

Nihat Ay · Paolo Gibilisco
František Matúš *Editors*

Information Geometry and Its Applications

IGAIA IV, Liblice, Czech Republic,
June 2016

Springer Proceedings in Mathematics & Statistics

Volume 252

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at <http://www.springer.com/series/10533>

Nihat Ay · Paolo Gibilisco
František Matúš
Editors

Information Geometry and Its Applications

On the Occasion of
Shun-ichi Amari's 80th Birthday,
IGAIA IV, Liblice, Czech Republic, June 2016



Springer

Editors

Nihat Ay

Information Theory of Cognitive Systems
MPI for Mathematics in the Sciences
Leipzig, Germany

and

Santa Fe Institute
Santa Fe, NM, USA

Paolo Gibilisco

Department of Economics and Finance
University of Rome “Tor Vergata”
Rome, Italy

František Matúš

Institute of Information Theory and
Automation
Academy of Sciences of the Czech Republic
Prague, Czech Republic

ISSN 2194-1009

ISSN 2194-1017 (electronic)

Springer Proceedings in Mathematics & Statistics

ISBN 978-3-319-97797-3

ISBN 978-3-319-97798-0 (eBook)

<https://doi.org/10.1007/978-3-319-97798-0>

Library of Congress Control Number: 2018950804

Mathematics Subject Classification (2010): 60A10, 62B05, 62B10, 62G05, 53B21, 53B05, 46B20, 94A15, 94A17, 94B27

© Springer Nature Switzerland AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland



Shun-ichi Amari

Contents

Part I Applications of Information Geometry

Geometry of Information Integration	3
Shun-ichi Amari, Naotsugu Tsuchiya and Masafumi Oizumi	
Information Geometry and Game Theory	19
Jürgen Jost, Nils Bertschinger, Eckehard Olbrich and David Wolpert	
Higher Order Equivalence of Bayes Cross Validation and WAIC	47
Sumio Watanabe	
Restricted Boltzmann Machines: Introduction and Review	75
Guido Montúfar	

Part II Infinite-Dimensional Information Geometry

Information Geometry of the Gaussian Space	119
Giovanni Pistone	
Congruent Families and Invariant Tensors	157
Lorenz Schwachhöfer, Nihat Ay, Jürgen Jost and Hông Vân Lê	
Nonlinear Filtering and Information Geometry: A Hilbert Manifold Approach	189
Nigel J. Newton	
Infinite-Dimensional Log-Determinant Divergences III: Log-Euclidean and Log-Hilbert–Schmidt Divergences	209
Hà Quang Minh	

Part III Theoretical Aspects of Information Geometry

Entropy on Spin Factors	247
Peter Harremoës	

Information Geometry Associated with Generalized Means	279
Shinto Eguchi, Osamu Komori and Atsumi Ohara	
Information Geometry with (Para-)Kähler Structures	297
Jun Zhang and Teng Fei	
Doubly Autoparallel Structure on the Probability Simplex	323
Atsumi Ohara and Hideyuki Ishi	
Complementing Chentsov's Characterization	335
Akio Fujiwara	
Constant Curvature Connections On Statistical Models	349
Alexander Rylov	
Relation Between the Kantorovich–Wasserstein Metric and the Kullback–Leibler Divergence	363
Roman V. Belavkin	
Part IV Quantum Information Geometry	
Some Inequalities for Wigner–Yanase Skew Information	377
Shunlong Luo and Yuan Sun	
Information Geometry of Quantum Resources	399
Davide Girolami	
Characterising Two-Sided Quantum Correlations Beyond Entanglement via Metric-Adjusted f-Correlations	411
Marco Cianciaruso, Irénée Frérot, Tommaso Tufarelli and Gerardo Adesso	
The Effects of Random Qubit-Qubit Quantum Channels to Entropy Gain, Fidelity and Trace Distance	431
Attila Andai	
Robertson-Type Uncertainty Principles and Generalized Symmetric and Antisymmetric Covariances	445
Attila Lovas	

Introduction

This volume contains contributions from the participants of the fourth international conference on Information Geometry and Its Applications (IGAIA IV) which took place at the Liblice Castle in Czech Republic, a beautiful and ideal location for this very special event. Indeed, in the list of the IGAIA conferences (Pescara 2002, Tokyo 2005, Leipzig 2010) this event was organised in order to honour the numerous scientific achievements of Shun-ichi Amari on the occasion of his 80th birthday. Amari has pioneered the field of information geometry (IG) and contributed to a variety of its applications. Moreover, he is the one who found the perfect and powerfully evocative name for this branch of mathematics.

