Information Geometry and its Applications IV
Liblice, Czech Republic

Václav Kratochvíl, ÚTIA AV ČR

June 12-17, 2016

Published by:
MatfyzPress,
Publishing House of the Faculty of Mathematics and Physics Charles University
Sokolovská 83, 186 75 Praha 8, Czech Republic
as the 510. publication.
Printed by Reprostředisko MFF UK.

The text hasn't passed the review or lecturer control of the publishing company MatfyzPress.
The publication has been issued for the purposes of the IGAIA IV conference.
The publishing house Matfyzpress is not responsible for the quality and content of the text.

Printed in Prague — June 2016

ISBN: 978-80-7378-316-7

Organized by:
Institute of Information Theory and Automation, Czech Academy of Sciences

Sponsored by:
Max Planck Institute for Mathematics in the Sciences
Institute of Information Theory and Automation of the Czech Academy of Sciences
Department of Economics and Finance, Università di Roma "Tor Vergata"

Credits:
Editor: Václav Kratochvíl
LaTeXeditor: Václav Kratochvíl
Logo photo: Castle hotel Liblice
using LaTeX's 'confproc' package, version 0.8 (optional: by V. Verfaille)

INTERNATIONAL CONFERENCE ON INFORMATION GEOMETRY AND ITS APPLICATIONS IV

in honor of Shun-ichi Amari

Organized by:
Institute of Information Theory and Automation,
Czech Academy of Sciences

Liblice

June 12-17, 2016

Programme and Conference Committee:
Nihat Ay *Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
Paolo Gibilisco *Dipartimento di Economia e Finanza, Università di Roma "Tor Vergata", Italy*
František Matúš, *Institute of Information Theory and Automation, Czech Academy of Sciences*

Organizing Committee:
Václav Kratochvíl, *Czech Academy of Sciences, Prague*
František Matúš, *Czech Academy of Sciences, Prague*
Milan Studený, *Czech Academy of Sciences, Prague*

Welcome letter

Welcome to the fourth international conference on Information Geometry and its Applications (IGAIA IV)!

This conference is special, as it honors the numerous scientific achievements of Shun-ichi Amari on the occasion of his 80th birthday. Amari has pioneered the field of information geometry and contributed to a variety of its applications, in particular to mathematical neuroscience. Information geometry is a quickly growing field which has attracted many scientists from mathematics, physics, neuroscience, cognitive systems, robotics, and machine learning.

This conference is special for us also for another reason. The three of us met for the first time in 2002, at the first IGAIA conference, which took place in Pescara, Italy. Staying in touch over the years, later we jointly organised the third IGAIA conference in Leipzig, Germany, in 2010. Now, for the first time IGAIA has cycled through all our home countries, making a detour with IGAIA II through Tokyo, Japan, in 2005, and coming to this beautiful location at Liblice Castle, Czech Republic.

The aim of the conference is to highlight recent developments within the field of information geometry and to identify new directions of research. We are confident that we have put together an exciting program consisting of invited talks and contributed poster presentations. All titles and abstracts are included in this book. We thank Václav Kratochvíl for his terrific work with the local organisation! Special thanks go to Antje Vandenberg who managed all the administrative work in Leipzig. We are also grateful for the support that we received from Milan Studený.

The conference is financially supported by the Max Planck Institute for Mathematics in the Sciences (Information Theory of Cognitive Systems Group), the Institute of Information Theory and Automation of the Czech Academy of Sciences, and the Department of Economics and Finance of the Università degli Studi di Roma "Tor Vergata".

We look forward to many stimulating contributions, and, last but not least, we wish you, Amari Sensei, a wonderful future, full of happiness and information-geometric enchantment!

Nihat Ay, Paolo Gibilisco, and František Matúš

Conference Program

Day 1: June 13, 2016

Morning session

Afternoon session

Day 2: June 14, 2016

Morning session

Afternoon session

Day 3: June 15, 2016

Morning session

Day 4: June 16, 2016

Morning session

Afternoon session

Day 5: June 17, 2016

Morning session

Poster sessions

Information Geometry: Historical Episodes and Recent Developments

Shun-ichi Amari
RIKEN Brain Science Institute
Japan
e-mail: amari@brain.riken.jp

I begin with personal and historical episodes concerning early periods of information geometry. Then, I will touch upon topics of my recent interests: They are

1) geometrical aspects of system complexity, information integration and consciousness,

2) topological aspects and natural gradient learning of singular statistical models, in particular multilayer perceptrons used in deep learning,

3) canonical divergence of statistical manifolds, and

4) geometry of score matching.

The Orlicz-Sobolev Exponential Manifold

Giovanni Pistone
deCastro Statistics
Collegio Carlo Alberto
Moncalieri, Italy
e-mail: giovanni.pistone@gmail.com

One of the possible framework for Information Geometry is the nonparametric generalization of exponential families leading to a Banack manifold modeled on the exponential Orlicz space, called exponential manifold. This leads to a generalization of Amari's setup, including dual affine bundles and second order calculus. However, a nonparametric Information Geometry should be able to discuss nonparametric problems about probability measures as they appear outside Statistics, e.g. in Mathematical Analysis and Statistical Physics. In particular we will discuss the issue of differentiability of densities under Orlicz-Sobolev assumptions. The talk will present some new developments following from

References

[1] B. Lods, G. Pistone (2015). Information Geometry Formalism for the Spatially Homogeneous Boltzmann Equation. *Entropy*, 17(6), 4323-4363

[2] L. Malag, G. Pistone (2015). Second-order Optimization over the Multivariate Gaussian Distribution. In *Geometric Science of Information* F. Barbaresco and F. Nielsen ed.s, Springer LNCS 9389, 349-358.

[3] D. Brigo and G. Pistone (2016). Projection based dimensionality reduction for measure valued evolution equations in statistical manifolds. arXiv:1601.04189

[4] D. Brigo and G. Pistone (2016). Eigenfunctions based Maximum Likelihood estimation of the Fokker Planck Equation and Hellinger projection. Submitted.

Kantorovich optimal transport problem and Shannon's optimal channel problem

Roman Belavkin
Middlesex University
United Kingdom
e-mail: R.Belavkin@mdx.ac.uk

We show that the optimal transport problem (OTP) in the Kantorovich formulation is equivalent to Shannon's variational problem on optimal channel with one additional constraint, which fixes a specific output measure. Without this constraint, a solution to Shannon's problem generally achieves smaller expected cost than the optimal transport. Therefore, from game theoretic point of view, solutions to the Shannon's problem should be always preferred to the optimal transport maps. This result is a consequence of the geometry of the information divergence. Specifically, we show that strict convexity and differentiability of the divergence implies the fact that optimal joint probability measures are always in the interior of the simplex of all joint measures. We discuss these results in the context of OTP on discrete and continuous domains.

Riemannian interpretation of the Wasserstein geometry

Felix Otto
Max Planck Institute for Mathematics in the Sciences
Germany
e-mail: Felix.Otto@mis.mpg.de

We review the infinite-dimensional Riemannian interpretation of the space of probability measures endowed with the Wasserstein metric. In particular, we shall address the relation between the differential geometry of the base space and that of the space of probability measures.

Information Geometry in Multiple Priors Models, Worst Case and Almost Worst Case Distributions

Imre Csiszár
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences, Hungary
e-mail: csiszar.imre@renyi.mta.hu

Thomas Breuer
PPE Research Centre
FH Vorarlberg, Austria
e-mail: thomas.breuer@fhv.at

Abstract

Minimisation of the expectation $E_{\mathbb{P}}(X)$ of a random variable X over a family Γ of plausible distributions $\mathbb{P}$ is addressed, when Γ is a level set of some entropy functional. It is shown, using a generalized Pythagorean identity, that whether or not a worst case distribution (minimising $E_{\mathbb{P}}(X)$ subject to $\mathbb{P} \in \Gamma$) exists, the almost worst case distributions cluster around an explicitely specified, perhaps incomplete distribution. It is also analysed how the existence of a worst case distribution depends on the threshold defining the set Γ.

1 Entropy functionals

An *entropy functional* over nonnegative measurable functions pm a set Ω, equipped with a (finite or σ-finite) measure μ, is

$$H(p) = H_\beta(p) := \int_\Omega \beta(\omega, p(\omega))\mu(d\omega),$$

where β belongs to the collection $\mathbb{B}$ of functions $\beta(\omega, s)$ on $\Omega \times \mathbb{R}^+$, measurable in ω for each $s \in \mathbb{R}^+$, and strictly convex and differentiable in s on $(0, +\infty)$ for each $\omega \in \Omega$, with $\beta(\omega, 0) = \lim_{s\downarrow 0} \beta(\omega, s)$.

Entropy functionals are of main interest for densities $p = \frac{d\mathbb{P}}{d\mu}$ of distributions $\mathbb{P}$. Best known is I-divergence or relative entropy, also familiar are other f*-divergences*

$$D_f(\mathbb{P}||\mu)) = \int f(p(\omega))\mu(d\omega) \quad \text{if } \mu(\Omega) = 1,$$

and *Bregman distances*

$$B_f(p, q) = \int \Delta_f(p(\omega), q(\omega))\mu(d\omega).$$

Here $\Delta_f(s, r) = f(s) - f(r) - f'(r)(s - r)$. More general Bregman distances, with any $\beta \in \mathbb{B}$ in the role of f, will also be needed.

2 The problem

In Mathematical Finance, the monetary payoff or utility of some action, e.g. of a portfolio choice, is a function $X(\omega)$ of a collection ω of random risk factors governed by a distribution not known exactly that can often be assumed to belong to a set Γ of "plausible" distributions. Then the negative of the worst case expected payoff

$$\inf_{\mathbb{P}\in\Gamma} E_{\mathbb{P}}(X) = \inf_{\mathbb{P}\in\Gamma} \int_{\Omega} X(\omega)\mathbb{P}(d\omega)$$

is a measure of the risk of this action. Familiar choices for Γ are an I-divergence ball or some other f-divergence ball around a "default distribution".

More gererally, $\Gamma = \{\mathbb{P} : d\mathbb{P} = p d\mu,\ H(p) \le k\}$ will be considered with any entropy functional H and threshold k, addressing the corresponding infimum $V(k)$. This problem motivated by Mathematical Finance appears of independent interest, and the results obtained are expected to be relevant also in other fields. Study of the closely related problem of minimising entropy functionals subject to moment constraints has substantially contributed to the development of Information Geometry. Results of joint work of the first author with F. Matúš on that problem, including a generalised Pythagorean identity, will be essentially used in this talk.

3 Sketch of results

First, a known expression of $V(k)$ when H is an f-divergence is extended to general entropy functionals.

A main result (new even for I-divergence) is that the densities p with $H(p)$ close to k and $E_p(X)$ close to $V(k)$ cluster in Bregman distance around an explicitly specified function that equals the worst case density if it exists, but otherwise may even have integral less than 1.

Next, the function $V(k)$ is analysed, it is shown differentiable in typical cases but not always. Finally, the dependence on k of the existence of a worst case density is addressed. In case of f-divergences, it is shown to either exist for all k, or to exist/do not exist for k less/larger than a critical value. A conjecture is formulated about how this result might extend to general entropy functionals.

Acknowledgement

Imre Csiszár is supported by the Hungarian National Science Foundation, Grant 105840. Thomas Breuer is supported by the Josef Ressel Centre for Scientific Computing in Finance, Logistics, and Energy. Parts of this paper were presented at ISIT 2013 in Istanbul and at GSI 2015 in Paris. The full paper has been submitted to IEEE Transactions on Information Theory.

Information Geometry and Game Theory

Jürgen Jost
Max Planck Institute for Mathematics in the Sciences
Germany
e-mail: jost@mis.mpg.de

In economics, the values of parameters can influence utilities, and in order to quantify such effects, one needs to compute products between gradients. Therefore, one needs a Riemannian metric. When the parameters are of an information theoretical nature, like capacities of information channel, one is naturally lead to the Fisher metric. In this talk, I shall develop the corresponding mathematical framework. and describe applications and examples in game theory. The players need not be fully rational, but may rather play so-called quantal response equilibria, which are based on kinds of Gibbs distributions.

Entropy on convex sets

Peter Harremoës
Copenhagen Business College
Danemark
e-mail: harremoes@ieee.org

On the simplex of probability distributions information divergence is characterized as a Bregman divergence that satisfies a certain sufficiency condition. Bregman divergences are associated with convex optimization and the sufficiency condition means that the optimal decision is not influenced by irrelevant information. For quantum systems the simplex of probability distributions is replaced by a convex set of density matrices, but in some cases our knowledge of a system may be represented even more general convex sets. For any convex set it is possible to define a function that generalizes the well-known Shannon entropy defined on the simplex, but only special convex sets have the property that the corresponding Bregman divergence satisfies a generalized sufficiency condition. These problems lead to a strengthened version of Caratheodory's theorem and some open problem related to the foundation of quantum mechanics.

Estimation with infinite dimensional kernel exponential families

Kenji Fukumizu
The Institute of Statistical Mathematics
Japan
e-mail: fukumizu@ism.ac.jp

I will discuss infinite dimensional exponential families given by reproducing kernel Hilbert spaces, focusing estimation of the functional parameter with the score matching method. Some results of asymptotic theory are shown on the convergence of parameters for large sample size in both of the well-specified case, where the true density is in the model, and the miss-specified case, where the true density is not within the model but in a slightly larger function space. I will also show some practical applications of the model including parameter estimation with intractable likelihood.

