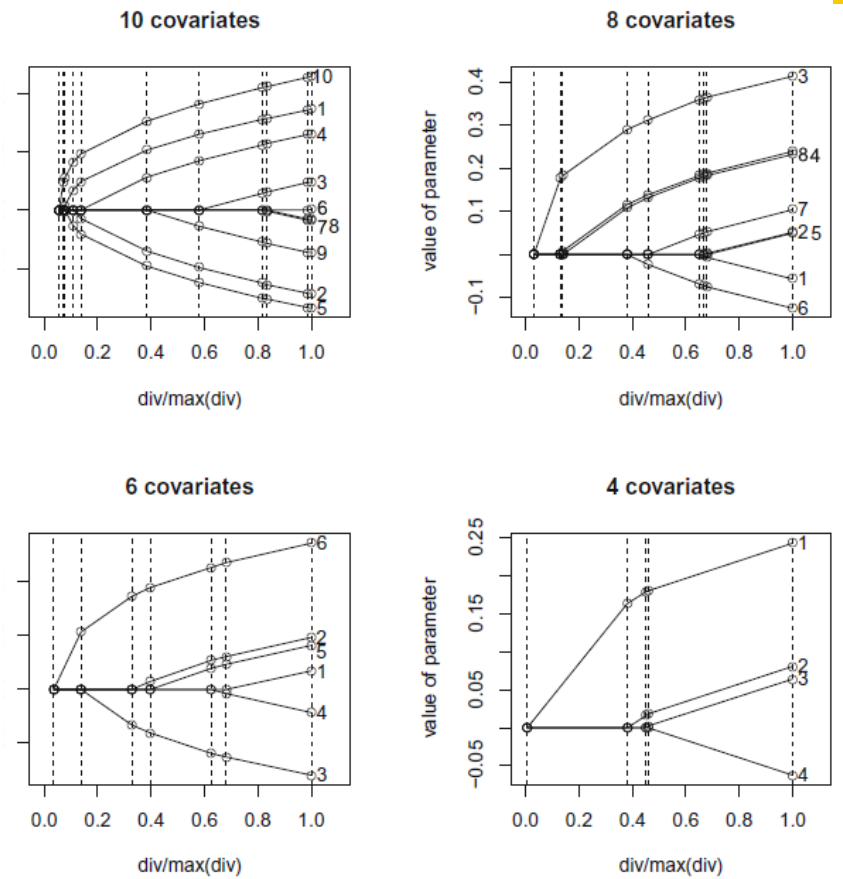
Abstract

One of the main problems studied in statistics is the fitting of models. Ideally, we would like to explain a large dataset with as few parameters as possible. There have been numerous attempts at automatizing this process. Most notably, the Least Angle Regression algorithm, or LARS, is a computationally efficient algorithm that ranks the covariates of a linear model. The algorithm is further extended to a class of distributions in the generalized linear model by using properties of the manifold of exponential families as dually flat manifolds. However this extension assumes that the normalizing constant of the joint distribution of observations is easy to compute. This is often not the case, for example the normalizing constant may contain a complicated integral. We circumvent this issue if the normalizing constant satisfies a holonomic system, a system of linear partial differential equations with a finite-dimensional space of solutions. In this paper we present a modification of the holonomic gradient method and add it to the extended LARS algorithm. We call this the holonomic extended least angle regression algorithm, or HELARS. The algorithm was implemented using the statistical software R, and was tested with real and simulated datasets.

Keywords Generalized linear model · Holonomic gradient method · Least angle regression



Volume 3
Issue 2
Article 24



imulation of 100 observations with uncorrelated covariates

RESEARCH PAPER

Geometry of Arimoto algorithm

Shoji Toyota[1]

Received: 11 July 2019 / Revised: 21 January 2020 / Published online: 30 October 2020

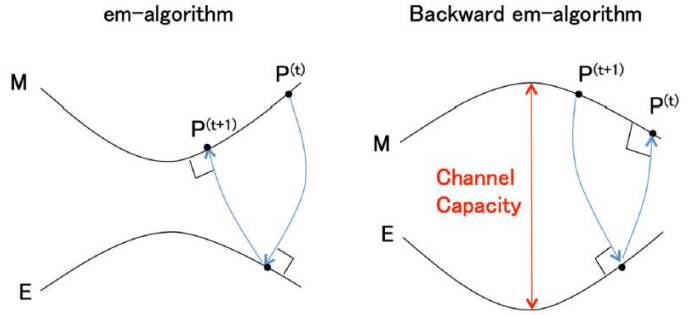
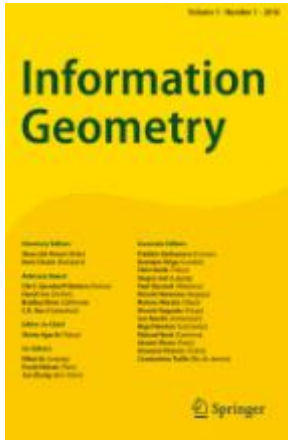


Fig. 1 Comparison between em-algorithm and Backward em-algorithm



Volume 3
Issue 2
Article 25

Abstract

In information theory, the channel capacity, which indicates how efficient a given channel is, plays an important role. The best-used algorithm for evaluating the channel capacity is Arimoto algorithm [3]. This paper aims to reveal an information geometric structure of Arimoto algorithm. In the process of trying to reveal an information geometric structure of Arimoto algorithm, a new algorithm that monotonically increases the Kullback-Leibler divergence is proposed, which is called "the Backward em-algorithm." Since the Backward em-algorithm is available in many cases where we need to increase the Kullback-Leibler divergence, it has a rich potential for application to many problems of statistics and information theory.

Keywords Information geometry · Arimoto algorithm · Channel capacity · Em-algorithm · Backword em-algorithm

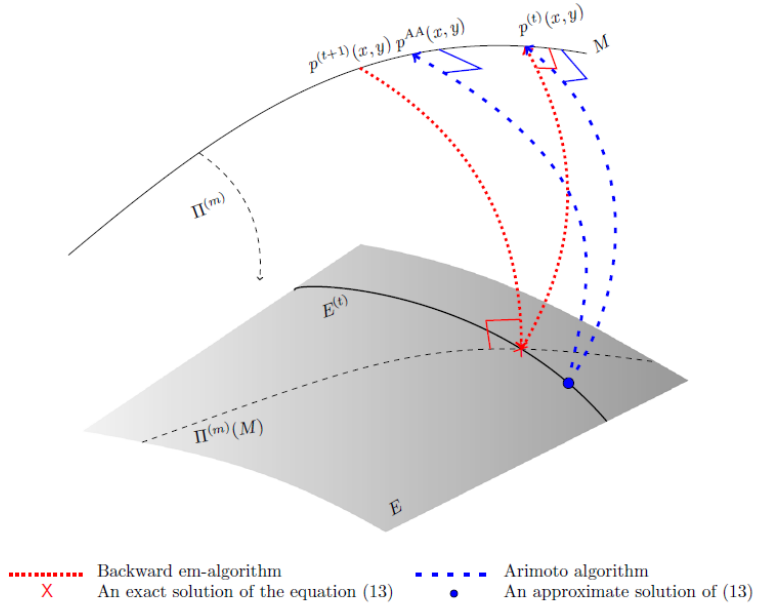


Fig. 2 Information geometric view of Arimoto algorithm. For the given probability distribution $p^{(t)}(x, y)$, $p^{(t+1)}(x, y)$ and $p^{AA}(x, y)$ is determined by the Backward em-algorithm and Arimoto algorithm respectively

RESEARCH PAPER

Natural alpha embeddings

Riccardo Volpi[1,2] · Luigi Malagò[1,2]

Received: 30 December 2019 / Revised: 23 November 2020 / Accepted: 2 February 2021 /
Published online: 18 April 2021

Abstract

Learning an embedding for a large collection of items is a popular approach to overcome the computational limitations associated to one-hot encodings. The aim of item embeddings is to learn a low dimensional space for the representations, able to capture with its geometry relevant features or relationships for the data at hand. This can be achieved for example by exploiting adjacencies among items in large sets of unlabelled data. In this paper we interpret in an Information Geometric framework the item embeddings obtained from conditional models. By exploiting the α-geometry of the exponential family, first introduced by Amari, we introduce a family of natural α-embeddings represented by vectors in the tangent space of the probability simplex, which includes as a special case standard approaches available in the literature. A typical example is given by word embeddings, commonly used in natural language processing, such as Word2Vec and GloVe. In our analysis, we show how the α-deformation parameter can impact on standard evaluation tasks.

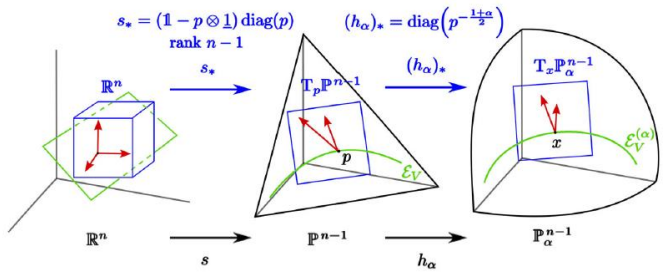


Fig. 1 Illustration of the mapping from the over-parametrization of the natural parameters in input to the softmax s, to the α-representation of the full simplex and of the exponential family $\mathcal{E}_V \subset \mathbb{P}^{n-1}$. Vectors in the tangent space are transported with the pushforward of the composite mapping

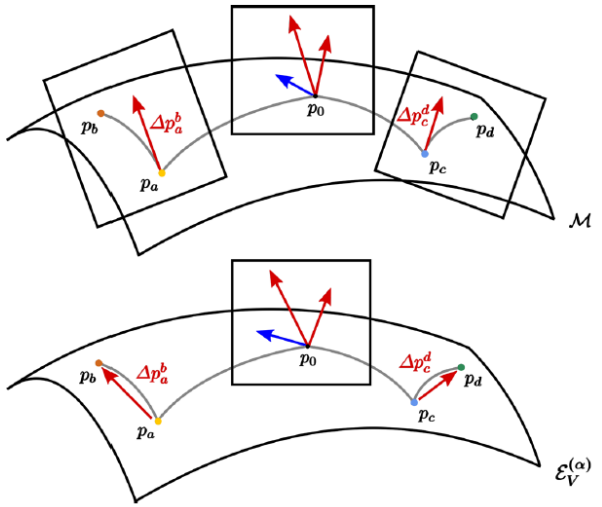
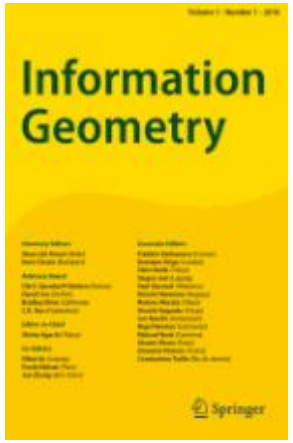


Fig. 2 (top) Geometric resolution of an analogy on a Riemannian manifold in a reference point p_0, by using a metric connection. (bottom) Resolution of an analogy in the α-embeddings framework. The points p should be intended as belonging to $\mathcal{E}_V^{(\alpha)}$, through the h_α-representation. In the bottom figure the vector $\Delta p_a^b = \mathrm{Log}_{p_a}^{(\alpha)} p_b$ does not belong to $\mathrm{T}_{p_a}\mathcal{E}_V^{(\alpha)}$, analogous is true for Δp_c^d



Volume 4
Issue 1
Article 26

RESEARCH PAPER

The dually flat structure for singular models

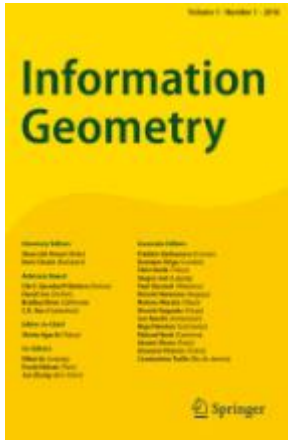
Naomichi Nakajima[1] · Toru Ohmoto[2]

Received: 13 November 2020 / Revised: 2 February 2021 / Accepted: 8 March 2021 /
Published online: 23 March 2021

Abstract

The *dually flat structure* introduced by Amari–Nagaoka is highlighted in information geometry and related fields. In practical applications, however, the underlying pseudo-Riemannian metric may often be degenerate, and such an excellent geometric structure is rarely defined on the entire space. To fix this trouble, in the present paper, we propose a novel generalization of the dually flat structure for a certain class of singular models from the viewpoint of *Lagrange and Legendre singularity theory*—we introduce a *quasi-Hessian manifold* endowed with a possibly degenerate metric and a particular symmetric cubic tensor, which exceeds the concept of statistical manifolds and is adapted to the theory of (weak) contrast functions. In particular, we establish Amari–Nagaoka's extended Pythagorean theorem and projection theorem in this general setup, and consequently, most of applications of these theorems are suitably justified even for such singular cases. This work is motivated by various interests with different backgrounds from Frobenius structure in mathematical physics to Deep Learning in data science.

Keywords Dually flat structure · Canonical divergence · Hessian geometry · Legendre duality · Wavefronts · Caustics · Singularity Theory



Volume 4
Issue 1
Article 27

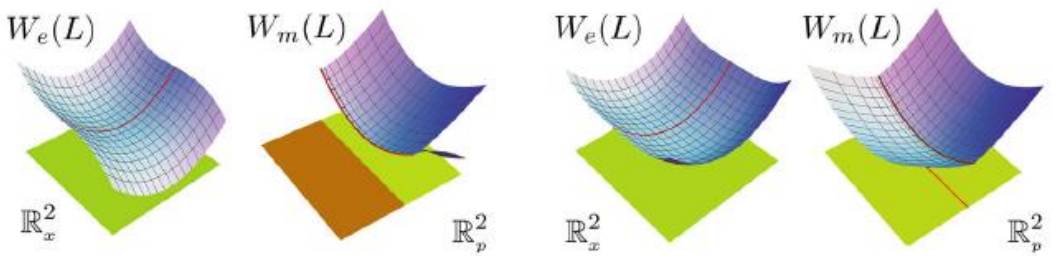
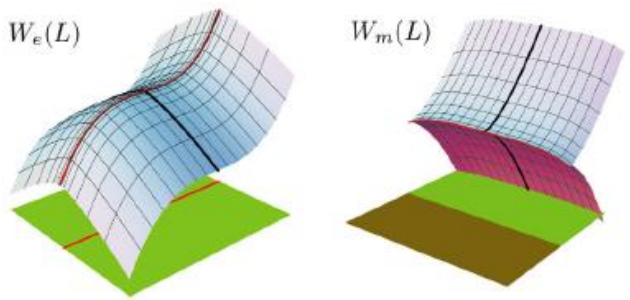


Fig. 1 The e/m-wavefronts and the e/m-caustics (Examples 3.4 and 3.5)

Fig. 2 Both e/m-wavefronts are singular (Example 3.6)



RESEARCH PAPER

Towards a canonical divergence within information geometry

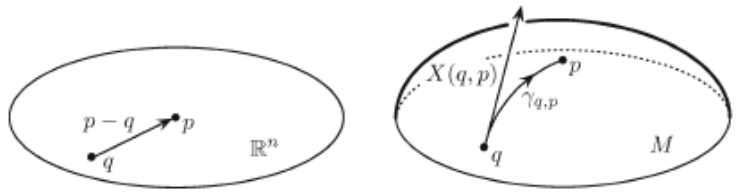
Domenico Felice[1,2] · Nihat Ay[2,3,4,5]

Received: 27 June 2018 / Revised: 22 May 2021 / Accepted: 27 May 2021 /
Published online: 22 June 2021

Abstract

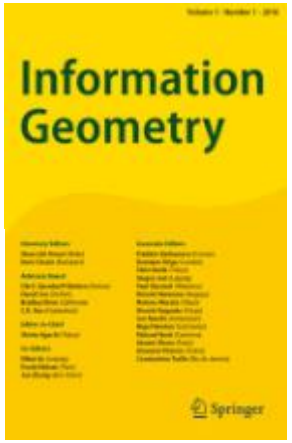
In Riemannian geometry geodesics are integral curves of the Riemannian distance gradient. We extend this classical result to the framework of Information Geometry. In particular, we prove that the rays of level-sets defined by a pseudo-distance are generated by the sum of two tangent vectors. By relying on these vectors, we propose a novel definition of a canonical divergence and its dual function. We prove that the new divergence allows to recover a given dual structure (g, ∇, ∇*) of a dually convex set on a smooth manifold M. Additionally, we show that this divergence coincides with the canonical divergence proposed by Ay and Amari in the case of: (a) self-duality, (b) dual flatness, (c) statistical geometric analogue of the concept of symmetric spaces in Riemannian geometry. For a dually convex set, the case (c) leads to a further comparison of the new divergence with the one introduced by Henmi and Kobayashi.

Keywords Classical differential geometry (02.40.Hw) · Riemannian geometries (02.40.Ky) · Inverse problems (02.30.Zz) · Information geometry · Divergence functions

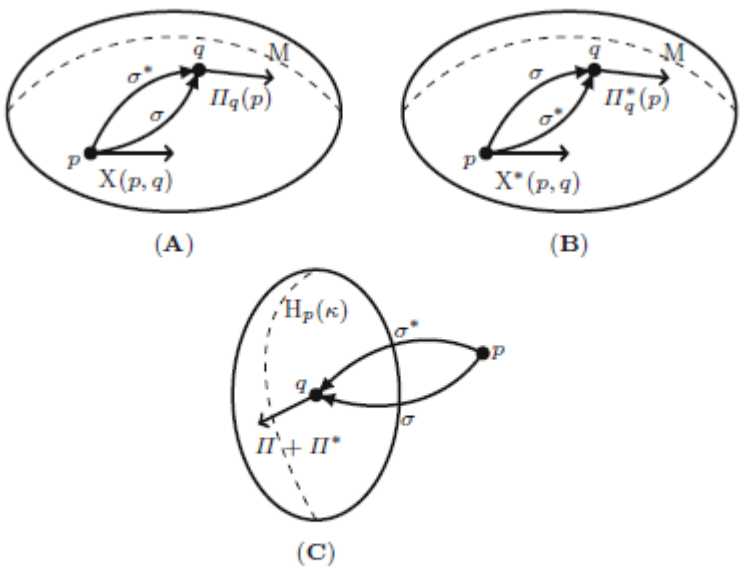


(A) (B)

Fig. 1 On the left, (**A**) illustrates the difference vector $p - q$ in the linear vector space $\mathbb{R}^n$; whereas, in (**B**) we can see the difference vector $X(q, p) = \dot{\gamma}_{q,p}(0)$ in M as the inverse of the exponential map at q (this figure comes from [4])



Volume 4
Issue 1
Article 28



(A) (B) (C)

Fig. 2 From the top to the bottom, (**A**) isllustrates the vector Π that is the ∇-parallel transport of $X(p, q) = \dot{\sigma}(0)$ along the ∇*-geodesic σ^* from p to q; while (**B**) illustrates the vector Π^* that is the ∇*-parallel transport of $X^*(p, q) = \dot{\sigma}^*(0)$ along the ∇-geodesic σ from p to q. Finally, (**C**) shows that the sum $\Pi + \Pi^*$ is orthogonal to the level-hypersurface $H_p(\kappa)$ of constant pseudo-squared-distance $r_p(q)$

RESEARCH PAPER

Information geometry and asymptotic geodesics on the space of normal distributions

Wolfgang Globke[1] · Raul Quiroga-Barranco[2]

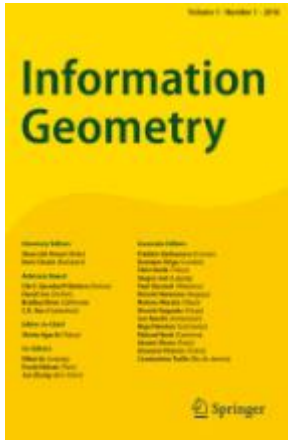
Received: 6 May 2019 / Revised: 10 June 2021 / Accepted: 17 June 2021 /
Published online: 25 June 2021

Abstract

The family $\mathcal{N}$ of n-variate normal distributions is parameterized by the cone of positive definite symmetric $n \times n$-matrices and the n-dimensional real vector space. Equipped with the Fisher information metric, $\mathcal{N}$ becomes a Riemannian manifold. As such, it is diffeomorphic, but not isometric, to the Riemannian symmetric space $\mathrm{Pos}_1(n+1, \mathbb{R})$ of unimodular positive definite symmetric $(n+1) \times (n+1)$-matrices. As the computation of distances in the Fisher metric for $n > 1$ presents some difficulties, Lovrič et al. (J Multivar Anal 74:36–48, 2000) proposed to use the Killing metric on $\mathrm{Pos}_1(n+1, \mathbb{R})$ as an alternative metric in which distances are easier to compute. In this work, we survey the geometric properties of the space $\mathcal{N}$ and provide a quantitative analysis of the defect of certain geodesics for the Killing metric to be geodesics for the Fisher metric. We find that for these geodesics the use of the Killing metric as an approximation for the Fisher metric is indeed justified for long distances.

Keywords Gaussian distributions · Fisher metric · Cone of positive definite matrices · Symmetric spaces · Geodesics.

Mathematics Subject Classification Primary 53C35; Secondary 53C30 · 62H05 · 62B10



Volume 4
Issue 1
Article 29

RESEARCH PAPER

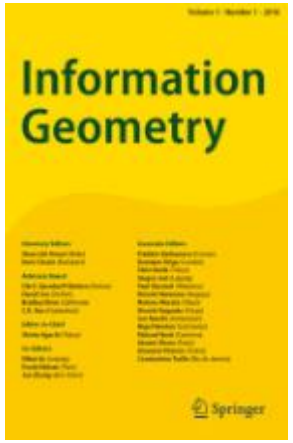
1-Conformal geometry of quasi statistical manifolds

Keisuke Haba[1]

Received: 13 July 2019 / Revised: 20 September 2020 / Accepted: 25 September 2020 /
Published online: 7 October 2020

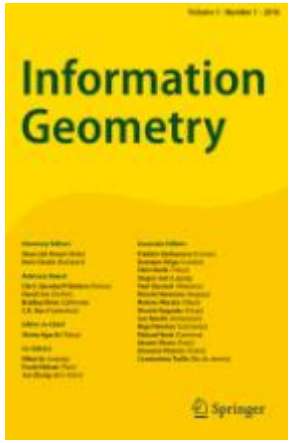
Abstract
A quasi statistical manifold is a generalization of a statistical manifold. The notion of quasi statistical manifolds was introduced to formulate the geometry of non-conservative estimating functions in statistics. Later, it was showed that quasi statistical manifolds are induced from affine distributions in the same way as statistical manifolds are induced from affine immersions. Here, an affine distribution is a non-integrable version of an affine immersion, and it is useful in quantum information geometry. On the other hand, it is known that generalized conformal geometry is useful for the study of statistical manifolds from the viewpoint of affine differential geometry. In particular, 1-conformal geometry of statistical manifolds gives a relation with the notion of affine immersions. Although generalized conformal geometry of quasi statistical manifolds is also expected to be useful, the geometry has not been cleared yet. The aim of this paper is to formulate 1-conformal geometry of quasi statistical manifolds. We research a relation between 1-conformal geometry of quasi statistical manifolds and the notion of affine distributions. As the main result, we show the fundamental theorems for affine distributions. We also formulate a hypersurface theory of quasi statistical manifolds.

Keywords Information geometry · Statistical manifold admitting torsion · Quasi statistical manifold · Affine distribution · 1-Conformal geometry



Volume 4
Issue 1
Article 30

RESEARCH PAPER



Volume 4
Issue 1
Article 31

A characterization of the alpha-connections on the statistical manifold of normal distributions

Hitoshi Furuhata[1] · Jun-ichi Inoguchi[2] · Shimpei Kobayashi[1]

Received: 3 September 2019 / Revised: 2 September 2020 / Accepted: 12 October 2020 /
Published online: 20 October 2020

Abstract

We show that the statistical manifold of normal distributions is homogeneous. In particular, it admits a 2-dimensional solvable Lie group structure. In addition, we give a geometric characterization of the Amari–Chentsov α-connections on the Lie group.

Keywords Statistical manifolds · The Amari–Chentsov α-connection · Lie groups

RESEARCH PAPER

Statistical submanifolds from a viewpoint of the Euler inequality

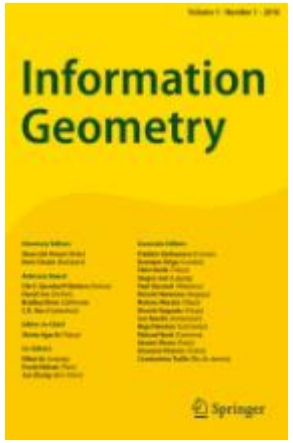
Naoto Satoh[1] · Hitoshi Furuhata[1] · Izumi Hasegawa[2] · Toshiyuki Nakane[4] · Yukihiko Okuyama[4] · Kimitake Sato[4] · Mohammad Hasan Shahid[3] · Aliya Naaz Siddiqui[3]

Received: 1 October 2019 / Revised: 15 April 2020 / Accepted: 21 August 2020 /
Published online: 4 September 2020

Abstract

We generalize the Euler inequality for statistical submanifolds. Several basic examples of doubly autoparallel statistical submanifolds in warped product spaces are described, for which the equality holds at each point. Besides, doubly totally-umbilical submanifolds are also illustrated.

Keywords Statistical manifolds · Warped product · The Euler inequality · Doubly autoparallel submanifolds · Doubly totally-umbilical submanifolds



Volume 4
Issue 1
Article 32

RESEARCH PAPER

Harmonic exponential families on homogeneous spaces

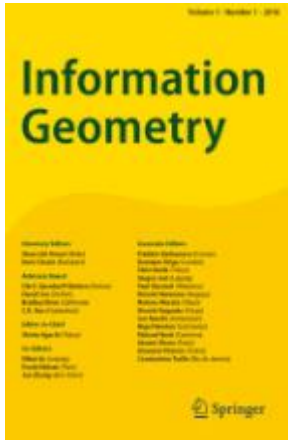
Koichi Tojo[1] · Taro Yoshino[2]

Received: 30 November 2019 / Revised: 27 August 2020 / Accepted: 4 September 2020 /
Published online: 6 October 2020

Abstract

Exponential families play an important role in the field of information geometry. By definition, there are infinitely many exponential families. However, only a small part of them are widely used. We want to give a framework to deal with these "good" families. In the light of the observation that the sample spaces of most of them are homogeneous spaces of certain Lie groups, we propose a method to construct exponential families on homogeneous spaces G/H by taking advantage of representation theory. Families obtained by this method are G-invariant exponential families. Then the following question naturally arises: are any G-invariant exponential families on G/H obtained by this method? We give an affirmative answer to this question. More precisely, any G-invariant exponential family on G/H can be realized as a subfamily of a family obtained by our method.