The aim of the IGAIA series was, from the very beginning, to bring together researchers working in the many different branches of IG, no matter if purely mathematical or applied, with the explicit goal to create unexpected links within the IG community. IGAIA IV was no exception to this rule, which is reflected by the variety of the contributions that the reader will find in this volume. As any mature mathematical field, IG has its own internal and purely mathematical problems. But it still remains very much linked to the applications in statistics, information theory, neurosciences, machine learning, and quantum physics, just to mention some areas. Correspondingly, we have in this volume two application-oriented parts (I and IV), and two parts of more general nature (II and III).

The conference was 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 certainly want to thank all the participants for their contribution to create such a beautiful and stimulating atmosphere during the conference. Furthermore, we are grateful for the terrific work with the local organisation of the conference

made by Václav Kratochvíl and the support that we received from Milan Studený. Special thanks go to Antje Vandenberg who managed all the administrative works in Leipzig.

July 2018

Nihat Ay
Paolo Gibilisco
František Matúš

Post Scriptum. While finalising the editorial work on this volume, our dear colleague and friend František Matúš passed away. There are no words to express this great loss of an exceptional person and a highly respected scientist. Our thoughts are with his family. We will miss him and dedicate this book to his memory.

Nihat Ay
Paolo Gibilisco

Part I

Applications of Information Geometry

Geometry of Information Integration



Shun-ichi Amari, Naotsugu Tsuchiya and Masafumi Oizumi

Abstract Information geometry is used to quantify the amount of information integration within multiple terminals of a causal dynamical system. Integrated information quantifies how much information is lost when a system is split into parts and information transmission between the parts is removed. Multiple measures have been proposed as a measure of integrated information. Here, we analyze four of these measures and elucidate their relations from the viewpoint of information geometry. Two of them use dually flat manifolds and the other two use curved manifolds to define a split model. We show that there are hierarchical structures among the measures. We provide explicit expressions of these measures.

Keywords Integrated information theory · Information geometry
Kullback-Leibler divergence · Mismatched decoding · Consciousness

1 Introduction

It is an interesting problem to quantify how much information is integrated in a multi-terminal causal system. The concept of information integration was introduced

S. Amari (✉) · M. Oizumi
RIKEN Brain Science Institute, Tokyo, Japan
e-mail: amari@brain.riken.jp

S. Amari · M. Oizumi
Araya Inc., Tokyo, Japan

N. Tsuchiya
School of Psychological Sciences, Monash University, Clayton, Australia

N. Tsuchiya
Monash Institute of Cognitive and Clinical Neurosciences, Monash University,
Clayton, Australia

by Tononi and colleagues in Integrated Information Theory (IIT), which attempts to quantify the levels and contents of consciousness [1–3]. Inspired by Tononi’s idea, many variants of integrated information have been proposed [4–7]. Independent from the perspective of IIT, Ay derived the measure proposed in [4] to quantify complexity in a system [8, 9].

In this paper, we use information geometry [10] to clarify the nature of various measures of integrated information as well as the relations among them. Consider a joint probability distribution $p(\mathbf{x}, \mathbf{y})$ of sender X and receiver Y , where \mathbf{x} and \mathbf{y} are vectors consisting of n components, denoting actual values of X and Y . Here, \mathbf{y} is stochastically generated depending on \mathbf{x} . That is, information is sent from the sender X to the receiver Y . We consider a Markov model, where $\mathbf{x}_{t+1} (= \mathbf{y})$ is generated from $\mathbf{x}_t (= \mathbf{x})$ stochastically by transition probability matrix $p(\mathbf{x}_{t+1} | \mathbf{x}_t)$. In this way, we quantify how much information is integrated within a system through one step of state transition.