A Monte Carlo approach to a divergence minimization problem

Michel Broniatowski
Université Pierre et Marie Curie
Paris, France
e-mail: michel.broniatowski@upmc.fr

Large deviation probabilities for conditional weighted empirical measures exhibit divergences as rate functions. Reciprocally, many divergences can be seen as such rates, for specific weights. The talk is based on this remark, and states various connections between natural exponential families, their variance functions, and classes of divergences. As a direct consequence, minimization of divergences over specific constraints can be performed using a simple Monte Carlo procedure.

Information geometry associated with two generalized means

Shinto Eguchi
The Institute of Statistical Mathematics
Japan
e-mail: eguchi@ism.ac.jp

We discuss a generalization of e-geodesic and m-geodesic, which provides a natural extension for the standard framework of information geometry. The main idea is to employ quasi-arithmetic and quasi-harmonic means for positive numbers. The generalized e-geodesic is defined by the quasi-arithmetic mean, which associates with the canonical divergence and the generalized expectation; the generalized m-geodesic is defined by the quasi-harmonic mean, which is characterized to preserve the generalized expectation. We elucidate that there is a variety of generalization for the standard framework in which the space of probability density functions is viewed as a dual Euclidean space in the sense that the Pythagoras theorem holds via the generalized e-geodesic and m-geodesic.

Quantum entropy derived from first principles

Frank Hansen
Institute for Excellence in Higher Education
Tohoku University
41 Kawauchi, Aoba-ku, Sendai, Japan
e-mail: frank.hansen@m.tohoku.ac.jp

The most fundamental properties of quantum entropy are derived by considering the union of two ensembles. We discuss the limits these properties put on any entropy measure and obtain, within reasonable interpretations, that they uniquely determine the form of the entropy functional up to normalisation. In particular, the result implies that all other properties of quantum entropy may be derived from these first principles.

Keywords: quantum entropy, ensembles.

References

[1] F. Hansen. (2014), Trace functions with applications in physics, *J. Stat. Phys.*, 154:807–818.

[2] F. Hansen (2016), Quantum entropy derived from first principles, *arXiv: 1604.05093.*

Information geometry of quantum resources

Davide Girolami
University of Oxford
United Kingdom
e-mail: davide.girolami@physics.ox.ac.uk

Quantum systems exhibit peculiar properties which cannot be justified by classical physics, e.g. quantum coherence and quantum correlations. Once confined to thought experiments, they are nowadays created and manipulated by exerting an exquisite experimental control of atoms, molecules and photons. It is important to identify and quantify such quantum features, as they are deemed to be key resources to achieve supraclassical performances in computation and communication protocols.

I show that information geometry is a useful framework to characterize quantum resources. In particular, it elucidates the advantage provided by quantum systems in metrology tasks as phase estimation. Also, geometric measures of quantum resources are observable. Indeed, they can be evaluated in the laboratory by a limited number of interferometric measurements as well as alternative schemes.

Superadditivity of Fisher Information: Classical vs. Quantum

Shunlong Luo
Chinese Academy of Science
China
e-mail:

Superadditivity of Fisher information concerns the relation between the Fisher information in a composite system and those of its constituent parts, and is a surprisingly subtle issue: While the classical Fisher information is superadditive and thus is in accordance with our intuition, various versions of quantum Fisher information, which are natural generalizations of the classical one, may violate superadditivity, as illustrated by F. Hansen for the Wigner-Araki-Yanase skew information. This stands in sharp contrast to many authors belief. In this talk, we review several aspects of superadditivity, discuss its implications and applications, and highlight some related problems.

Revisit to the autoparallelity and the canonical divergence for dually flat spaces

Hiroshi Nagaoka
Graduate School of Informatics and Engineering
The University of Electro-Communications
Chofu, Tokyo 182-8585, Japan
e-mail: nagaoka@is.uec.ac.jp

Let (S, g, ∇, ∇^*) be a finite-dimensional dually flat space with the canonical divergence D, including the representative example where S is an exponential family, g is the Fisher metric, ∇ and ∇^* are the e, m-connections and D is the relative entropy (KL divergence). A submanifold M is said to be ∇-autoparallel (∇^*-autoparallel, resp.) when M forms an open subset of an affine subspace in the coordinate space of a ∇-affine (∇^*-affine, resp.) coordinate system. Note that, when M is either ∇-autoparallel or ∇^*-autoparallel, M itself becomes dually flat. For a submanifold M of S, the following three conditions are shown to be equivalent.

1. There exists a ∇-autoparallel submanifold K of S such that M is a ∇^*- autoparallel submanifold of K.

2. There exists a ∇^*-autoparallel submanifold K of S such that M is a ∇- autoparallel submanifold of K.

3. M is a dually flat space for which the canonical divergence is the restriction $D|_{M\times M}$ of D.

In addition, a submanifold M satisfying (1)-(3) can be represented as $M = K_1 \cap K_2$ by a ∇-autoparallel K_1 and a ∇^*-autoparallel K_2. An important example is given by $S = P(\mathcal{X}^n)$ (the set of positive n-joint distributions) with K_1 being the exponential family consisting of markovian distributions and K_2 being the mixture family consisting of stationary distributions, so that the set $M = K_1 \cap K_2$ of stationary markovian distributions becomes dually flat with the relative entropy as the canonical divergence.

Uniqueness of the Fisher-Rao metric on the space of smooth positive densities

Peter Michor
Universitt Wien
Austria
e-mail: peter.michor@univie.ac.at

For a smooth compact manifold M, any weak Riemannian metric on the space of smooth positive densities which is invariant under the right action of the diffeomorphism group $Diff(M)$ is of the form

$$G_\mu(\alpha, \beta) = C_1(\mu(M)) \int_M \frac{\alpha}{\mu}\frac{\beta}{\mu}\mu + C_2(\mu(M)) \int_M \alpha \cdot \int_M \beta$$

for smooth functions C_1, C_2 of the total volume $\mu(M) = \int_M \mu$. For more details, see `http://arxiv.org/abs/1411.5577`.

In this talk the result is extended to:

1. manifolds with boundary, possibly (there is still a gap) even for manifolds with corners and orbifolds
2. to tensor fields of the form $G_\mu(\alpha_1, \alpha_2, \ldots, \alpha_k)$ for any k which are invariant under $Diff(M)$.

Parametrized measure models and a generalization of Chentsov's theorem

Lorenz Schwachhöfer
Technische Universität Dortmund
Germany
e-mail: Lorenz.Schwachhoefer@math.uni-dortmund.de

Nihat Ay
Max Planck Institute for Mathematics in the Sciences
Germany
e-mail: nay@mis.mpg.de

Jürgen Jost
Max Planck Institute for Mathematics in the Sciences
Germany
e-mail: jjost@mis.mpg.de

Hông Vân Lê
Mathematical Institute, Czech Academy of Sciences
Prague, Czech Republic
e-mail: hvle@math.cas.cz

We review the infinite-dimensional Riemannian interpretation of the space of probability measures endowed with the Wasserstein metric. In particular, we shall address the relation between the differential geometry of the base space and that of the space of probability measures.

Nonlinear PDE in Information Geometry

Gerard Misiolek
University of Notre Dame
USA
e-mail: gmisiole@nd.edu

Using a geometric framework developed by Arnold in 1960's for the Euler equations of hydrodynamics I will describe a family of evolution equations associated with the Amari-Chentsov structure on the space of densities on the circle.

Information Geometric Nonlinear Filtering: a Hilbert Approach

Nigel Newton
University of Essex
United Kingdom
njn@essex.ac.uk

Nonlinear filtering is a branch of Bayesian estimation in which a "signal" process is progressively estimated from the history of a related "observations" process. Nonlinear filters are typically expressed in terms of stochastic differential equations for the posterior distribution of the signal, which is nearly always of infinite-dimension (in the sense that it cannot be represented by a finite number of statistics). The natural "state space" for a nonlinear filter is a suitably rich family of probability measures having an appropriate topology, and the statistical manifolds of Information Geometry are obvious candidates.

The talk will outline recent results on Hilbert manifold representations for nonlinear filters, concentrating on their information-theoretic properties. Finite-dimensional filters, and their role in approximation, will be briefly discussed.

Embeddings of statistical manifolds

Hông Vân Lê
Mathematical Institute, Czech Academy of Sciences
Prague, Czech Republic
e-mail: hvle@math.cas.cz

I shall present a theorem stating that any compact (possibly with boundary) statistical manifold (M, g, T) admits an isostatistical embedding into the statistical manifold $Cap_+(\Omega)$ of all positive probability measures on a finite sample space Ω provided with the Fisher metric and the Amari-Chentsov tensor. Furthermore, any finite dimensional noncompact statistical manifold (M, g, T) admits an embedding I into the space $Cap_+(N^+)$ of all positive probability measures on the set N^+ of all natural numbers such that g is equal to the Fisher metric defined on $I(M)$ and T is equal to the Amari-Chentsov tensor defined on $I(M)$. Hence any statistical manifold is a statistical model.

Knowledge modelling after Shannon

Flemming Topsøe
Department of Mathematical Sciences
University of Copenhagen
Denmark
e-mail: topsoe@math.ku.dk

All researchers – whatever field they represent – encounter situations where elements of cognition such as truth, belief, perception and knowledge are important. A main aim in such situations is to make inference based on sound rational considerations. Since uncertainty is an inevitable element in these situations, it is not surprising that models based on probabilistic thinking have had a dominating role. This is also true of information theory, especially having Shannon Theory in mind. However, models have been suggested which are more abstract in nature. Three types of such models are: *Geometric* models with a focus on geodesics (initiated by Amari), models based on *convexity* with a focus on duality (with Csiszár and Matús as central researchers) and then models based on *game theory* with a focus on notions of equilibrium (models promoted by the author).

The talk will focus on game theoretical models. It will be consistent with previous research of the author but more abstract in nature and, as a pronounced element, motivated to a large extent by philosophical considerations. As examples of the kind of philosophy that enters we mention two: Firstly, the mantra going back to Good that *belief is a tendency to act* and then the view – emerging from discussions with Harremoës – that there are limits to what can be known, indeed, *you can only know what you can describe.* The technical modelling takes as starting point a bivariate function, *description effort* depending on elements of truth and of belief. Features which may be touched upon: A notion of *robustness* to ease inference; discussion of the role of convexity and affinity and, related to this, the introduction of a notion of *control*; introduction of the simplest type of models from the theory developed, emphasizing the close relation to *Bregman divergencies*; examples to show that both geometric problems – in casu *Sylvesters problem* from location theory – and well known problems from information theory – such as *capacity problems* – can be discussed conveniently based on the game theoretical modelling.

Higher Order Analysis of Bayesian Cross Validation in Regular Asymptotic Theory

Sumio Watanabe
Tokyo Institute of Technology
Japan
e-mail: swatanab@dis.titech.ac.jp

In Bayesian estimation, optimization of the hyperparameter is one of the most important problems. Since the effect of the hyperparameter choice to the generalization loss can not be derived by the lower order asymptotic theory, we need the higher order statistics using tensor analysis. In this talk, we introduce the higher order analysis of the Bayesian cross validation loss and prove the following results. Firstly, the cross validation loss and the average generalization loss are minimized by the common hyperparameter asymptotically. Secondly, such a hyperparameter does not minimizes the random generalization loss even asymptotically. And lastly, the information criterion WAIC has the same higher order asymptotic behavior as the Bayesian cross validation loss. We also show that minimizing the cross validation loss is different from maximizing the marginal likelihood.

Cosmological parameter estimation and Fisher information matrix

Shiro Ikeda
The Institute of Statistical Mathematics
Japan
e-mail:shiro@ism.ac.jp

In cosmology, estimation of so called cosmological parameters is the ultimate goal in order to understand of our universe. With modern tech- nology, computer can randomly simulate a universe according to a set of cosmological parameters, and we observe the real universe through telescopes. Thorough comparison between them will provide a good esti- mate of the cosmological parameters. But before going to the estimation, we should understand how difficult the problem is. We simulated a lot of universe with different cosmological parameters and estimate the Fisher information matrix with a non-parametric method. This will show how difficult the problem is, and how much sample points we need in order to have a good estimate.

Geometry of affine immersions and construction of geometric divergences

Hiroshi Matsuzoe
Nagoya Institute of Technology
Japan
e-mail: matsuzoe@nitech.ac.jp

Affine differential geometry (ADG) is to study hypersurfaces or immersions which are affinely congruent in an affine space. It is known that dual affine connections and statistical manifold structures naturally arise in this framework. In particular, generalized conformal transformations of statistical manifolds have important roles in ADG. Originally, such generalized conformal transformations were introduced in asymptotic theory of sequential estimations in information geometry. In addition, by latest developments of Tsallis nonextensive statistical physics, importance of these conformal structures are rapidly increasing. In this presentation, we summarize geometry of generalized conformal structures on statistical manifolds from the viewpoint of ADG. After that, we apply ADG for construction of divergence functions. Recently, generalized means and divergence functions of non KL-type have been discussed in Tsallis statistics. Therefore, we elucidate geometric meanings of generalized means using generalized conformal structures, and we consider generalization of canonical divergence from the viewpoint of ADG.

Geometry of Boltzmann Machines

Guido Montúfar
Max Planck Institute for Mathematics in the Sciences
Germany
e-mail: montufar@mis.mpg.de

A Boltzmann machine is a network of stochastic units. It defines an exponential family of probability distributions over the joint states of all network units, with natural parameters given by pair interaction weights and biases. When some of the units are hidden, the observable probability distributions form an interesting geometric object, which has been studied in information geometry, algebraic statistics, and machine learning. In this talk I give an overview on these investigations and present new results regarding the representational power of deep Boltzmann machines and the identifiability of parameters in restricted Boltzmann machines.