Keywords Exponential family · Representation theory · Homogeneous space · Harmonic analysis



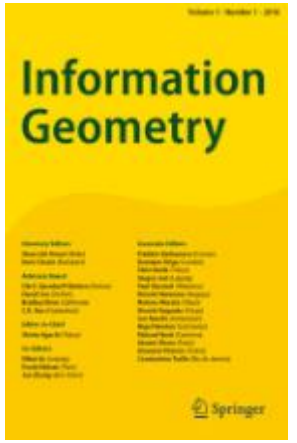
Volume 4
Issue 1
Article 33

SURVEY PAPER

Koszul lecture related to geometric and analytic mechanics, Souriau's Lie group thermodynamics and information geometry

Frédéric Barbaresco[1]

Received: 30 November 2019 / Revised: 6 November 2020 / Accepted: 8 November 2020 /
Published online: 13 January 2021



Volume 4
Issue 1
Article 34

Abstract

This paper deals with Jean-Louis Koszul's works related to Geometric and Analytic Mechanics, and to Souriau's Lie Group Thermodynamics that have appeared over time as elementary structures of Information Geometry. The 2nd Koszul form has been extended by Jean-Marie Souriau in his Symplectic model of Statistical Physics called "Lie Groups Thermodynamics" providing an extension of Fisher metric for homogeneous Symplectic manifolds, associated to KKS (Kirillov-Kostant-Souriau) 2-form in case of non-null cohomology. Jean-Louis Koszul has developed mathematical foundation of Souriau model in the Lecture "Introduction to Symplectic Geometry".

Keywords Koszul forms · Affine representation of lie algebra and lie group · Lie groups thermodynamics

Fig. 1 Jean-Louis Koszul and Hirihiko Shima at GSI'13 "Geometric Science of Information" conference in Ecole des Mines ParisTech in Paris, October 2013

RESEARCH PAPER

The last formula of Jean-Louis Koszul

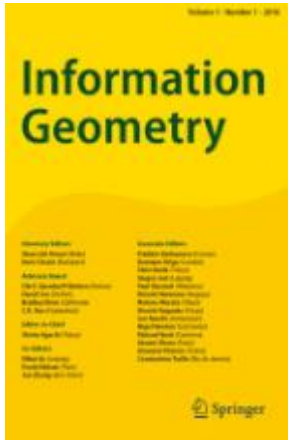
Michel Nguiffo Boyom[1]

Received: 6 April 2020 / Revised: 1 September 2020 / Accepted: 11 September 2020 /
Published online: 29 September 2020

Abstract

Since the first international conference in France on the Information Geometry, GSI2013, Jean-Louis Koszul's interest in Information Geometry went increasing. Motivated by the impact of the cohomology of Koszul-Vinberg algebras on the Information Geometry and moved by some issues raised by Albert Nijenhuis, Jean-Louis Koszul undertook another rewriting of the **Brut formula** of the coboundary operator of the KV complex. The source of other motivations of Jean-Louis Koszul was the relationships between the theory of KV cohomology and the theory of deformation of locally flat manifolds.In 2015 Jean-Louis Koszul sent me his *Last formula* of the KV boundary operator. In that *Last formula* Jean-Louis Koszul dealt with the case where spaces of coefficients are trivial modules of KV algebras. A part of the present work is devoted to extending the *Last formula* of Jean-Louis Koszul to KV cochain complexes whose spaces of coefficients are non-trivial two-sided modules of KV algebras. At another side, I also aim to highlight other significant impacts of the theory of KV cohomology of Koszul-Vinberg algebras. In particular I will use the KV cohomology to widely revisit the theory of statistical models of measurable sets. The reader will see why the source of the theory of statistical models is of homological nature.I also intend to highlight several impacts of the KV cohomology on the quantitative differential topology. I am particularly concerned with problems regarding the existence of Riemannian foliations, the existence of symplectic foliations as well as the existence of multi-dimensional webs. The homological theory of statistical models is presented as branches of rooted trees whose roots are *weakly Jensen* random cohomology classes.

Keywords Lie algebroids · KV cohomology · Canonical characteristic class · Koszul geometry · Functor of Amari · Locally flat manifolds · Complex systems



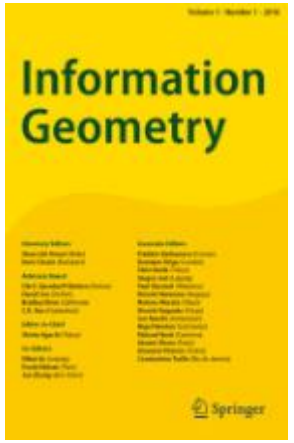
Volume 4
Issue 1
Article 35

RESEARCH PAPER

Self-similar solutions to the Hesse flow

Shun Maeta[1]

Received: 26 January 2021 / Revised: 27 May 2021 / Accepted: 9 August 2021 /
Published online: 17 August 2021



Volume 4
Issue 2
Article 36

Abstract

We define a Hesse soliton, that is, a self-similar solution to the Hesse flow on Hessian manifolds. On information geometry, the e-connection is important, which does not coincide with the Levi–Civita one. Therefore, it is interesting to consider a Hessian manifold with a flat connection which does not coincide with the Levi–Civita one. We call it a proper Hessian manifold. In this paper, we show that any compact proper Hesse soliton is expanding and any non-trivial compact gradient Hesse soliton is proper. Furthermore, we show that the dual space of a Hesse–Einstein manifold can be understood as a Hesse soliton.

Keywords Hesse flow · Hesse solitons · Hessian manifolds · Information geometry

RESEARCH PAPER

Riemannian barycentres of Gibbs distributions: new results on concentration and convexity in compact symmetric spaces

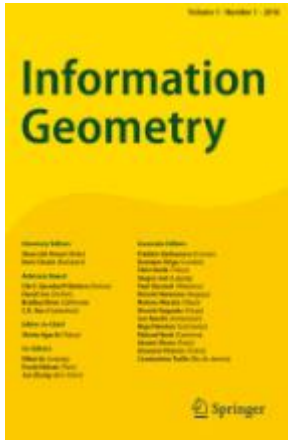
Salem Said[1] · Jonathan H. Manton[2]

Received: 3 June 2019 / Revised: 17 August 2021 / Accepted: 23 August 2021 /
Published online: 1 September 2021

Abstract

The Riemannian barycentre (or Fréchet mean) is the workhorse of data analysis for data taking values in Riemannian manifolds. The Riemannian barycentre of a probability distribution P on a Riemannian manifold M is a possible generalisation of the concept of expected value, at least when the barycentre is unique. Knowing when the barycentre of P is unique is of fundamental importance for its interpretation and computation. Existing results can only guarantee this uniqueness by assuming P is supported inside a convex geodesic ball $B(x^*, \delta) \subset M$. This assumption is overly restrictive since many distributions have support equal to M yet are sufficiently concentrated within a convex geodesic ball that they nevertheless have a unique barycentre. This paper studies the concentration of Gibbs distributions on Riemannian manifolds and gives conditions for the barycentre to be unique. Specifically, consider the Gibbs distribution $P = P_T$ with unnormalised density $\exp(-U/T)$ for some potential $U : M \to \mathbb{R}$ and some temperature $T > 0$. If M is a simply connected compact Riemannian symmetric space, and U has a unique global minimum at x^*, then for each $\delta < \frac{1}{2} r_{cx}$ (r_{cx} the convexity radius of M), there exists a critical temperature T_δ such that $T < T_\delta$ implies P_T has a unique Riemannian barycentre $\bar{x}_T$ and this $\bar{x}_T$ belongs to the geodesic ball $B(x^*, \delta)$. Moreover, if U is invariant by geodesic symmetry about x^*, then $\bar{x}_T = x^*$. Remarkably, this conclusion does not require the potential U to be smooth and therefore serves as the foundation of a new general algorithm for black-box optimisation. This algorithm is briefly illustrated with two numerical experiments.

Keywords Gibbs distribution · Riemannian barycentre · Wasserstein distance · Symmetric space · Black-box optimisation



Volume 4
Issue 2
Article 37

RESEARCH PAPER

The Bonnet theorem for statistical manifolds

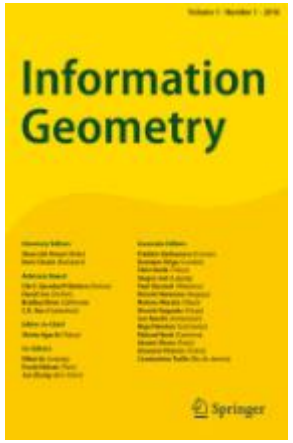
Taiji Marugame[1,2]

Received: 18 March 2021 / Revised: 18 July 2021 / Accepted: 16 September 2021 /
Published online: 29 September 2021

Abstract

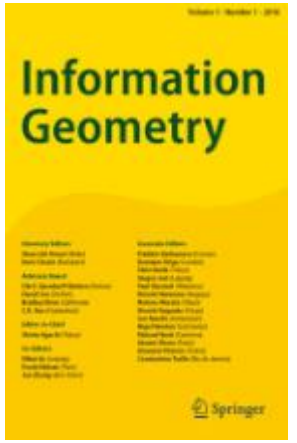
We prove the Bonnet theorem for statistical manifolds, which states that if a statistical manifold admits tensors satisfying the Gauss–Codazzi–Ricci equations, then it is locally embeddable to a flat statistical manifold (or a Hessian manifold). The proof is based on the notion of statistical embedding to the product of a vector space and its dual space introduced by Lauritzen. As another application of Lauritzen's embedding, we show that a statistical manifold admitting an affine embedding of codimension 1 or 2 is locally embeddable to a flat statistical manifold of the same codimension.

Keywords Statistical manifolds · Hessian manifolds · The Gauss–Codazzi–Ricci equations · The Bonnet theorem



Volume 4
Issue 2
Article 38

RESEARCH PAPER



Volume 4
Issue 2
Article 39

The Banach manifold of measures and the Lagrange multipliers of statistical mechanics

S. Dostoglou[1] · A. M. Hughes[1]

Received: 26 March 2021 / Revised: 6 September 2021 / Accepted: 18 October 2021 /
Published online: 3 November 2021

Abstract

Using a Banach manifold structure on the space of finite positive measures it is shown that all critical points of the Gibbs/information entropy are grand canonical equilibria when the constraints are scalar, and local equilibria when the constraints are integrable functions. This provides a rigorous derivation of equilibrium and local equilibrium Gibbs measures via Lagrange multipliers.

Keywords Gibbs ensembles · Lagrange multipliers · Local equilibrium · Banach manifold

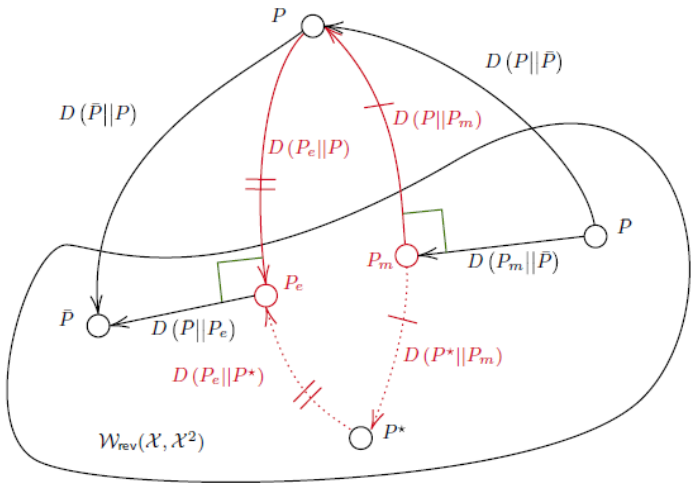


Fig. 1 Information projections P_e and P_m of P onto $\mathcal{W}_{\text{rev}}(\mathcal{X}, \mathcal{X}^2)$ in the full support case ($\mathcal{E} = \mathcal{X}^2$) (Theorem 7), Pythagorean identities (Theorem 7), and the bisection property (Proposition 2)

RESEARCH PAPER

Information Geometry of Reversible Markov Chains

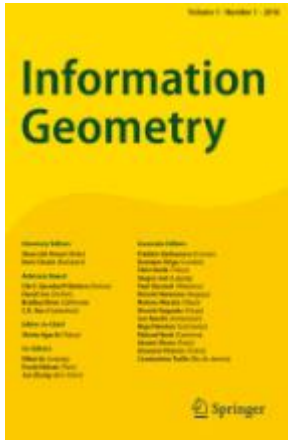
Geoffrey Wolfer[1] · Shun Watanabe[1]

Received: 11 June 2021 / Revised: 1 November 2021 / Accepted: 2 November 2021 /
Published online: 22 November 2021

Abstract

We analyze the information geometric structure of time reversibility for parametric families of irreducible transition kernels of Markov chains. We define and characterize reversible exponential families of Markov kernels, and show that irreducible and reversible Markov kernels form both a mixture family and, perhaps surprisingly, an exponential family in the set of all stochastic kernels. We propose a parametrization of the entire manifold of reversible kernels, and inspect reversible geodesics. We define information projections onto the reversible manifold, and derive closed-form expressions for the e-projection and m-projection, along with Pythagorean identities with respect to information divergence, leading to some new notion of reversiblization of Markov kernels. We show the family of edge measures pertaining to irreducible and reversible kernels also forms an exponential family among distributions over pairs. We further explore geometric properties of the reversible family, by comparing them with other remarkable families of stochastic matrices. Finally, we show that reversible kernels are, in a sense we define, the minimal exponential family generated by the m-family of symmetric kernels, and the smallest mixture family that comprises the e-family of memoryless kernels.

Keywords Irreducible Markov chain · Reversible Markov chain · Exponential family · Mixture family · Information projection



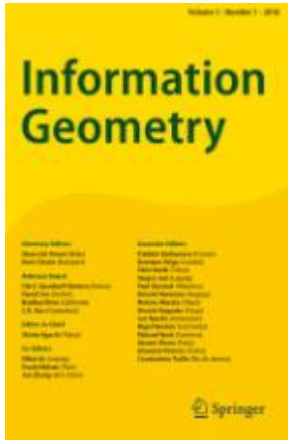
Volume 4
Issue 2
Article 40

RESEARCH PAPER

Transport information Bregman divergences

Wuchen Li[1]

Received: 29 January 2021 / Revised: 3 November 2021 / Accepted: 3 November 2021 /
Published online: 15 November 2021



Volume 4
Issue 2
Article 41

Abstract

We study Bregman divergences in probability density space embedded with the L^2–Wasserstein metric. Several properties and dualities of transport Bregman divergences are provided. In particular, we derive the transport Kullback–Leibler (KL) divergence by a Bregman divergence of negative Boltzmann–Shannon entropy in L^2–Wasserstein space. We also derive analytical formulas and generalizations of transport KL divergence for one-dimensional probability densities and Gaussian families.

Keywords Transport Bregman divergence · Transport KL divergence · Transport Jensen- Shannon divergence

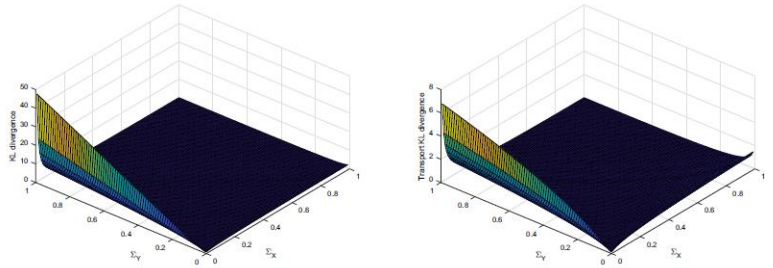


Fig. 1 A comparison between KL divergence and transport KL divergence for one dimensional Gaussian distributions. Left represents the KL divergence. Right represents the transport KL divergence

RESEARCH PAPER

Exponential arcs in the manifold of vector states on a σ-finite von Neumann algebra

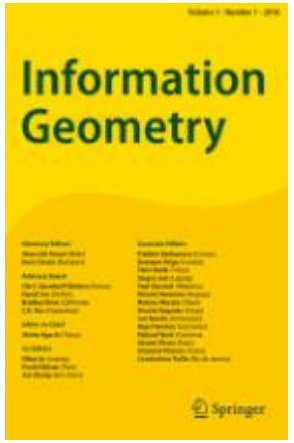
Jan Naudts[1]

Received: 23 April 2020 / Revised: 7 December 2021 / Accepted: 15 December 2021 /
Published online: 5 January 2022

Abstract

This paper introduces the notion of exponential arcs in Hilbert spaces and of exponential arcs connecting vector states on a sigma-finite von Neumann algebra in its standard representation. Results from Tomita–Takesaki theory form an essential ingredient. Starting point is a non-commutative Radon–Nikodym theorem that involves positive operators affiliated with the commutant algebra. It is shown that exponential arcs are differentiable and that parts of an exponential arc are again exponential arcs. Special cases of probability theory and of quantum probability are used to illustrate the approach.

Keywords Exponential arc · Exponential family · Tomita–Takesaki theory · Information geometry · Probability theory · Quantum probability



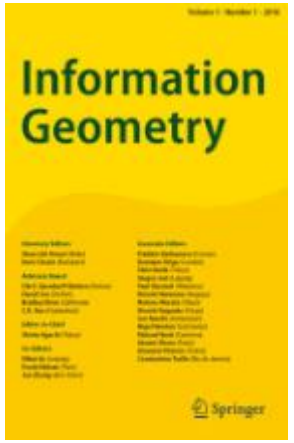
Volume 5
Issue 1
Article 42

RESEARCH PAPER

On a constant curvature statistical manifold

Shimpei Kobayashi[1] · Yu Ohno[1]

Received: 19 July 2021 / Revised: 31 January 2022 / Accepted: 6 February 2022 /
Published online: 15 February 2022



Volume 5
Issue 1
Article 43

Abstract

We will show that a statistical manifold (M, g, ∇) has a constant curvature if and only if it is a projectively flat conjugate symmetric manifold, that is, the affine connection ∇ is projectively flat and the curvatures satisfies $R = R^*$, where R^* is the curvature of the dual connection ∇^*. Moreover, we will show that properly convex structures on a projectively flat compact manifold induces constant curvature -1 statistical structures and vice versa.

Keywords Statistical manifolds · constant curvatures · Conjugate symmetries · Projective flatness · Properly convex structures

SURVEY PAPER

When optimal transport meets information geometry

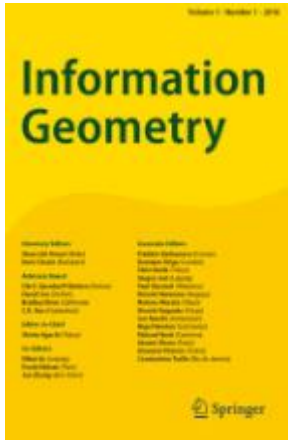
Gabriel Khan[1] · Jun Zhang[2]

Received: 22 March 2022 / Revised: 16 May 2022 / Accepted: 19 May 2022 /
Published online: 30 June 2022

Abstract

Information geometry and optimal transport are two distinct geometric frameworks for modeling families of probability measures. During the recent years, there has been a surge of research endeavors that cut across these two areas and explore their links and interactions. This paper is intended to provide an (incomplete) survey of these works, including entropy-regularized transport, divergence functions arising from c-duality, density manifolds and transport information geometry, the para-Kähler and Kähler geometries underlying optimal transport and the regularity theory for its solutions. Some outstanding questions that would be of interest to audience of both these two disciplines are posed. Our piece also serves as an introduction to the Special Issue on Optimal Transport of the journal *Information Geometry*.

Keywords Entropy-regulated transport · c-duality and divergence function · Kähler and para-Kähler geometries



Volume 5
Issue 1
Article 44

RESEARCH PAPER

Optimal transportation plans with escort entropy regularization

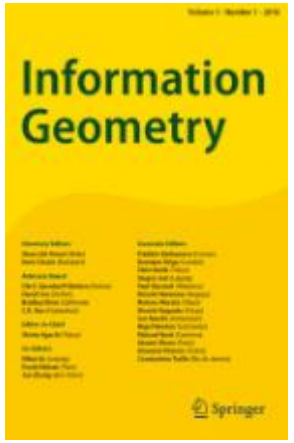
Takashi Kurose[1] · Shintaro Yoshizawa[2] · Shun-ichi Amari[3,4]

Received: 4 December 2019 / Revised: 26 September 2021 / Accepted: 21 October 2021 /
Published online: 9 November 2021

Abstract

The entropy-regularized Wasserstein transportation problem is useful for solving lots of problems in various applications. We study the deformed escort entropy regularization including the q-escort regularization and prove that the family of the inverse escort distributions of the optimal transportation plans forms a deformed exponential family, which has dually flat information-geometric structure. This elucidates the role of the escort transformation and its inverse in the theory of deformed exponential families. We further prove that the regularized cost function gives the dual potential of the flat manifold. We derive a new divergence function between two probability distributions in a probability simplex and a related Riemannian metric based on the regularized cost.

Keywords Wasserstein geometry · Deformed exponential family · Generalized entropy regularization · Dually flat manifold



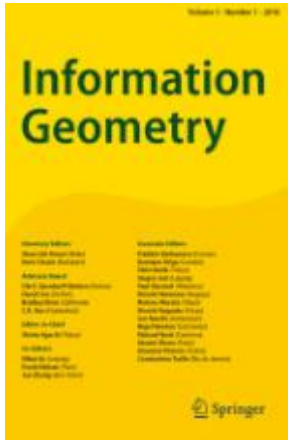
Volume 5
Issue 1
Article 45

RESEARCH PAPER

Optimal transport problems regularized by generic convex functions: a geometric and algorithmic approach

Daiji Tsutsui[1]

Received: 28 November 2020 / Revised: 31 October 2021 / Accepted: 3 November 2021 /
Published online: 25 November 2021



Volume 5
Issue 1
Article 46

Abstract

In order to circumvent the difficulties in solving numerically the discrete optimal transport problem, in which one minimizes the linear target function $P \mapsto \langle C, P\rangle := \sum_{i,j} C_{ij} P_{ij}$, Cuturi introduced a variant of the problem in which the target function is altered by a convex one $\Phi(P) = \langle C, P\rangle - \lambda \mathcal{H}(P)$, where $\mathcal{H}$ is the Shannon entropy and λ is a positive constant. We herein generalize their formulation to a target function of the form $\Phi(P) = \langle C, P\rangle + \lambda f(P)$, where f is a generic strictly convex smooth function. We also propose an iterative method for finding a numerical solution, and clarify that the proposed method is particularly efficient when $f(P) = \frac{1}{2}\|P\|^2$.

Keywords Information geometry · Discrete optimal transport · Entropic regularization · Convex optimization · Wasserstein barycenter

RESEARCH PAPER

Pseudo-Riemannian geometry encodes information geometry in optimal transport

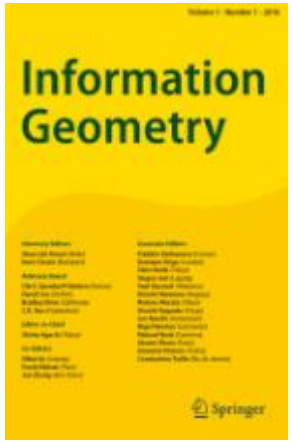
Ting-Kam Leonard Wong[1] · Jiaowen Yang[2]

Received: 28 September 2020 / Revised: 10 July 2021 / Accepted: 19 July 2021 /
Published online: 30 July 2021

Abstract

Optimal transport and information geometry both study geometric structures on spaces of probability distributions. Optimal transport characterizes the cost-minimizing movement from one distribution to another, while information geometry originates from coordinate invariant properties of statistical inference. Their relations and applications in statistics and machine learning have started to gain more attention. In this paper we give a new differential-geometric relation between the two fields. Namely, the pseudo-Riemannian framework of Kim and McCann, which provides a geometric perspective on the fundamental Ma–Trudinger–Wang (MTW) condition in the regularity theory of optimal transport maps, encodes the dualistic structure of statistical manifold. This general relation is described using the framework of c-divergence under which divergences are defined by optimal transport maps. As a by-product, we obtain a new information-geometric interpretation of the MTW tensor on the graph of the transport map. This relation sheds light on old and new aspects of information geometry. The dually flat geometry of Bregman divergence corresponds to the quadratic cost and the pseudo-Euclidean space, and the logarithmic $L^{(\alpha)}$-divergence introduced by Pal and the first author has constant sectional curvature in a sense to be made precise. In these cases we give a geometric interpretation of the information-geometric curvature in terms of the divergence between a primal-dual pair of geodesics.

Keywords Optimal transport · Information geometry · Pseudo-Riemannian geometry · c-Divergence · Logarithmic divergence · Bregman divergence · Ma–Trudinger–Wang tensor



Volume 5
Issue 1
Article 47

RESEARCH PAPER

Transport information geometry: Riemannian calculus on probability simplex

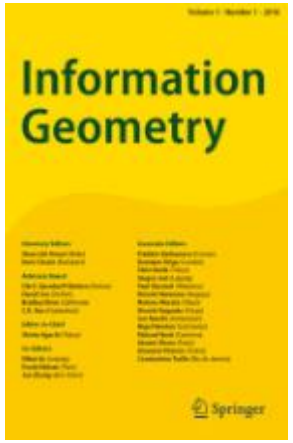
Wuchen Li[1]

Received: 8 April 2020 / Revised: 12 November 2020 / Accepted: 28 October 2021 /
Published online: 8 November 2021

Abstract

We formulate the Riemannian calculus of the probability set embedded with L^2-Wasserstein metric. This is an initial work of transport information geometry. Our investigation starts with the probability simplex (probability manifold) supported on vertices of a finite graph. The main idea is to embed the probability manifold as a submanifold of the positive measure space with a weighted graph Laplacian operator. By this viewpoint, we establish torsion–free Christoffel symbols, Levi–Civita connections, curvature tensors and volume forms in the probability manifold by Euclidean coordinates. As a consequence, the Jacobi equation, Laplace-Beltrami, Hessian operators and diffusion processes on the probability manifold are derived. These geometric computations are also provided in the infinite-dimensional density space (density manifold) supported on a finite-dimensional manifold. In particular, we present an identity connecting among Baker–Émery Γ_2 operator (carré du champ itéré), Fisher–Rao metric and optimal transport metric. Several examples are demonstrated.