To quantify the amount of integrated information, we need to consider a split version of the system in which information transmission between different elements are removed, so that we can compare the original joint probability with the split one. The joint probability distribution of a split model is denoted by $q(\mathbf{x}, \mathbf{y})$. We define the amount of information integration by the minimized Kullback–Leibler (KL) divergence between the original distribution $p(\mathbf{x}, \mathbf{y})$ and the split distribution $q(\mathbf{x}, \mathbf{y})$,

$$\Phi = \min_q D_{KL} [p(\mathbf{x}, \mathbf{y}) : q(\mathbf{x}, \mathbf{y})], \quad (1)$$

which quantifies to what extent $p(\mathbf{x}, \mathbf{y})$ and $q(\mathbf{x}, \mathbf{y})$ are different. Minimizing KL-divergence means selecting the best approximation of the original distribution $p(\mathbf{x}, \mathbf{y})$ among the split distributions $q(\mathbf{x}, \mathbf{y})$.

We need to search for a reasonable split model. For each distinct version of split models, corresponding measure of integrated information can be derived [4, 6–9]. The present paper studies four reasonable split models and the respective measures of integrated information. Among the four integrated information, Φ_G , the geometric Φ defined in [7], is what we believe the most reasonable measure for information integration in a sense that it purely quantifies causal influences between parts, although the others have their own meanings and useful characteristics. See a recent paper [11] for the detailed comparisons of various measures of integrated information by using the randomly connected Boltzmann machine.

2 Markovian Dynamical Systems

We consider a Markovian dynamical system

$$\mathbf{x}_{t+1} = \mathcal{T}(\mathbf{x}_t) \quad (2)$$

where \mathbf{x}_t and \mathbf{x}_{t+1} are the states of the system at time t and at the next time step $t + 1$, respectively, which are vectors consisting of n elements. \mathcal{T} is a stochastic state transition operator, which symbolically denotes the stochastic generation of the next state \mathbf{x}_{t+1} from the current state \mathbf{x}_t and is represented by the conditional probability distribution of \mathbf{x}_{t+1} given \mathbf{x}_t , $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$. $\mathcal{T} = \{p(\mathbf{x}_{t+1}|\mathbf{x}_t)\}$ is called a transition probability matrix. Throughout this paper, we will use \mathbf{x} for \mathbf{x}_t and \mathbf{y} for \mathbf{x}_{t+1} for the ease of notation.

Given the probability distribution of \mathbf{x} at time t , $p(\mathbf{x})$, the probability distribution of the next state, $p(\mathbf{y})$, is given by

$$p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}), \quad (3)$$

and the joint probability distribution is given by

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}). \quad (4)$$

Throughout the paper, we will use $p(\mathbf{x})$ and $p(\mathbf{y})$ instead of $p_X(\mathbf{x})$ and $p_Y(\mathbf{y})$, which explicitly denote X and Y .

The state \mathbf{x} is supported by n terminals and information at terminals x_1, x_2, \dots, x_n are integrated to give information in the next state $\mathbf{y} = (y_1, \dots, y_n)$, so that each y_i stochastically depends on all of x_1, \dots, x_n . We quantify how much information is integrated among different terminals through state transition. All such information is contained in the form of the joint probability distribution $p(\mathbf{x}, \mathbf{y})$. We use a general model \mathcal{M} to represent $p(\mathbf{x}, \mathbf{y})$, called a full model, which is a graphical model where all the terminals of sender X and receiver Y are fully connected. We consider the discrete case, in particular the binary case, in which x_i and y_i are binary taking values of 0 or 1, although generalization to other cases (e.g., continuous, more discretization steps than binary) is not difficult. We also study the case where random continuous variables are subject to Gaussian distributions.

In order to quantify the amount of information integration, we consider a “split model” \mathcal{M}_S , where information transmission from one terminal x_i to the other terminals y_j ($j \neq i$) is removed. Let $q(\mathbf{x}, \mathbf{y})$ be the joint probability distribution of \mathbf{x} and \mathbf{y} in a split model. The amount of information integration in $p(\mathbf{x}, \mathbf{y})$ is measured by the KL-divergence from $p(\mathbf{x}, \mathbf{y})$ to \mathcal{M}_S , that is, the KL-divergence from $p(\mathbf{x}, \mathbf{y})$ to $q^*(\mathbf{x}, \mathbf{y})$, which is a particular instantiation of the split model and is the one that is closest to $p(\mathbf{x}, \mathbf{y})$. Integrated information is defined as the minimized KL-divergence between the full model p and the split model q [7],

$$\begin{aligned} \Phi &= \min_{q \in \mathcal{M}_S} D_{KL} [p(\mathbf{x}, \mathbf{y}) : q(\mathbf{x}, \mathbf{y})], \\ &= D_{KL} [p(\mathbf{x}, \mathbf{y}) : q^*(\mathbf{x}, \mathbf{y})]. \end{aligned}$$

Depending on various definitions of “split” model \mathcal{M}_S , different measures of integrated information can be defined. Below, we elucidate the nature of four candidates of integrated information and their relations.