A PROJECTION ALGORITHM BASED ON THE PYTHAGORIAN THEOREM AND ITS APPLICATIONS

Shotaro Akaho
AIST
Tsukuba, Ibaraki 305-8568 Japan
e-mail: s.akaho@aist.go.jp

Hideitsu Hino
University of Tsukuba
Tsukuba, Ibaraki 305-8573, Japan
e-mail: hinohide@cs.tsukuba.ac.jp

Neneka Nara Ken Takano Noboru Murata
Waseda University
Shinjuku, Tokyo 169-8555, Japan
e-mail: {extraterrestrial@moegi, ken.takano@toki}.waseda.jp,
noboru.murata@eb.waseda.ac.jp

We consider the α-projection from a point p on a dually flat manifold $\mathcal{S}$ to a submanifold $\mathcal{M} \subset \mathcal{S}$, which is a fundamental procedure in statistical inference. Since the α-projection can be found by minimizing an α-divergence[1], gradient descent type algorithms are often used. However, in some applications, the derivative of divergence is not available or numerically unstable. In this poster, we propose a simple and robust algorithm without calculating the derivative of divergence.

The algorithm is based on the Pythagorian theorem for dually flat manifold. Suppose $\{p_i\}_{i=1,\ldots,k} \in \mathcal{S}$ are represented by $-\alpha$-affine coordinate system, they define the $-\alpha$-flat submanifold $\mathcal{M}$ by their affine combinations, $\mathcal{M} = \{\sum_{i=1}^{k} \theta_i p_i \mid \sum_{i=1}^{k} \theta_i = 1\}$. Let $q \in \mathcal{M}$ be a candidate of the α-projection of $p \in \mathcal{S}$. When q is actually the α-projection, the Pythagorian theorem holds

$$r_i = D^{(\alpha)}(p, q) + D^{(\alpha)}(q, p_i) - D^{(\alpha)}(p, p_i) = 0. \tag{1}$$

If r_i is more than or less than zero, it means that the α-geodesic connecting p and q does not intersect orthogonally to $\mathcal{M}$.

Based on this fact, the proposed algorithm increases θ_i when $r_i > 0$ while it decreases θ_i when $r_i < 0$. In particular when we can assume all θ_i's are nonnegative, θ_i can be updated by $\theta_i^{(t+1)} = \theta_i^{(t)} f(r_i)$, where $f(r)$ is a positive and monotonically increasing function such that $f(0) = 1$. After the update, θ_i's are normalized so that $\sum_{i=1}^{k} \theta_i = 1$.

As applications of the proposed algorithm, we consider two problems: nonparametric e-mixture estimation and nonnegative matrix factorization.

The e-mixture is defined as an exponential mixture of k distributions $\{p_i(x)\}$,

$$p(x;\theta) = \exp\left(\sum_{i=1}^{k} \theta_i \log p_i(x) - b(\theta)\right), \quad \sum_{i=1}^{k} \theta_i = 1, \quad \theta_i \geq 0, \tag{2}$$

where $b(\theta)$ is a normalization factor. Compared to an ordinary mixture $\sum \theta_i p_i(x)$, the e-mixture has advantages that it belongs to exponential families and it satisfies the maximum entropy principle. We applied the e-mixture modeling to a transfer learning problem, where we have only a small number of samples for a target task while a lot of samples are given for similar tasks. The problem is to find the m-projection ($\alpha = -1$) of $p(x)$ representing the target data to an e-flat submanifold ($\alpha = 1$) defined by a set of e-mixtures of data distributions $\{p_i(x)\}_{i=1,\ldots,k}$ corresponding to the data of similar tasks. We consider the problem in a nonparametric setting, where $p(x)$ and $p_i(x)$'s are empirical distributions. However, since the derivative of divergence is not available in the nonparametric setting, we apply the proposed algorithm to estimate θ_i's by using a characterization of e-mixture[2] and a nonparametric estimation of divergence[3].

Nonnegative matrix factorization (NMF) is a method for dimension reduction, where data matrix X is approximated by a product of low rank matrices W and H, and all components of X, W, H are nonnegative. Letting Π be the column-wise L_1 normalization operator, $\Pi(X) = \Pi(W)\Pi(H)$ holds if $X = WH$. The normalized version of NMF is known as a topic model used in natural language processing. Since the normalized column can be regarded as a probability vector, the NMF is formulated as a fitting problem of an m-flat submanifold[4]. This problem can be solved by alternating e-projections. Exising methods of NMF[5] are numerically unstable when zero components are included in W or H because of the logarithm of zero. To avoid the unstability, we apply the proposed algorithm to estimate the matrices W and H.

Keywords: Pythagorian theorem, α-projection, mixture models, topic models

References

[1] Amari, S. (1985) *Differential-Geometrical Methods in Statistics*, Springer

[2] Murata, N., Fujimoto, Y. (2009) Bregman divergence and density integration, *Journal of Math for Industry*, 1, 97-104

[3] Hino, H., Murata, N. (2013) Information estimators for weighted observations, *Neural Networks*, 1, 260-275

[4] Akaho, S. (2004) The e-PCA and m-PCA: dimension reduction of parameters by information geometry, In *Proc. of IJCNN*, 129-134

[5] Sra, S., Dhillon, I. S. (2005) Generalized nonnegative matrix approximations with Bregman divergences. In *NIPS*, 283-290

ON ROBERTSON-TYPE UNCERTAINTY PRINCIPLES

Attila Andai, Attila Lovas
Department for Mathematical Analysis
Budapest University of Technology and Economics
H-1521 Budapest XI. Stoczek u. 2, Hungary
e-mail: andaia@math.bme.hu, lovas@math.bme.hu

A generalisation of the classical covariance for quantum mechanical observables has previously been presented by Gibilisco, Hiai and Petz. Gibilisco and Isola has proved that the usual quantum covariance gives the sharpest inequalities for the determinants of covariance matrices. We introduce new generalisations of the classical covariance which gives better inequalities, furthermore it has a direct geometric interpretation.

Keywords: uncertainty principle, quantum Fisher information

In quantum mechanics the notion of the (symmetrized) covariance of the observables A and B at a given state D was defined as

$$\mathrm{Cov}_D(A,B) = \frac{1}{2}\left(\mathrm{Tr}(DAB) + \mathrm{Tr}(DBA)\right) - \mathrm{Tr}(DA)\,\mathrm{Tr}(DB)$$

and the variance as $\mathrm{Var}_D(A) = \mathrm{Cov}_D(A,A)$. In 1930 Scrödinger proved the uncertainty relation

$$\mathrm{Var}_D(A)\,\mathrm{Var}_D(B) - \mathrm{Cov}_D(A,B)^2 \geq \frac{1}{4}\left|\mathrm{Tr}(D\,[A,B])\right|^2$$

which was generalized by Robertson in 1934 [6] for the set of observables $(A_i)_{1,\dots,N}$ as

$$\det\left(\left[\mathrm{Cov}_D(A_h,A_j)\right]_{h,j=1,\dots,N}\right) \geq \det\left(\left[-\frac{\mathrm{i}}{2}\,\mathrm{Tr}(D\,[A_h,A_j])\right]_{h,j=1,\dots,N}\right).$$

Later several notions of covariance occured, such as the quantum f-covariance [5]

$$\mathrm{Cov}_D^f(A,B) = \mathrm{Tr}\left(Af(L_{n,D}R_{n,D}^{-1})R_{n,D}(B)\right),$$

the antisymmetric f-covariance and the symmetric f-covariance

$$\mathrm{qCov}^{as}_{D,f}(A,B) = \frac{f(0)}{2}\left\langle \mathrm{i}\,[D,A]\,, \mathrm{i}\,[D,B]\right\rangle_{D,f}$$
$$\mathrm{qCov}^{s}_{D,f}(A,B) = \frac{f(0)}{2}\left\langle \{D,A\}\,, \{D,B\}\right\rangle_{D,f}\,,$$

where $[.,.]$ is the commutator, $\{.,.\}$ the anticommutator and the scalar product

$$\langle A,B\rangle_{D,f} = \mathrm{Tr}\left(A\left(R_D^{1/2} f(L_D R_D) R_D^{1/2}\right)^{-1}(B)\right)$$

is induced by an operator monotone function f, according to Petz classification theorem [4]. Gibilisco and Isola in 2006 conjectured that

$$\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}^{as}_{D,f})$$

holds [3]. The conjecture was proved by Andai [1] and Gibilisco, Imparato and Isola [2] in 2008. We show the more accurate inequality

$$\det(\mathrm{Cov}_D) \geq \det(\mathrm{qCov}^{s}_{D,f}) \geq \det(\mathrm{qCov}^{as}_{D,f})$$

and the following estimation for the gap between the symmetric and antisymmetric covariance

$$\det(\mathrm{qCov}^{s}_{D,f}) - \det(\mathrm{qCov}^{as}_{D,f}) \geq (2f(0))^N \det(\mathrm{Cov}^{f_{RLD}}_D),$$

where $f_{RLD}(x) = \frac{2x}{1+x}$. Moreover we show that the symmetric covariance generated by the function $f_{opt} = \frac{1}{2}\left(\frac{1+x}{2} + \frac{2x}{1+x}\right)$ is universal in the following sense. For every function g the inequality $\det(\mathrm{qCov}^{s}_{D,f_{opt}}) \geq \det(\mathrm{qCov}^{as}_{D,g})$ holds.

References

[1] A. Andai. Uncertainty principle with quantum Fisher information. *J. Math. Phys.*, 49(1), JAN 2008.

[2] P. Gibilisco, D. Imparato, and T. Isola. A Robertson-type uncertainty principle and quantum Fisher information. *Linear Algebra Appl.*, 428(7):1706 – 1724, 2008.

[3] P. Gibilisco and T. Isola. Uncertainty principle and quantum Fisher information. *Ann. Inst. Statist. Math.*, 59(1):147–159, 2007.

[4] D. Petz and Cs. Sudár. Geometries of quantum states. *J. Math. Phys.*, 37(6):2662–2673, 1996.

[5] Dénes Petz. Covariance and Fisher information in quantum mechanics. *J. Phys. A*, 35(4):929–939, 2002.

[6] H. P. Robertson. An indeterminacy relation for several observables and its classical interpretation. *Phys. Rev.*, 46:794–801, Nov 1934.

A Novel Approach to Canonical Divergences within Information Geometry

Nihat Ay[1,2,3] Shun-ichi Amari[4]

[1] Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany
[2] Faculty of Mathematics and Computer Science, University of Leipzig, PF 100920, 04009 Leipzig, Germany
[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[4] RIKEN Brain Science Institute, Laboratory for Mathematical Neuroscience Wako-shi Hirosawa 2-1, Saitama 351-0198, Japan

Following Eguchi [3], a divergence $D : M \times M \to \mathbb{R}$ on a manifold M defines a Riemannian metric g and a pair of dually coupled torsion-free affine connections ∇ and ∇^* in terms of

$$g_p(X,Y) := -X_p Y_q \, D(p \,\|\, q)_{|q=p} \,, \tag{1}$$

and

$$\begin{aligned} g_p(\nabla_X Y, Z) &:= -X_p Y_p Z_q \, D(p \,\|\, q)_{|q=p} \,, \\ g_p(\nabla^*_X Y, Z) &:= -Z_p X_q Y_q \, D(p \,\|\, q)_{|q=p} \,, \end{aligned} \tag{2}$$

where X, Y, Z are smooth vector fields on M. We have the following inverse problem: Given such a structure (M, g, ∇, ∇^*), can we find a divergence D that induces this structure in the sense of (1) and (2)? Matumoto [5] has shown that this is indeed always the case, where, however, D is highly non-unique. On the other hand, when M is dually flat, that is flat with respect to ∇ and ∇^*, a canonical divergence with particularly nice properties has been defined [2]. We propose a natural definition of a canonical divergence for a general, not necessarily flat, M by using the geodesic integration of the (locally defined) inverse exponential map [1] (see also the related work [4]). The new canonical divergence is consistent with Eguchi's approach in the sense that it satisfies (1) and (2). Furthermore, it reduces to the known canonical divergence in the case of dual flatness. Finally, our approach allows us to recover the Kullback-Leibler divergence as well as the α-divergence as special cases.

References

[1] N. Ay and S. Amari. *A Novel Approach to Canonical Divergences within Information Geometry*. Entropy **17** (2015), 8111–8129.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry.* American Mathematical Society: Providence, RI, USA; Oxford University Press: Oxford, UK, 2000.

[3] S. Egutchi. *Second order efficiency of minimum contrast estimators in a curved exponential family.* Ann. Stat. **11** (1983), 793–803.

[4] M. Henmi and R. Kobayashi. *Hooke's law in statistical manifolds and divergences.* Nagoya Math. J. **159** (2000), 1–24.

[5] T. Matumoto. *Any statistical manifold has a contrast functionOn the C3-functions taking the minimum at the diagonal of the product manifold.* Hiroshima Math. J. **23** (1993), 327–332.

Operational Characterization of Divisiblity of Dynamical Maps [arXiv:1601.05522]

Joonwoo Bae
Hanyang University (ERICA), Korea
Freiburg Institute for Advanced Studies (FRIAS), Germany
e-mail: joonwoobae@hanyang.ac.kr

Dariusz Chruściński
Faculty of Physics, Astronomy, and Informatics
Nicolaus Copernicus University Grudziadzka 5, 87-100 Torun, Poland
e-mail: darch@fizyka.umk.pl

In this work, we show the operational characterization to divisibility of dynamical maps in terms of distinguishability of quantum channels. It is proven that distinguishability of any pair of quantum channels does not increase under divisible maps, in which the full hierarchy of divisibility is isomorphic to the structure of entanglement between system and environment. This shows that i) channel distinguishability is the operational quantity signifying (detecting) divisibility (indivisibility) of dynamical maps and ii) the decision problem for divisibility of maps is as hard as the separability problem in entanglement theory. We also provide the information-theoretic characterisation to divisibility of maps with conditional min-entropy.