Keywords Optimal transport · Information geometry · Probability manifold · Linear weighted Laplacian · Graph



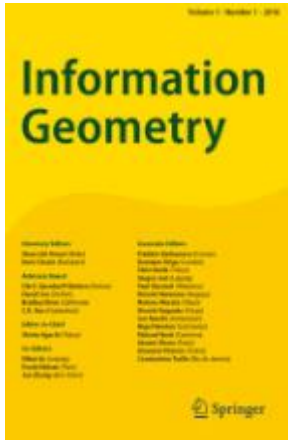
Volume 5
Issue 1
Article 48

RESEARCH PAPER

Kantorovich distance on finite metric spaces: Arens–Eells norm and CUT norms

Luigi Montrucchio[1] · Giovanni Pistone[2]

Received: 11 November 2019 / Revised: 18 May 2021 / Accepted: 17 June 2021 /
Published online: 1 July 2021



Volume 5
Issue 1
Article 49

Abstract

We study the possible closed-form representations of the K-distance arising in the Kantorovich transport problems on finite metric spaces. Weighted graphs, ℓ_1-embeddable metrics, and the related CUT norms receive special attention. Following an in-depth analysis of weighted trees and the tree-like spaces, we treat general metric spaces through their spanning trees.

Keywords Optimal transport · Kantorovich distance · Arens–Eells space · Finite metric space · Tree · Tree-like space · Spanning tree · Cut metric · ℓ_1-Embeddable space · Quotient map

RESEARCH PAPER

Tropical optimal transport and Wasserstein distances

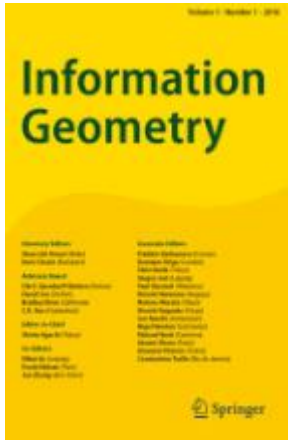
Wonjun Lee[1] · Wuchen Li[2] · Bo Lin[3] · Anthea Monod[4]

Received: 7 January 2020 / Revised: 18 May 2021 / Accepted: 20 May 2021 / Published online: 7 June 2021

Abstract

We study the problem of optimal transport in tropical geometry and define the Wasserstein-p distances in the continuous metric measure space setting of the tropical projective torus. We specify the tropical metric—a combinatorial metric that has been used to study of the tropical geometric space of phylogenetic trees—as the ground metric and study the cases of $p = 1, 2$ in detail. The case of $p = 1$ gives an efficient computation of the infinitely-many geodesics on the tropical projective torus, while the case of $p = 2$ gives a form for Fréchet means and a general inner product structure. Our results also provide theoretical foundations for geometric insight a statistical framework in a tropical geometric setting. We construct explicit algorithms for the computation of the tropical Wasserstein-1 and 2 distances and prove their convergence. Our results provide the first study of the Wasserstein distances and optimal transport in tropical geometry. Several numerical examples are provided.

Keywords Optimal transport · Tropical geometry · Tropical metric · Tropical projective torus · Wasserstein distances



Volume 5
Issue 1
Article 50

RESEARCH PAPER

Entropy-regularized 2-Wasserstein distance between Gaussian measures

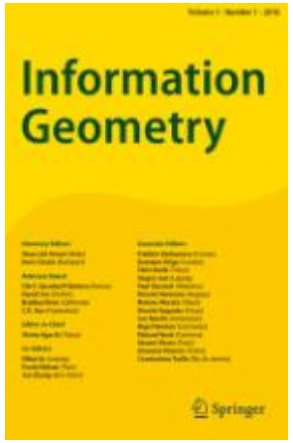
Anton Mallasto[1] · Augusto Gerolin[2] · Hà Quang Minh[3]

Received: 24 December 2019 / Revised: 16 July 2021 / Accepted: 17 July 2021 /
Published online: 16 August 2021

Abstract

Gaussian distributions are plentiful in applications dealing in uncertainty quantification and diffusivity. They furthermore stand as important special cases for frameworks providing geometries for probability measures, as the resulting geometry on Gaussians is often expressible in closed-form under the frameworks. In this work, we study the Gaussian geometry under the entropy-regularized 2-Wasserstein distance, by providing closed-form solutions for the distance and interpolations between elements. Furthermore, we provide a fixed-point characterization of a population barycenter when restricted to the manifold of Gaussians, which allows computations through the fixed-point iteration algorithm. As a consequence, the results yield closed-form expressions for the 2-Sinkhorn divergence. As the geometries change by varying the regularization magnitude, we study the limiting cases of vanishing and infinite magnitudes, reconfirming well-known results on the limits of the Sinkhorn divergence. Finally, we illustrate the resulting geometries with a numerical study.

Keywords Sinkhorn divergences · Multivariate Gaussian measures · Optimal transportation theory



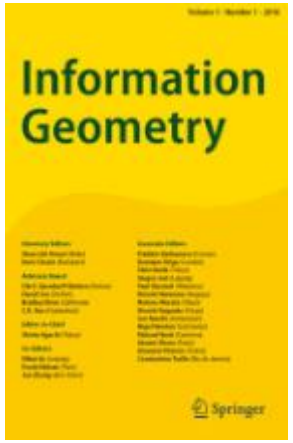
Volume 5
Issue 1
Article 51

RESEARCH PAPER

Coordinate-wise transformation of probability distributions to achieve a Stein-type identity

Tomonari Sei[1]

Received: 20 December 2019 / Revised: 1 March 2021 / Accepted: 17 June 2021 /
Published online: 28 June 2021



Volume 5
Issue 1
Article 52

Abstract

It is shown that for any given multi-dimensional probability distribution with regularity conditions, there exists a unique coordinate-wise transformation such that the transformed distribution satisfies a Stein-type identity. A sufficient condition for the existence is referred to as copositivity of distributions. The proof is based on an energy minimization problem over a totally geodesic subset of the Wasserstein space. The result is considered as an alternative to Sklar's theorem regarding copulas, and is also interpreted as a generalization of a diagonal scaling theorem. The Stein-type identity is applied to a rating problem of multivariate data. A numerical procedure for piece-wise uniform densities is provided. Some open problems are also discussed.

Keywords Copositive distribution · Copula · Energy minimization · Optimal transportation · Stein-type distribution · Wasserstein space

RESEARCH PAPER

Assignment flows for data labeling on graphs: convergence and stability

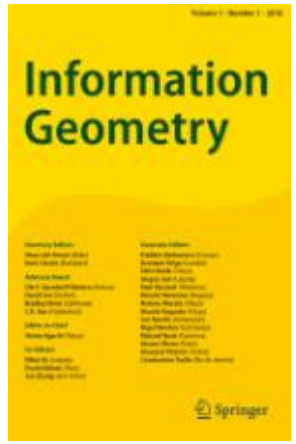
Artjom Zern[1] · Alexander Zeilmann[1] · Christoph Schnörr[1]

Received: 27 February 2020 / Revised: 7 October 2021 / Accepted: 31 October 2021 /
Published online: 18 November 2021

Abstract

The assignment flow recently introduced in the *J. Math. Imaging and Vision* 58/2 (2017) constitutes a high-dimensional dynamical system that evolves on a statistical product manifold and performs contextual labeling (classification) of data given in a metric space. Vertices of an underlying corresponding graph index the data points and define a system of neighborhoods. These neighborhoods together with nonnegative weight parameters define the regularization of the evolution of label assignments to data points, through geometric averaging induced by the affine e-connection of information geometry. From the point of view of evolutionary game dynamics, the assignment flow may be characterized as a large system of replicator equations that are coupled by geometric averaging. This paper establishes conditions on the weight parameters that guarantee convergence of the continuous-time assignment flow to integral assignments (labelings), up to a negligible subset of situations that will not be encountered when working with real data in practice. Furthermore, we classify attractors of the flow and quantify corresponding basins of attraction. This provides convergence guarantees for the assignment flow which are extended to the discrete-time assignment flow that results from applying a Runge–Kutta–Munthe–Kaas scheme for the numerical geometric integration of the assignment flow. Several counterexamples illustrate that violating the conditions may entail unfavorable behavior of the assignment flow regarding contextual data classification.

Keywords Assignment flow · Image and data labeling · Replicator equation · Evolutionary game dynamics · Information geometry



Volume 5
Issue 2
Article 53

RESEARCH PAPER

Bures–Wasserstein geometry for positive-definite Hermitian matrices and their trace-one subset

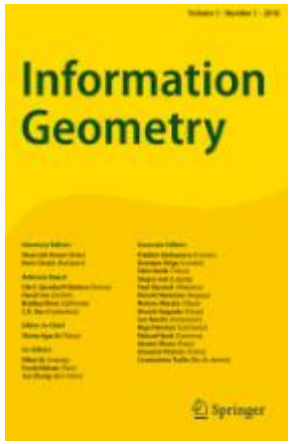
Jesse van Oostrum[1]

Received: 10 November 2021 / Revised: 17 June 2022 / Accepted: 10 August 2022

Abstract

In his classical argument, Rao derives the Riemannian distance corresponding to the Fisher metric using a mapping between the space of positive measures and Euclidean space. He obtains the Hellinger distance on the full space of measures and the Fisher distance on the subset of probability measures. In order to highlight the interplay between Fisher theory and quantum information theory, we extend this construction to the space of positive-definite Hermitian matrices using Riemannian submersions and quotient manifolds. The analog of the Hellinger distance turns out to be the Bures–Wasserstein (BW) distance, a distance measure appearing in optimal transport, quantum information, and optimisation theory. First we present an existing derivation of the Riemannian metric and geodesics associated with this distance. Subsequently, we present a novel derivation of the Riemannian distance and geodesics for this metric on the subset of trace-one matrices, analogous to the Fisher distance for probability measures.

Keywords Information geometry · Positive-definite matrices · Bures distance · Wasserstein metric · Optimal transport · Quantum information



Volume 5
Issue 2
Article 54

RESEARCH PAPER

The information geometry of two-field functional integrals

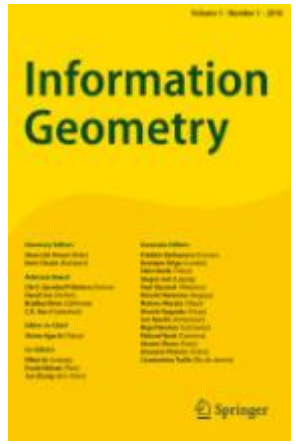
Eric Smith[1,2,3,4]

Received: 20 June 2019 / Revised: 12 September 2022 / Accepted: 2 October 2022 /
Published online: 19 October 2022

Abstract

Two-field functional integrals (2FFI) are an important class of solution methods for generating functions of dissipative processes, including discrete-state stochastic processes, dissipative dynamical systems, and decohering quantum densities. The stationary trajectories of these integrals describe a conserved current by Liouville's theorem, despite the absence of a conserved kinematic phase space current in the underlying stochastic process. We develop the information geometry of generating functions for discrete-state classical stochastic processes in the Doi-Peliti 2FFI form, and exhibit two quantities conserved along stationary trajectories. One is a Wigner function, familiar as a semiclassical density from quantum-mechanical time-dependent density-matrix methods. The second is an overlap function, between directions of variation in an underlying distribution and those in the directions of relative large-deviation probability that can be used to interrogate the distribution, and expressed as an inner product of vector fields in the Fisher information metric. To give an interpretation to the time invertibility implied by current conservation, we use generating functions to represent importance sampling protocols, and show that the conserved Fisher information is the differential of a sample volume under deformations of the nominal distribution and the likelihood ratio. We derive a pair of dual affine connections particular to Doi-Peliti theory for the way they separate the roles of the nominal distribution and likelihood ratio, distinguishing them from the standard dually-flat connection of Nagaoka and Amari defined on the importance distribution, and show that dual flatness in the affine coordinates of the coherent-state basis captures the special role played by coherent states in Doi-Peliti theory.

Keywords Information geometry · Doi-Peliti theory · Liouville's theorem · Fisher information · importance sampling · duality



Volume 5
Issue 2
Article 55

RESEARCH PAPER

A method to construct exponential families by representation theory

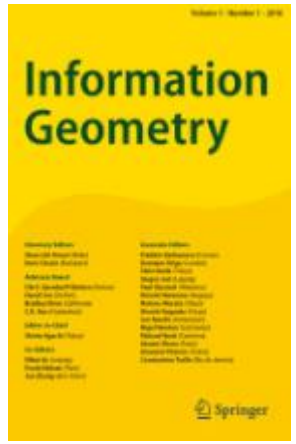
Koichi Tojo[1] · Taro Yoshino[2]

Received: 13 May 2022 / Revised: 30 September 2022 / Accepted: 8 October 2022 /
Published online: 23 October 2022

Abstract

In this paper, we give a method to construct "good" exponential families systematically by representation theory. More precisely, we consider a homogeneous space G/H as a sample space and construct an exponential family invariant under the transformation group G by using a representation of G. The method generates widely used exponential families such as normal, gamma, Bernoulli, categorical, Wishart, von Mises, Fisher–Bingham and hyperboloid distributions.

Keywords Exponential family · Representation theory · Homogeneous space · Transformation model · Harmonic exponential family



Volume 5
Issue 2
Article 56

RESEARCH PAPER

One-dimensional exponential families with constant Hessian sectional curvature

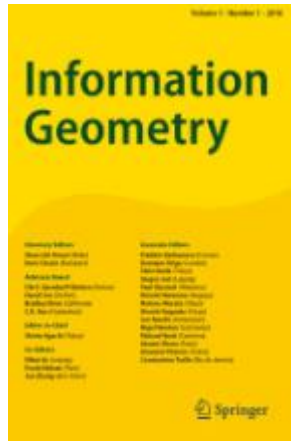
Mathieu Molitor[1]

Received: 22 September 2021 / Revised: 6 October 2022 / Accepted: 13 October 2022 /
Published online: 27 October 2022

Abstract

The concept of Hessian sectional curvature of a Hessian manifold M was introduced by Shima as a real-analogue of the holomorphic sectional curvature in Kähler geometry [15]. The former, unlike the latter, is also well-defined when the real dimension of M is 1. In this case, the Hessian sectional curvature is just a real-valued function on M. In this paper, we give a complete classification of 1-dimensional exponential families $\mathscr{E}$ defined over a finite set $\Omega = \{x_0, \ldots, x_m\}$ whose Hessian sectional curvature is constant. We observe an interesting phenomenon: if $\mathscr{E}$ has constant Hessian sectional curvature, say λ, then $\lambda = -\frac{1}{k}$ for some integer $1 \leq k \leq m$. We show that the family of Binomial distributions plays a central role in this classification.

Keywords Exponential families · Constant Hessian sectional curvature



Volume 5
Issue 2
Article 57

RESEARCH PAPER

On the locality of the natural gradient for learning in deep Bayesian networks

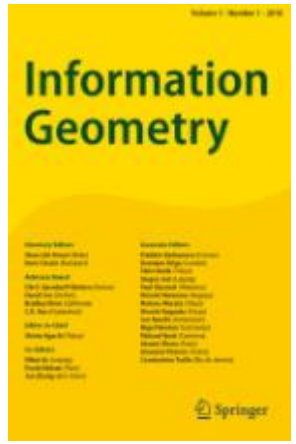
Nihat Ay[1,2,3]

Received: 29 June 2020 / Revised: 29 October 2020 / Accepted: 31 October 2020 / Published online: 24 November 2020

Abstract

We study the natural gradient method for learning in deep Bayesian networks, including neural networks. There are two natural geometries associated with such learning systems consisting of visible and hidden units. One geometry is related to the full system, the other one to the visible sub-system. These two geometries imply different natural gradients. In a first step, we demonstrate a great simplification of the natural gradient with respect to the first geometry, due to locality properties of the Fisher information matrix. This simplification does not directly translate to a corresponding simplification with respect to the second geometry. We develop the theory for studying the relation between the two versions of the natural gradient and outline a method for the simplification of the natural gradient with respect to the second geometry based on the first one. This method suggests to incorporate a recognition model as an auxiliary model for the efficient application of the natural gradient method in deep networks.

Keywords Natural gradient · Fisher–Rao metric · Deep learning · Helmholtz machines · Wake–sleep algorithm



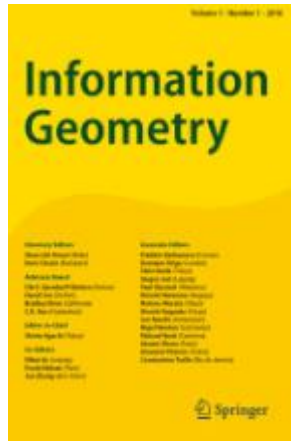
Volume 6
Issue 1
Article 58

RESEARCH PAPER

Invariance properties of the natural gradient in overparametrised systems

Jesse van Oostrum[1] · Johannes Müller[2] · Nihat Ay[1,3,4]

Received: 5 October 2021 / Revised: 31 March 2022 / Accepted: 18 May 2022



Volume 6
Issue 1
Article 59

Abstract

The natural gradient field is a vector field that lives on a model equipped with a distinguished Riemannian metric, e.g. the Fisher–Rao metric, and represents the direction of steepest ascent of an objective function on the model with respect to this metric. In practice, one tries to obtain the corresponding direction on the parameter space by multiplying the ordinary gradient by the inverse of the Gram matrix associated with the metric. We refer to this vector on the parameter space as the natural parameter gradient. In this paper we study when the pushforward of the natural parameter gradient is equal to the natural gradient. Furthermore we investigate the invariance properties of the natural parameter gradient. Both questions are addressed in an overparametrised setting.

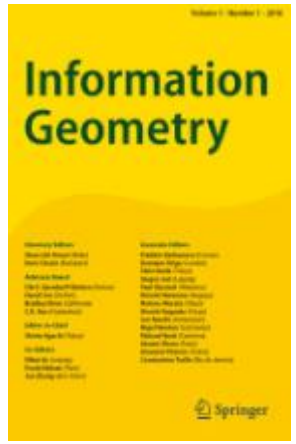
Keywords Natural gradient · Riemannian metric · Deep learning · Information geometry

RESEARCH PAPER

Laplacian operator on statistical manifold

Ruichao Jiang[1] · Javad Tavakoli[2] · Yiqiang Zhao[1]

Received: 24 February 2022 / Revised: 21 June 2022 / Accepted: 12 August 2022 /
Published online: 8 September 2022



Volume 6
Issue 1
Article 60

Abstract

In this paper, we define a Laplacian operator on a statistical manifold, called the vector Laplacian. This vector Laplacian incorporates information from the Amari–Chentsov tensor. We derive a formula for the vector Laplacian. We also give two applications using the heat kernel associated with the vector Laplacian.

RESEARCH PAPER

Active learning by query by committee with robust divergences

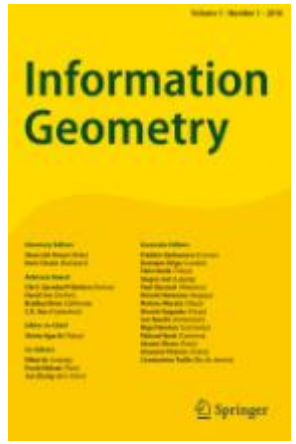
Hideitsu Hino[1,2] · Shinto Eguchi[1]

Received: 1 June 2022 / Revised: 13 November 2022 / Accepted: 15 November 2022 /
Published online: 21 November 2022

Abstract

Active learning is a widely used methodology for various problems with high measurement costs. In active learning, the next object to be measured is selected by an acquisition function, and measurements are performed sequentially. The query by committee is a well-known acquisition function. In conventional methods, committee disagreement is quantified by the Kullback–Leibler divergence. In this paper, the measure of disagreement is defined by the Bregman divergence, which includes the Kullback–Leibler divergence as an instance, and the dual γ-power divergence. As a particular class of the Bregman divergence, the β-divergence is considered. By deriving the influence function, we show that the proposed method using β-divergence and dual γ-power divergence are more robust than the conventional method in which the measure of disagreement is defined by the Kullback–Leibler divergence. Experimental results show that the proposed method performs as well as or better than the conventional method.

Keywords Information geometry · Bregman divergence · Power divergence · Active learning · Query by committee · Robust statistics



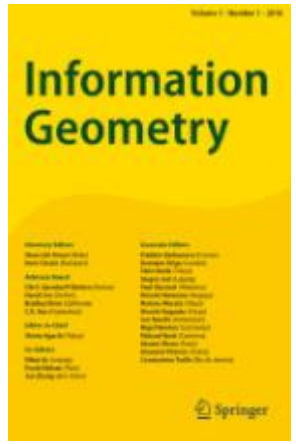
Volume 6
Issue 1
Article 61

RESEARCH PAPER

The Fisher–Rao loss for learning under label noise

Henrique K. Miyamoto[1] · Fábio C. C. Meneghetti[1] · Sueli I. R. Costa[1]

Received: 17 July 2022 / Revised: 28 October 2022 / Accepted: 2 November 2022 /
Published online: 2 December 2022



Volume 6
Issue 1
Article 62

Abstract

Choosing a suitable loss function is essential when learning by empirical risk minimisation. In many practical cases, the datasets used for training a classifier may contain incorrect labels, which prompts the interest for using loss functions that are inherently robust to label noise. In this paper, we study the Fisher–Rao loss function, which emerges from the Fisher–Rao distance in the statistical manifold of discrete distributions. We derive an upper bound for the performance degradation in the presence of label noise, and analyse the learning speed of this loss. Comparing with other commonly used losses, we argue that the Fisher–Rao loss provides a natural trade-off between robustness and training dynamics. Numerical experiments with synthetic and MNIST datasets illustrate this performance.

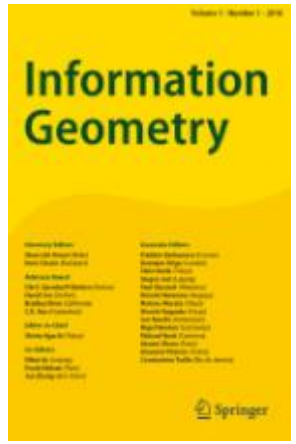
Keywords Classification · Fisher–Rao distance · Information geometry · Label noise · Loss functions · Neural networks

RESEARCH PAPER

Information geometry of warped product spaces

Yasuaki Fujitani[1]

Received: 25 February 2022 / Revised: 19 November 2022 / Accepted: 6 December 2022 /
Published online: 16 December 2022



Volume 6
Issue 1
Article 63

Abstract

Information geometry is an important tool to study statistical models. There are some important examples in statistical models which are regarded as warped products. In this paper, we study information geometry of warped products. We consider the case where the warped product and its fiber space are equipped with dually flat connections and, in the particular case of a cone, characterize the connections on the base space $\mathbb{R}_{>0}$. The resulting connections turn out to be the α-connections with $\alpha = \pm 1$.

Keywords Information geometry · Warped product · α-connection

RESEARCH PAPER

Coarse geometric kernels for networks embedding

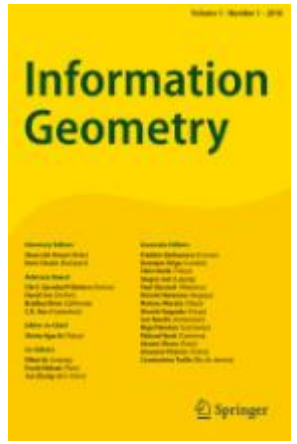
Emil Saucan[1] · Vladislav Barkanass[1] · Jürgen Jost[2]

Received: 18 July 2022 / Revised: 5 November 2022 / Accepted: 18 December 2022 /
Published online: 26 January 2023

Abstract

We develop embedding kernels based on the Forman–Ricci curvature and intertwined Bochner–Laplacian and employ them for the detection of the coarse structure of networks, as well as for network visualization with applications to support-vector machines (SVMs).

Keywords Coarse geometry · Embedding kernels · Forman graph Ricci curvature · Forman Bochner–Laplacian



Volume 6
Issue 1
Article 64

RESEARCH PAPER

Plücker coordinates of the best-fit Stiefel tropical linear space to a mixture of Gaussian distributions

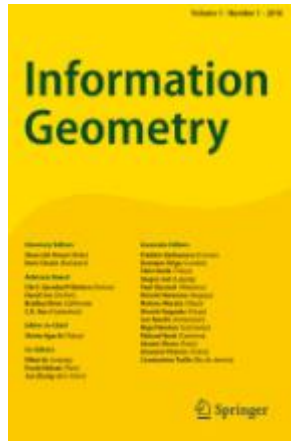
Keiji Miura[1] · Ruriko Yoshida[2]

Received: 22 December 2021 / Revised: 16 December 2022 / Accepted: 22 January 2023 /
Published online: 31 January 2023

Abstract

In this research, we investigate a tropical principal component analysis (PCA) as a best-fit Stiefel tropical linear space to a given sample over the tropical projective torus for its dimensionality reduction and visualization. Especially, we characterize the best-fit Stiefel tropical linear space to a sample generated from a mixture of Gaussian distributions as the variances of the Gaussians go to zero. For a single Gaussian distribution, we show that the sum of residuals in terms of the tropical metric with the max-plus algebra over a given sample to a fitted Stiefel tropical linear space converges to zero by giving an upper bound for its convergence rate. Meanwhile, for a mixtures of Gaussian distribution, we show that the best-fit tropical linear space can be determined uniquely when we send variances to zero. We briefly consider the best-fit topical polynomial as an extension for the mixture of more than two Gaussians over the tropical projective space of dimension three. We show some geometric properties of these tropical linear spaces and polynomials.