3 Stochastic Models of Causal Systems

3.1 Full Model

A full model \mathcal{M} , $p(\mathbf{x}, \mathbf{y})$, is a graphical model in which all the nodes (terminals) are connected (Fig. 1). We consider the binary case. In that case, $p(\mathbf{x}, \mathbf{y})$ is an element of an exponential family and can be expanded as

$$p(\mathbf{x}, \mathbf{y}) = \exp \left\{ \sum \theta_i^X x_i + \sum \theta_j^Y y_j + \sum \theta_{ij}^{XX} x_i x_j + \sum \theta_{ij}^{YY} y_i y_j + \sum \theta_{ij}^{XY} x_i y_j + h(\mathbf{x}, \mathbf{y}) - \psi \right\}, \quad (5)$$

where we show linear and quadratic terms explicitly by using parameters $\theta_i^X, \theta_j^Y, \theta_{ij}^{XX}, \theta_{ij}^{YY}, \theta_{ij}^{XY}$. $h(\mathbf{x}, \mathbf{y})$ is the higher order terms of \mathbf{x} and \mathbf{y} and the last term ψ is the free energy term (or cumulant generating function) corresponding to the normalizing factor. The set of distributions in the full model form a dually flat statistical manifold [10].

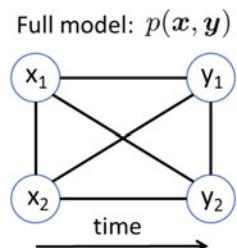
We hereafter neglect higher-order terms in order to make discussions simple. This means that the full model is a Boltzmann machine described by parameters

$$\boldsymbol{\theta} = (\theta_i^X, \theta_j^Y, \theta_{ij}^{XX}, \theta_{ij}^{YY}, \theta_{ij}^{XY}). \quad (6)$$

They form an e -coordinate system to specify a distribution $p(\mathbf{x}, \mathbf{y})$. The dual coordinate system, m -coordinate system, is denoted by $\boldsymbol{\eta}$,

$$\boldsymbol{\eta} = (\eta_i^X, \eta_j^Y, \eta_{ij}^{XX}, \eta_{ij}^{YY}, \eta_{ij}^{XY}). \quad (7)$$

Fig. 1 Full model $p(\mathbf{x}, \mathbf{y})$



The components of η are expectations of corresponding random variables. For example,

$$\eta_{ij}^{XX} = E[x_i x_j], \quad (8)$$

$$\eta_{ij}^{XY} = E[x_i y_j], \quad (9)$$

where E is the expectation. In the followings, we consider the case where the number of elements is 2 ($n = 2$) for the explanatory purpose, but generalization for larger n is straightforward.

3.2 Fully Split Model

Ay considered a split model from the viewpoint of complexity of a system [8, 9]. The split model $q(\mathbf{y}|\mathbf{x})$ is given by

$$q(\mathbf{y}|\mathbf{x}) = \prod_i q(y_i|x_i), \quad (10)$$

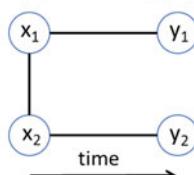
where the conditional probability distribution of the whole system $q(\mathbf{y}|\mathbf{x})$ is fully split into that of each part. We call this model “fully split model” \mathcal{M}_{FS} . The corresponding measure Φ_{FS} was also introduced by Barrett and Seth [4] following the measure of integrated information proposed by Balduzzi and Tononi [2].

This split model deletes branches connecting X_i and Y_j ($i \neq j$) and also deletes the branches connecting different Y_i and Y_j ($i \neq j$). (Here, we use capital letters X and Y to emphasize random variables, not their values.) This split model is reasonable because when terminals Y_i are split, all the branches connecting Y_i and the other nodes should be deleted except for branches connecting X_i and Y_i . Branches connecting X_i and X_j remain as they are (Fig. 2). However, even though branches connecting Y_i and Y_j are deleted, this does not imply that Y_i and Y_j ($i \neq j$) are independent, because when input X_i and X_j are correlated, Y_i and Y_j are also correlated even though no branches exist connecting X_i and Y_j and Y_i and Y_j .