Dynamics of open quantum systems shares some similarity with entanglement in that the characterisation does have a classical counterpart. In recent years, there has been much progress in understanding, characterising, and detecting divisibility of dynamical maps, the fundamental property that encapsulates and generalises Markovianity of quantum evolution [1, 2, 5, 3], together with careful analysis and classification of non-Markovian quantum evolution (see the collection of papers in [6]). These are of general importance for the study of the interaction between a quantum system and its environment, that is, fundamental phenomena such as dissipation, decay, and decoherence. In the view of quantum information theory, they corresponds to quantum channels conveying information between parties and the properties are connected to information capabilities. In fact, recently it has been shown that complexity of the divisibility problem is computationally intractable, NP-hard [7].

In this work, we present the operational characterisation to divisibility of dynamical maps, more precisely to the refined and general notion k-divisibility, with the unifying idea of quantum channel discrimination. This also merges different approaches of Markovianity in an operational way. Namely, we identify distinguishability of a pair of quantum channels as *information flow* that signifies divisibility of dynamical maps. We show that an infinitesimal increase of distinguishability for an arbitrary pair of channels, hence interpreted as *information backflow*, implies indivisibility, and vice versa. This therefore provides schemes of, both theoretically and practically, detecting indivisible maps including non-Markov processes, similarly to entanglement detection schemes such as entanglement witnesses while the separability problem itself also remains intractable. Our results imply that min-entropy, by which distinguishability is quantified, is hence the information-theoretic tool for the characterisation of divisibility of quantum channels.

References

[1] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford Univ. Press, Oxford, 2007).

[2] U. Weiss, *Quantum Dissipative Systems*, (World Scientific, Singapore, 2000).

[3] R. Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications* (Springer, Berlin, 1987).

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).

[5] Á. Rivas, S. F. Huelga, and M. B. Plenio, Rep. Prog. Phys. **77**, 094001 (2014).

[6] Special issue on Loss of Coherence and Memory Effects in Quantum Dynamics, edited by F. Benatti, R. Floreanini, and G. Scholes [J. Phys. B **45** (2012)].

[7] T. S. Cubitt, J. Eisert, and M. M. Wolf, Phys. Rev. Lett. **108**, 120503 (2012).

[8] R. Koenig, R. Renner, and C. Schaffner, IEEE Trans. Inf. Th., vol. **55**, no. 9 (2009).

INFORMATION DECOMPOSITION BASED ON COMMON INFORMATION

Pradeep Kr. Banerjee
Max Planck Institute for Mathematics in the Sciences,
Leipzig, Germany,
e-mail: pradeep@mis.mpg.de

The total mutual information (MI) that a pair of predictor random variables (RVs) (X_1,X_2) convey about a target RV Y can have aspects of *synergistic information* (conveyed only by the joint RV (X_1X_2), denoted $SI(\{X_1X_2\};Y)$), of *redundant information* (conveyed identically by both X_1 and X_2, denoted $I_\cap(\{X_1,X_2\};Y)$), and of *unique information* (conveyed exclusively by either X_1 or X_2, denoted resp. $UI(\{X_1\};Y)$ and $UI(\{X_2\};Y)$). We have [1]

$$\begin{aligned} I(X_1X_2;Y) &= I_\cap(\{X_1,X_2\};Y) + SI(\{X_1X_2\};Y) + UI(\{X_1\};Y) + UI(\{X_2\};Y) \\ I(X_i;Y) &= I_\cap(\{X_1,X_2\};Y) + UI(\{X_i\};Y),\ i = 1,2 \end{aligned} \tag{1}$$

In this note, we show that a recently proposed measure of $I_\cap$ inspired by the information-theoretic notion of *common information* (due to Gács and Körner) cannot induce a nonnegative decomposition of $I(X_1X_2;Y)$.

Consider the AND mechanism, $Y = \text{AND}(X_1,X_2)$, where $X_i = \text{Bernoulli}(\frac{1}{2})$, $i = 1,2$ and joint pmf $p_{X_1X_2Y}$ is such that $p_{(000)} = p_{(010)} = p_{(100)} = p_{(111)} = \frac{1}{4}$. The decomposition evinces both synergistic and redundant contributions to the total MI. First note that $X_1 \perp X_2$, but $X_1 \not\perp X_2|Y$ since $I(X_1;X_2|Y) = +.189 \neq 0$. Fixing Y induces correlations between X_1 and X_2 when there was none to start with. The induced correlations are the source of positive synergy. The redundancy can be explained by noting that if either $X_1 = 0$ or $X_2 = 0$, then both X_1 and X_2 can exclude the possibility of $Y = 1$ with probability of agreement one. Hence the latter is nontrivial information shared between X_1 and X_2. For independent X_1 and X_2, when one can attribute any nonzero redundancy entirely to the mechanism, there is some consensus that $I_\cap(\{X_1,X_2\};Y) = \frac{3}{4}\log\frac{4}{3} = +.311$ and $SI(\{X_1X_2\};Y) = +.5$ [2].

Given two RVs (X,Y), Gács and Körner (GK) defined the notion of a *common RV* to capture the dependence between X and Y and showed that in general, common information does not account for all the mutual information between X and Y. A measure of $I_\cap$ was defined in [2] to measure how well the redundancy that X_1 and X_2 share about Y can be captured by a RV.

$$I_\cap(\{X_1,X_2\};Y) := \max_{\substack{Q:\ Q-X_1-Y \\ Q-X_2-Y}} I(Q;Y), \tag{2}$$

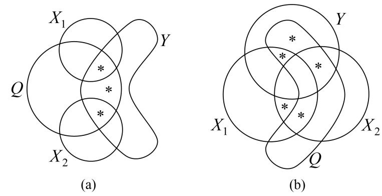

Figure 1: I-diagrams for proof of Lemma 1. Denoting the I-Measure of RVs (Q,X_1,X_2,Y) by μ^*, the atoms on which μ^* vanishes are marked by an asterisk.

where we consider only X_1, X_2 and Y with finite alphabets $\mathcal{X}_1$,$\mathcal{X}_2$ and $\mathcal{Y}$ resp. and it suffices to restrict ourselves to $p_{Q|X_1X_2Y}$ with alphabet $\mathcal{Q}$ such that $|\mathcal{Q}| \leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|+3$. Intuitively, if Q specifies the optimal redundant RV, then conditioning on any predictor X_i should remove all the redundant information about Y, i.e., $I(Q;Y|X_i)=0$, $i=1,2$. We show that for independent X_1 and X_2, if Y is a function of X_1 and X_2 when any positive redundancy can be attributed solely to the function (e.g., as in the AND mechanism above), $I_\cap$ defined as per (2) fails to capture a nonnegative decomposition.

For finite RVs, there is a one-to-one correspondence between Shannon's information measures and a signed measure μ^* over sets, called the I-measure. We denote the I-Measure of RVs (Q,X_1,X_2,Y) by μ^*. We use X to also label the corresponding set in the Information or I-diagram. The I-diagrams in Fig. 1 are valid diagrams since the sets Q,X_1,X_2,Y intersect each other generically and the region representing the set Q splits each atom into two smaller ones.

Lemma 1. *(a) If $X_1 \perp X_2$, then $I_\cap(\{X_1,X_2\};Y)=0$. (b) If $X_1 - Y - X_2$, then $SI(\{X_1X_2\};Y)\leq 0$.*

Proof. (a)The atoms on which μ^* vanishes when the Markov chains $Q-X_1-Y$ and $Q-X_2-Y$ hold and $X_1 \perp X_2$ are shown in the generic I-diagram in Fig. 1(a); $\mu^*(Q\cap Y)=0$ which gives (a).
(b)The atoms on which μ^* vanishes when the Markov chains $Q-X_1-Y$, $Q-X_2-Y$ and X_1-Y-X_2 hold are shown in the I-diagram in Fig. 1(b). In general, for the atom $X_1\cap X_2\cap Y$, μ^* can be negative. However, since X_1-Y-X_2 is a Markov chain by assumption, we have $\mu^*(X_1\cap X_2\cap Y)=\mu^*(X_1\cap X_2)\geq 0$. Then $\mu^*(Q\cap Y)\leq \mu^*(X_1\cap X_2)$, which gives $I_\cap(\{X_1,X_2\};Y)\leq I(X_1;X_2)$. From (1), if X_1-Y-X_2, then the derived synergy measure is $SI(\{X_1X_2\};Y)=I_\cap(\{X_1,X_2\};Y)-I(X_1;X_2)\leq 0$ which gives the desired claim. □

References

[1] P. L. Williams and R. D. Beer (2010), Nonnegative decomposition of multivariate information, *arXiv:1004.2515*.

[2] V. Griffith, T. Ho (2015), Quantifying redundant information in predicting a target random variable, *Entropy*, vol. 17, no. 7, pp. 4644-4653.

KÄHLER AND PARA-KÄHLER STRUCTURES FOR INFORMATION GEOMETRY

Teng Fei
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
e-mail: tfei@mit.edu

Jun Zhang*
University of Michigan
Ann Arbor, Michigan, USA
e-mail: junz@umich.edu

Let $\mathcal{M}$ be a smooth (real) manifold of even dimensions and ∇ be a (not necessarily torsion-free) connection on it. We study the interaction of ∇ with three compatible geometric structures on $\mathcal{M}$: a pseudo-Riemannian metric g, a nondegenerate two-form ω, and a tangent bundle isomorphism $L : T\mathcal{M} \to T\mathcal{M}$. Two special cases of L are: almost complex structure $L^2 = -id$ and almost para-complex structure $L^2 = id$, which will be treated in a unified fashion. When both g and ω are parallel under a torsion-free ∇, it is well known that L and ω are integrable, turning an almost (para-)Hermitian manifold $(\mathcal{M}, g, L)$ into an (para-)Kähler manifold $(\mathcal{M}, g, \omega, L)$, where (g, ω, L) forms "compatible triple". We relax the condition of parallelism under ∇ to the condition of Codazzi coupling with ∇, for each member of the triple.

To this end, we define an *almost Codazzi-(para-)Kähler manifold* $(\mathcal{M}, g, L, \nabla)$ to be an almost (para-)Hermitian manifold $(\mathcal{M}, g, L)$ with an affine connection ∇ (not necessarily torsion-free) which is Codazzi-coupled to both g and L. We prove that if ∇ is torsion-free, then L is automatically integrable and ω is parallel. In this case, $(\mathcal{M}, g, L, \nabla)$ is said to be a *Codazzi-(para-)Kähler manifold*.

Definitions. Let ∇ be a torsion-free connection on $\mathcal{M}$, g and ω be symmetric and skew-symmetric non-degenerate (0,2)-tensor fields respectively, and L be an almost (para-)complex structure. Consider the following relations (for arbitrary vector fields X, Y, Z on $\mathcal{M}$):

(i) $\omega(X, Y) = g(LX, Y)$;

(ii) $g(LX, Y) + g(X, LY) = 0$;

(iii) $\omega(LX, Y) = \omega(LY, X)$;

(iv) $(\nabla_X L)Y = (\nabla_Y L)X$;

(v) $(\nabla_X g)(Y, Z) = (\nabla_Y g)(X, Z)$;

(vi) $(\nabla_X \omega)(Y, Z) = 0$.

Conditions (i)-(iii) define a *compatible triple* (g, ω, L) – any two of the three specifies the third. Condition (iv), (v), and (vi) defines *Codazzi coupling* with ∇ for L, g, and ω, respectively. We call (g, ω, L, ∇) a *compatible quadruple* on $\mathcal{M}$ if Conditions (i)-(vi) are all satisfied.

Our results are shown as the following two main Theorems.

Theorem 1. Let $\mathcal{M}$ admit a torsion-free connection ∇, along with any two of the three tensor fields: g, ω, L. Then $\mathcal{M}$ is a Codazzi-(para-)Kähler manifold if and only if *any* of the following conditions holds (which then implies the rest):

1. (g, L, ∇) satisfy (ii), (iv) and (v);
2. (ω, L, ∇) satisfy (iii), (iv) and (vi);
3. (g, ω, ∇) satisfy (v) and (vi), in which case L is determined by (i).

Furthermore, (g, ω, L, ∇) forms a compatible quadruple on $\mathcal{M}$.

An alternative characterization of the above finding is through relationships among the three transformations of a (not necessarily torsion-free) connection ∇: its g-conjugate ∇^*, its ω-conjugate $\nabla^\dagger$, and its L-gauge transform ∇^L.

Theorem 2. Let (g, ω, L) be a compatible triple. Then $(id, *, \dagger, L)$ act as the 4-element Klein group on the space of affine connections:

$$\begin{aligned}
(\nabla^*)^* = (\nabla^\dagger)^\dagger = (\nabla^L)^L = \nabla;\\
\nabla^* = (\nabla^\dagger)^L = (\nabla^L)^\dagger;\\
\nabla^\dagger = (\nabla^*)^L = (\nabla^L)^*;\\
\nabla^L = (\nabla^*)^\dagger = (\nabla^\dagger)^*.
\end{aligned}$$

It follows that any Codazzi-(para-)Kähler manifold admits a Codazzi dual ∇^C of ∇, defined as $\nabla^* = \nabla^L$, satisfying

(iv) $(\nabla^C_X L)Y = (\nabla^C_Y L)X$;

(v) $(\nabla^C_X g)(Y, Z) = (\nabla^C_Y g)(X, Z)$;

(vi) $(\nabla^C_X \omega)(Y, Z) = 0$.