Keywords Max-plus algebra · Principal component analysis · Tropical geometry · Tropical metric · Mixtures of Gaussian distributions · Tropical polynomials



Volume 6
Issue 1
Article 65

RESEARCH PAPER

Wasserstein information matrix

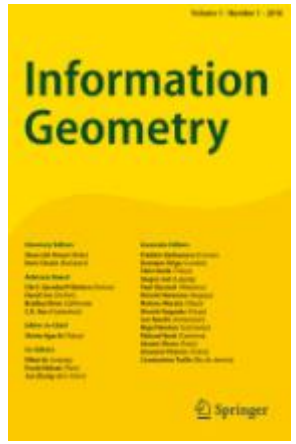
Wuchen Li[1] · Jiaxi Zhao[1]

Received: 29 November 2020 / Revised: 19 September 2022 / Accepted: 22 January 2023 /
Published online: 14 February 2023

Abstract

We study information matrices for statistical models by the L^2-Wasserstein metric. We call them Wasserstein information matrices (WIMs), which are analogs of classical Fisher information matrices. We introduce Wasserstein score functions and study covariance operators in statistical models. Using them, we establish Wasserstein–Cramer–Rao bounds for estimations and explore their comparisons with classical results. We next consider the asymptotic behaviors and efficiency of estimators. We derive the online asymptotic efficiency for Wasserstein natural gradient. Besides, we establish a Poincaré efficiency for Wasserstein natural gradient of maximal likelihood estimation. Several analytical examples of WIMs are presented, including location-scale families, independent families, rectified linear unit (ReLU) generative models.

Keywords Wasserstein information matrix · Wasserstein score function · Wasserstein–Cramer–Rao inequality · Wasserstein online efficiency · Poincaré efficiency



Volume 6
Issue 1
Article 66

RESEARCH PAPER

Non-negative low-rank approximations for multi-dimensional arrays on statistical manifold

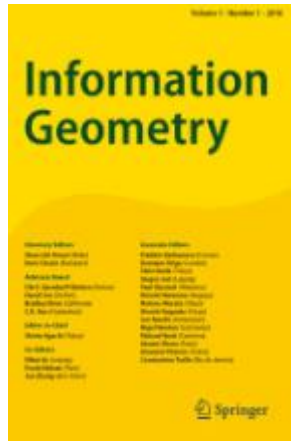
Kazu Ghalamkari[1,2] · Mahito Sugiyama[1,2]

Received: 17 July 2022 / Revised: 16 December 2022 / Accepted: 26 January 2023 /
Published online: 24 February 2023

Abstract

Although *low-rank approximation* of multi-dimensional arrays has been widely discussed in linear algebra, its statistical properties remain unclear. In this paper, we use information geometry to uncover a statistical picture of non-negative low-rank approximations. First, we treat each input array as a probability distribution using a log-linear model on a poset, where a structure of an input array is realized as a partial order. We then describe the low-rank condition of arrays as constraints on parameters of the model and formulate the low-rank approximation as a projection onto a subspace that satisfies such constraints, where parameters correspond to coordinate systems of a statistical manifold. Second, based on information-geometric analysis of low-rank approximation, we point out the unexpected relationship between the rank-1 non-negative low-rank approximation and *mean-field approximation*, a well-established method in physics that uses a one-body problem to approximate a many-body problem. Third, our theoretical discussion leads to a novel optimization method of non-negative low-rank approximation, called Legendre Tucker rank reduction. Because the proposed method does not use the gradient method, it does not require tuning parameters such as initial position, learning rate, and stopping criteria. In addition, the flexibility of the log-linear model enables us to treat the problem of non-negative multiple matrix factorization (NMMF), a variant of low-rank approximation with shared factors. We find the best rank-1 NMMF formula as a closed form and develop a rapid rank-1 NMF method for arrays with missing entries based on the closed form, called A1GM.

Keywords Low-rank approximation · Information geometry · Tucker rank reduction



Volume 6
Issue 1
Article 67

RESEARCH PAPER

The face lattice of the set of reduced density matrices and its coatoms

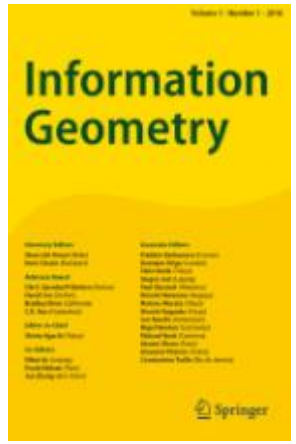
Stephan Weis[1] · João Gouveia[2]

Received: 21 February 2022 / Revised: 1 January 2023 / Accepted: 19 February 2023 /
Published online: 9 March 2023

Abstract

The lattice of faces of the convex set of reduced density matrices is essential for the construction of the information projection to a hierarchical model. The lattice of faces is also important in quantum state tomography. Yet, the description and computation of these faces is elusive in the simplest examples. Here, we study the face lattice of the set of two-body reduced density matrices: We show that the three-qubit lattice has no elements of rank seven and that it has a family of coatoms of rank five. This contrasts with the three-bit lattice, where every coatom has rank six. We discovered the coatoms of rank five using a novel experimental method, which employs convex duality, semidefinite programming, and algebra. We also discuss nonexposed points for three and six qubits. Using frustration-free Hamiltonians, we provide a new characterization of probability distributions that factor.

Keywords Information projection · Reduced density matrices · Face lattice · Exposed face · Joint numerical range · Spectrahedron · Semidefinite programming · Local Hamiltonian · Frustration-free



Volume 6
Issue 1
Article 68

RESEARCH PAPER

Power transformations of relative count data as a shrinkage problem

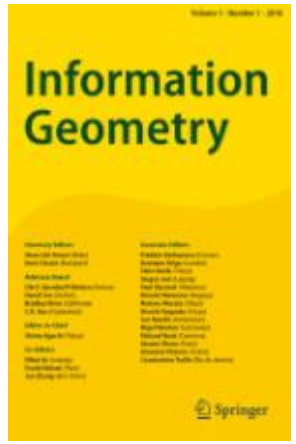
Ionas Erb[1]

Received: 25 May 2022 / Revised: 13 February 2023 / Accepted: 14 March 2023 /
Published online: 13 April 2023

Abstract

Here we show an application of our recently proposed information-geometric approach to compositional data analysis (CoDA). This application regards relative count data, which are, e.g., obtained from sequencing experiments. First we review in some detail a variety of necessary concepts ranging from basic count distributions and their information-geometric description over the link between Bayesian statistics and shrinkage to the use of power transformations in CoDA. We then show that *powering*, i.e., the equivalent to scalar multiplication on the simplex, can be understood as a shrinkage problem on the tangent space of the simplex. In information-geometric terms, traditional shrinkage corresponds to an optimization along a mixture (or m-) geodesic, while powering (or, as we call it, *exponential* shrinkage) can be optimized along an exponential (or e-) geodesic. While the m-geodesic corresponds to the posterior mean of the multinomial counts using a conjugate prior, the e-geodesic corresponds to an alternative parametrization of the posterior where prior and data contributions are weighted by geometric rather than arithmetic means. To optimize the exponential shrinkage parameter, we use mean-squared error as a cost function on the tangent space. This is just the expected squared Aitchison distance from the true parameter. We derive an analytic solution for its minimum based on the delta method and test it via simulations. We also discuss exponential shrinkage as an alternative to zero imputation for dimension reduction and data normalization.

Keywords Compositional data · Information geometry · Dual geodesics · Multinomial distribution · Box–Cox transformation · Zero handling · James–Stein shrinkage · Empirical Bayes



Volume 6
Issue 1
Article 69

RESEARCH PAPER

Bregman dynamics, contact transformations and convex optimization

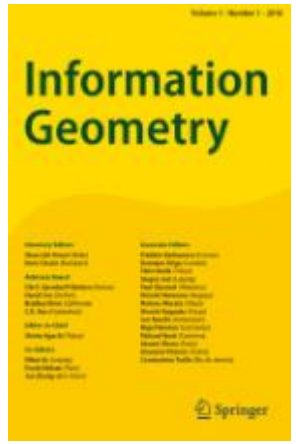
Alessandro Bravetti[1] · Maria L. Daza-Torres[2] · Hugo Flores-Arguedas[3] · Michael Betancourt[4]

Received: 16 July 2022 / Revised: 27 January 2023 / Accepted: 10 April 2023 /
Published online: 30 April 2023

Abstract

Recent research on accelerated gradient methods of use in optimization has demonstrated that these methods can be derived as discretizations of dynamical systems. This, in turn, has provided a basis for more systematic investigations, especially into the geometric structure of those dynamical systems and their structure-preserving discretizations. In this work, we introduce dynamical systems defined through a *contact geometry* which are not only naturally suited to the optimization goal but also subsume all previous methods based on geometric dynamical systems. As a consequence, all the deterministic flows used in optimization share an extremely interesting geometric property: they are invariant under contact transformations. In our main result, we exploit this observation to show that the celebrated Bregman Hamiltonian system can always be transformed into an equivalent but separable Hamiltonian by means of a contact transformation. This in turn enables the development of fast and robust discretizations through geometric *contact splitting integrators*. As an illustration, we propose the Relativistic Bregman algorithm, and show in some paradigmatic examples that it compares favorably with respect to standard optimization algorithms such as classical momentum and Nesterov's accelerated gradient.

Keywords Convex optimization · Bregman Hamiltonian · Contact geometry



Volume 6
Issue 1
Article 70

RESEARCH PAPER

Anomaly detection in the probability simplex under different geometries

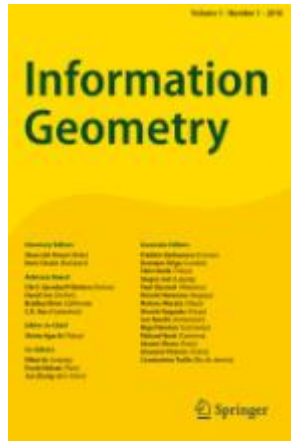
Uriel Legaria[1] · Sergio Mota[1] · Sergio Martinez[1] · Alfredo Cobá[2] · Argenis Chable[2] · Antonio Neme[3]

Received: 4 August 2022 / Revised: 24 April 2023 / Accepted: 13 May 2023 /
Published online: 30 May 2023

Abstract

An open problem in data science is that of anomaly detection. Anomalies are instances that do not maintain a certain property that is present in the remaining observations in a dataset. Several anomaly detection algorithms exist, since the process itself is ill-posed mainly because the criteria that separates common or expected vectors from anomalies are not unique. In the most extreme case, data is not labelled and the algorithm has to identify the vectors that are anomalous, or assign a degree of anomaly to each vector. The majority of anomaly detection algorithms do not make any assumptions about the properties of the feature space in which observations are embedded, which may affect the results when those spaces present certain properties. For instance, compositional data such as normalized histograms, that can be embedded in a probability simplex, constitute a particularly relevant case. In this contribution, we address the problem of detecting anomalies in the probability simplex, relying on concepts from Information Geometry, mainly by focusing our efforts in the distance functions commonly applied in that context. We report the results of a series of experiments and conclude that when a specific distance-based anomaly detection algorithm relies on Information Geometry-related distance functions instead of the Euclidean distance, the performance is significantly improved.

Keywords Anomaly detection · Probability simplex · Information geometry



Volume 6
Issue 2
Article 71

RESEARCH PAPER

Differential geometry for the optimal design of the contingent valuation method

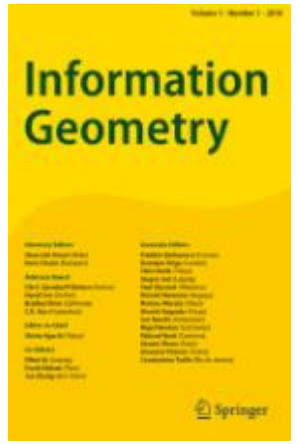
Hisatoshi Tanaka[1]

Received: 9 August 2022 / Revised: 30 April 2023 / Accepted: 13 May 2023 /
Published online: 31 May 2023

Abstract

The contingent valuation method (CVM) is a widely used experimental method to measure the monetary value of goods. However, CVM estimates are sensitive to experiment design. In this study, we formulated the optimal design problem as a minimization problem of the Fisher information metric of a gradient vector field generated by using the statistical model of the CVM. Furthermore, a necessary and sufficient condition of the optimal design was proven.

Keywords Optimal design · Contingent valuation method · Binary response model · Efficiency bound



Volume 6
Issue 2
Article 72

RESEARCH PAPER

Object embedding using an information geometrical perspective

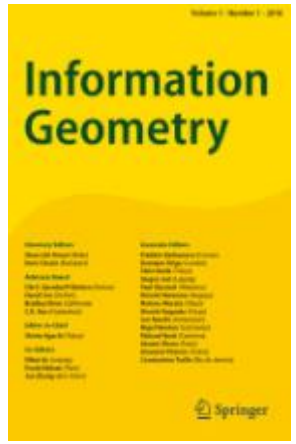
Taiki Sugiura[1] · Noboru Murata[1]

Received: 2 May 2022 / Revised: 9 June 2023 / Accepted: 20 June 2023 /
Published online: 27 July 2023

Abstract

Acquiring vector representations of objects is essential for applying machine learning, statistical inference, and visualization. Although various vector acquisition methods have been proposed considering the relationship between objects in target data, most of them are supposed to use only a specific relevance level. In real-world data, however, there are cases where multiple relationships are contained between objects, such as time-varying similarity in time-series data or various weighted edges on graph-structured data. In this paper, a vector acquisition method which assigns vectors in a single coordinate system to objects preserving the information given by multiple relations between objects is proposed. In the proposed method, a logarithmic bilinear model parameterized by representation vectors is utilized for approximating relations between objects based on a stochastic embedding idea. The inference algorithm proposed in this study is interpreted in terms of information geometry: the m-projection from the probability distribution constructed from observed relations on the model manifold and the e-mixture in the model manifold are alternately repeated to estimate the parameters. Finally, the performance of the proposed method is evaluated using artificial data, and a case study is conducted using real data.

Keywords Constellation · Relevance · Mapping · Stochastic embedding · Mixture distribution



Volume 6
Issue 2
Article 73

RESEARCH PAPER

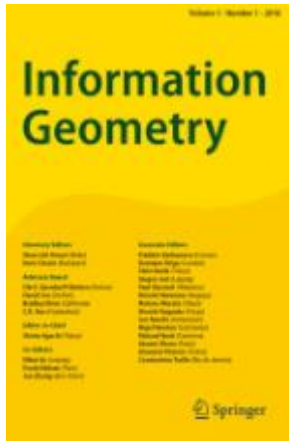
Conelike radiant structures

Daniel J. F. Fox[1]

Received: 11 December 2022 / Revised: 9 June 2023 / Accepted: 29 June 2023 /
Published online: 4 October 2023

Abstract
Analogues of the classical affine-projective correspondence are developed in the context of statistical manifolds compatible with a radiant vector field. These utilize a formulation of Einstein equations for special statistical structures that generalizes the usual Einstein equations for pseudo-Riemannian metrics and is of independent interest. A conelike radiant structure is a not necessarily flat affine connection equipped with a family of surfaces that behave like the intersections of the planes through the origin with a convex cone in a real vector space. A radiant structure is a torsion-free affine connection and a vector field whose covariant derivative is the identity endomorphism. A radiant structure is conelike if for every point and every two-dimensional subspace containing the radiant vector field there is a totally geodesic surface passing through the point and tangent to the subspace. Such structures exist on the total space of any principal bundle with one-dimensional fiber and on any Lie group with a quadratic structure on its Lie algebra. The affine connection of a conelike radiant structure can be normalized in a canonical way to have antisymmetric Ricci tensor. Applied to a conelike radiant structure on the total space of a principal bundle with one-dimensional fiber this yields a generalization of the classical Thomas connection of a projective structure. The compatibility of radiant and conelike structures with metrics is investigated and yields a construction of connections for which the symmetrized Ricci curvature is a constant multiple of a compatible metric that generalizes well-known constructions of Riemannian and Lorentzian Einstein–Weyl structures over Kähler–Einstein manifolds having nonzero scalar curvature. A formulation of Einstein equations for special statistical manifolds is given that generalizes the Einstein–Weyl equations and encompasses these more general examples. There are constructed left-invariant conelike radiant structures on a Lie group endowed with a left-invariant nondegenerate bilinear form, and the case of three-dimensional unimodular Lie groups is described in detail.

Keywords Statistical structures · Einstein equations · Radiant manifolds · Projective structures



Volume 6
Issue 2
Article 74

RESEARCH PAPER

On the geometric mechanics of assignment flows for metric data labeling

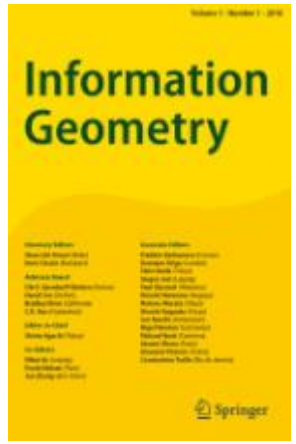
Fabrizio Savarino[1] · Peter Albers[2] · Christoph Schnörr[2]

Received: 17 December 2021 / Revised: 20 September 2023 / Accepted: 4 October 2023 /
Published online: 3 November 2023

Abstract

Metric data labeling refers to the task of assigning one of multiple predefined labels to every given datapoint based on the metric distance between label and data. This assignment of labels typically takes place in a spatial or spatio-temporal context. Assignment flows are a class of dynamical models for metric data labeling that evolve on a basic statistical manifold, the so called assignment manifold, governed by a system of coupled replicator equations. In this paper we generalize the result of a recent paper for uncoupled replicator equations and adopting the viewpoint of geometric mechanics, relate assignment flows to critical points of an action functional via the associated Euler–Lagrange equation. We also show that not every assignment flow is a critical point and characterize precisely the class of coupled replicator equations fulfilling this relation, a condition that has been missing in recent related work. Finally, some consequences of this connection to Lagrangian mechanics are investigated including the fact that assignment flows are, up to initial conditions of measure zero, reparametrized geodesics of the so-called Jacobi metric.

Keywords Assignment flows · Replicator equation · Information geometry · Geometric mechanics · Metric data labeling



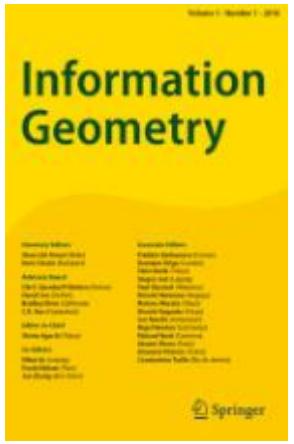
Volume 7
Issue 1
Article 75

RESEARCH PAPER

The generalized Pythagorean theorem on the compactifications of certain dually flat spaces via toric geometry

Hajime Fujita[1]

Received: 31 May 2023 / Revised: 18 October 2023 / Accepted: 30 October 2023 /
Published online: 24 November 2023



Volume 7
Issue 1
Article 76

Abstract

In this paper we study dually flat spaces arising from Delzant polytopes equipped with a symplectic potential together with their corresponding toric Kähler manifolds as their torifications. We introduce a dually flat structure and the associated Bregman divergence on the boundary from the viewpoint of toric Kähler geometry. We show a continuity and a generalized Pythagorean theorem for the divergence on the boundary. We also provide a characterization for a toric Kähler manifold to become a torification of a mixture family on a finite set.

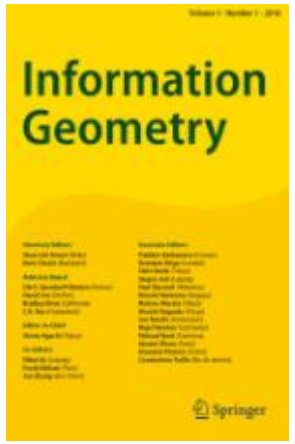
Keywords Dually flat space · Delzant polytope · Toric Kähler manifold · Bregman divergence · Pythagorean theorem

RESEARCH PAPER

Diffusion hypercontractivity via generalized density manifold

Wuchen Li[1]

Received: 4 February 2020 / Revised: 12 October 2023 / Accepted: 29 October 2023 /
Published online: 27 November 2023



Volume 7
Issue 1
Article 77

Abstract

We prove a one-parameter family of diffusion hypercontractivity from a class of drift-diffusion processes. We next derive the related log–Sobolev, Poincare, and Talagrand inequalities. The derivation is based on the calculation of Hessian operators along generalized gradient flows in Dolbeault–Nazaret–Savare metric spaces (Dolbeault et al., Calc Var Partial Differ 2:193–231, 2010). In this direction, a mean-field type Bakry–Emery iterative calculus is presented. In particular, an inequality among Pearson divergence (P), negative Sobolev metric (H^{-1}), and generalized Fisher information functional (I), named $PH^{-1}I$ inequality, is presented.

Keywords Information theory · Mean-field Bakry–Emery calculus · Generalized log–Sobolev inequality · Generalized Poincare inequality · Generalized Talagrand inequality · Generalized Yano's formula.

RESEARCH PAPER

Information geometry of dynamics on graphs and hypergraphs

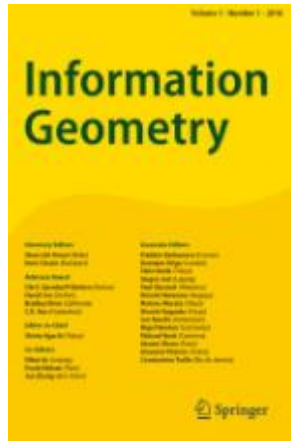
Tetsuya J. Kobayashi[1,2,3] · Dimitri Loutchko[1] · Atsushi Kamimura[1] · Shuhei A. Horiguchi[2] · Yuki Sughiyama[1]

Received: 1 December 2022 / Revised: 22 September 2023 / Accepted: 5 November 2023 /
Published online: 22 December 2023

Abstract

We introduce a new information-geometric structure associated with the dynamics on discrete objects such as graphs and hypergraphs. The presented setup consists of two dually flat structures built on the vertex and edge spaces, respectively. The former is the conventional duality between density and potential, e.g., the probability density and its logarithmic form induced by a convex thermodynamic function. The latter is the duality between flux and force induced by a convex and symmetric dissipation function, which drives the dynamics of the density. These two are connected topologically by the homological algebraic relation induced by the underlying discrete objects. The generalized gradient flow in this doubly dual flat structure is an extension of the gradient flows on Riemannian manifolds, which include Markov jump processes and nonlinear chemical reaction dynamics as well as the natural gradient. The information-geometric projections on this doubly dual flat structure lead to information-geometric extensions of the Helmholtz–Hodge decomposition and the Otto structure in L^2-Wasserstein geometry. The structure can be extended to non-gradient nonequilibrium flows, from which we also obtain the induced dually flat structure on cycle spaces. This abstract but general framework can broaden the applicability of information geometry to various problems of linear and nonlinear dynamics.

Keywords Dually flat structure · Thermodynamics · Homological Algebra · Discrete calculus · Helmholtz decomposition



Volume 7
Issue 1
Article 78

RESEARCH PAPER

On the complete integrability of gradient systems on manifold of the beta family of the first kind

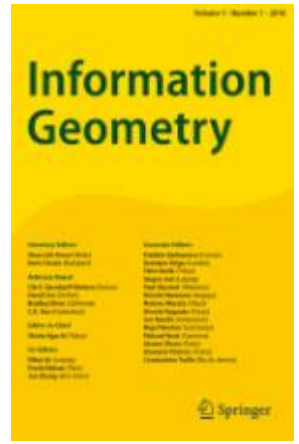
Prosper Rosaire Mama Assandje[1] · Joseph Dongho[1] · Thomas Bouetou Bouetou[2]

Received: 4 August 2023 / Revised: 27 December 2023 / Accepted: 28 December 2023 /
Published online: 29 January 2024

Abstract

In this article, it is shown that there exists a gradient system that is Hamiltonian and completely integrable on the beta manifold of the first kind with two parameters and whose existence depends on a potential function with duality on the manifold and which is the solution of the Legendre equation. It is shown that, by making a statistical borrowing from the gamma function, the gradient system remains Hamiltonian and completely integrable. It is shown that the gradient system constructed on the first kind of two-parameter beta manifold admits a Lax pair representation. This also makes it possible to prove its complete integrability and to determine its Hamiltonian function.

Keywords Hamiltonian system · Gradient system · Lax pair representation · Hamiltonian



Volume 7
Issue 1
Article 79

RESEARCH PAPER

Variational representations of annealing paths: Bregman information under monotonic embedding

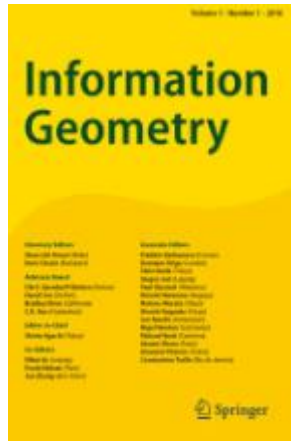
Rob Brekelmans[1] · Frank Nielsen[2]

Received: 22 September 2022 / Revised: 22 December 2023 / Accepted: 27 December 2023 /
Published online: 2 February 2024

Abstract

Markov chain Monte Carlo methods for sampling from complex distributions and estimating normalization constants often simulate samples from a sequence of intermediate distributions along an *annealing path*, which bridges between a tractable initial distribution and a target density of interest. Prior works have constructed annealing paths using quasi-arithmetic means, and interpreted the resulting intermediate densities as minimizing an expected divergence to the endpoints. To analyze these variational representations of annealing paths, we extend known results showing that the arithmetic mean over arguments minimizes the expected Bregman divergence to a single representative point. In particular, we obtain an analogous result for quasi-arithmetic means, when the inputs to the Bregman divergence are transformed under a monotonic embedding function. Our analysis highlights the interplay between quasi-arithmetic means, parametric families, and divergence functionals using the rho-tau representational Bregman divergence framework, and associates common divergence functionals with intermediate densities along an annealing path.