When $n = 2$, the random variables X_i and Y_j have a Markovian structure,

Fig. 2 Fully split model
 $q(\mathbf{x}, \mathbf{y})$

Fully split model: $q(\mathbf{x}, \mathbf{y})$



$$Y_1 - X_1 - X_2 - Y_2, \quad (11)$$

so that Y_1 and Y_2 are conditionally independent when (X_1, X_2) is fixed. Also X_2 and Y_1 (or X_1 and Y_2) are conditionally independent when X_1 (or X_2) are fixed. These constraints correspond to putting

$$\theta_{12}^{XY} = \theta_{21}^{XY} = \theta_{12}^{YY} = 0 \quad (12)$$

in the $\boldsymbol{\theta}$ -coordinates. They are linear constraints in the $\boldsymbol{\theta}$ -coordinates. Thus, the fully split model \mathcal{M}_{FS} is an exponential family. It is an e -flat submanifold of \mathcal{M} . Given $p(\mathbf{x}, \mathbf{y}) \in \mathcal{M}$, let $q^*(\mathbf{x}, \mathbf{y})$ be the m -projection of p to \mathcal{M}_{FS} . Then, $q^*(\mathbf{x}, \mathbf{y})$ is given by the minimizer of KL-divergence,

$$q^*(\mathbf{x}, \mathbf{y}) = \arg \min_{q(\mathbf{x}, \mathbf{y}) \in \mathcal{M}_{FS}} D_{KL}[p(\mathbf{x}, \mathbf{y}) : q(\mathbf{x}, \mathbf{y})]. \quad (13)$$

We use the mixed coordinate system of \mathcal{M} ,

$$\boldsymbol{\xi} = (\eta_i^X, \eta_j^Y, \eta_{ij}^{XX}, \eta_{ij}^{YY}, \eta_{11}^{XY}, \eta_{22}^{XY}; \theta_{12}^{XY}, \theta_{21}^{XY}, \theta_{12}^{YY}). \quad (14)$$

Then \mathcal{M}_{FS} is specified by (12).

We use the Pythagorean theorem in \mathcal{M} ([10, 12]): When the m -geodesic connecting p and q is orthogonal to the e -geodesic connecting q and r ,

$$D_{KL}[p : r] = D_{KL}[p : q] + D_{KL}[q : r]. \quad (15)$$

For $p \in \mathcal{M}$, when q and r belong to \mathcal{M}_{FS} , the minimizer of $D_{KL}[p : r]$ is q^* , which is the m -projection of p to \mathcal{M}_{FS} . q^* is explicitly given by

$$\boldsymbol{\xi}^* = (\eta_i^X, \eta_j^Y, \eta_{ij}^{XX}, \eta_{ij}^{YY}, \eta_{11}^{XY}, \eta_{22}^{XY}; 0) \quad (16)$$

in the $\boldsymbol{\xi}$ -coordinate system, where $\boldsymbol{\eta}$ -part is the same as that of the mixed coordinates of $p(\mathbf{x}, \mathbf{y})$.

By simple calculations, we obtain

$$q^*(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}) p(y_1|x_1) p(y_2|x_2), \quad (17)$$

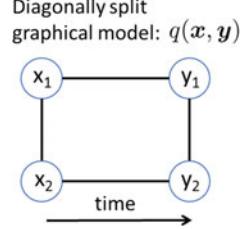
which means

$$q^*(\mathbf{x}) = p(\mathbf{x}), \quad (18)$$

$$q^*(\mathbf{y}|\mathbf{x}) = \prod p(y_i|x_i). \quad (19)$$

The corresponding measure of integrated information is given by

Fig. 3 Diagonally split graphical model $q(\mathbf{x}, \mathbf{y})$



$$\Phi_{FS} = \sum H[Y_i|X_i] - H[Y|X], \quad (20)$$

where $H[Y_i|X_i]$ and $H[Y|X]$ are the conditional entropies corresponding to the random variables. This measure was termed “stochastic interaction” by Ay [8, 9].