To summarize: Codazzi-(para-)Kähler manifold is a (para-)Kähler manifold that is simultaneously a statistical manifold. A statistical structure (g, ∇) can be enhanced to a Codazzi-(para-)Kähler structure, which is a special kind of (para-)Kähler manifold, with the introduction of a "nice enough" L in the sense that L is compatible with g and Codazzi coupled to ∇. When ∇ is dually flat (i.e., Hessian statistical structure), we get the so-called "special Kähler geometry."
Keywords: symplectic, Codazzi dual, compatible triple, compatible quadruple

References

[1] T. Fei and J. Zhang (preprint). Interaction of Codazzi coupling and (para)-Kähler geometry.

Information geometry induced from sandwiched Rényi α-divergence

Akio Fujiwara and Kaito Takahashi
Department of Mathematics
Osaka University
Toyonaka, Osaka 560-0043, Japan
e-mail: fujiwara@math.sci.osaka-u.ac.jp

Recently, Müller-Lennert *et al.* [9] and Wilde *et al.* [11] independently proposed an extension of the Rényi relative entropy [10] to the quantum domain. Let $\mathcal{L}(\mathcal{H})$ and $\mathcal{L}_{\rm sa}(\mathcal{H})$ denote the set of linear operators and selfadjoint operators on a finite dimensional complex Hilbert space $\mathcal{H}$, and let $\mathcal{L}_+(\mathcal{H})$ and $\mathcal{L}_{++}(\mathcal{H})$ denote the subset of $\mathcal{L}_{\rm sa}(\mathcal{H})$ comprising positive operators and strictly positive operators. Given $\rho, \sigma \in \mathcal{L}_+(\mathcal{H})$ with $\rho \neq 0$, let,

$$\tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log\,\mathrm{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha - \frac{1}{\alpha-1}\log\,\mathrm{Tr}\,\rho \tag{1}$$

for $\alpha \in (0,1)\cup(1,\infty)$, with the convention that $\tilde{D}_\alpha(\rho\|\sigma) = \infty$ if $\alpha > 1$ and $\ker\sigma \not\subset \ker\rho$. The quantity (1) is called the *quantum Rényi divergence* in [9] or the *sandwiched Rényi relative entropy* in [11], and is extended to $\alpha = 1$ by continuity, to obtain the von Neumann relative entropy. The limiting cases $\alpha \downarrow 0$ and $\alpha \to \infty$ have also been studied in [4, 2] and [9], respectively. The sandwiched Rényi relative entropy has several desirable properties: amongst others, if $\alpha \geq \frac{1}{2}$, it is monotone under completely positive trace preserving maps [9, 11, 3, 6]. This property was successfully used in studying the strong converse properties of the channel capacity [11, 8] and the quantum hypothesis testing problem [7].

Now we confine our attention to the case when both ρ and σ are faithful density operators that belong to the *quantum state space* $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{L}_{++}(\mathcal{H}) \mid \mathrm{Tr}\,\rho = 1\}$. In this case there is no difficulty in extending the quantity (1) to the region $\alpha < 0$. However, it does not seem to give a reasonable measure of information for $\alpha < 0$ [10], since it takes negative values. Motivated by this fact, we study the "rescaled" sandwiched Rényi relative entropy:

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha}\tilde{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha(\alpha-1)}\log\,\mathrm{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\,\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha \tag{2}$$

for $\alpha \in \mathbb{R}\backslash\{0,1\}$ and $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, and is extended to $\alpha = 1$ by continuity. We shall call the quantity (2) as the *sandwiched Rényi α-divergence*. In particular, we are interested in the information geometrical structure [1, 5] induced from (2) on the quantum state space $\mathcal{S}(\mathcal{H})$.

Theorem 1. *The induced Riemannian metric* $g^{(D_\alpha)}$ *is monotone under completely positive trace preserving maps if and only if* $\alpha \in (-\infty, -1] \cup [\frac{1}{2}, \infty)$.

As a by-product, we arrive at the following corollary, the latter part of which was first observed by numerical evaluation [9].

Corollary 2. *The sandwiched Rényi* α*-divergence* $D_\alpha(\rho||\sigma)$ *is not monotone under completely positive trace preserving maps if* $\alpha \in (-1, 0) \cup (0, \frac{1}{2})$. *Consequently, the original sandwiched Rényi relative entropy* $\tilde{D}_\alpha(\rho||\sigma)$ *is not monotone if* $\alpha \in (0, \frac{1}{2})$.

We also studied the dualistic structure $(g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ on the quantum state space $\mathcal{S}(\mathcal{H})$, and obtained the following

Theorem 3. *The quantum statistical manifold* $(\mathcal{S}(\mathcal{H}), g^{(D_\alpha)}, \nabla^{(D_\alpha)}, \nabla^{(D_\alpha)*})$ *is dually flat if and only if* $\alpha = 1$.

Keywords: quantum Rényi α-divergence, monotone metric, dually flatness

References

[1] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs 191 (AMS and Oxford, RI, 2000).

[2] K. M. R. Audenaert and N. Datta, J. Math. Phys., **56**, 022202 (2015); arXiv:1310.7178.

[3] S. Beigi, J. Math. Phys., **54**, 122202 (2013); arXiv:1306.5920.

[4] N. Datta and F. Leditzky, J. Phys. A: Math. Theor. **47**, 045304 (2014); arXiv:1308.5961.

[5] S. Eguchi, Hiroshima Math. J., **22**, 631-647 (1992).

[6] R. L. Frank and E. H. Lieb, arXiv:1306.5358.

[7] M. Mosonyi and T. Ogawa, Commun. Math. Phys., **334**, 1617-1648 (2015); arXiv:1309.3228.

[8] M. Mosonyi and T. Ogawa, arXiv:1409.3562.

[9] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel, J. Math. Phys., **54**, 122203 (2013): arXiv:1306.3142.

[10] A. Rényi, in *Proc. Symp. on Math. Stat. and Probability*, 547-561 (University of California Press, Berkeley, 1961).

[11] M. M. Wilde, A. Winter, and D. Yang, Commun. Math. Phys., **331**, 593-622 (2014): arXiv:1306.1586.

An Inequality for Expectation of Means of Positive Random Variables

Paolo Gibilisco
Universita di Roma "Tor Vergata"
Italy
e-mail: gibilisco@volterra.uniroma2.it

Frank Hansen
Tohoku University, Sendai
Japan
e-mail: frank.hansen@m.tohoku.ac.jp

Let x, y be positive real numbers. The arithmetic, geometric, harmonic, and logarithmic means are defined by

$$\begin{aligned} m_a(x,y) &= \frac{x+y}{2} & \qquad m_g(x,y) &= \sqrt{xy} \\ m_h(x,y) &= \frac{2}{x^{-1}+y^{-1}} & \qquad m_l(x,y) &= \frac{x-y}{\log x - \log y}\,. \end{aligned}$$

Suppose $X, Y\colon \Omega \to (0,+\infty)$ are positive random variables. Linearity of the expectation operator trivially implies $\mathrm{E}(m_a(X,Y)) = m_a(\mathrm{E}(X),\mathrm{E}(Y))$. On the other hand the Cauchy-Schwartz inequality implies $\mathrm{E}(m_g(X,Y)) \le m_g(\mathrm{E}(X),\mathrm{E}(Y))$. Working on a result by Fisher on ancillary statistics Rao [12, 13] obtained the following proposition by an application of Hölder's inequality together with the harmonic-geometric mean inequality.

Proposition 0.1.

$$\mathrm{E}(m_h(X,Y)) \le m_h(\mathrm{E}(X),\mathrm{E}(Y)). \tag{1}$$

It is natural to ask about the generality of this result. For example, does it hold also for the logarithmic mean? To properly answer this question it is better to choose one of the many axiomatic approaches to the notion of a mean.

The best way to face the above problem is to recall the notion of *perspective* of a function [1] [2] [3] [8]; in this way it is possible to see that any mean of pairs of positive numbers may be represented as the perspective of a certain *representing function* [10]. In the paper [4] we prove that inequality (1) holds for a mean m_f if and only if the representing function f is concave.

Once this is done it becomes natural to address the analog question in the non-commutative setting. A positive answer to the case of the matrix harmonic mean was given by Prakasa Rao in [11] and by C.R. Rao himself in [14]. But also in this case the inequality holds in a much wider generality. Using the notion of non-commutative perspectives one can see that also for the Kubo-Ando operator means there is bijection with a class of representing function for which the non-commutative means are just, indeed, the perspectives [9] [6] [5]. The inequality (1) holds true also in the non-commutative case but in a full generality. This follows from the fact that operator means are generated by operator monotone functions; indeed operator monotonicity of a function defined in the positive half-line implies operator concavity [7, Corollary 2.2]; rendering the non-commutative setting completely different from the commutative counter part.

One can also discuss the random matrix case which, to some extent, encompasses the previous results.

References

[1] E.G. Effros. *A matrix convexity approach to some celebrated quantum inequalities.* Proc. Natl. Acad. Sci. USA, **106** (2009), no. 4, 1006-1008.

[2] E.G. Effros and F. Hansen. *Non-commutative perspective.* Ann. Funct. Anal., **5**, (2014), no. 2, 74-79.

[3] A. Ebadian, I. Nikoufar and M.E. Gordji. *Perspective of matrix convex functions.* Proc. Natl. Acad. Sci. USA, **108** (2011), no. 18, 7313 – 7314.

[4] P. Gibilisco and F. Hansen. An Inequality for Expectation of Means of Positive Random Variables. *Preprint* (2016).

[5] P. Gibilisco, F. Hansen and T. Isola. *On a correspondence between regular and non-regular operator monotone functions.* Lin. Alg. Appl., **430** (2009), no. 8-9.

[6] F. Hansen. *Means and concave products of positive semidefinite matrices.* Math Ann., **264** (1983), 119 – 128.

[7] F. Hansen. *The fast track to Löwner's theorem.* Linear Algebra Appl., **438** (2013).

[8] F. Hansen. *Perspectives and partial traces.* ArXiv 1601.06723 (2016).

[9] F. Kubo and T.Ando. *Means of positive linear operators.* Math. Ann. **246** (1979/80), no. 3, 205224.

[10] D. Petz and R. Temesi. *Means of positive numbers.* SIAM J. Matrix Anal. Appl., **27** (2005), 712–720.

[11] B.L.S. Prakasa Rao. *An inequality for the expectation of harmonic mean of random matrices.* Technical Report, Indian Statistical Institute, Delhi, (1998).

[12] C.R. Rao. *R.A. Fisher: the founder of modern statistics.* Statistical Science, **7** (1992), 34–48.

[13] C.R. Rao. *Seven inequalities in statistical estimation theory.* Student, **1** (1996), no. 3, 149–158.

[14] C.R. Rao. *Statistical proofs of some matrix inequalities.* Lin. Alg. Appl., **321** (2000), 307 – 320.

Estimation in a Deformed Exponential Family

Harsha K V *
kv.harsha@yahoo.com

Subrahamanian Moosath K S
Department of Mathematics
Indian Institute of Space Science and Technology
Thiruvananthapuram-695547, Kerala, India
smoosath@iist.ac.in

In this paper we discuss about certain generalized notions of maximum likelihood estimator and estimation problem in a deformed exponential family. A deformed exponential family has two dually flat structures, U-geometry by Naudts [1] and the χ-geometry by Amari et al. [2]. First we recall the U-estimator defined by Eguchi et al. [3] in a deformed exponential family and its properties. A proof of the generalized Cramer-Rao bound defined by Naudts [1] is given. Then we give a proof of the result that in a deformed exponential family the U-estimator for the dual coordinate in the U-geometry is optimal with respect to the generalized Cramer-Rao bound defined by Naudts.

A generalized MLE called the maximum F-likelihood estimator (F-MLE) is defined in a deformed exponential family. Then we show that F-MLE is given in terms of the dual coordinate in the χ-geometry. Finally we pose an open problem regarding the consistency and efficiency of the F-MLE in a deformed exponential family.

Keywords: Deformed exponential family, U-estimator, F-MLE, U-geometry, χ-geometry
Acknowledgments
We express our sincere gratitude to Prof. Shun-ichi Amari for the fruitful discussions.

References

[1] Naudts J. (2004), Estimators, escort probabilities, and ϕ-exponential families in statistical physics. *Journal of Inequalities in Pure and Applied Mathematics*, 5(4), Article 102.

[2] Amari S., Ohara A. and Matsuzoe H. (2012), Geometry of deformed exponential families: Invariant, dually flat and conformal geometries. *Physica A: Statistical Mechanics and its Applications*, 391: 4308-4319.

* *Visiting* Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai-400076, Maharashtra, India

[3] Eguchi S., Komori o. and Ohara A. (2014), Duality of Maximum Entropy and Minimum Divergence. *Entropy*, 16: 3552-3572.

Selecting Moment Conditions in the Generalized Method of Moments for Mixture Models

Zhiyue Huang
MRC-Biostatistics Unit
Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
e-mail: robin.huang@mrc-bsu.cam.ac.uk

The generalized method of moments for mixture models is proposed in [1]. This method involves a convex optimization problem: weighted projecting the sample generalized moments onto a generalized moment space. It has been shown that mixture models can be consistently fitted in point-wise by the generalized method of moments; see [1]. When the generalized moment conditions are carefully selected, the fitted models are robust to the outliers in the data at the cost of losing efficiency. However, it remains unclear that how to choose the generalized moments to balance the trade-off between the efficiency and the robustness. In this poster, we are going to investigate and discuss this problem through a few numerical examples.