Keywords Bregman divergence · Bregman information · Monotone embedding · Quasi-arithmetic means · Non-parametric information geometry · Gauge freedom · Annealing paths · Markov chain Monte Carlo



Volume 7
Issue 1
Article 80

RESEARCH

Statistical manifold with degenerate metric

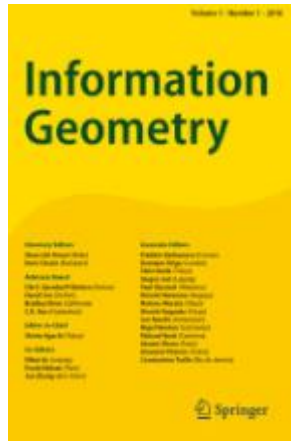
Kaito Kayo[1]

Received: 15 November 2023 / Revised: 15 January 2024 / Accepted: 31 January 2024 /
Published online: 8 March 2024

Abstract

A *statistical manifold* is a pseudo-Riemannian manifold endowed with a *Codazzi structure*. This structure plays an important role in Information Geometry and its related fields, e.g., a statistical model admits this structure with the Fisher–Rao metric. In practical application, however, the metric may be degenerate, and then this geometric structure is not fully adapted. In the present paper, for such cases, we introduce the notion of *quasi-Codazzi structure* which consists of a possibly degenerate metric (i.e., symmetric (0,2)-tensor) and a pair of coherent tangent bundles with affine connections. This is thought of as an affine differential geometry of Lagrange subbundles of para-Hermitian vector bundles and also as a submanifold theory of para-Hermitian space-form. As a special case, the quasi-Codazzi structure with flat connections coincides with the *quasi-Hessian structure* previously studied by Nakajima–Ohmoto. The relation among our quasi-Codazzi structure, quasi-Hessian structure and weak contrast functions generalizes the relation among Codazzi structure, dually flat (i.e., Hessian) structure and contrast functions.

Keywords Codazzi structure · Statistical manifold · Para-complex geometry · Dually flat structure



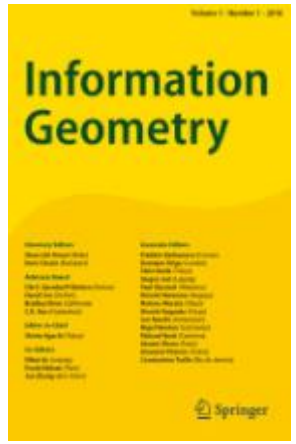
Volume 7
Issue 1
Article 81

RESEARCH PAPER

Codivergences and information matrices

Alexis Derumigny[1] · Johannes Schmidt-Hieber[2]

Received: 22 March 2023 / Revised: 24 April 2024 / Accepted: 25 April 2024 /
Published online: 22 May 2024



Volume 7
Issue 1
Article 82

Abstract

We propose a new concept of codivergence, which quantifies the similarity between two probability measures P_1, P_2 relative to a reference probability measure P_0. In the neighborhood of the reference measure P_0, a codivergence behaves like an inner product between the measures $P_1 - P_0$ and $P_2 - P_0$. Codivergences of covariance-type and correlation-type are introduced and studied with a focus on two specific correlation-type codivergences, the χ^2-codivergence and the Hellinger codivergence. We derive explicit expressions for several common parametric families of probability distributions. For a codivergence, we introduce moreover the divergence matrix as an analogue of the Gram matrix. It is shown that the χ^2-divergence matrix satisfies a data-processing inequality.

Keywords Divergence · Chi-square divergence · Hellinger affinity · Gram matrix

RESEARCH PAPER

Information geometry of Wasserstein statistics on shapes and affine deformations

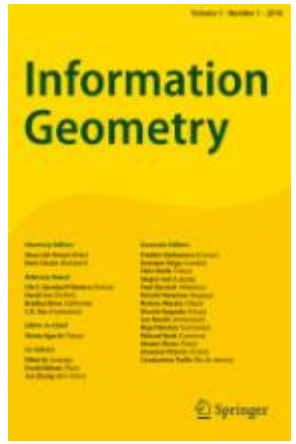
Shun-ichi Amari[1,2] · Takeru Matsuda[2,3]

Received: 31 July 2023 / Revised: 13 February 2024 / Accepted: 16 June 2024 /
Published online: 15 July 2024

Abstract

Information geometry and Wasserstein geometry are two main structures introduced in a manifold of probability distributions, and they capture its different characteristics. We study characteristics of Wasserstein geometry in the framework of [32] for the affine deformation statistical model, which is a multi-dimensional generalization of the location-scale model. We compare merits and demerits of estimators based on information geometry and Wasserstein geometry. The shape of a probability distribution and its affine deformation are separated in the Wasserstein geometry, showing its robustness against the waveform perturbation in exchange for the loss in Fisher efficiency. We show that the Wasserstein estimator is the moment estimator in the case of the elliptically symmetric affine deformation model. It coincides with the information-geometrical estimator (maximum-likelihood estimator) when the waveform is Gaussian. The role of the Wasserstein efficiency is elucidated in terms of robustness against waveform change.

Keywords Elliptically symmetric · Information geometry · Optimal transport · Robustness · Wasserstein distance



Volume 7
Issue 2
Article 83

RESEARCH

On closed-form expressions for the Fisher–Rao distance

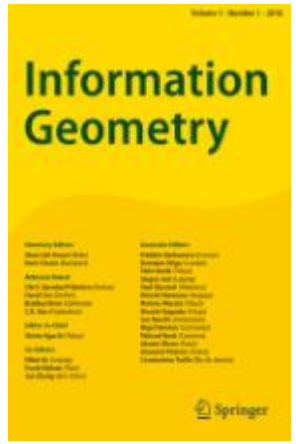
Henrique K. Miyamoto[1] · Fábio C. C. Meneghetti[2] · Julianna Pinele[3] · Sueli I. R. Costa[2]

Received: 29 February 2024 / Revised: 31 July 2024 / Accepted: 1 August 2024 /
Published online: 16 September 2024

Abstract

The Fisher–Rao distance is the geodesic distance between probability distributions in a statistical manifold equipped with the Fisher metric, which is a natural choice of Riemannian metric on such manifolds. It has recently been applied to supervised and unsupervised problems in machine learning, in various contexts. Finding closed-form expressions for the Fisher–Rao distance is generally a non-trivial task, and those are only available for a few families of probability distributions. In this survey, we collect examples of closed-form expressions for the Fisher–Rao distance of both discrete and continuous distributions, aiming to present them in a unified and accessible language. In doing so, we also: illustrate the relation between negative multinomial distributions and the hyperbolic model, include a few new examples, and write a few more in the standard form of elliptical distributions.

Keywords Fisher metric · Fisher–Rao distance · Geodesic distance · Parametric distributions · Statistical manifolds



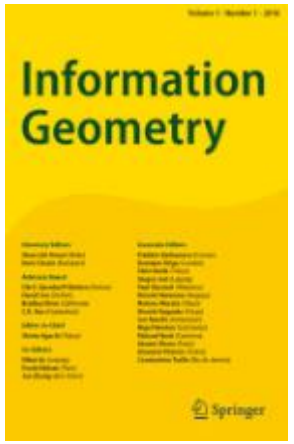
Volume 7
Issue 2
Article 84

PERSPECTIVE

Maximal co-ancillarity and maximal co-sufficiency

H. S. Battey[1]

Received: 10 June 2024 / Revised: 10 August 2024 / Accepted: 12 August 2024 /
Published online: 17 September 2024



Volume 7
Issue 2
Article 85

Abstract

The purpose of this exposition is to provide some new perspectives on conditional inference through a notional idealised separation within the minimal sufficient statistic, allowing a geometric account of key ideas from the Fisherian position. The notional idealised separation, in terms of an ancillary statistic and what I call a maximal co-ancillary statistic, provides conceptual insight and clarifies what is sought from an approximate conditional analysis, where exact calculations may not be available. A parallel framework applies in the Fisherian assessment of model adequacy. Both aspects are discussed and illustrated geometrically through examples.

Keywords Ancillary · Conditional inference · Inferential separations · Information · Minimal sufficiency · Model adequacy

RESEARCH PAPER

Reverse em-problem based on Bregman divergence and its application to classical and quantum information theory

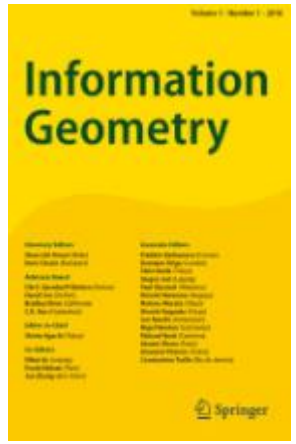
Masahito Hayashi[1,2,3]

Received: 12 April 2023 / Revised: 15 September 2024 / Accepted: 17 September 2024 /
Published online: 3 October 2024

Abstract

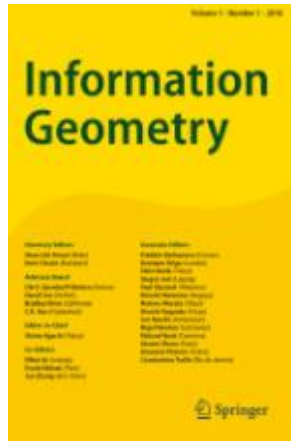
The recent paper (IEEE Trans. IT 69, 1680) proposed an analytical calculation method for the channel capacity without iteration. However, it has restrictions for its application. Also, it does not explain why the channel capacity can be solved analytically in this special case. To expand the applicability of this method, we recall the reverse em-problem, which was proposed by Toyota (Information Geometry, 3, 1355 (2020)) as the repetition of the inverse map of the em iteration to calculate the channel capacity, which is given as the maximum of the mutual information. However, it left several open problems. We formulate the reverse em-problem based on Bregman divergence, and solve these open problems. Using these results, we convert the reverse em-problem into em-problems, and derive a non-iterative formula for the reverse em-problem, which can be considered as a generalization of the above analytical calculation method. Additionally, this derivation explains the information geometrical structure of this special case.

Keywords Maximization · Bregman divergence · Information geometry · Channel capacity



Volume 7
Issue 2
Article 86

BRIEF REPORT



Volume 7
Issue 2
Article 87

On the statistical Lie groups of normal distributions

Jun-ichi Inoguchi[1]

Received: 31 May 2024 / Revised: 11 September 2024 / Accepted: 22 September 2024 /
Published online: 5 October 2024

Abstract

The α-connections (other than the Levi-Civita connection of the Fisher metric) on the statistical Lie group of the normal distributions can not be the Levi-Civita connection of any left invariant semi-Riemannian metrics.

Keywords Normal distribution · Statistical Lie groups · Lorentz metrics

RESEARCH PAPER

Simple variational inference based on minimizing Kullback–Leibler divergence

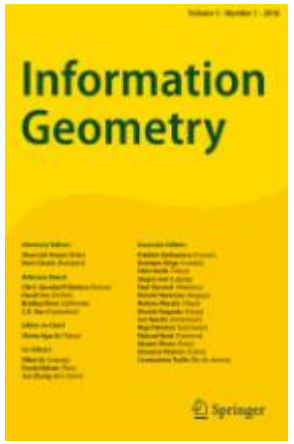
Ryo Nakamura[1] · Tomooki Yuasa[2] · Takafumi Amaba[3] · Jun Fujiki[3]

Received: 14 July 2023 / Revised: 13 May 2024 / Accepted: 22 September 2024 /
Published online: 13 October 2024

Abstract

We introduce a new methodology of estimation of the true distribution. The procedure of getting estimated distribution is constructed from Bayesian statistical models in which the statistical model is fixed and prior distributions for the parameters are varied. Then we consider the Kullback–Leibler divergence of the true distribution from the estimated one and derive variational formulae of the Kullback–Leibler divergence over prior distributions. Next, we propose a Newton–Raphson method for simulating the prior distribution, which is the critical point, based on the Riemannian geometry of the probability simplex. The method can run once a sample from the true distribution is obtained without any other knowledge of the true distribution. For the geometry, we employ the Riemannian metric induced from the characteristic function, which appears in Vinberg's theory of homogeneous convex cones. As a by-product of the geometry, we derive an interpretation that the Kullback–Leibler divergence is the logarithm of a gauge transformation. From this, we obtain a viewpoint that envisaging the true distribution is nothing but gauge fixing, and this depiction as gauge theory seems to match the context of statistics. Also, we show some numerical results on how well our methodology works.

Keywords Prior distribution · Kullback–Leibler divergence · Cross entropy · Variational formula · Vinberg's Riemannian metric · Gauge



Volume 7
Issue 2
Article 88

RESEARCH

The Bayesian central limit theorem for exponential family distributions: a geometric approach

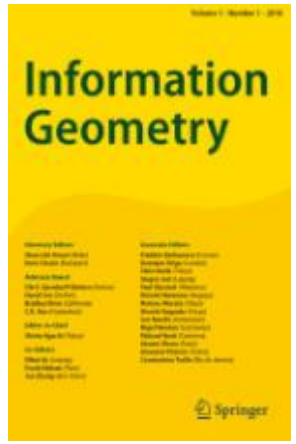
Geoff Goehle[1]

Received: 23 April 2024 / Revised: 23 September 2024 / Accepted: 30 September 2024 /
Published online: 15 October 2024

Abstract

The Bernstein-von Mises theorem, also known as the Bayesian Central Limit Theorem (BCLT), states that under certain assumptions a posterior distribution can be approximated as a multivariate normal distribution as long as the precision parameter is large. We derive a special case of the BCLT for the canonical conjugate prior of a regular exponential family distribution using the machinery of information geometry. Our approach applies the core approximation for the BCLT, Laplace's method, to the free entropy (i.e., log-normalizer) of an exponential family distribution. Additionally, we formulate approximations for the Kullback–Leibler divergence and Fisher-Rao metric on the conjugate prior manifold in terms of corresponding quantities from the likelihood manifold. We also include an application to the categorical distribution and show that the free entropy derived approximations are related to various series expansions of the gamma function and its derivatives. Furthermore, for the categorical distribution, the free entropy approximation produces higher order expansions than the BCLT alone.

Keywords Information geometry · Bayesian inference · Probability · Differential geometry



Volume 7
Issue 2
Article 89

RESEARCH PAPER

An algorithm for learning representations of models with scarce data

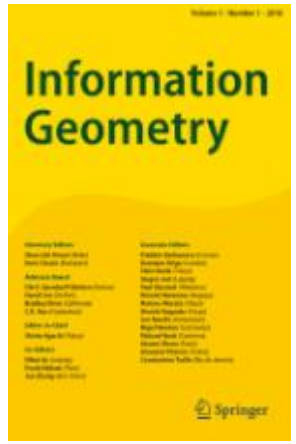
Adrian de Wynter[1]

Received: 29 March 2021 / Revised: 30 March 2024 / Accepted: 11 October 2024 /
Published online: 30 October 2024

Abstract

We present an algorithm for solving binary classification problems when the dataset is not fully representative of the problem being solved, and obtaining more data is not possible. It relies on a trained model with loose accuracy constraints, an iterative hyperparameter searching-and-pruning procedure over a search space Θ, and a data-generating function. Our algorithm works by reconstructing up to homology the manifold on which lies the support of the underlying distribution. We provide an analysis on correctness and runtime complexity under ideal conditions and an extension to deep neural networks. In the former case, if $|\Theta|$ is the number of hyperparameter sets in the search space, this algorithm returns a solution that is up to $2(1 - 2^{-|\Theta|})$ times better than simply training with an enumeration of Θ and picking the best model. As part of our analysis we also prove that an open cover of a dataset has the same homology as the manifold on which lies the support of the underlying probability distribution, if and only said dataset is learnable. This latter result acts as a formal argument to explain the effectiveness of contemporary data expansion techniques.

Keywords Data augmentation · Semi-supervised learning



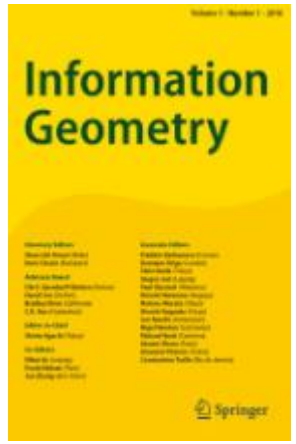
Volume 7
Issue 2
Article 90

RESEARCH

An embedding structure of determinantal point process

Hideitsu Hino[1] · Keisuke Yano[1]

Received: 2 July 2024 / Revised: 8 October 2024 / Accepted: 18 October 2024 /
Published online: 7 November 2024



Volume 7
Issue 2
Article 91

Abstract

This paper investigates the information geometrical structure of a determinantal point process (DPP). It demonstrates that a DPP is embedded in the exponential family of log-linear models. The extent of deviation from an exponential family is analyzed using the e-embedding curvature tensor, which identifies partially flat parameters of a DPP. On the basis of this embedding structure, an information-geometrical relationship between a marginal kernel and an L-ensemble kernel is discovered.

Keywords Curved exponential family · Discrete statistical model · L-ensemble kernel · Partially ordered set · Statistical curvature

RESEARCH

On statistics which are almost sufficient from the viewpoint of the Fisher metrics

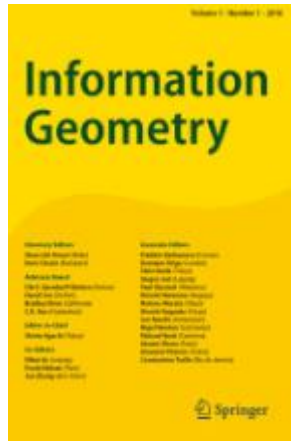
Kaori Yamaguchi[1] · Hiraku Nozawa[1]

Received: 24 August 2024 / Revised: 1 November 2024 / Accepted: 10 November 2024 /
Published online: 22 November 2024

Abstract

Given a statistical model, a statistic on the model is sufficient if the Fisher metric of the induced model coincides with the original Fisher metric, according to the definition by Ay-Jost-Lê-Schwachhöfer. We introduce and study its quantitative version: for $0 < \delta \leq 1$, we call a statistic δ-almost sufficient if $\delta^2 \mathfrak{g}(v, v) \leq \mathfrak{g}'(v, v)$ for every tangent vector v of the parameter space, where $\mathfrak{g}$ and $\mathfrak{g}'$ are the Fisher metric of the original and the induced model, respectively. By the monotonicity theorem due to Amari-Nagaoka and Ay-Jost-Lê-Schwachhöfer, the Fisher metric $\mathfrak{g}'$ of the induced model for such a statistic is bi-Lipschitz equivalent to the original one $\mathfrak{g}$, which means that the information loss of the statistic is uniformly bounded. We characterize such statistics in terms of the conditional probability or by the existence of a certain decomposition of the density function in a way similar to the characterizations of sufficient statistics due to Ay-Jost-Lê-Schwachhöfer and Fisher-Neyman.

Keywords Information geometry · Sufficient statistics · Fisher metrics · Statistical manifolds · Binomial distribution



Volume 7
Issue 2
Article 92

RESEARCH PAPER

On partial likelihood and the construction of factorisable transformations

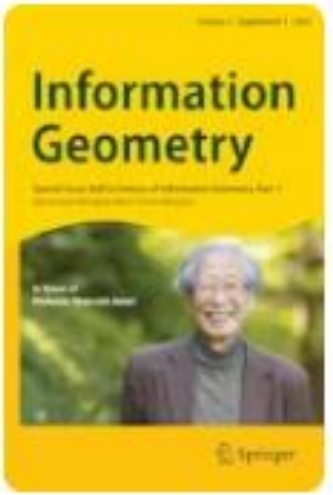
H. S. Battey[1] · D. R. Cox[2] · Su Hyeong Lee[3]

Received: 15 April 2022 / Revised: 11 June 2022 / Accepted: 15 June 2022 /
Published online: 28 June 2022

Abstract

Models whose associated likelihood functions fruitfully factorise are an important minority allowing elimination of nuisance parameters via partial likelihood, an operation that is valuable in both Bayesian and frequentist inferences, particularly when the number of nuisance parameters is not small. After some general discussion of partial likelihood, we focus on marginal likelihood factorisations, which are particularly difficult to ascertain from elementary calculations. We suggest a systematic approach for deducing transformations of the data, if they exist, whose marginal likelihood functions are free of the nuisance parameters. This is based on the solution to an integro-differential equation constructed from aspects of the Laplace transform of the probability density function, for which candidate solutions solve a simpler first-order linear homogeneous differential equation. The approach is generalised to the situation in which such factorisable structure is not exactly present. Examples are used in illustration. Although motivated by inferential problems in statistics, the proposed construction is of independent interest and may find application elsewhere.

Keywords Inferential separation · Marginal likelihood · Matched comparisons · Method of characteristics · Partial differential equations · Nuisance parameters



Volume 7
Special Issue 1
Article 1
Number 93

PERSPECTIVE

Independent component analysis in the light of Information Geometry

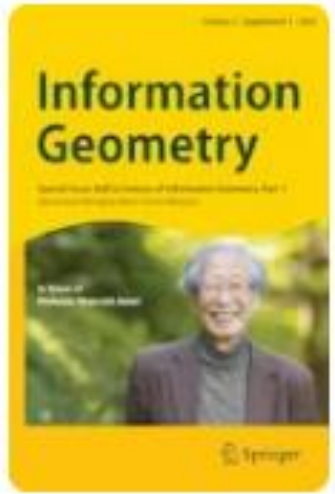
Jean-François Cardoso[1]

Received: 21 September 2022 / Revised: 10 October 2022 / Accepted: 12 October 2022 /
Published online: 24 October 2022

Abstract

I recall my first encounter with Professor Shun-ichi Amari who, once upon a time in Las Vegas, gave me a precious hint about connecting independent component analysis (ICA) to Information Geometry. The paper sketches, rather informally, some of the insights gained in following this lead.

Keywords Independent component analysis · Information Geometry · Pythagoras theorem · Non Gaussianity



Volume 7
Special Issue 1
Article 2
Number 94

SURVEY PAPER

Geometry of EM and related iterative algorithms

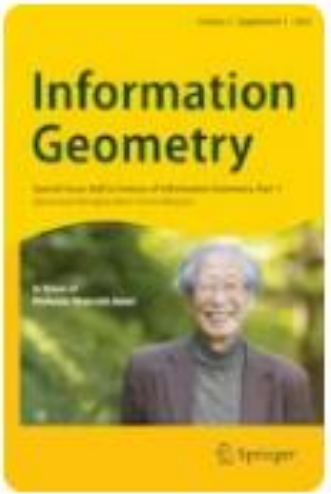
Hideitsu Hino[1,2] · Shotaro Akaho[3] · Noboru Murata[4]

Received: 29 August 2022 / Revised: 12 November 2022 / Accepted: 13 November 2022 /
Published online: 19 November 2022

Abstract

The Expectation–Maximization (EM) algorithm is a simple meta-algorithm that has been used for many years as a methodology for statistical inference when there are missing measurements in the observed data or when the data is composed of observables and unobservables. Its general properties are well studied, and also, there are countless ways to apply it to individual problems. In this paper, we introduce the *em* algorithm, an information geometric formulation of the EM algorithm, and its extensions and applications to various problems. Specifically, we will see that it is possible to formulate an outlier–robust inference algorithm, an algorithm for calculating channel capacity, parameter estimation methods on probability simplex, particular multivariate analysis methods such as principal component analysis in a space of probability models and modal regression, matrix factorization, and learning generative models, which have recently attracted attention in deep learning, from the geometric perspective provided by Amari.

Keywords Information geometry · EM algorithm · *em* algorithm · Bregman divergence · Information theory · Robust statistics · Generative models



Volume 7
Special Issue 1
Article 3
Number 95

SURVEY PAPER

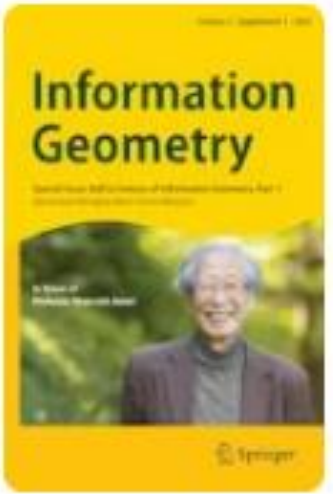
Hommage to Chentsov's theorem

Akio Fujiwara[1]

Received: 26 September 2022 / Revised: 6 November 2022 / Accepted: 12 November 2022 /
Published online: 21 November 2022

Abstract
Chentsov's theorem, which characterises Markov invariant Riemannian metric and affine connections of manifolds of probability distributions on finite sample spaces, is undoubtedly a cornerstone of information geometry. This article aims at providing a comprehensible survey of Chentsov's theorem as well as its modest extensions to generic tensor fields and to parametric models comprising continuous probability densities on $\mathbb{R}^k$.

Keywords Chentsov's theorem · Markov invariance · Fisher metric · α-connections · Amari–Chentsov tensor · Information geometry



Volume 7
Special Issue 1
Article 4
Number 96

PERSPECTIVE

Toward differential geometry of statistical submanifolds

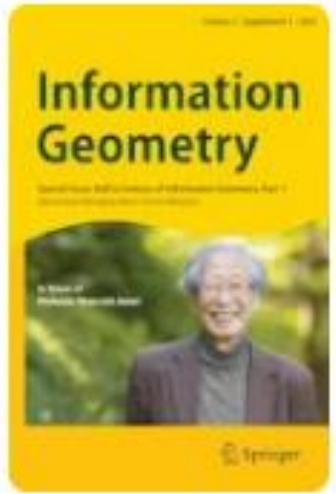
Hitoshi Furuhata[1]

Received: 31 August 2022 / Revised: 30 October 2022 / Accepted: 3 November 2022 /
Published online: 22 November 2022

Abstract

A brief introduction of doubly minimal submanifolds of statistical manifolds is given. A complex submanifold of a holomorphic statistical manifold is doubly minimal. Similar properties are obtained in the case where the ambient space is a Sasakian statistical manifold.