While Φ_{FS} is straightforward in derivation and its concept, it has an undesirable property as a measure of integrated information. Specifically, as we proposed in [6, 7], any measure of integrated information Φ , is expected to satisfy the following constraint,

$$0 \leq \Phi \leq I(X; Y), \quad (21)$$

where $I(X; Y)$ is the mutual information between X and Y . This requirement is natural because Φ should quantify the “loss of information” caused by splitting a system into parts, i.e., removing information transmission between parts. The loss of information should not exceed the total amount of information in the whole system, $I(X; Y)$, and should be always positive or 0. Φ should be 0 only when X and Y are independent. However, Φ_{FS} does not satisfy the requirement of the upper bound, as was pointed by [6, 7]. This is because \mathcal{M}_{FS} does not include the submanifold \mathcal{M}_I consisting of the independent distributions of X and Y ,

$$\mathcal{M}_I = \{q(\mathbf{x})q(\mathbf{y})\}. \quad (22)$$

\mathcal{M}_I is characterized by

$$\theta_{ij}^{XY} = 0 \quad (\text{for } \forall i, j). \quad (23)$$

It is an e -flat submanifold of \mathcal{M} . The minimized KL-divergence between $p(\mathbf{x}, \mathbf{y})$ and \mathcal{M}_I is mutual information,

$$I(X; Y) = \min_{q \in \mathcal{M}_I} D_{KL}(p(\mathbf{x}, \mathbf{y}) : q(\mathbf{x}, \mathbf{y})) \quad (24)$$

Thus, while stochastic interaction, derived from the submanifold \mathcal{M}_{FS} , has a simple expression Eq. 20 and nice properties on its own, it may not be an ideal measure of integrated information due to its violation of the upper-bound requirement.

3.3 Diagonally Split Graphical Model

In order to overcome the above difficulties, we consider an undirected graphical model of pairwise interactions in which all the branches connecting x_i and y_j ($i \neq j$) are deleted but all the other branches remain as shown in Fig. 3. We call this model “diagonally split graphical model” \mathcal{M}_{DS} .

The model is defined by

$$\theta_{12}^{XY} = \theta_{21}^{XY} = 0. \quad (25)$$

It is also an e -flat submanifold of \mathcal{M} . The branches connecting different y_i 's exist because $\theta_{ij}^{YY} \neq 0$. The model does not remove direct interactions among y_i 's, which can be caused by correlated noises directly applied to the output nodes (not through causal influences from \mathbf{x}). The fully split model \mathcal{M}_{FS} introduced in the previous section is an e -flat submanifold of \mathcal{M}_{DS} , since $\theta_{ij}^{YY} = 0$ ($i \neq j$) is further required for \mathcal{M}_{FS} .

In the case of $n = 2$, the full model \mathcal{M} is 10-dimensional (excluding higher-order interactions), \mathcal{M}_{FS} is 7-dimensional and \mathcal{M}_{DS} is 8-dimensional. \mathcal{M}_{DS} satisfies the conditions that \mathbf{x}_1 and \mathbf{y}_2 as well as \mathbf{x}_2 and \mathbf{y}_1 are conditionally independent when $(\mathbf{x}_2, \mathbf{y}_1)$ and $(\mathbf{x}_1, \mathbf{y}_2)$ are fixed, respectively. The model is characterized by

$$q(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})g(\mathbf{y}) \prod_i h(x_i, y_i). \quad (26)$$

We use the following mixed coordinates

$$\xi = (\eta_i^X, \eta_j^Y, \eta_{ij}^{XX}, \eta_{ij}^{YY}; \theta_{12}^{XY}, \theta_{21}^{XY}). \quad (27)$$

Then, the m -projection of $p(\mathbf{x}, \mathbf{y})$ to \mathcal{M}_{SG} is given by

$$\xi^* = (\eta_i^X, \eta_j^Y, \eta_{ij}^{XX}, \eta_{ij}^{YY}; 0, 0) \quad (28)$$

in these coordinates. This implies that

$$q^*(\mathbf{x}) = p(\mathbf{x}), \quad (29)$$

$$q^*(\mathbf{y}) = p(\mathbf{y}), \quad (30)$$

$$q^*(y_i|x_i) = p(y_i|x_i), \quad \forall i. \quad (31)$$

The corresponding measure of integrated information is

$$\Phi_{DS} = D_{KL}[p(\mathbf{x}, \mathbf{y}) : q^*(\mathbf{x}, \mathbf{y})]. \quad (32)$$

It satisfies the natural requirement for integrated information Eq. 21. Thus, it resolves the shortcomings of Φ_{FS} .