Keywords: Generalized method of moments, Mixture models, Spectral decomposition.

References

[1] Huang Z. (2016), Fitting nonparametric mixture models with generalized moments, *to be submitted.*

A simple mixture model for probability density estimation based on a quasi divergence

Osamu Komori
Department of Electrical and Electronics Engineering
University of Fukui
3-9-1 Bunkyo, Fukui 910-8507, Japan
e-mail: o_komori@u-fukui.ac.jp

Shinto Eguchi
Department of Mathematical Analysis and Statistical Inference
The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa,Tokyo, Japan
e-mail: eguchi@ism.ac.jp

A simple mixture model for probability density estimation is proposed based on a loss function associated with a generator function Φ. The function Φ is assumed to be a strictly increasing and concave function, leading to a class of quasi divergence that does not require a bias correction term to obtain a consistent probability density estimator. This property enables us to conduct the probability density estimation in a computationally efficient way, which is in clear contrast with the property of the original U-divergence. The statistical as well as information geometric properties are investigated. Some simulation studies are conducted to demonstrate the performance of the proposed method.

Keywords: Probability density estimation, Quasi divergence, U-divergence References

References

[1] Murata, N., Takenouchi, T., Kanamori, T. and Eguchi, S. (2004). Information geometry of U-boost and Bregman divergence. *Neural Computation 16*, 1437?1481

Symmetry condition for partially factorizable discrete distributions

Aleš Kuběna
ÚTIA, Czech Academy of Sciences
Prague, Czech Republic
e-mail: kubena@utia.cas.cz

Definition

For a hypergraph $\mathcal{A} \subseteq 2^I$, $|I| = n$, we define a probability manifold of everywhere-nonzero $\mathcal{A}$-factorizable distributions on $\Omega_I = \{0,1\}^I$,

$$\mathcal{P}_{\mathcal{A}} = \{\mathbf{p} : \exists_{\theta_Z; Z \in \mathcal{A}} \forall_{X \subseteq I} \ \mathbf{p}(X = (1,\ldots,1)) = \prod_{Z \in \mathcal{A} \cap 2^X} \theta_Z\}$$

Let $\mathcal{P}_k = \mathcal{P}_{\{Z : |Z| = k\}}$ Any ordering $\varphi : I \to \{0,\ldots,n-1\}$ define a projection operator

$$m_\varphi : \mathcal{P}_{\mathcal{A}} \to \mathcal{P}_1, \quad m_\varphi(\mathbf{p})|_k = \frac{\mathbf{p}(\{i : \varphi(i) \leqq k\})}{\mathbf{p}(\{i : \varphi(i) < k\})}.$$

For a permutation group $G \leq S_n$ let $m_{\varphi G}(\mathbf{p})|_k = GeoMean_{\pi \in G}(m_{\pi \cdot \varphi}(\mathbf{p})|_k)$.

Results

We proved that there exist an algebraic condition for the permutation group G, for which the operator $m_{\varphi G}$ is defined unambiguously on $\mathcal{P}_{\mathcal{A}}$, that is, $m_{\varphi G}$ is independent of the original order φ:

Let $sOrb(G)$ be the set of all orbits of all subgroups of G. We claim that $m_{\varphi G} : \mathcal{P}_{\mathcal{A}} \to \mathcal{P}_1$ is independent of the original order φ iff $\mathcal{A} \subseteq sOrb(G)$. Such permutation groups G can be classified for $\mathcal{A}_k = \begin{pmatrix} I \\ k \end{pmatrix}$, $k > 2$. The case $k = 2$ is hard.

In the special case $\mathcal{A} = 2^I$, the condition $\mathcal{A} \subseteq sOrb(G)$ is satisfied only for the full symmetric group S_n for all n, alternating group A_n for $n \geq 4$, and the image of a non-standard embedding $S_5 \to S_6$ for $n = 6$. If $\mathcal{A} = \mathcal{A}_k$, then a subgroup $G \leq S_n$ different from S_n and A_n that satisfies $\mathcal{A} \subseteq sOrb(G)$ only exists for $k \leq 6$ or $k \geq n-3$.

On the other hand, the trivial group $G = \{e\}$ guarantees φ-independency only for a system of n independent Bernoulli distributions.

For a general permutation group $G \leq S_n$, the set of all operators $\{m_{\varphi G} : \varphi \in I^n\}$ forms a convex set with dimension

$$\left[[u^k]\left(e^{\sum_{i=1}^n u^i \frac{\partial}{\partial x_i}} u \frac{\partial}{\partial x_1} - e^{\sum_{i=1}^n u^i \frac{\partial}{\partial x_i}}\right)\right] Z_G(1, 1, \dots, 1)$$

where Z_G is the cycle index of the permutation group G [1] and $e^{\sum_{i=1}^n u^i \frac{\partial}{\partial x_i}}$ is applied as a formal operator.

References

[1] Cameron, P. (1999) Permutation Groups, Cambridge University Press, ISBN 0-521-65379

[2] Giblisco, P. & all (ed.), (2009) Algebraic and geometric method in statistic, Cambridge University Press, ISBN 9780511642401

Volume of the space of qubit channels and the distribution of some scalar quantities on it

Attila Lovas, Attila Andai
Department of Mathematical Analysis
Budapest University of Technology and Economics
H-1521 Budapest XI. Stoczek str. 2., Hungary
e-mail: lovas@math.bme.hu

May 26, 2016

In quantum information theory, a qubit channel is the simplest quantum analogue of a probability transition matrix known from Kolmogorovian probability theory. The space of qubit channels can be identified with a convex submanifold of $\mathbb{R}^{12}$ via Choi representation [2]. To each qubit channel a classical channel can be associated which is called the underlying classical channel.

Our main goal is to investigate the distribution of scalar quantities, which are interesting in information geometrical point of view, on the space of qubit channels.

Our approach based on the positivity criterion for self-adjoint matrices by means of the left upper submatrices. This method was previously succesfully applied by A. Andai to compute the volume of density matrices [1].

The volume of the space of qubit channels with respect to the canonical Eucledian measure is computed, and explicit formulas are presented for the distribution of the volume over classical channels. We have constructed an efficient algorithm for generating uniformly distributed points in the space of qubit channels which enables us to investigate numerically the distribution of scalar quantities on the whole space or over a fixed classical channel. Distribution of trace-distance contraction coefficient (η^{Tr}) was investigated numerically by Monte-Carlo simulations. We computed the distribution of the Hilbert-Schmidt distance between the indentity and its image under the action of a qubit channel.

The range of possible values of η^{Tr} over an arbitrary fixed classical channel was determined explicitely and the mode of η^{Tr} was calculated numerically. We have found that the distribution of trace-distance contraction coefficient shows drammaticaly different behaviour over real and complex unital qubit channels.

Keywords: Qubit channel, trace-distance contraction coefficient, Choi matrix, volume, Monte-Carlo methods.

References

[1] A. Andai (2006), Volume of the quantum mechanical state space, *Journal of Phiysics A: Mathematical and Theoretical*, 39:13641–13657.

[2] Man Duen Choi (1975), Completely positive maps on complex matrices, *Linear Algebra and Applications*, 10:285–290.

Second-order Information Geometry of Hessian Manifolds

Luigi Malagò and Giovanni Pistone
Collegio Carlo Alberto
Address: Via Real Collegio, 30, 10024 Moncalieri, Italy
e-mail: luigi.malago@gmail.com; giovanni.pistone@carloalberto.org

Information Geometry is an interdisciplinary and expanding research field at the intersection of statistics and differential geometry, which studies the geometry of statistical models, represented as manifolds of probability distributions. Notably, Information Geometry provides a principled framework for the analysis and design of natural Riemannian gradient descent algorithms for the optimization of functions defined over statistical models, with applications in machine learning, statistical inference, information theory, stochastic optimization, and several fields in computer science, such as robotics and computer vision.

The task of optimizing a function whose variables are the parameters of a statistical model is widespread in data science, think for example to the optimization of the expected value of a function with respect to a distribution in a statistical model, the maximization of the likelihood, or more in general the minimization of a loss function. Whenever the closed formula for the solution of the problem is unknown, gradient descent methods constitute a classical approach to optimization. However, it is a well-known result in statistics that the geometry of a statistical model is not Euclidean, instead the unique metric which is invariant to reparameterization is the Fisher information metric. It follows that the direction of maximum decrement of a function over a statistical model is given by the Riemannian natural gradient, first proposed by Amari. Despite the directness of first-order methods, there are situations where taking into account the information on the Hessian of the function to be optimized gives an advantage, for instance for ill-conditions problems for which gradient methods may converge too slowly. Similarly to the natural gradient, also the definition of the Hessian of a function depends on the metric, so that second-order methods over statistical manifolds need to be generalized to the Riemannian geometry of the search space.

When we move to the second-order geometry of a differentiable manifold, the notion of covariant derivative is required for the parallel transport between tangent spaces, in particular to compute directional derivatives of vector fields over a manifold. However, an important result in Information Geometry affirms that exponential families, and more in general Hessian manifolds, have a dually-flat nature, which implies the existence of at least two other relevant geometries for statistical models: the mixture and the exponential geometries. Differently from the Riemannian geometry, the exponential and mixture geometries are independent from the notion of metric, and they are defined by two dual affine connections, the mixture and the exponential connections. The dual connections, which are equivalently specified by the dual covariant derivatives, allow to define dual parallel transports, dual geodetics, and ultimately the exponential and mixture Hessians. What is specific of Hessian manifolds, is that the combination of dual Hessians and geodetics allows to define alternative second-order Taylor approximations of a function, without the explicit computation of the Riemannian Hessian and the Riemannian geodetic, which are computationally expensive operations in general. Compared to Riemannian manifolds, dually-flat manifolds have a richer geometry that can be exploited in the design of more sophistical second-order optimization algorithms.

Second-order methods, such as the Newton method, conjugate gradient, and trust region methods, are popular algorithms in mathematical optimization, known for their super-linear convergence rates. The application of such methods to the optimization over statistical manifolds using second-order Riemannian optimization algorithms is a novel and promising area of research, indeed even if Information Geometry and second-order manifold optimization are well consolidated fields, surprisingly little work has been done at the intersection of the two. The optimization methods developed for statistical models based on dual geometries can be adapted to the larger class of Hessian manifolds. Indeed, all Hessian manifolds admit a dual geometrical structure analogous to that of statistical manifolds, given by the dual affine connections. Hessian manifolds include matrix manifolds, such as the cone of positive-definite matrices, and several other convex cones, with applications in robotics, computer vision, pattern recognition, signal processing, conic optimization, and many others.

In this work, after a description of the general theory behind second-order Information Geometry, we present an application to the optimization over the multivariate Gaussian distribution, and analogously, over the cone of positive definite matrices.

Keywords: Information geometry, natural gradient, optimization over manifolds, Hessian manifolds, affine connections, exponential and mixture Hessians.

Log-Hilbert-Schmidt distance between covariance operators and its approximation

Hà Quang Minh
Pattern Analysis and Computer Vision
Istituto Italiano di Tecnologia (IIT)
Via Morego 30, Genova, Italy
e-mail: minh.haquang@iit.it

One of the most commonly used Riemannian metrics on the set of symmetric, positive definite (SPD) matrices is the Log-Euclidean metric [1]. In this metric, the geodesic distance between two SPD matrices A and B is given by

$$d_{\mathrm{logE}}(A,B) = ||\log(A) - \log(B)||_F, \tag{1}$$

where log denotes the matrix principal logarithm.

Log-Hilbert-Schmidt distance. The generalization of the Log-Euclidean metric to the infinite-dimensional manifold $\Sigma(\mathcal{H})$ of positive definite Hilbert-Schmidt operators on a Hilbert space $\mathcal{H}$ has recently been given by [2]. In this metric, termed *Log-Hilbert-Schmidt (Log-HS) metric*, the distance between two positive definite Hilbert-Schmidt operators $A + \gamma I > 0$ and $B + \mu I > 0$, $A, B \in \mathrm{HS}(\mathcal{H})$, $\gamma, \mu > 0$, is given by

$$d_{\mathrm{logHS}}[(A+\gamma I), (B+\mu I)] = ||\log(A+\gamma I) - \log(B+\mu I)||_{\mathrm{eHS}}, \tag{2}$$

with the *extended Hilbert-Schmidt* norm defined by $||A+\gamma I||^2_{\mathrm{eHS}} = ||A||^2_{\mathrm{HS}} + \gamma^2$.