Keywords Statistical submanifolds · Doubly minimal · Doubly totally umbilical



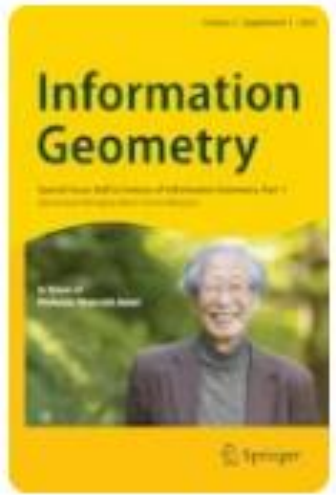
Volume 7
Special Issue 1
Article 5
Number 97

SURVEY PAPER

Affine statistical bundle modeled on a Gaussian Orlicz–Sobolev space

Giovanni Pistone[1]

Received: 14 October 2022 / Revised: 10 November 2022 / Accepted: 12 November 2022 /
Published online: 24 November 2022



Volume 7
Special Issue 1
Article 6
Number 98

Abstract

The dually flat structure of statistical manifolds can be derived in a non-parametric way from a particular case of affine space defined on a qualified set of probability measures. The statistically natural displacement mapping of the affine space depends on the notion of Fisher's score. The model space must be carefully defined if the state space is not finite. Among various options, we discuss how to use Orlicz–Sobolev spaces with Gaussian weight. Such a fully non-parametric set-up provides tools to discuss intrinsically infinite-dimensional evolution problems

Keywords Information geometry · Gaussian Orlicz–Sobolev space · Statistical bundle · Exponential manifold · Dually flat affine manifold

SURVEY PAPER

Conditional inference of Poisson models and information geometry: an ancillary review

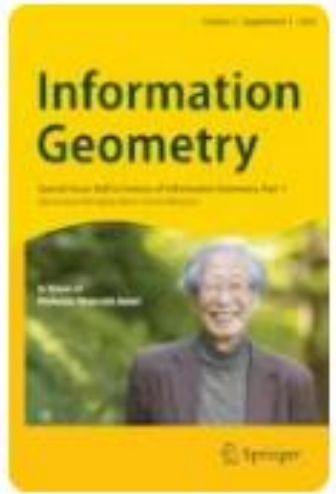
Tomonari Sei[1]

Received: 31 August 2022 / Revised: 5 November 2022 / Accepted: 15 November 2022 /
Published online: 27 November 2022

Abstract

The Poisson distribution is a fundamental tool in categorical data analysis. This paper reviews conditional inference for the independent Poisson model. It is noted that the conditioning variable is not an ancillary statistic in the exact sense except in the case of the product multinomial sampling scheme, whereas two versions of the ancillary property hold in general. The ancillary properties justify the use of conditional inference, as first proposed by R. A. Fisher and subsequently discussed by many researchers. The mixed coordinate system developed in information geometry is emphasized as effective for the description of facts.

Keywords A-hypergeometric distribution · Ancillary statistic · Asymptotic theory · Fisher information · Mixed coordinate system



Volume 7
Special Issue 1
Article 7
Number 99

RESEARCH PAPER

Generalized estimators, slope, efficiency, and fisher information bounds

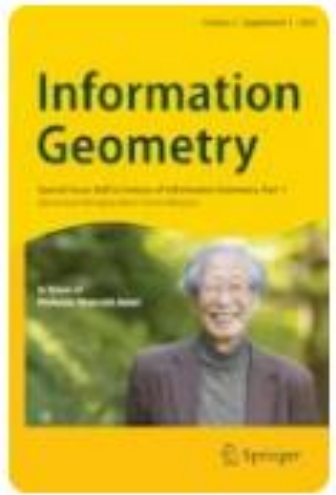
Paul Vos[1]

Received: 10 August 2022 / Revised: 22 November 2022 / Accepted: 22 November 2022 /
Published online: 4 December 2022

Abstract

Point estimators may not exist, need not be unique, and their distributions are not parameter invariant. Generalized estimators provide distributions that are parameter invariant, unique, and exist when point estimates do not. Comparing point estimators using variance is less useful when estimators are biased. A squared slope Λ is defined that can be used to compare both point and generalized estimators and is unaffected by bias. Fisher information I and variance are fundamentally different quantities: the latter is defined at a distribution that need not belong to a family, while the former cannot be defined without a family of distributions, M. Fisher information and Λ are similar quantities as both are defined on the tangent bundle TM and I provides an upper bound, $\Lambda \leq I$, that holds for all sample sizes—asymptotics are not required. Comparing estimators using Λ rather than variance supports Fisher's claim that I provides a bound even in small samples. Λ-efficiency is defined that extends the efficiency of unbiased estimators based on variance. While defined by the slope, Λ-efficiency is simply ρ^2, the square of the correlation between estimator and score function.

Keywords Nonasymptotics · Generalized estimators · Slope · Fisher information · Efficiency



Volume 7
Special Issue 1
Article 8
Number 100

RESEARCH PAPER

A two-parameter family of non-parametric, deformed exponential manifolds

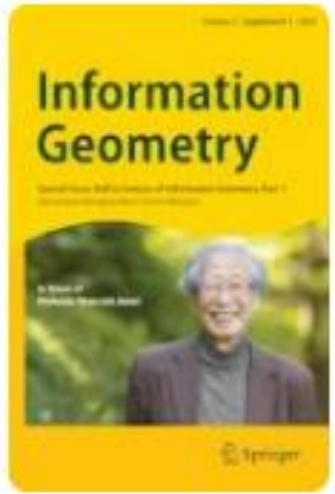
Nigel J. Newton[1]

Received: 1 September 2022 / Revised: 1 September 2022 / Accepted: 13 November 2022 /
Published online: 6 December 2022

Abstract

We construct a new family of non-parametric statistical manifolds by means of a two-parameter class of deformed exponential functions, that includes functions with power-law, linear and sublinear rates of growth. The manifolds are modelled on weighted, mixed-norm Sobolev spaces that are especially suited to this purpose, in the sense that an important class of nonlinear superposition operators (those used in the construction of divergences and tensors) act continuously on them. We analyse variants of these operators, that map into "subordinate" Sobolev spaces, and evaluate the associated gain in regularity. With appropriate choice of parameter values, the manifolds support a large variety of the statistical divergences and entropies appearing in the literature, as well as their associated tensors, eg. the Fisher-Rao metric. Manifolds of finite measures and probability measures are constructed; the latter are shown to be smoothly embedded submanifolds of the former.

Keywords Banach manifold · Fisher-Rao metric · Information Theory · Log-Sobolev inequality · Non-parametric statistics · Sobolev spaces



Volume 7
Special Issue 1
Article 9
Number 101

SURVEY PAPER

Recent advances in algebraic geometry and Bayesian statistics

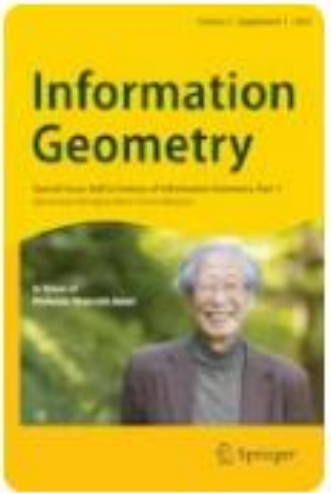
Sumio Watanabe[1]

Received: 30 August 2022 / Revised: 2 November 2022 / Accepted: 18 November 2022 /
Published online: 6 December 2022

Abstract

This article is a review of theoretical advances in the research field of algebraic geometry and Bayesian statistics in the last two decades. Many statistical models and learning machines which contain hierarchical structures or latent variables are called nonidentifiable, because the map from a parameter to a statistical model is not one-to-one. In nonidentifiable models, both the likelihood function and the posterior distribution have singularities in general, hence it was difficult to analyze their statistical properties. However, from the end of the 20th century, new theory and methodology based on algebraic geometry have been established which enable us to investigate such models and machines in the real world. In this article, the following results in recent advances are reported. First, we explain the framework of Bayesian statistics and introduce a new perspective from the birational geometry. Second, two mathematical solutions are derived based on algebraic geometry. An appropriate parameter space can be found by a resolution map, which makes the posterior distribution be normal crossing and the log likelihood ratio function be well-defined. Third, three applications to statistics are introduced. The posterior distribution is represented by the renormalized form, the asymptotic free energy is derived, and the universal formula among the generalization loss, the cross validation, and the information criterion is established. Two mathematical solutions and three applications to statistics based on algebraic geometry reported in this article are now being used in many practical fields in data science and artificial intelligence.

Keywords Birational geometry · Resolution of singularities · Bayesian statistics · Real log canonical threshold



Volume 7
Special Issue 1
Article 10
Number 102

SURVEY PAPER

Geometry and applied statistics

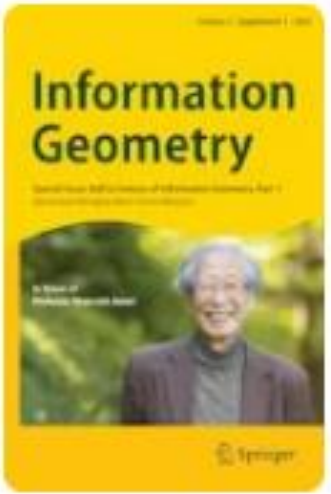
Paul Marriott[1]

Received: 5 September 2022 / Revised: 25 November 2022 / Accepted: 26 November 2022 /
Published online: 6 December 2022

Abstract

We take a very high level overview of the relationship between Geometry and Applied Statistics 50 years from the birth of Information Geometry. From that date we look both backwards and forwards. We show that Geometry has always been part of the statistician's toolbox and how it played a vital role in the evolution of Statistics in the last 50 years.

Keywords Applied statistics · Information geometry



Volume 7
Special Issue 1
Article 11
Number 103

RESEARCH PAPER

Minimum information divergence of Q-functions for dynamic treatment resumes

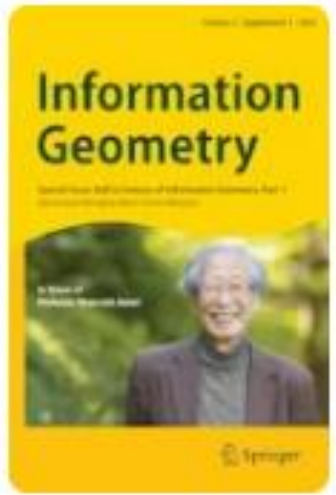
Shinto Eguchi[1]

Received: 3 September 2022 / Revised: 17 November 2022 / Accepted: 18 November 2022 /
Published online: 9 December 2022

Abstract

This paper aims at presenting a new application of information geometry to reinforcement learning focusing on dynamic treatment resumes. In a standard framework of reinforcement learning, a Q-function is defined as the conditional expectation of a reward given a state and an action for a single-stage situation. We introduce an equivalence relation, called the policy equivalence, in the space of all the Q-functions. A class of information divergence is defined in the Q-function space for every stage. The main objective is to propose an estimator of the optimal policy function by a method of minimum information divergence based on a dataset of trajectories. In particular, we discuss the γ-power divergence that is shown to have an advantageous property such that the γ-power divergence between policy-equivalent Q-functions vanishes. This property essentially works to seek the optimal policy, which is discussed in a framework of a semiparametric model for the Q-function. The specific choices of power index γ give interesting relationships of the value function, and the geometric and harmonic means of the Q-function. A numerical experiment demonstrates the performance of the minimum γ-power divergence method in the context of dynamic treatment regimes.

Keywords Information geometry · Geometric mean · Minimum divergence · Optimal policy · Semiparametric model · Value function



Volume 7
Special Issue 1
Article 12
Number 104

RESEARCH PAPER

Hierarchy of deformations in concavity

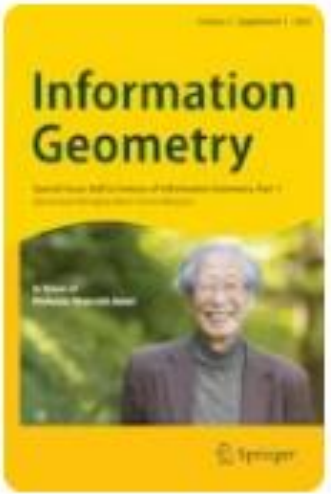
Kazuhiro Ishige[1] · **Paolo Salani**[2] · **Asuka Takatsu**[3,4]

Received: 1 September 2022 / Revised: 25 November 2022 / Accepted: 28 November 2022 /
Published online: 12 December 2022

Abstract

A *deformation* is a positive continuous function defined on an appropriate interval. Through deformations, we generalize the notion of concavity for functions. We introduce the *order function* of a deformation, which permits to determine precisely the ranking of a deformation by taking account of the corresponding concavities. In the hierarchy, the action of positive constant multiples provides an equivalence relation and, if we focus on C^1-deformations, a one-to-one correspondence between the equivalence classes and the order functions is determined. Deformations having a constant valued order function play a fundamental role, and this is only the case of power functions. We show that the concavity associated to a deformation whose order function is nonincreasing and uniformly bounded from above by $q \in \mathbb{R}$ can approximate the concavity associated to the power function of exponent q. Finally, we review three examples of deformations whose order function is nonincreasing and uniformly bounded from above. One is a power function, and the others are related to a concavity preserved by the Dirichlet heat flow in convex domains of Euclidean space.

Keywords Deformation · Deformed logarithmic function · Generalized concavity · Hierarchy



Volume 7
Special Issue 1
Article 13
Number 105

SURVEY PAPER

Natural differentiable structures on statistical models and the Fisher metric

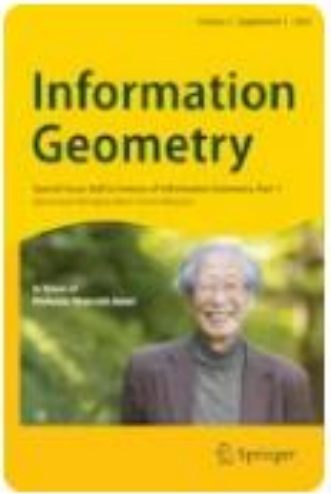
Hông Vân Lê[1]

Received: 24 August 2022 / Revised: 20 November 2022 / Accepted: 5 December 2022 /
Published online: 13 December 2022

Abstract

In this paper I discuss the relation between the concept of the Fisher metric and the concept of differentiability of a family of probability measures. I compare the concepts of smooth statistical manifolds, differentiable families of measures, k-integrable parameterized measure models, diffeological statistical models, differentiable measures, which arise in Information Geometry, mathematical statistics and measure theory, and discuss some related problems.

Keywords Smooth statistical model · Differentiable family of measures · Fisher–Rao metric · Amari–Chentsov tensor · Parameterized measure model · Diffeological statistical model · Cramér–Rao inequality



Volume 7
Special Issue 1
Article 14
Number 106

PERSPECTIVE

Uncertainty and Quantum Variance at the light of Quantum Information Geometry

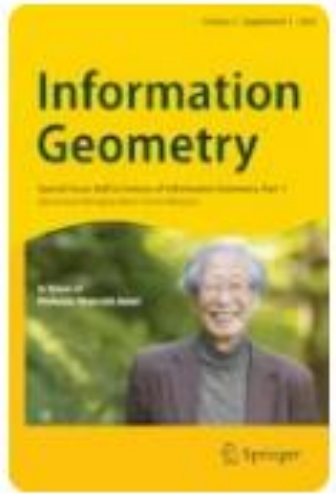
Paolo Gibilisco[1]

Received: 30 October 2022 / Revised: 22 November 2022 / Accepted: 28 November 2022 /
Published online: 14 December 2022

Abstract

Using the Kubo-Ando operator means Denes Petz was able to explain what Quantum Covariance(s) and Quantum Fisher Information(s) are. In this paper the relation between QC and QFI is explained and some of its consequences, notably the Dynamical Uncertainty Principle, are described. Finally some new directions for Classical and Quantum Information Geometry are suggested.

Keywords Quantum Covariance · Quantum Fisher Information · Metric Adjusted Skew Information · Dynamical Uncertainty Principle



Volume 7
Special Issue 1
Article 15
Number 107

RESEARCH PAPER

Conformal mirror descent with logarithmic divergences

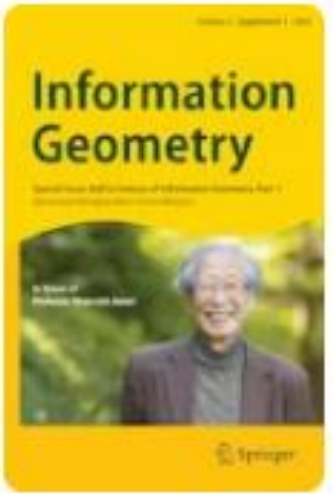
Amanjit Singh Kainth[1,3] · Ting-Kam Leonard Wong[2] · Frank Rudzicz[1,3]

Received: 1 September 2022 / Revised: 21 November 2022 / Accepted: 30 November 2022 /
Published online: 14 December 2022

Abstract

The logarithmic divergence is an extension of the Bregman divergence motivated by optimal transport and a generalized convex duality, and satisfies many remarkable properties. Using the geometry induced by the logarithmic divergence, we introduce a generalization of continuous time mirror descent that we term the conformal mirror descent. We derive its dynamics under a generalized mirror map, and show that it is a time change of a corresponding Hessian gradient flow. We also prove convergence results in continuous time. We apply the conformal mirror descent to online estimation of a generalized exponential family, and construct a family of gradient flows on the unit simplex via the Dirichlet optimal transport problem.

Keywords Mirror descent · Gradient flow · Logarithmic divergence · Conformal Hessian metric · λ-duality · λ-exponential family



Volume 7
Special Issue 1
Article 16
Number 107

RESEARCH PAPER

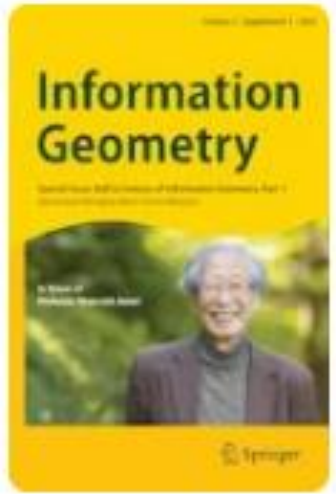
Parametric models and information geometry on W*-algebras

F. M. Ciaglia[1] · F. Di Nocera[2] · J. Jost[2] · L. Schwachhöfer[3]

Received: 5 September 2022 / Revised: 20 December 2022 / Accepted: 21 December 2022 /
Published online: 12 January 2023

Abstract
We introduce the notion of smooth parametric model of normal positive linear functionals on possibly infinite-dimensional $W^\star$-algebras generalizing the notions of parametric models used in classical and quantum information geometry. We then use the Jordan product naturally available in this context in order to define a Riemannian metric tensor on parametric models satsfying suitable regularity conditions. This Riemannian metric tensor reduces to the Fisher–Rao metric tensor, or to the Fubini-Study metric tensor, or to the Bures–Helstrom metric tensor when suitable choices for the $W^\star$-algebra and the models are made.

Keywords Probability distributions · Quantum states · C^*-algebras · Banach manifolds



Volume 7
Special Issue 1
Article 17
Number 108

RESEARCH PAPER

Gram matrices of quantum channels via quantum Fisher information with applications to decoherence and uncertainty

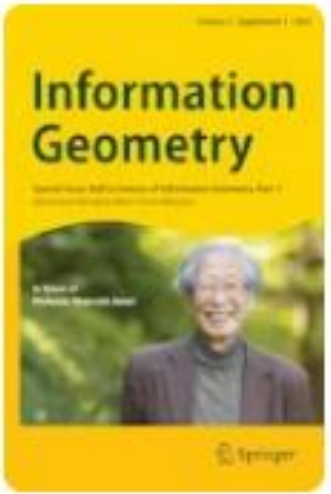
Shunlong Luo[1,2] · **Yuan Sun**[3]

Received: 22 August 2022 / Revised: 21 December 2022 / Accepted: 5 January 2023 /
Published online: 22 January 2023

Abstract

By use of Gram matrices associated with various versions of quantum Fisher information (related to monotone metrics and metric-adjusted skew information), we study information and geometry of quantum channels by introducing several quantities involving commutators and anticommutators for operators and quantum states. We reveal their basic features and evaluate these quantities for some important channels. As applications, we employ Gram matrices to quantify decoherence of quantum channels caused on quantum states, and establish some uncertainty relations, which refine the conventional Heisenberg uncertainty relations involving variance.

Keywords Quantum channels · Quantum Fisher information · Commutator · Anticommutator · Gram matrices



Volume 7
Special Issue 1
Article 18
Number 109

RESEARCH PAPER

The exponential Orlicz space in quantum information geometry

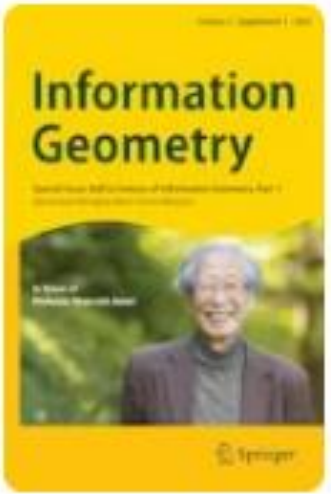
Anna Jenčová[1]

Received: 10 October 2022 / Revised: 2 January 2023 / Accepted: 5 January 2023 /
Published online: 24 January 2023

Abstract

We review the construction of a quantum version of the exponential statistical manifold over the set of all faithful normal positive functionals on a von Neumann algebra. The construction is based on the relative entropy approach to state perturbation. We construct a quantum version of the exponential Orlicz space and discuss the properties of this space and its dual with respect to Kosaki L_p-spaces. We show that the constructed manifold admits a canonical divergence satisfying a Pythagorean relation. We also prove that the manifold structure is invariant under sufficient channels.

Keywords Quantum exponential manifold · Quantum relative entropy · Perturbation of states · Canonical divergence



Volume 7
Special Issue 1
Article 19
Number 110

RESEARCH PAPER

Projections of SDEs onto submanifolds

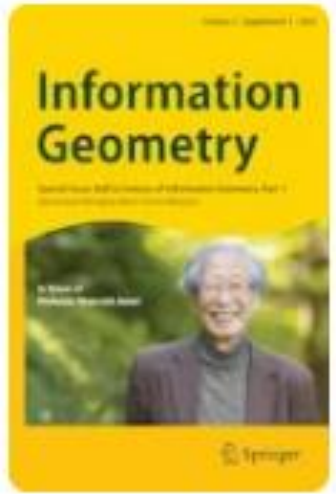
John Armstrong[1] · Damiano Brigo[2] · Emilio Ferrucci[3]

Received: 2 September 2022 / Revised: 13 December 2022 / Accepted: 20 December 2022 /
Published online: 2 February 2023

Abstract

In Armstrong et al. (Proc Lond Math Soc (3) 119(1):176–213, 2019) the authors define three projections of $\mathbb{R}^d$-valued stochastic differential equations (SDEs) onto submanifolds: the Stratonovich, Itô-vector and Itô-jet projections. In this paper, after a brief survey of SDEs on manifolds, we begin by giving these projections a natural, coordinate-free description, each in terms of a specific representation of manifold-valued SDEs. We proceed by deriving formulae for the three projections in ambient $\mathbb{R}^d$-coordinates. We use these to show that the Itô-vector and Itô-jet projections satisfy respectively a weak and mean-square optimality criterion "for small t": this is achieved by solving constrained optimisation problems. These results confirm, but do not rely on the approach taken in Armstrong et al. (Proc Lond Math Soc (3) 119(1):176–213, 2019), which is formulated in terms of weak and strong Itô–Taylor expansions. In the final section we exhibit examples showing how the three projections can differ, and explore alternative notions of optimality.

Keywords Ito · Stratonovich · Projection



Volume 7
Special Issue 1
Article 20
Number 111

RESEARCH PAPER

A parameterisation-invariant modification of the score test

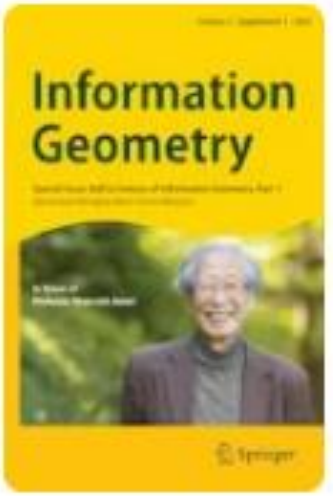
P. E. Jupp[1]

Received: 31 August 2022 / Revised: 10 December 2022 / Accepted: 4 February 2023 /
Published online: 7 March 2023

Abstract

The null distribution of the score test statistic is asymptotically chi-squared for large samples. The error in this approximation is improved greatly by a cubic modification. The coefficients of this cubic that are given in the literature depend on the parameterisation. This paper provides parameterisation-invariant versions of the coefficients, expresses them in terms of appropriate tensors, and provides geometric interpretations.

Keywords Bartlett correction · Generalised Bartlett correction · Interest parameter · Invariant Taylor expansion · Large-sample asymptotics · Likelihood yoke · Tensor



Volume 7
Special Issue 1
Article 21
Number 112

RESEARCH PAPER

Geometric thermodynamics for the Fokker–Planck equation: stochastic thermodynamic links between information geometry and optimal transport

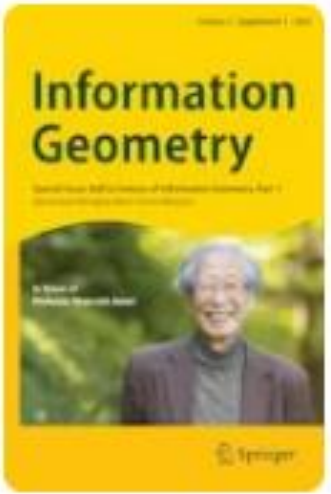
Sosuke Ito[1]

Received: 7 September 2022 / Revised: 31 January 2023 / Accepted: 6 February 2023 /
Published online: 9 March 2023

Abstract

We propose a geometric theory of non-equilibrium thermodynamics, namely geometric thermodynamics, using our recent developments of differential-geometric aspects of entropy production rate in non-equilibrium thermodynamics. By revisiting our recent results on geometrical aspects of entropy production rate in stochastic thermodynamics for the Fokker–Planck equation, we introduce a geometric framework of non-equilibrium thermodynamics in terms of information geometry and optimal transport theory. We show that the proposed geometric framework is useful for obtaining several non-equilibrium thermodynamic relations, such as thermodynamic trade-off relations between the thermodynamic cost and the fluctuation of the observable, optimal protocols for the minimum thermodynamic cost and the decomposition of the entropy production rate for the non-equilibrium system. We clarify several stochastic-thermodynamic links between information geometry and optimal transport theory via the excess entropy production rate based on a relation between the gradient flow expression and information geometry in the space of probability densities and a relation between the velocity field in optimal transport and information geometry in the space of path probability densities.