However, there still remains a problem to take into consideration. To illustrate it, let us consider the two-terminal Gaussian case (autoregressive (AR) model), in which \mathbf{x} is linearly transformed to \mathbf{y} by the connectivity matrix A and the Gaussian noise $\boldsymbol{\epsilon}$ is added,

$$\mathbf{y} = A\mathbf{x} + \boldsymbol{\epsilon}. \quad (33)$$

Here, in the two terminals case, A is given by,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad (34)$$

and $\boldsymbol{\epsilon}$ is zero mean Gaussian noise whose covariance matrix is given by

$$\Sigma(E) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}. \quad (35)$$

Let $\Sigma(X)$ be the covariance matrix of \mathbf{x} . Then, the joint probability distribution is written as

$$p(\mathbf{x}, \mathbf{y}) = \exp \left\{ -\frac{1}{2} (\mathbf{x}^T \Sigma(X)^{-1} \mathbf{x}) + (\mathbf{y} - A\mathbf{x})^T \Sigma(E)^{-1} (\mathbf{y} - A\mathbf{x}) - \psi \right\}, \quad (36)$$

where the means of all random variables are assumed to be equal to 0. The θ -coordinates consist of three matrices,

$$\theta = (\theta_{XX}, \theta_{YY}, \theta_{XY}), \quad (37)$$

$$\theta_{XX} = \Sigma(X)^{-1}, \quad \theta_{YY} = \Sigma(E)^{-1}, \quad \theta_{XY} = -A\Sigma(E)^{-1} \quad (38)$$

and the corresponding η -coordinates are

$$\eta = (\eta_{XX}, \eta_{YY}, \eta_{XY}), \quad (39)$$

$$\eta_{XX} = \Sigma(X), \quad \eta_{YY} = A\Sigma(X)A, \quad \eta_{XY} = A\Sigma(X). \quad (40)$$

We project $p(\mathbf{x}, \mathbf{y})$ Eq. 36 to \mathcal{M}_{DS} . The closest point $q^*(\mathbf{x}, \mathbf{y})$ is again given by an AR model,

$$\mathbf{y} = A^*\mathbf{x} + \boldsymbol{\epsilon}^*. \quad (41)$$

where A^* and the covariance matrix of $\boldsymbol{\epsilon}^*$, $\Sigma(E^*)$, are determined from A , $\Sigma(X)$ and $\Sigma(E)$. However, the off-diagonal elements of A^* is not zero. Therefore, the deletion of the diagonal branches in a graphical model is not equivalent to the deletion of the off-diagonal elements of A in the Gaussian case.

The off-diagonal elements of A , A_{ij} , determines causal influences from x_i to y_j . In the diagonal split model \mathcal{M}_{DS} , the causal influences are non-zero because the off-diagonal elements of A^* , A_{ij}^* , are non-zero. Thus, the corresponding measure

of integrated information Φ_{DS} Eq. 32 does not purely quantify causal influences between the elements. In IIT, integrated information is designed to quantify causal influences [2, 3]. In this sense, it is desirable to have a split model, which results in a diagonal connectivity matrix A .

3.4 Causally Split Model (Geometric Model)

To derive a split model where only causal influences between elements are removed, we consider that the essential part is to remove branches connecting x_i and y_j ($i \neq j$), without destroying other constituents. The minimal requirement to remove the effect of the branch (i, j) is to let x_i and y_j be conditionally independent, when all the other elements are fixed. In our case of $n = 2$, we should have two Markovian conditions

$$X_1 — X_2 — Y_2, \quad (42)$$

$$X_2 — X_1 — Y_1. \quad (43)$$

Note that \mathcal{M}_{DS} is characterized by the two Markovian conditions

$$X_1 — (X_2, Y_1) — Y_2, \quad (44)$$

$$X_2 — (X_1, Y_2) — Y_1. \quad (45)$$

which are different from Eqs. 42 and 43.

The split model that satisfies the above conditions Eqs. 42 and 43 was introduced by Oizumi, Tsuchiya and Amari [7] and was called “geometric model” \mathcal{M}_G , because information geometry was used as a guiding principle to obtain the model. We can also call it “causally split model” because causal influences between elements are removed.