RKHS covariance operators. As examples of positive Hilbert-Schmidt operators, consider covariance operators in reproducing kernel Hilbert spaces (RKHS), which play an important role in machine learning and statistics. Let $\mathcal{X}$ be any nonempty set. Let K be a positive definite kernel on $\mathcal{X}\times\mathcal{X}$ and $\mathcal{H}_K$ its induced RKHS. Let $\mathcal{H}$ be any Hilbert feature space for K, assumed to be separable, which we identify with $\mathcal{H}_K$, with the corresponding feature map $\Phi : \mathcal{X} \to \mathcal{H}$, so that $K(x,y) = \langle \Phi(x), \Phi(y)\rangle_{\mathcal{H}}$ $\forall (x,y) \in \mathcal{X}\times\mathcal{X}$. Let $\mathbf{x} = [x_1, \ldots, x_m]$ be a data matrix randomly sampled from $\mathcal{X}$ according to some probability distribution. The feature map Φ gives the (potentially infinite) data matrix $\Phi(\mathbf{x}) = [\Phi(x_1), \ldots, \Phi(x_m)]$ in $\mathcal{H}$. Formally, $\Phi(\mathbf{x})$ is a bounded linear operator $\Phi(\mathbf{x}) : \mathbb{R}^m \to \mathcal{H}$, defined by $\Phi(\mathbf{x})\mathbf{b} = \sum_{j=1}^m b_j \Phi(x_j)$, $\mathbf{b} \in \mathbb{R}^m$. The covariance operator for $\Phi(\mathbf{x})$ is defined by

$$C_{\Phi(\mathbf{x})} = \frac{1}{m}\Phi(\mathbf{x}) J_m \Phi(\mathbf{x})^* : \mathcal{H} \to \mathcal{H}, \quad J_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T. \tag{3}$$

For $\gamma > 0, \mu > 0$, the Log-HS distance $d_{\mathrm{logHS}}[(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}), (C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})]$ between two regularized covariance operators $(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}})$ and $(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})$

$$d_{\mathrm{logHS}} = ||\log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})||_{\mathrm{eHS}} \tag{4}$$

has a closed form in terms of the corresponding Gram matrices [2]. This distance is generally computationally intensive for large m, however.

Approximation by finite-dimensional Log-Euclidean distances. To reduce the computational cost, we consider computing an *explicit approximate feature map* $\hat{\Phi}_D : \mathcal{X} \to \mathbb{R}^D$, where D is finite and $D << \dim(\mathcal{H})$, so that

$$\langle \hat{\Phi}_D(x), \hat{\Phi}_D(y) \rangle_{\mathbb{R}^D} = \hat{K}_D(x,y) \approx K(x,y), \quad \text{with} \lim_{D\to\infty} \hat{K}_D(x,y) = K(x,y), \tag{5}$$

$\forall (x,y) \in \mathcal{X} \times \mathcal{X}$. With the approximate feature map $\hat{\Phi}_D$, we have the matrix $\hat{\Phi}_D(\mathbf{x}) = [\hat{\Phi}_D(x_1), \ldots, \hat{\Phi}_D(x_m)] \in \mathbb{R}^{D\times m}$ and the approximate covariance operator

$$C_{\hat{\Phi}_D(\mathbf{x})} = \frac{1}{m} \hat{\Phi}_D(\mathbf{x}) J_m \hat{\Phi}_D(\mathbf{x})^T : \mathbb{R}^D \to \mathbb{R}^D. \tag{6}$$

We then consider the following as an approximate version of the Log-HS distance given in Formula (4):

$$\left\| \log\left(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D \right) - \log\left(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D \right) \right\|_F. \tag{7}$$

Key theoretical question. We need to determine whether Formula (7) is truly a finite-dimensional approximation of Formula (4), in the sense that

$$\begin{aligned} &\lim_{D\to\infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D) \right\|_F \\ &= || \log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})||_{\mathrm{eHS}}. \end{aligned} \tag{8}$$

The following results shows that in general, this is *not* possible.

Theorem 1. *Assume that* $\gamma \neq \mu$, $\gamma > 0$, $\mu > 0$. *Then*

$$\lim_{D\to\infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D) \right\|_F = \infty.$$

In practice, however, it is reasonable to assume that we can use the same regularization parameter for both $C_{\hat{\Phi}_D(\mathbf{x})}$ and $C_{\hat{\Phi}_D(\mathbf{y})}$, that is to set $\gamma = \mu$. In this setting, we obtain the necessary convergence, as follows.

Theorem 2. *Assume that* $\gamma = \mu > 0$. *Then*

$$\begin{aligned} &\lim_{D\to\infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \gamma I_D) \right\|_F \\ &= || \log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \gamma I_{\mathcal{H}})||_{\mathrm{eHS}}. \end{aligned} \tag{9}$$

References

[1] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N., 2007. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1), pp.328-347.

[2] Minh, H.Q., San Biagio, M. and Murino, V., 2014. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces. In *Advances in Neural Information Processing Systems* (pp. 388-396).

Correlation functions of hippocampal place cells in open field environments

Mauro M. Monsalve-Mercado, Christian Leibold
Department Biology II
Ludwig-Maximilians-Universitaet
Munich, Germany
e-mail: monsalve@bio.lmu.de

Hippocampal place cells' spiking patterns have been extensively studied on linear tracks. The direct correspondence between space and time in this case, provides an accesible framework on which to study temporal coding strategies, with implications for spatial navigation and memory formation. On the population level, the precise temporal relationship between spiking neurons is readily encoded in the correlation functions, and in the linear track case, can be linked in a straightforward manner to spatial properties characterizing the place fields by means of phase precession. There is, however, little work concerning the temporal structure in an open field exploratory task. Indeed, undersampling of the area shared among two place fields hinders a clear observation of the associated correlation function.

In this work, we develop an analytical framework in which to explore the temporal relationship between two dimensional place fields, which are modeled to undergo phase precession and theta modulation. As our main result, we provide a concise mathematical description of the correlation function, and highlight its connection to spatial parameters shaping the place fields. We contrast our findings with a numerical simulation of place cell activity following a random walk on a circular arena, whose firing patterns arise form a Poisson point process that generate noisy spike patterns.

Keywords: correlation functions, spatial navigation, place fields

Mismatched Estimation in an Exponential Family

Subrahamanian Moosath K S
Department of Mathematics
Indian Institute of Space Science and Technology
Thiruvananthapuram-695547, Kerala, India
smoosath@iist.ac.in

Harsha K V *
kv.harsha@yahoo.com

In this paper we discuss the information geometric framework for the mismatched estimation problem in an exponential family. In the context of population decoding in neuroscience, one often uses a mismatched model or an unfaithful model instead of the original model for a computational convenience or to quantify the correlated activities of neurons etc., see [1], [2] for more details. Oizumi et al. [1] studied the maximum likelihood estimation problem based on a mismatched model in the case of an exponential family from an information geometric point of view. In this paper we discuss the information geometric approach to the general estimation problem based on a mismatched model in an exponential family. We describe the necessary and sufficient conditions for an estimator based on mismatched model to be consistent and efficient. Then we consider the maximum likelihood estimator based on a mismatched model. Oizumi et al. [1] stated certain conditions for the maximum likelihood estimator based on a mismatched model to be consistent and efficient. We give a theoretical formulation of their results in a curved exponential family and a detailed proof of the same.

Keywords: Exponential family, Mismatched model, deformed exponential family
Acknowledgments
We express our sincere gratitude to Prof. Shun-ichi Amari for the fruitful discussions.

References

[1] Oizumi M., Okada M. and Amari S. (2011), Information loss associated with imperfect observation and mismatched decoding. *Frontiers in Computational Neuroscience*, 5(9).

[2] Wu S., Nakahara H. and Amari S. (2001), Population Coding with Correlation and an Unfaithful Model. *Neural Computation*, 13(4): 775-797.

* *Visiting* Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai-400076, Maharashtra, India

QUANTUM MEASUREMENT AS A CONDITIONAL EXPECTATION

Jan Naudts
Physics Department
University of Antwerp
Universiteitsplein 1, 2610 Antwerpen Wilrijk, Belgium
jan.naudts@uantwerpen.be

The state of a quantum system is described by a wave function, this is a normalized element ψ of a Hilbert space, or by a density matrix ρ, which is a trace-class operator with trace 1 and non-negative eigenvalues. Typical for a quantum measurement is that it does *not* reveal the state of the system up to some measurement error. This is a major deviation from the situation in classical systems. If one measures the speed of say a car then the result is a number which usually is a good approximation of the actual speed of the car. This is not the case for quantum measurements. The measurement process itself must be described in quantum-mechanical terms. The coupling between the system at hand and the measurement apparatus fixes an orthonormal basis $(e_n)_n$ in the Hilbert space of quantum states. The outcome of the experiment is then that the state of the system is e_n with probability $|\langle e_n|\psi\rangle|^2$, where $\langle\cdot|\cdot\rangle$ denotes the inner product of the Hilbert space. This phenomenon is known as the *collapse of the wave function*. From these probabilities one can then try to reconstruct the original wave function ψ by repeating the measurement under identical initial conditions.

The point of view here is that the experimental setup necessarily introduces a condition on the outcomes of the experiment. Conditional expectations have been studied in quantum probability theory [4]. The origin of these studies is the discovery [1] of a link with the Tomita-Takesaki theory, which describes one-parameter groups of automorphisms of von Neumann algebras. The condition that the outcome of the experiment is an element of an orthonormal basis is a conditional expectation in this mathematical sense.

The probabilities $|\langle e_n|\psi\rangle|^2$ can now be explained in geometrical terms. The quantum analogue of the Kullback-Leibler divergence is given by

$$D(\sigma||\rho) = \mathrm{Tr}\sigma\ln\sigma - \mathrm{Tr}\sigma\ln\rho. \tag{1}$$

Here, σ and ρ are density matrices. Assume now that σ is the orthogonal projection onto the multiples of the wave function ψ and that ρ is conditioned to be diagonal in the given basis $(e_n)_n$. Then the divergence $D(\sigma||\rho)$ is minimal when the diagonal elements of the matrix ρ equal the probabilities $|\langle e_n|\psi\rangle|^2$. One concludes that the experimental outcome is given by the density matrix which minimizes the divergence $D(\sigma||\rho)$ under the condition that the matrix ρ is diagonal.This minimization process is equivalent to an orthogonal projection onto the manifold of diagonal density matrices.

At the end of the previous century physicists succeeded in devising experiments which avoid the collapse of the wave function. In some cases, thousands of consecutive measurements are possible [3] before the collapse of the wave function is reached. Such measurements are now referred to as *weak measurements*. In a typical setup the system under study is weakly coupled to a second quantum system. On the latter strong measurements are performed as usual. By keeping the coupling between the two subsystems very weak the system of interest is not too much disturbed by the measurements. In addition, by a proper choice of basis vectors the sensitivity of the experiment can be increased [2]. The latter can be understood from the fact that the divergence function diverges at the borders of the manifold [5].

Keywords: Quantum probability theory, quantum conditional expectations, quantum divergence, weak measurements.

References

[1] Accardi L. and Cecchini C. (1982), *Conditional expectations in von Neumann algebras and a theorem of Takesaki,* J. Funct. Anal. **45**, 245–273.

[2] Aharonov Y., Albert D. Z. and Vaidman L. (1988), *How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100,* Phys. Rev. Lett. **60**, 1351–1354.

[3] Guerlin Ch., Bernu J., Deléglise S., Sayrin C., Gleyzes S., Kuhr S., Brune M., Raimond J.-M., Haroche S. (2007), *Progressive field-state collapse and quantum non-demolition photon counting,* Nature **448**, 06057.

[4] Petz D. (2008) *Quantum Information Theory and Quantum Statistics* (Springer)

[5] Naudts J. and Anthonis B. (2015), *Extension of Information Geometry to Non-statistical Systems: Some Examples,* in: Geometric Science of Information, GSI 2015 LNCS proceedings, F. Nielsen and F. Barbaresco eds., (Springer)

Doubly autoparallel structure on statistical manifolds and its applications

Atsumi Ohara
Department of Electrical and Electrnics Engineering
University of Fukui
3-9-1 Bunkyo Fukui, Fukui 910-8507, JAPAN
e-mail: ohara@fuee.u-fukui.ac.jp

One of the important feature of information geometry [1] is a pair of mutually dual affine connections. In statistical manifold there exists a submanifold that is simultaneously autoparallel with respect to both of the affine connections. Such submanifolds, which we call *doubly autoparallel*, play interesting and important roles in applications, e.g., MLE of structured covariance matrices, semidefinite program (SDP) [2, 3], the self-similar solutions to the porous medium equation [4] and so on.

We discuss several properties of doubly autoparallel submanifolds and show a characterization in terms of Jordan algebra when their ambient space is a symmetric cone.

This is a joint work with Prof. Hideyuki Ishi in Nagoya University.

Keywords: mutually dual affine connections, doubly autoparallel submanifold, symmetric cone

References

[1] S. Amari and H. Nagaoka (2000) Methods of information geometry, AMS & Oxford University Press.

[2] A. Ohara, (1999) Information Geometric Analysis of an Interior Point Method for Semidefinite Programming, Proc. of Geometry in Present Day Science (O. Barndorf-Nielsen and V. Jensen eds.) World Scientific, 49-74.

[3] S. Kakihara, A. Ohara and T. Tsuchiya (2014), Curvature integrals and iteration complexities in SDP and symmetric cone programs, Computational Optimization and Applications, Vol. 57, No.3, 623-665.

[4] A. Ohara and T. Wada (2010), Information Geometry of q-Gaussian Densities and Behaviors of Solutions to Related Diffusion Equations, Journal of Physics A: Mathematical and Theoretical, Vol.43, 035002 (18pp.).