Keywords Stochastic thermodynamics · Entropy production · Information geometry · Optimal transport theory · Fokker–Planck equation



Volume 7
Special Issue 1
Article 22
Number 113

RESEARCH PAPER

Geometric thermodynamics for the Fokker–Planck equation: stochastic thermodynamic links between information geometry and optimal transport

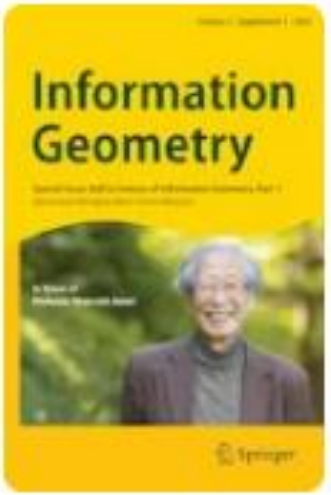
Sosuke Ito[1]

Received: 7 September 2022 / Revised: 31 January 2023 / Accepted: 6 February 2023 /
Published online: 9 March 2023

Abstract

We propose a geometric theory of non-equilibrium thermodynamics, namely geometric thermodynamics, using our recent developments of differential-geometric aspects of entropy production rate in non-equilibrium thermodynamics. By revisiting our recent results on geometrical aspects of entropy production rate in stochastic thermodynamics for the Fokker–Planck equation, we introduce a geometric framework of non-equilibrium thermodynamics in terms of information geometry and optimal transport theory. We show that the proposed geometric framework is useful for obtaining several non-equilibrium thermodynamic relations, such as thermodynamic trade-off relations between the thermodynamic cost and the fluctuation of the observable, optimal protocols for the minimum thermodynamic cost and the decomposition of the entropy production rate for the non-equilibrium system. We clarify several stochastic-thermodynamic links between information geometry and optimal transport theory via the excess entropy production rate based on a relation between the gradient flow expression and information geometry in the space of probability densities and a relation between the velocity field in optimal transport and information geometry in the space of path probability densities.

Keywords Stochastic thermodynamics · Entropy production · Information geometry · Optimal transport theory · Fokker–Planck equation



Volume 7
Special Issue 1
Article 23
Number 114

SURVEY PAPER

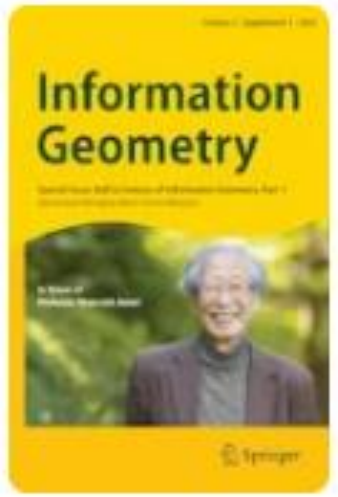
Optimal projection filters with information geometry

John Armstrong[1] · Damiano Brigo[2] · Bernard Hanzon[3]

Received: 18 February 2023 / Revised: 3 May 2023 / Accepted: 13 May 2023 /
Published online: 2 June 2023

Abstract
We review the introduction of several types of projection filters. Projection structures coming from information geometry are used to obtain a finite dimensional filter in the form of a stochastic differential equation (SDE), starting from the exact infinite-dimensional stochastic partial differential equation (SPDE) for the optimal filter. We start with the Stratonovich projection filters based on the Hellinger distance as introduced and developed in Brigo et al. (IEEE Trans Autom Control 43(2):247–252, 1998, Bernoulli 5(3):495–534, 1999), where the SPDE is put in Stratonovich form before projection, hence the term "Stratonovich projection". The correction step of the filtering algorithm can be made exact by choosing a suitable exponential family as manifold, there is equivalence with assumed density filters and numerical examples have been studied. Other authors further developed these projection filters and we present a brief literature review. A second type of Stratonovich projection filters was introduced in Armstrong and Brigo (Math Control Signals Syst 28(1):1–33, 2016) where a direct L^2 metric is used for projection. Projecting on mixtures of densities as a manifold coincides with Galerkin methods. All the above projection filters lack optimality, as the single vector fields of the Stratonovich SPDE are projected optimally but the SPDE solution as a whole is not approximated optimally by the projected SDE solution according to a clear criterion. This led to the optimal projection filters in Armstrong et al. (Proc Lond Math Soc 119(1):176–213, 2019, Projection of SDEs onto submanifolds. "Information Geometry", 2023 special issue on half a century of information geometry, 2018), based on the Ito vector and Ito jet projections, where several types of mean square distances between the optimal filter SPDE solution and the sought finite dimensional SDE approximations are minimized, with numerical examples. After reviewing the above developments, we conclude with the remaining challenges.

Keywords Stochastic partial differential equations · Stochastic differential equations · SPDEs projection on a submanifold · Stratonovich projection · Itô-vector projection · Itô-jet projection · Nonlinear filtering · Projection filters · Stratonovich projection filters · Optimal projection filters



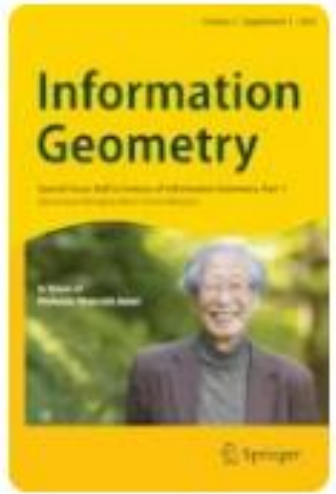
Volume 7
Special Issue 1
Article 24
Number 115

RESEARCH PAPER

A certain ODE-system defining the geometric divergence

Takashi Kurose[1]

Received: 13 December 2022 / Revised: 27 April 2023 / Accepted: 25 May 2023 /
Published online: 5 June 2023



Volume 7
Special Issue 1
Article 25
Number 116

Abstract

A system of ordinary differential equations satisfied by the geometric divergence, a generalization of the canonical divergence on a dually flat space, is introduced. By using the system, it is shown that a family of contrast functions can be constructed on an arbitrary statistical manifold in a certain canonical manner.

Keywords Statistical manifolds · Contrast functions · Geometric divergences

RESEARCH PAPER

Doubly autoparallel structure and curvature integrals

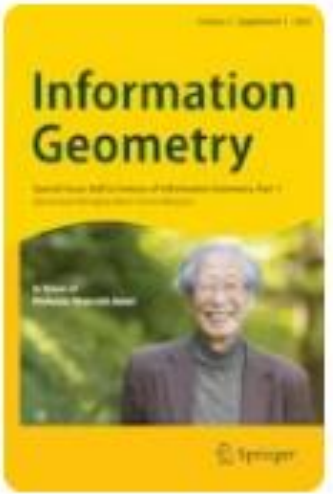
Applications to iteration complexity for solving convex programs

Atsumi Ohara[1] · Hideyuki Ishi[2] · Takashi Tsuchiya[3]

Received: 9 September 2022 / Revised: 4 July 2023 / Accepted: 9 July 2023 /
Published online: 27 July 2023

Abstract
On a statistical manifold, we can define autoparallel submanifolds and path integrals of the second fundamental forms (*curvature integrals*) for its primal and dual affine connections, respectively. A submanifold is called *doubly autoparallel* if it is simultaneously autoparallel with respect to the both connections. In this paper we first discuss common properties of such submanifolds. In particular we next give an algebraic characterization of them in Jordan algebras and show their applications. Further, we exhibit that both curvature integrals induced from dually flat structure are interestingly related to an unexpected quantity, i.e., iteration-complexity of the interior-point algorithms for convex optimization defined on a submanifold that is *not* doubly autoparallel.

Keywords Doubly autoparallel structure · Jordan algebra · Curvature integral · Interior point method · Structured covariance estimation



Volume 7
Special Issue 1
Article 26
Number 117

RESEARCH PAPER

G-dual Teleparallel Connections in Information Geometry

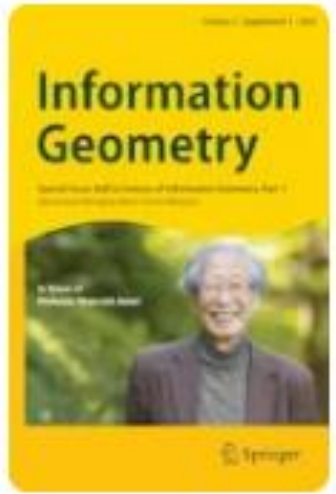
F. M. Ciaglia[1] · F. Di Cosmo[1,2] · A. Ibort[1,2] · G. Marmo[3,4]

Received: 26 July 2022 / Revised: 13 June 2023 / Accepted: 26 July 2023 /
Published online: 25 August 2023

Abstract

Given a real, finite-dimensional, smooth parallelizable Riemannian manifold $(\mathcal{N}, G)$ endowed with a teleparallel connection ∇ determined by a choice of a global basis of vector fields on $\mathcal{N}$, we show that the G-dual connection ∇^* of ∇ in the sense of Information Geometry must be the teleparallel connection determined by the basis of G-gradient vector fields associated with a basis of differential one-forms which is (almost) dual to the basis of vector fields determining ∇. We call any such pair (∇, ∇^*) a ***G-dual teleparallel pair***. Then, after defining a covariant $(0, 3)$ tensor T uniquely determined by $(\mathcal{N}, G, \nabla, \nabla^*)$, we show that T being symmetric in the first two entries is equivalent to ∇ being torsion-free, that T being symmetric in the first and third entry is equivalent to ∇^* being torsion free, and that T being symmetric in the second and third entries is equivalent to the basis vectors determining ∇ (∇^*) being parallel-transported by ∇^* (∇). Therefore, G-dual teleparallel pairs provide a generalization of the notion of Statistical Manifolds usually employed in Information Geometry, and we present explicit examples of G-dual teleparallel pairs arising both in the context of both Classical and Quantum Information Geometry.

Keywords Probability distributions · Quantum states · Dual connections



Volume 7
Special Issue 1
Article 27
Number 118

RESEARCH PAPER

Estimation with infinite-dimensional exponential family and Fisher divergence

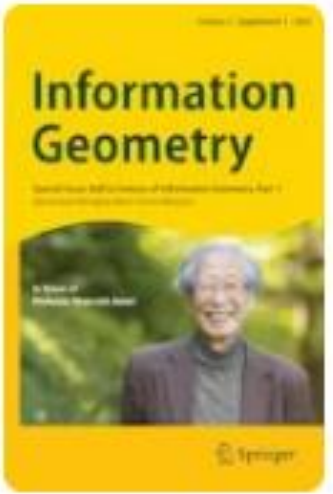
Kenji Fukumizu[1]

Received: 28 February 2023 / Revised: 25 October 2023 / Accepted: 29 October 2023 /
Published online: 13 November 2023

Abstract

Infinite dimensional exponential families have been theoretically studied, but their practical applications are still limited because empirical estimation is not straightforward. This paper first gives a brief survey of studies on the estimation method for infinite-dimensional exponential families. The method uses score matching, which is based on the Fisher divergence. The second topic is to investigate the Fisher divergence as a member of an extended family of divergences, which employ operators in defining divergences.

Keywords Infinite-dimensional exponential family · Fisher divergence · Score matching



Volume 7
Special Issue 1
Article 28
Number 119

SURVEY PAPER

Legendre duality: from thermodynamics to information geometry

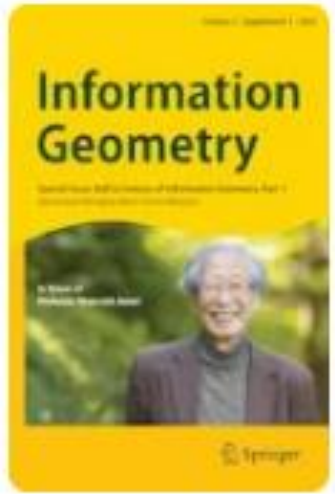
Jan Naudts[1] · Jun Zhang[2]

Received: 29 March 2023 / Revised: 19 October 2023 / Accepted: 27 October 2023 /
Published online: 8 November 2023

Abstract

This paper reviews the role of convex duality in Information Geometry. It clarifies the notion of bi-orthogonal coordinates associated with Legendre duality by treating its two underlying aspects separately: as a dual coordinate system and as a bi-orthogonal frame. It addresses the deformation of exponential families in a way that stills preserves the dually-flat geometry of 1- and (-1)-connections. The deformation involves a metric which generalizes the Fisher–Rao metric controlled by one degree of freedom and a pair of connections controlled by an additional degree of freedom.

Keywords Convex duality · Dually-flat geometry · Deformed exponential and logarithmic functions · Rho–tau connections



Volume 7
Special Issue 1
Article 29
Number 120

RESEARCH PAPER

The Fisher metric as a metric on the cotangent bundle

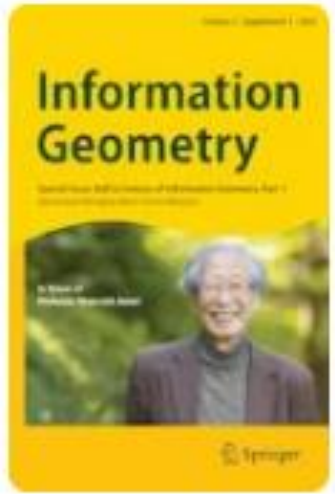
Hiroshi Nagaoka[1]

Received: 11 September 2023 / Revised: 10 November 2023 / Accepted: 12 November 2023 /
Published online: 21 December 2023

Abstract

The Fisher metric on a manifold of probability distributions is usually treated as a metric on the tangent bundle. In this paper, we focus on the metric on the cotangent bundle induced from the Fisher metric with calling it the Fisher co-metric. We show that the Fisher co-metric can be defined directly without going through the Fisher metric by establishing a natural correspondence between cotangent vectors and random variables. This definition clarifies a close relation between the Fisher co-metric and the variance/covariance of random variables, whereby the Cramér-Rao inequality is trivialized. We also discuss the monotonicity and the invariance of the Fisher co-metric with respect to Markov maps, and present a theorem characterizing the co-metric by the invariance, which can be regarded as a cotangent version of Čencov's characterization theorem for the Fisher metric. The obtained theorem can also be viewed as giving a characterization of the variance/covariance.

Keywords Information geometry · Fisher metric · Cotangent space · Čencov's (Chentsov's) theorem



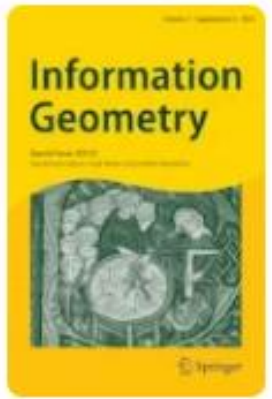
Volume 7
Special Issue 1
Article 30
Number 121

RESEARCH PAPER

Gibbs manifolds

Dmitrii Pavlov[1] · **Bernd Sturmfels**[1,2] · **Simon Telen**[1,3]

Received: 13 February 2023 / Revised: 26 May 2023 / Accepted: 31 May 2023 /
Published online: 3 July 2023



Volume 7
Special Issue 2
Article 1
Number 122

Abstract

Gibbs manifolds are images of affine spaces of symmetric matrices under the exponential map. They arise in applications such as optimization, statistics and quantum physics, where they extend the ubiquitous role of toric geometry. The Gibbs variety is the zero locus of all polynomials that vanish on the Gibbs manifold. We compute these polynomials and show that the Gibbs variety is low-dimensional. Our theory is applied to a wide range of scenarios, including matrix pencils and quantum optimal transport.

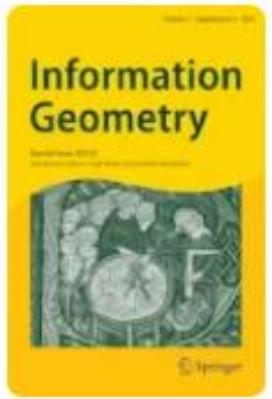
Keywords Gibbs variety · Toric geometry · Semidefinite programming · Quantum optimal transport

RESEARCH

Dually flat structure of binary choice models

Hisatoshi Tanaka[1]

Received: 25 December 2023 / Revised: 30 April 2024 / Accepted: 1 May 2024 /
Published online: 18 May 2024



Volume 7
Special Issue 2
Article 2
Number 123

Abstract

In this study, we consider parametric binary choice models from the perspective of information geometry. The set of models is a dually flat manifold with dual connections, naturally derived from the Fisher information metric. Under the dual connections, the canonical divergence and the Kullback–Leibler divergence of the binary choice model coincide if and only if the model is a logit model.

Keywords Discrete choice models · Logit model · Single-index models · Hessian manifolds · Maximum likelihood estimation.

RESEARCH

A historical perspective on Schützenberger-Pinsker inequalities (extended version)

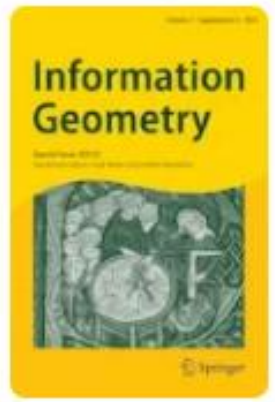
Olivier Rioul[1]

Received: 20 February 2024 / Revised: 11 May 2024 / Accepted: 18 May 2024 /
Published online: 30 May 2024

Abstract

This paper presents a tutorial overview of so-called Pinsker inequalities which establish a precise relationship between information and statistics, and whose use have become ubiquitous in many applications. According to Stigler's law of eponymy, no scientific discovery is named after its original discoverer. Pinsker's inequality is no exception: Years before the publication of Pinsker's book in 1960, the French medical doctor, geneticist, epidemiologist, and mathematician Marcel-Paul (Marco) Schützenberger, in his 1953 doctoral thesis, not only proved what is now called Pinsker's inequality (with the optimal constant that Pinsker himself did not establish) but also the optimal second-order improvement, more than a decade before Kullback's derivation of the same inequality. We review Schützenberger and Pinsker contributions as well as those of Volkonskii and Rozanov, Sakaguchi, McKean, Csiszár, Kullback, Kemperman, Vajda, Bretagnolle and Huber, Krafft and Schmitz, Toussaint, Reid and Williamson, Gilardoni, as well as the optimal derivation of Fedotov, Harremoës, and Topsøe. We also present some historical elements on the life and work of Schützenberger, and discuss an interesting problem of an erroneous constant in the Schützenberger-Pinsker inequality.

Keywords Pinsker inequality · Total variation · Kullback–Leibler divergence · Statistical distance · Mutual information · Data processing inequality



Volume 7
Special Issue 2
Article 3
Number 124

RESEARCH

Fisher–Rao geometry of equivalent Gaussian measures on infinite-dimensional Hilbert spaces

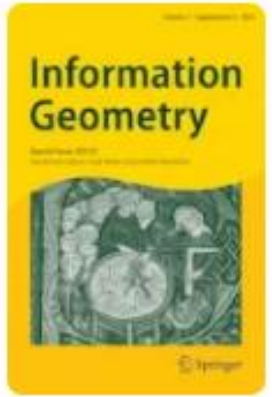
Hà Quang Minh[1]

Received: 18 October 2023 / Revised: 6 March 2024 / Accepted: 14 May 2024 /
Published online: 5 June 2024

Abstract

This work presents an explicit description of the Fisher–Rao Riemannian metric on the Hilbert manifold of equivalent centered Gaussian measures on an infinite-dimensional Hilbert space. We show that the corresponding quantities from the finite-dimensional setting of Gaussian densities on Euclidean space, including the Riemannian metric, Levi–Civita connection, curvature, geodesic curve, and Riemannian distance, when properly formulated, directly generalize to this setting. Furthermore, we discuss the connection with the Riemannian geometry of positive definite unitized Hilbert–Schmidt operators on Hilbert space, which can be viewed as a regularized version of the current setting.

Keywords Fisher–Rao metric · Gaussian measures · Hilbert space · Positive Hilbert–Schmidt operators



Volume 7
Special Issue 2
Article 4
Number 125

RESEARCH

Unveiling cellular morphology: statistical analysis using a Riemannian elastic metric in cancer cell image datasets

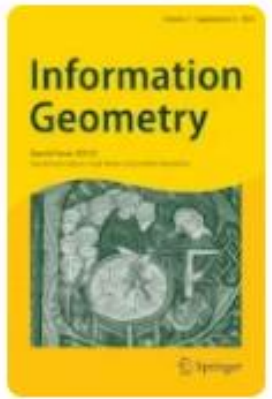
Wanxin Li[1] · Ashok Prasad[2] · Nina Miolane[3] · Khanh Dao Duc[4]

Received: 1 January 2024 / Revised: 13 June 2024 / Accepted: 15 August 2024 /
Published online: 27 August 2024

Abstract

Elastic metrics can provide a powerful tool to study the heterogeneity arising from cellular morphology. To assess their potential application (e.g. classifying cancer treated cells), we consider a specific instance of the elastic metric, the Square Root Velocity (SRV) metric and evaluate its performance against the linear metric for two datasets of osteosarcoma (bone cancer) cells including pharmacological treatments, and normal and cancerous breast cells. Our comparative statistical analysis shows superior performance of the SRV at capturing cell shape heterogeneity when comparing distance to the mean shapes, with better separation and interpretation between different cell groups. Secondly, when using multidimensional scaling (MDS) to find a low-dimensional embedding for unrescaled contours, we observe that while the linear metric better preserves original pairwise distances, the SRV yields better classification.

Keywords Elastic metric · Shape analysis · Cell morphology · Dimensionality reduction



Volume 7
Special Issue 2
Article 5
Number 126

RESEARCH

A variational principle of minimum for Navier–Stokes equation and Bingham fluids based on the symplectic formalism

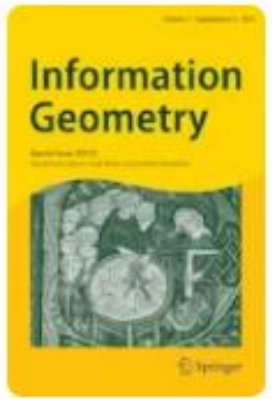
G. de Saxcé[1]

Received: 17 July 2024 / Revised: 17 July 2024 / Accepted: 22 October 2024 /
Published online: 4 November 2024

Abstract

In a previous paper, we proposed a symplectic version of Brezis-Ekeland-Nayroles principle based on the concepts of Hamiltonian inclusions and symplectic polar functions. We illustrated it by application to the standard plasticity in small deformations. The object of this work is to generalize the previous formalism to dissipative media in large deformations and Eulerian description. This aim is reached in three steps. Firstly, we develop a Lagrangian formalism for the reversible media based on the calculus of variation by jet theory. Next, we propose a corresponding Hamiltonian formalism for such media. Finally, we deduce from it a symplectic minimum principle for dissipative media and we show how to obtain a minimum principle for unstationary compressible and incompressible Navier–Stokes equation and Bingham fluids.

Keywords Dynamical dissipative systems · Hamiltonian methods · Brezis-Ekeland-Nayroles principle · Convex dissipation · Navier–Stokes equation · Bingham fluids



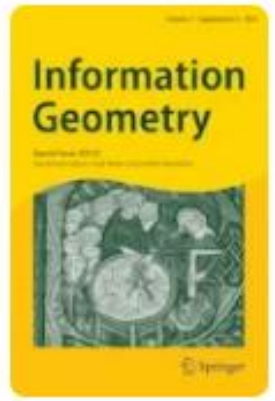
Volume 7
Special Issue 2
Article 6
Number 127

RESEARCH

Cartan moving frames and the data manifolds

Eliot Tron[1] · **Rita Fioresi**[3] · **Nicolas Couëllan**[1,2] · **Stéphane Puechmorel**[1,2]

Received: 13 September 2024 / Revised: 6 November 2024 / Accepted: 7 November 2024 /
Published online: 20 November 2024



Volume 7
Special Issue 2
Article 7
Number 128

Abstract

The purpose of this paper is to employ the language of Cartan moving frames to study the geometry of the data manifolds and its Riemannian structure, via the data information metric and its curvature at data points. Using this framework and through experiments, explanations on the response of a neural network are given by pointing out the output classes that are easily reachable from a given input. This emphasizes how the proposed mathematical relationship between the output of the network and the geometry of its inputs can be exploited as an explainable artificial intelligence tool.

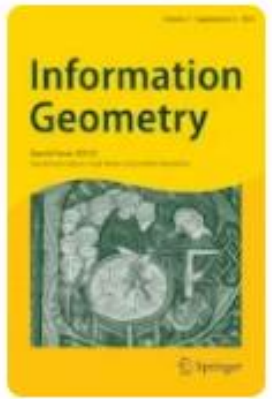
Keywords Neural Networks · Data Manifolds · Moving Frames · Curvature · Explainable AI

RESEARCH

Totally geodesic submanifolds in the manifold SPD of symmetric positive-definite real matrices

Alice Barbara Tumpach[1,2,3] · **Gabriel Larotonda**[4,5]

Received: 5 June 2024 / Revised: 5 June 2024 / Accepted: 27 August 2024 /
Published online: 21 November 2024



Volume 7
Special Issue 2
Article 8
Number 129

Abstract

This paper is a self-contained exposition of the geometry of symmetric positive-definite real $n \times n$ matrices $\mathrm{SPD}(n)$, including necessary and sufficent conditions for a submanifold $\mathcal{N} \subset \mathrm{SPD}(n)$ to be totally geodesic for the affine-invariant Riemannian metric. A non-linear projection $x \mapsto \pi(x)$ on a totally geodesic submanifold is defined. This projection has the minimizing property with respect to the Riemannian metric: it maps an arbitrary point $x \in \mathrm{SPD}(n)$ to the unique closest element $\pi(x)$ in the totally geodesic submanifold for the distance defined by the affine-invariant Riemannian metric. Decompositions of the space $\mathrm{SPD}(n)$ follow, as well as variants of the polar decomposition of non-singular matrices known as Mostow's decompositions. Applications to decompositions of covariant matrices are mentioned.

