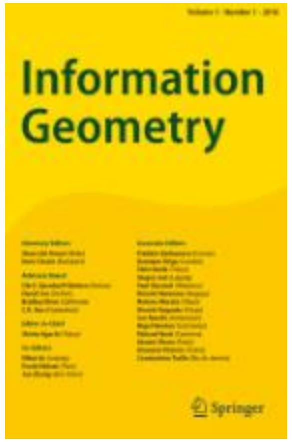
RESEARCH PAPER



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

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

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

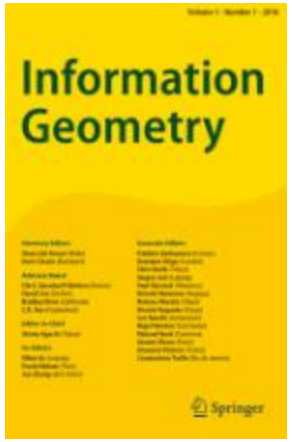
Hà Quang Minh[1]

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

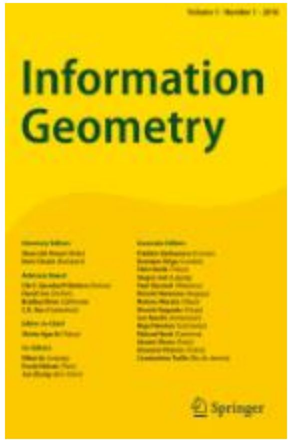
Robust estimation for kernel exponential families with smoothed total variation distances

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

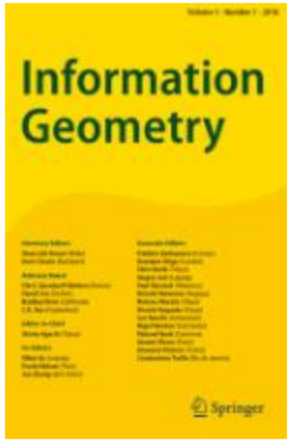
Feature learning and generalization error analysis of two-layer linear neural networks for high-dimensional inputs

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

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

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



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

RESEARCH PAPER

Iterative minimization algorithm on a mixture family

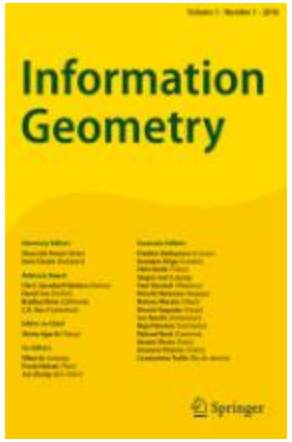
Masahito Hayashi[1,2,3]

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

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



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

RESEARCH PAPER

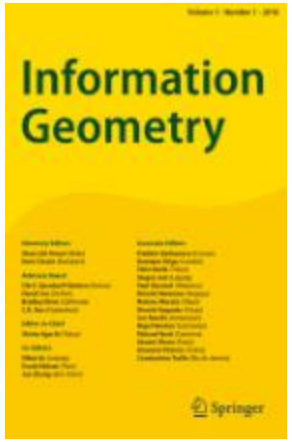
Langevin dynamics for the probability of finite state Markov processes

Wuchen Li[1]

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

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



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

RESEARCH

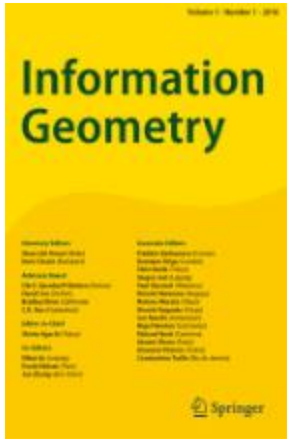
Information geometry of the Otto metric

Nihat Ay[1,2,3]

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

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



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

RESEARCH

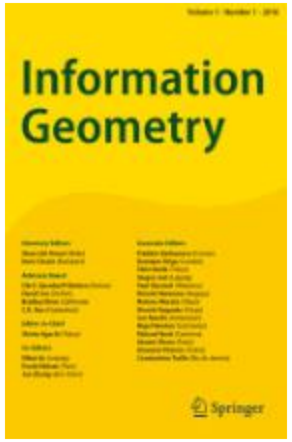
A geometric modeling of Occam's razor in deep learning

Ke Sun[1] · Frank Nielsen[2]

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

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



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

RESEARCH PAPER

Non degeneracy of affinelike lie algebra

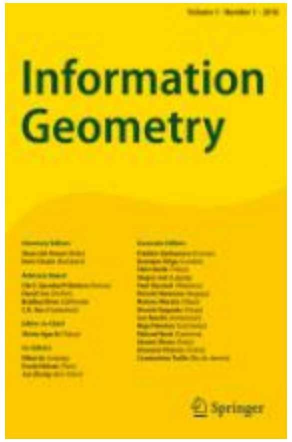
Michel Nguifo Boyom[1] · Stephane Puechmorel[2]

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

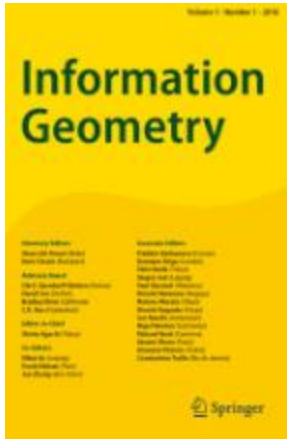
Abstract
The aim of this note is to prove some cohomological vanishing theorems for non solvable affinelike Lie algebras, say ALLA. There are some relevant consequences of our vanishing theorems:

(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 ??



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