Generalized Geometric Quantum Speed Limits

Diego P. Pires
Instituto de Física de São Carlos
Universidade de São Paulo
13560-970, São Carlos, SP, Brasil
e-mail: diegopaivapires@gmail.com

Marco Cianciaruso
School of Mathematical Sciences
The University of Nottingham
University Park, Nottingham NG7 2RD, United Kingdom
e-mail: cianciaruso.marco@gmail.com

Lucas C. Céleri
Instituto de Física
Universidade Federal de Goiás
74.001-970 Goiânia, Goiás, Brazil
e-mail: lucas@chibebe.org

Gerardo Adesso
School of Mathematical Sciences
The University of Nottingham
University Park, Nottingham NG7 2RD, United Kingdom
e-mail: gerardo.adesso@nottingham.ac.uk

Diogo O. Soares-Pinto
Instituto de Física de São Carlos
Universidade de São Paulo
13560-970, São Carlos, SP, Brasil
e-mail: dosp@ifsc.usp.br

The attempt to gain a theoretical understanding of the concept of time in quantum mechanics has triggered significant progress towards the search for faster and more efficient quantum technologies. One of such advances consists in the interpretation of the time-energy uncertainty relations as lower bounds for the minimal evolution time between two distinguishable states of a quantum system, also known as quantum speed limits (QSLs)[1].

Distinguishing between two states of a system being described by a probabilistic model stands as the paradigmatic task of information theory. Information geometry, in particular, applies methods of differential geometry in order to achieve this task [2]. The set of states of both classical and quantum systems is indeed a Riemannian manifold, that is the set of probability distributions over the system phase space and the set of density operators over the system Hilbert space, respectively. Therefore it seems natural to use any of the possible Riemannian metrics defined on such sets of states in order to distinguish any two of its points. However, it is also natural to assume that for a Riemannian metric to be bona fide in quantifying the distinguishability between two states, it must be contractive under the physical maps that represent the mathematical counterpart of noise, i.e. stochastic maps in the classical settings and completely positive trace preserving maps in the quantum one. Interestingly, Čencov's theorem states that the Fisher information metric is the only Riemannian metric on the set of probability distributions that is contractive under stochastic maps, thus leaving us with only one choice of bona fide Riemannian geometric measure of distinguishability within the classical setting. On the contrary, it turns out that the quantum Fisher information metric, also known as Bures-Uhlmann metric, is not the only contractive Riemannian metric on the set of density operators, but rather there exists an infinite family of such metrics, as characterized by the Morozova, Čencov and Petz theorem.

We construct a new fundamental family of geometric QSLs which is in one to one correspondence with the family of contractive Riemannian metrics characterized by the Morozova, Čencov and Petz theorem. We demonstrate how such non uniqueness of a bona fide measure of distinguishability defined on the quantum state space affects the QSLs and can be exploited in order to look for tighter bounds. Our approach is general enough to provide a unified picture, encompassing both unitary and nonunitary dynamics. We show explicit instances of QSLs which make use of some particular contractive Riemannian metric such as the Wigner-Yanase skew information, and can be provably tighter than the corresponding QSLs obtained with the conventional quantum Fisher information, thus highlighting the power of our general approach to reach beyond the state of the art.

Keywords: Quantum Speed Limits, Information Geometry.

References

[1] M. M. Taddei (2014), *Quantum speed limits for general physical processes*, PhD Thesis (Universitade Federal do Rio de Janeiro), arXiv e-prints:1407.4343v1

[2] S. Amari and H. Nagaoka (2000) *Methods of Information Geometry*, Vol. 191, Oxford University Press

CONSTANT CURVATURE CONNECTIONS ON STATISTICAL MODELS

Alexander Rylov
Department of Mathematics
Financial University
125993, Russia, Moscow, Leningradskiy Ave., 49
e-mail: Alexander_Rylov@mail.ru

We discuss statistical manifolds [1] with connections of constant α-curvature. The Pareto two-dimensional statistical model has such a structure that each of its α-connections has a constant curvature $(-2\alpha - 2)$. It is known that if the statistical manifold has an α-connection of constant curvature then it is a conjugate symmetric manifold [2]. The Weibull two-dimensional statistical model has the following structure. Its 1-connection has the constant curvature

$$k^{(1)} = \frac{12\pi^2\gamma - 144\gamma + 72}{\pi^4}$$

where $\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{\infty} \frac{1}{k} - \log n\right)$ is Euler-Mascheroni constant. We compare this model with some known statistical models like normal and logistic ones [3].

Keywords: Statistical manifold, constant α-curvature, conjugate symmetric manifold, Pareto statistical model, Weibull statistical model.

References

[1] Amari S.-i. and Nagaoka H. (2000), *Methods of information geometry*, Transl. Math. Monogr, 191, AMS, Oxford Univ. Press, Providence, Rhode Island.

[2] Noguchi M. (1992), Geometry of statistical manifolds, *Diff. Geom. Appl.*, 2, 197-222.

[3] Ivanova R. (2010), A geometric observation on four statistical parameter spaces, *Tensor, N.S.*, 72, 188-195.

Binary information geometry: some theory and applications

Radka Sabolová, Germain Van Bever, Frank Critchley
The Open University
Walton Hall, United Kingdom
e-mail: [radka.sabolova, germain.van-bever, f.critchley]@open.ac.uk

Paul Marriott
University of Waterloo
Ontario, Canada
e-mail: pmarriott@uwaterloo.ca

Binary information geometry (BIG) is a particular case of computational information geometry (CIG) which provides and exploits a universal space of models for binary random vectors, an exponentially high-dimensional extended multinomial family. Overall, BIG finds natural and fruitful application in a range of important areas, including notably: binary graphical models, logistic regression and Boltzmann machines. A variety of results are presented and illustrated by examples.

On Curvature Computations of Multivariate Gaussian Distribution

Fatma Muazzez Şimşir *
Faculty of Arts and Sciences, Department of Mathematics
Hitit University
Ulukavak Mahallesi, 19040, Çorum, TURKEY
e-mail: fmuazzezsimsir@hitit.edu.tr

Recep Özkan
Department of Mathematics
Middle East Technical University
Üniversiteler Mahallesi, Dumlupınar Bulvarı No:1, 06800, Ankara, TURKEY
e-mail: recep.ozkan@metu.edu.tr

The geometry of the multivariate Gaussian distribution has been widely used in literature, [2, 3]. Considering Gaussian distribution as a Riemannian manifold equipped with the Fisher information metric one may compute curvatures of the induced geometry and use it for the purposes of inference in its usual mean and covariance parameters. On the other hand, the Gaussian distribution belongs to the exponential family, and indeed exponential family can be naturally regarded as a Kähler affine manifold with a pair of dually flat affine connections that play an important role in geometric theory of statistical inference. For more information about Kähler affine manifolds and their relation to information geometry apply to [1]. Ones you equip multivariate Gaussian distribution with this structure then affine curvature takes an important role because it carries more information compared to the Riemannian curvature. One can find a detailed discussion about the affine curvature in Shima [4]. It is remarkable that the Riemannian curvature of a Kähler affine metric depends only on the derivatives of the potential function to order at most three, whereas one would expect fourth derivatives of it to appear. Duistermaat gives some explanation for this phenomenon [5]. This property of the Fisher information metric allows us to avoid prolix curvature computations in case of multivariate Gaussian distributions. On contrary to Riemannian curvature, affine curvature does not have this property which make its computation lengthy.

In this work, we develop a simple method to compute different curvatures of multivariate Gaussian distribution and illustrate it for the bivariate case. Then, we

*The authors are supported by TÜBİTAK 1001-113F296

discuss the importance of affine curvature.

Keywords: Fisher information metric, curvature, Gaussian distributions, affine structure.

References

[1] J. Jost, F. M. Şimşir (2009), *Affine harmonic maps*, Analysis, **29**, 185–197.

[2] S. I. Amari, H. Nagaoka (2000), *Methods of information geometry*, Transl.Math. Monogr. 191, AMS & Oxford Univ. Press.

[3] L. T. Skovgaard (1984), *A Riemannian Geometry of the Multivariate Normal Model*, Scandinavian Journal of Statistics, **11(4)**, 211-223.

[4] H. G. Shima (2007), *The Geometry of Hessian Structures*, World Scientific.

[5] J. J. Duistermaat (2001), *On Hessian Riemannian structures*, The Asian Journal of Mathematics, **5(1)**, 79-91.

Parameter Estimation with Deformed Bregman Divergence

Takashi Takenouchi
Faculty of Systems Information Science, Department of Complex and Intelligent Systems
Future University Hakodate
116-2 Kamedanakano, Hakodate, Hokkaido 041-8655, Japan
e-mail: ttakashi@fun.ac.jp

Probabilistic models on discrete space are useful and parameter estimation of probabilistic models on discrete space is a popular and important issue. For example, the restricted Boltzmann machine (RBM) attracts increasing attention in the context of Deep learning [1]. The Maximum Likelihood Estimation (MLE) is popular method for parameter estimation, but constructions for probabilistic models on the discrete space are often difficult because of the normalization constant which sometimes requires exponential order computation. To avoid the problem, various kinds of methods have been proposed. The contrastive divergence [2] avoids the exponential order calculation using the Markov Chain Monte Carlo (MCMC) sampling. The score matching method [3] and the proper local scoring rules [4] utilize information of "neighbor" and estimate parameter without calculation of the normalization constant. [5] avoids the calculation of normalization constant by employing homogeneous divergence and a technique of empirical localization for unnormalized model.

In this paper, we focus on a deformed Bregman divergence [6] to estimate parameters of probabilistic models on discrete space. By combining the deformed Bregman divergence and the technique of the empirical localization, we propose an estimator, which can be constructed without calculation of the normalization constant and is asymptotically efficient as the MLE. Some experiments show that the proposed estimator attains comparable performance to the MLE with drastically lower computational cost.

Keywords: Fisher efficiency, Bregman divergence, Normalization constant

References

[1] G. E Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.

[2] G. E. Hinton. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[3] A. Hyvärinen. Some extensions of score matching. *Computational statistics & data analysis*, 51(5):2499–2512, 2007.

[4] A. P. Dawid, S. Lauritzen, and M. Parry. Proper local scoring rules on discrete sample spaces. *The Annals of Statistics*, 40(1):593–608, 2012.

[5] T. Takenouchi and T. Kanamori. Empirical localization of homogeneous divergences on discrete sample spaces. In *Advances in Neural Information Processing Systems*, pages 820–828, 2015.

[6] N. Murata, T. Takenouchi, T. Kanamori, and S. Eguchi. Information geometry of U-boost and bregman divergence. *Neural Computation*, 16(7):1437–1481, 2004.

Information Geometry in Population Genetics

Tat Dat Tran (joint work with Jürgen Jost and Julian Hofrichter)
Max Planck Institute for Mathematics in the Sciences
Inselstrasse 22, 04103-Leipzig
e-mail: trandat@mis.mpg.de

In this talk/poster, I will give out a brief overview of beautiful mathematical structures in population genetics by using information geometry techniques.

Keywords: information geometry, mathematical population genetics, random genetic drift, selection, mutation, recombination

References

[1] Akin, E. (1979), *The geometry of population genetics*, Lecture Notes in Biomathematics, Springer-Verlag, Berlin-New York, Volume **31**.

[2] Amari, S. (1985), Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics, Springer, Berlin., Volume **28**.

[3] Amari S. (2001), Information geometry on hierarchy of probability distributions, *IEEE Transactions on Information Theory*, Vol. **47**, No. **5**, 1701-1711.

[4] Antonelli P. L., Strobeck C. (1977), The Geometry of random drift I. Stochastic Distance and Diffusion, *Adv. Appl. Probab.*, **9**(2):238–249.

[5] Antonelli P. L., Morgan K., Lathrop G. M. (1977), The geometry of random drift III. Recombination and diffusion, *Adv. Appl. Probab.*, **9**(2):260–267.

[6] Shahshahani, S. (1979), A new mathematical framework for the study of linkage and selection, *Mem. Amer. Math. Soc.*, **17**.

[7] Tran T.D. (2012) *Information Geometry and the Wright-Fisher model of mathematical population genetics*, PhD Thesis.

[8] Hofrichter J., Jost J., Tran T.D. (2016) *Information geometry and population genetics*, submitted for Lecture Notes in Mathematical Biology.

Simulating the effect of Riemannian curvature on statistical inference using Brownian motion

Nicholas Wisniewski
Department of Medicine
University of California, Los Angeles
10833 Le Conte Ave, A2-237 CHS, Los Angeles, CA 90095-1679
e-mail: nw@ucla.edu

A number of information geometric distance measures, when used as test statistics, are known to asymptotically follow a χ^2 distribution. Examples of such statistics are the likelihood ratio, Wald statistic, Kullback-Leibler divergence, and geodesic distance. The asymptotic distribution is derived using Euclidean geometry, since as sample size $n \to \infty$, deviations are confined to the tangent plane of the hypothesis being tested. However, their non-asymptotic properties are much less understood; as deviations become large, the effect of the Riemannian curvature becomes apparent, and exact calculation becomes prohibitively difficult.

We investigated the effect of curvature on statistical inference in the family of normal distributions by numerically simulating Brownian motion as a generalization of the central limit theorem. Brownian motion on Riemannian manifolds differs from Euclidean diffusion due to Christoffel forces that arise from the curvature. Because the family of normal distributions forms a manifold of constant negative curvature, Brownian motion accelerates at farther distances, leading to a thicker tail than the χ^2 distribution. Additionally, because the Riemannian curvature of Gaussian models becomes more negative with higher dimensionality, this effect increases as more parameters are being estimated simultaneously.

The result shows how curvature effects can lead to significant deviations from asymptotic theory, which comparatively underestimates the potential for large fluctuations, and therefore overestimates statistical significance in hypothesis testing. To illustrate the effect on different test statistics, we computed the empirical distribution of several infomation geometric distance measures commonly used in diffusion tensor imaging, using simulated Brownian motion on Gaussian manifolds of increasing dimension. We show the least distortion is experienced by the geodesic and log-Euclidean distances, followed by J-divergence, and lastly the Frobenius norm.

Keywords: Brownian motion, Riemannian curvature, diffusion tensor

List of Authors