Keywords Covariance matrices · Reductive symmetric spaces · Decompositions of Lie groups · Symmetric positive-definite matrices

RESEARCH

Information measures and geometry of the hyperbolic exponential families of Poincaré and hyperboloid distributions

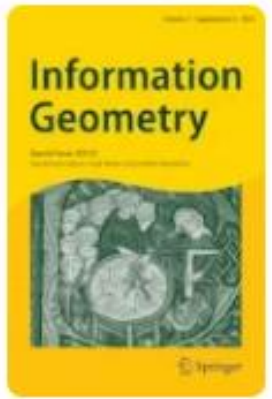
Frank Nielsen[1] · **Kazuki Okamura**[2]

Received: 21 November 2023 / Revised: 20 October 2024 / Accepted: 2 November 2024 /
Published online: 27 November 2024

Abstract

Hyperbolic geometry has become popular in machine learning due to its capacity to embed hierarchical graph structures with low distortions for further downstream processing. It has thus become important to consider statistical models and inference methods for data sets grounded in hyperbolic spaces. In this paper, we study various information-theoretic measures and the information geometry of the Poincaré distributions and the related hyperboloid distributions, and prove that their statistical mixture models are universal density estimators of smooth densities in hyperbolic spaces. The Poincaré and the hyperboloid distributions are two types of hyperbolic probability distributions defined using different models of hyperbolic geometry. Namely, the Poincaré distributions form a triparametric bivariate exponential family whose sample space is the hyperbolic Poincaré upper-half plane and natural parameter space is the open 3D convex cone of two-by-two positive-definite matrices. The family of hyperboloid distributions form another exponential family which has sample space the forward sheet of the two-sheeted unit hyperboloid modeling hyperbolic geometry. In the first part, we prove that all Ali–Silvey–Csiszár's f-divergences between Poincaré distributions can be expressed using three canonical terms using the framework of maximal group invariance. We also show that the f-divergences between any two Poincaré distributions are asymmetric except when those distributions belong to a same leaf of a particular foliation of the parameter space. We report a closed-form formula for the Fisher information matrix, the Shannon's differential entropy and the Kullback–Leibler divergence between such distributions using the framework of exponential families. In the second part, we state the corresponding results for the exponential family of hyperboloid distributions by highlighting a parameter correspondence between the Poincaré and the hyperboloid distributions. Finally, we describe a random generator to draw variates and present two Monte Carlo methods to estimate numerically f-divergences between hyperbolic distributions.

Keywords Exponential family · Group action · Maximal invariant · Csiszár's f-divergence · Poincaré hyperbolic upper plane · Foliation · Minkowski hyperboloid sheet · Information geometry · Statistical mixture models · Statistical inference · Clustering · Expectation-maximization



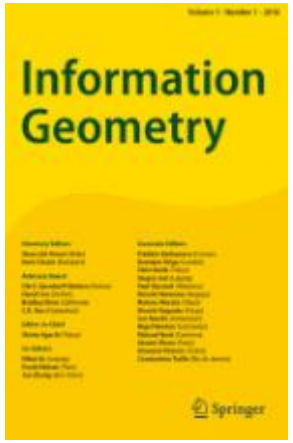
Volume 7
Special Issue 2
Article 9
Number 130

RESEARCH

Some computational results on Koszul–Vinberg cochain complexes

Hanwen Liu [1,3] · Jun Zhang[2,3]

Received: 28 April 2024 / Revised: 6 September 2024 / Accepted: 13 October 2024 /
Published online: 23 December 2024



Volume 8
Issue 1
Article 1
Number 131

Abstract

An affine connection is said to be flat if its curvature tensor vanishes identically. Koszul–Vinberg (KV for abbreviation) cohomology has been invoked to study the deformation theory of flat and torsion-free affine connections on tangent bundle. In this Note, we compute explicitly the differentials of various specific KV cochains, and study their relation to classical objects in information geometry, including deformations associated with projective and dual-projective transformations of a flat and torsion-free affine connection. As an application, we also give a simple yet non-trivial example of a left-symmetric algebra of which the second cohomology group does not vanish.

Keywords Affine structure · Koszul–Vinberg cochain complexes · Conformal and projective transform · Exterior covariant derivative

RESEARCH PAPER

Information geometry of system spaces

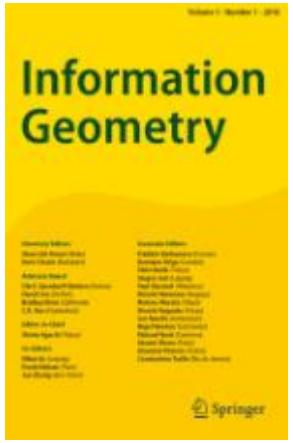
Masayuki Kumon[1]

Received: 16 September 2022 / Revised: 23 December 2024 / Accepted: 26 December 2024 /
Published online: 1 February 2025

Abstract

For the totality of square and regular transfer function matrices, basic differential geometrical structures such as Riemannian metrics and linear connections are introduced from the viewpoint of left and right invariance. These structures are successively introduced into the outer and inner system spaces, and then into the second and higher order cumulant spectral density spaces. The dualities of the derived geometrical structures with torsions and curvatures are investigated, further the possibilities for the existences of divergences are explored for the respective spaces. As an application of the information geometrical structures, we analyze the block oriented nonlinear output feedback mechanisms, and clarify the roles of the associated sensitivity operators. We also obtain the conditions with desired output feedback rules for which the divergences attain the ultimate zero values. Illustrative examples are given to the readers for the block oriented nonlinear output feedback expressions.

Keywords Block oriented model · Covariant derivative · Cumulant spectral density · Divergence · Information geometry · Lie group · Mutual information quantity · Nonlinear output feedback · Riemannian metric · Riemann–Christoffel curvature · Sensitivity operator · Torsion · Volterra series · Wiener series



Volume 8
Issue 1
Article 2
Number 132

RESEARCH

F-t joint distributions on real Siegel domains

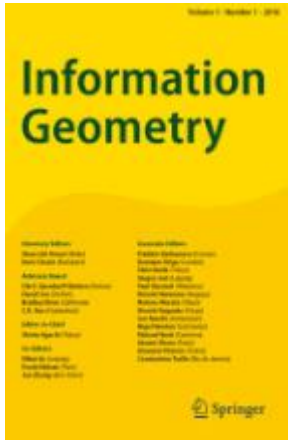
Hiroto Inoue[1]

Received: 31 July 2024 / Revised: 26 January 2025 / Accepted: 31 January 2025 /
Published online: 21 February 2025

Abstract

In statistics, the homogeneous cones are considered as the general parameter spaces for variance matrices related with the graphical models. For the argument of probability distributions, the Wishart distributions are defined based on the analysis of homogeneous cones, as the image of the normal distributions by quadratic maps. Now, the parameter space of pairs of variance matrix and mean vector is identified with more general homogeneous domain, called the real Siegel domain. In the present work, we consider random variables valued in the real Siegel domain associated with the quadratic map of homogeneous cone, and derive their probability density functions via the transformation group. This derivation is also based on the analysis on homogeneous cones including the gamma and beta functions due to Gindikin. We call our result the F-t joint distribution on the real Siegel domain, since its marginal distributions give an extension of the F and t-distributions in 1-dimensional case. Especially, it enables us to describe a joint estimation and a simultaneous hypothesis testing for the pair of parameters of normal distribution.

Keywords Homogeneous cone · Quadratic map · Wishart distribution · Simultaneous hypothesis testing



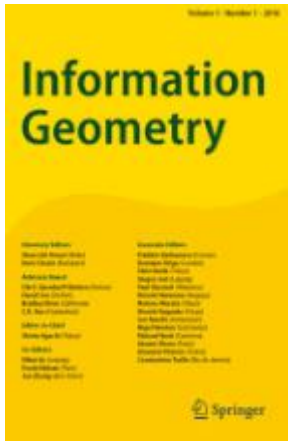
Volume 8
Issue 1
Article 3
Number 133

RESEARCH

Generalized estimation and information

Paul W. Vos[1] · Qiang Wu[1]

Received: 3 September 2024 / Revised: 17 December 2024 / Accepted: 9 February 2025 /
Published online: 22 February 2025



Volume 8
Issue 1
Article 4
Number 134

Abstract

This paper extends the idea of a generalized estimator for a scalar parameter (Vos in Inform Geom 7:151–170, 2022) to multi-dimensional parameters both with and without nuisance parameters. The title reflects the fact that generalized estimators provide more than simply another method to find point estimators, and that the methods to assess generalized estimators differ from those for point estimators. By *generalized estimation* we mean the use of generalized estimators together with an extended definition of *information* to assess their inferential properties. We show that Fisher information provides an upper bound for the information utilized by an estimator and that the score attains this bound. This optimality result provides theoretical justification for likelihood-based inference, effectively narrowing the search for optimal estimators when the score is computationally feasible.

Keywords Cramér–Rao bound · Fisher information · Geometry · Score · Slope

RESEARCH PAPER

Dual stochastic natural gradient descent

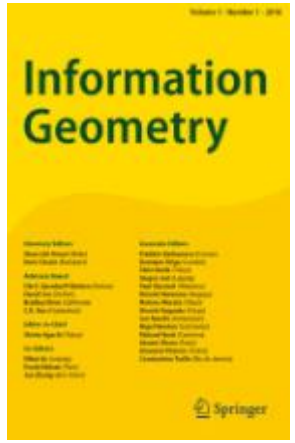
and convergence of interior half-space gradient approximations

Borja Sánchez-López[1] · Jesus Cerquides[1]

Received: 1 April 2020 / Revised: 13 December 2024 / Accepted: 2 March 2025 /
Published online: 4 April 2025

Abstract
The multinomial logistic regression (MLR) model is widely used in statistics and machine learning. On the one hand, stochastic gradient descent (SGD) is the most common approach for determining the parameters of a such model in big data scenarios, due to its simplicity and low computational complexity property. Furthermore, SGD has proven convergence under reasonable conditions. However, SGD has slow sublinear rates of convergence and it often reduces convergence speed due to the plateau phenomenon. On the other hand, stochastic natural gradient descent (SNGD), proposed by Amari, is a manifold optimization method shown to be Fisher efficient when it converges, but its convergence properties remain unproven and it is often computationally prohibitive for models with a large number of parameters. Here, we propose dual stochastic natural gradient descent (DSNGD), a stochastic optimization method for MLR based on manifold optimization concepts. In the discrete scenario, DSNGD (i) has linear per-iteration computational complexity in the number of parameters, and (ii) is proven to converge. To achieve (i) we leverage the dual flatness of the family of joint distributions for MLR to simplify computations. To ensure (ii) DSNGD builds on the foundational ideas of convergent stochastic natural gradient descent (CSNGD), a variant of SNGD with guaranteed convergence, using an independent sequence to construct a bounded approximation of the natural gradient. By generalizing a result from Sunehag et al., we prove that DSNGD converges in the discrete case and maintains linear computational complexity per iteration. Beyond its convergence property and linear computational complexity, DSNGD empirically demonstrates fast convergence comparable to SNGD, improves upon SGD performance, and exhibits stability where SNGD does not.

Keywords Multinomial logistic regression · Stochastic gradient descent · Natural gradient · Convergence · Riemannian manifold · Computational complexity



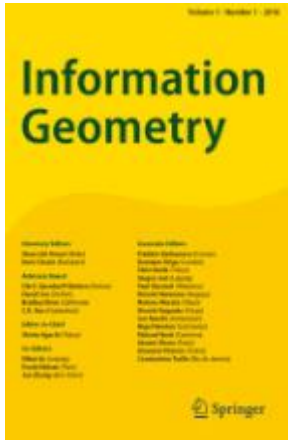
Volume 8
Issue 1
Article 5
Number 135

RESEARCH

Maximum information divergence from linear and toric models

Yulia Alexandr[1] · Serkan Hoşten[2]

Received: 24 April 2024 / Revised: 15 March 2025 / Accepted: 30 March 2025 /
Published online: 21 April 2025



Volume 8
Issue 1
Article 6
Number 136

Abstract

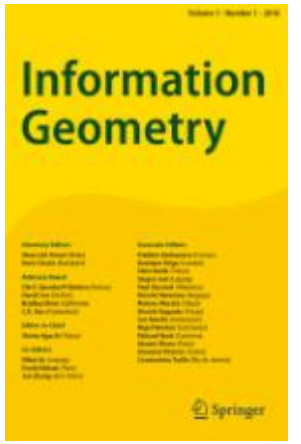
We study the problem of maximizing information divergence from a new perspective using logarithmic Voronoi polytopes. We show that for linear models, the maximum is always achieved at the boundary of the probability simplex. For toric models, we present an algorithm that combines the combinatorics of the chamber complex with numerical algebraic geometry. We pay special attention to reducible models and models of maximum likelihood degree one.

Keywords KL divergence · Information divergence · Maximum likelihood estimation · Toric models

RESEARCH PAPER

A harmonic property of right invariant priors

Tomonari Sei[1] · Fumiyasu Komaki[1,2,3]



Received: 16 July 2023 / Revised: 5 April 2024 / Accepted: 8 April 2024

Volume ??
Issue ??
Article ??
Number ??

Abstract

It is shown that, on any Lie group, the density ratio of the right invariant measure to the left invariant measure is harmonic with respect to the left invariant Riemannian metric. This result is applied to the Bayesian prediction theory on group invariant statistical models. A method of constructing Bayesian prior distributions that asymptotically dominate the right invariant priors is provided.

Keywords Bayesian prediction · Fisher metric · Group invariant model · Laplacian · Superharmonic prior

SURVEY PAPER

Infinite-dimensional distances and divergences between positive definite operators, Gaussian measures, and Gaussian processes

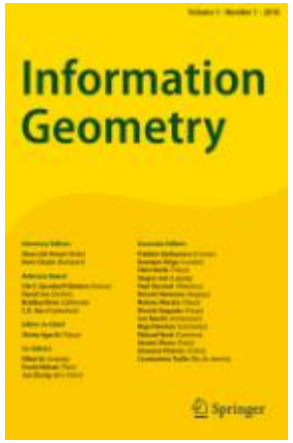
Hà Quang Minh[1]

Received: 16 July 2023 / Revised: 24 March 2024 / Accepted: 15 April 2024

Abstract

This paper presents a survey of recent results on the generalization of distances and divergences on the set of symmetric, positive definite (SPD) matrices to the infinite-dimensional setting of positive definite Hilbert–Schmidt operators on a Hilbert space. Our focus here is on the affine-invariant Riemannian metric and the Log–Determinant divergences. Key components in the proper formulation of the infinite-dimensional distances and divergences include the concepts of extended Hilbert–Schmidt and trace class operators, extended Hilbert–Schmidt inner product and norm, and extended Fredholm and Hilbert–Carleman determinants. On the set of positive trace class operators, the resulting affine-invariant Riemannian distance and Alpha Log–Det divergences can be viewed as regularized versions of the exact Fisher–Rao distance and Rényi divergences, respectively, between equivalent centered Gaussian measures on a Hilbert space. In the case of Gaussian measures corresponding to Gaussian processes with squared integrable paths, the regularized infinite-dimensional distances and divergences can be consistently estimated from finite-dimensional versions, with dimension-independent sample complexities, via the methodology of reproducing kernel Hilbert spaces (RKHS). We also discuss the practical applications of this framework in machine learning and computer vision in the setting of RKHS covariance operators.

Keywords Fisher-Rao metric · Log-Det divergences · positive definite operators · Gaussian measures · Gaussian processes



Volume ??
Issue ??
Article ??
Number ??

RESEARCH PAPER

Robust estimation for kernel exponential families with smoothed total variation distances

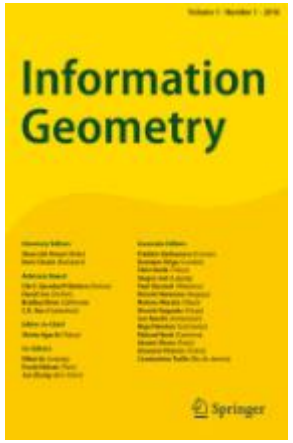
Takafumi Kanamori[1,2] · Kodai Yokoyama[1] · Takayuki Kawashima[1]

Received: 13 May 2023 / Revised: 3 May 2024 / Accepted: 12 July 2024

Abstract

In statistical inference, we commonly assume that samples are independent and identically distributed from a probability distribution included in a pre-specified statistical model. However, such an assumption is often violated in practice. Even an unexpected extreme sample called an *outlier* can significantly impact classical estimators. Robust statistics studies how to construct reliable statistical methods that efficiently work even when the ideal assumption is violated. Recently, some works revealed that robust estimators such as Tukey's median are well approximated by the generative adversarial net (GAN), a popular learning method for complex generative models using neural networks. GAN is regarded as a learning method using integral probability metrics (IPM), which is a discrepancy measure for probability distributions. In most theoretical analyses of Tukey's median and its GAN-based approximation, however, the Gaussian or elliptical distribution is assumed as the statistical model. In this paper, we explore the application of GAN-like estimators to a general class of statistical models. As the statistical model, we consider the kernel exponential family that includes both finite and infinite-dimensional models. To construct a robust estimator, we propose the smoothed total variation (STV) distance as a class of IPMs. Then, we theoretically investigate the robustness properties of the STV-based estimators. Our analysis reveals that the STV-based estimator is robust against the distribution contamination for the kernel exponential family. Furthermore, we analyze the prediction accuracy of a Monte Carlo approximation method, which circumvents the computational difficulty of the normalization constant.

Keywords Robust estimation · Integral probability metrics · Kernel exponential family



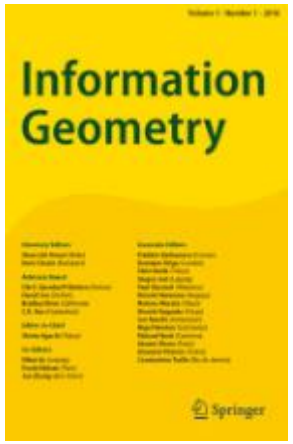
Volume ??
Issue ??
Article ??
Number ??

RESEARCH PAPER

Feature learning and generalization error analysis of two-layer linear neural networks for high-dimensional inputs

Hayato Nishimori[1] · Taiji Suzuki[1,2]

Received: 10 September 2023 / Revised: 15 June 2024 / Accepted: 12 July 2024



Volume ??
Issue ??
Article ??
Number ??

Abstract

It is well known that a model can generalize even when it completely interpolates the training data, which is known as the benign overfitting. Indeed, several work have theoretically revealed that the minimum-norm interpolator can exhibit the benign overfitting. On the other hand, deep learning models such as two-layer neural networks have been reported to outperform "shallow" learning models such as kernel methods under appropriate model sizes by adaptively learning the basis functions to the data. This mechanism is called feature learning, and it is known empirically to be beneficial even when the model size is large. However, it is generally difficult to show that benign overfitting occurs in learning models with feature learning especially for regression problems. In this study, we then analyze the predictive error of the estimator after one step feature learning in a two-layer linear neural network optimized by gradient descent methods and study the effect of feature learning on benign overfitting. The results show that feature learning reduces bias compared to a one-layer linear regression model without feature learning, especially when the eigenvalues of the covariance of input decay slowly. On the other hand, we clarify that the variance is hardly changed by feature learning. This differs significantly from the results for benign overfitting in the situation without feature learning and indicates the usefulness of feature learning.

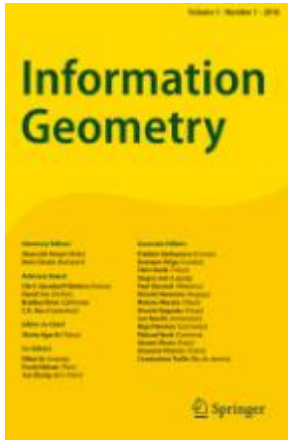
Keywords Neural network · Feature learning · Benign overfitting · Error analysis

RESEARCH PAPER

Iterative minimization algorithm on a mixture family

Masahito Hayashi[1,2,3]

Received: 9 May 2022 / Revised: 10 May 2024 / Accepted: 13 July 2024



Volume ??
Issue ??
Article ??
Number ??

Abstract

Iterative minimization algorithms appear in various areas including machine learning, neural networks, and information theory. The em algorithm is one of the famous iterative minimization algorithms in the area of machine learning, and the Arimoto–Blahut algorithm is a typical iterative algorithm in the area of information theory. However, these two topics had been separately studied for a long time. In this paper, we generalize an algorithm that was recently proposed in the context of the Arimoto–Blahut algorithm. Then, we show various convergence theorems, one of which covers the case when each iterative step is done approximately. Also, we apply this algorithm to the target problem of the em algorithm, and propose its improvement. In addition, we apply it to other various problems in information theory.

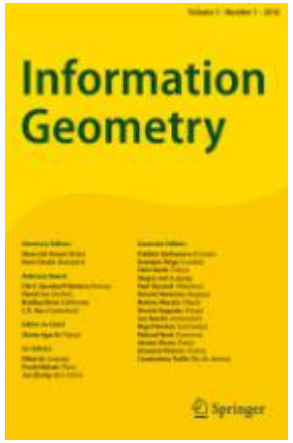
Keywords Minimization · Em algorithm · Mixture family · Channel capacity · Divergence

RESEARCH PAPER

Langevin dynamics for the probability of finite state Markov processes

Wuchen Li[1]

Received: 16 July 2023 / Revised: 18 September 2024 / Accepted: 11 October 2024



Volume ??
Issue ??
Article ??
Number ??

Abstract

We study gradient drift-diffusion processes on a probability simplex set with finite state Wasserstein metrics, namely *finite state Wasserstein common noises*. A fact is that the Kolmogorov transition equation of finite reversible Markov processes satisfies the gradient flow of entropy in finite state Wasserstein space. This paper proposes to perturb finite state Markov processes with Wasserstein common noises. In this way, we introduce a class of stochastic reversible Markov processes. We also define stochastic transition rate matrices, namely Wasserstein Q-matrices, for the proposed stochastic Markov processes. We then derive the functional Fokker–Planck equation in the probability simplex, whose stationary distribution is a Gibbs distribution of entropy functional in a simplex set. Several examples of Wasserstein drift-diffusion processes on a two-point state space are presented.

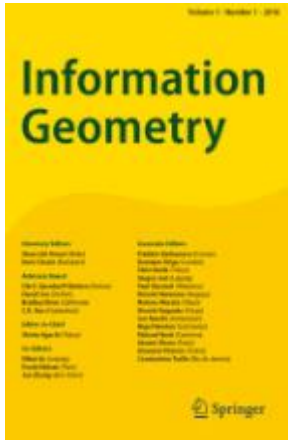
Keywords Optimal transport · Markov process · Wasserstein common noises

RESEARCH

Information geometry of the Otto metric

Nihat Ay[1,2,3]

Received: 29 June 2024 / Revised: 14 September 2024 / Accepted: 23 September 2024



Volume ??
Issue ??
Article ??
Number ??

Abstract

We introduce the dual of the mixture connection with respect to the Otto metric which represents a new kind of exponential connection. This provides a dual structure consisting of the mixture connection, the Otto metric as a Riemannian metric, and the new exponential connection. We derive the geodesic equation of this exponential connection, which coincides with the Kolmogorov forward equation of a gradient flow. We then derive the canonical contrast function of the introduced dual structure.

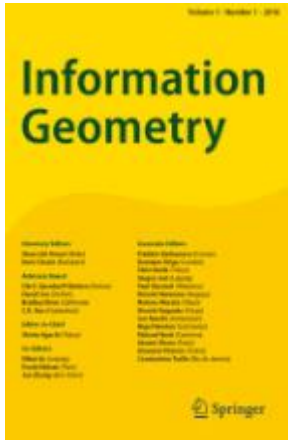
Keywords Wsserstein geometry · Otto metric · Exponential connection · Dual structure · Canonical contrast function

RESEARCH

A geometric modeling of Occam's razor in deep learning

Ke Sun[1] · Frank Nielsen[2]

Received: 17 December 2023 / Revised: 26 March 2025 / Accepted: 24 May 2025



Volume ??
Issue ??
Article ??
Number ??

Abstract

Why do deep neural networks (DNNs) benefit from very high dimensional parameter spaces? Their huge parameter complexities *vs*. stunning performance in practice is all the more intriguing and not explainable using the standard theory of model selection for regular models. In this work, we propose a geometrically flavored information-theoretic approach to study this phenomenon. With the belief that simplicity is linked to better generalization, as grounded in the theory of minimum description length, the objective of our analysis is to examine and bound the complexity of DNNs. We introduce the locally varying dimensionality of the parameter space of neural network models by considering the number of significant dimensions of the Fisher information matrix, and model the parameter space as a manifold using the framework of singular semi-Riemannian geometry. We derive model complexity measures which yield short description lengths for deep neural network models based on their singularity analysis thus explaining the good performance of DNNs despite their large number of parameters.

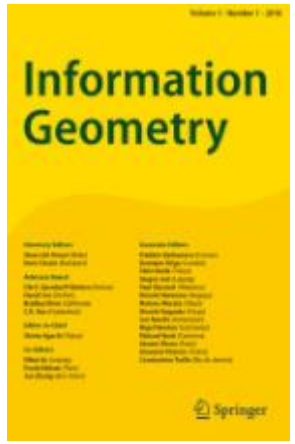
Keywords Information geometry · Deep learning · Minimum description length · Fisher information · Stochastic complexity

RESEARCH PAPER

Non degeneracy of affinelike lie algebra

Michel Nguifo Boyom[1] · Stephane Puechmorel[2]

Received: 29 April 2022 / Revised: 19 May 2025 / Accepted: 1 June 2025



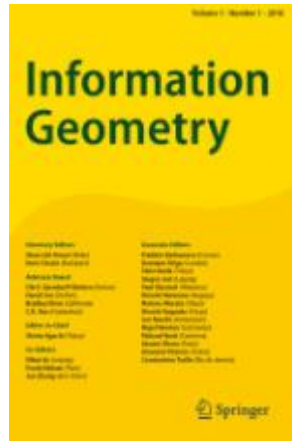
Volume ??
Issue ??
Article ??
Number ??

Abstract

The aim of this note is to prove some cohomological vanishing theorems for non solvable affinelike Lie algebras, say ALLA. There are some relevant consequences of our vanishing theorems:

(1) Every real non solvable affinelike Lie algebra is formally nondegenerate in the sense of A. Weinstein, (Theorem 1).
(2) Let G be a non solvable Lie group whose Lie algebra is an affinelike Lie algebra (g,e). If the radical of [ker(ad(e)),ker(ad(e))] is commutative, then G admits a left invariant symplectic structure if and only if it has an open coadjoint orbit, (Theorem 2).

In Section 6, we use our vanishing theorems to supply an algebraic proof of the normal form theorem for Lie non solvable a-algebroids.



Volume ??
Issue ??
Article ??
Number ??