The model \mathcal{M}_G is a 8-dimensional submanifold of \mathcal{M} in the case of $n = 2$, because there are two constraints Eqs. 42 and 43. These constraints are expressed as

$$q(x_1, y_2|x_2) = q(x_1|x_2)q(y_2|x_2), \quad (46)$$

$$q(x_2, y_1|x_1) = q(x_2|x_1)q(y_1|x_1). \quad (47)$$

We can write down the constraints in terms of θ -coordinates, but they are nonlinear. They are also nonlinear in the η -coordinates. Thus, \mathcal{M}_G is a curved submanifold and it is not easy to give an explicit solution of the m -projection of $p(\mathbf{x}, \mathbf{y})$ to \mathcal{M}_G .

We can solve the Gaussian case explicitly [7]. It is not difficult to prove that, when the Markovian conditions in Eqs. 42 and 43 are satisfied, the connectivity matrix A' of an AR model in \mathcal{M}_G ,

$$\mathbf{y} = A'\mathbf{x} + E', \quad (48)$$

is a diagonal matrix. From Eq. 38, we have

$$A' = -\theta_{YY}^{-1}\theta_{XY}. \quad (49)$$

Thus, the constraints in Eqs. 42 and 43 expressed in terms of θ -coordinates are equivalent to the off-diagonal elements of matrix $\theta_{YY}^{-1}\theta_{XY}$ being 0. Thus, the constraints are nonlinear in the θ -coordinates. The corresponding measure of integrated information, Φ_G (geometric integrated information), is given explicitly by

$$\Phi_G = \frac{1}{2} \log \frac{|\Sigma(E')|}{|\Sigma(E)|}, \quad (50)$$

where $\Sigma(E)$ is the noise covariance of $p(\mathbf{x}, \mathbf{y})$, $\Sigma(E')$ is that of projected $q^*(\mathbf{x}, \mathbf{y})$, $|\Sigma(E)|$ is the determinant of $\Sigma(E)$.

By construction, it is easy to see that Φ_G satisfies the requirements for integrated information,

$$0 \leq \Phi_G \leq I(X; Y), \quad (51)$$

because the causally split model \mathcal{M}_G includes the submanifold \mathcal{M}_I consisting of the independent distributions of X and Y Eq. 22. We believe that Φ_G is the best candidate measure in the sense that it is closest to the original philosophy of integrated information in IIT. In IIT, integrated information is designed to quantify causal influences between elements [2, 3]. Note that in IIT, “causal” influences are quantified by Pearl’s intervention framework [2, 13, 14] attempting to quantify the “actual” causation. On the other hand, causal influences quantified in this paper do not necessarily mean actual causation. Φ_G should be considered as an observational measure of causation, similar to Granger causality or Transfer entropy [7].

3.5 Mismatched Decoding Model

Starting from a completely distinct conceptual framework, we can consider another model, called a mismatched decoding model \mathcal{M}_{MD} . We use the concept of mismatched decoding in information theory proposed by Merhav et al. [15]. We have utilized this concept in the context of neuroscience [6, 16–19].

To introduce the decoding perspective, let us consider a situation where we try to estimate the input \mathbf{x} when the output \mathbf{y} is observed. When we know the correct joint probability distribution $p(\mathbf{x}, \mathbf{y})$, we can estimate \mathbf{x} by using the true distribution $p(\mathbf{x}|\mathbf{y})$. This is the optimal matched decoding. However, when we use a split model $q(\mathbf{x}, \mathbf{y})$ for decoding, there is always loss of information. This type of decoding is called mismatched decoding because the decoding model $q(\mathbf{x}, \mathbf{y})$ is different from the actual probability distribution $p(\mathbf{x}, \mathbf{y})$.

We previously considered the fully split model as a mismatched decoding model [6]

$$q(\mathbf{y}|\mathbf{x}) = \prod_i q(y_i|x_i). \quad (52)$$

By using Merhav's framework, the information loss when $q(\mathbf{y}|\mathbf{x})$ is used for decoding can be quantified by [6, 15]

$$\Phi_{MD} = \min_{\beta} D_{KL} [p(\mathbf{x}, \mathbf{y}) || q(\mathbf{x}, \mathbf{y}; \beta)], \quad (53)$$

where

$$q(\mathbf{x}, \mathbf{y}; \beta) = \frac{p(\mathbf{x})p(\mathbf{y}) \prod_i p(y_i|x_i)^\beta}{\sum_{\mathbf{x}'} p(\mathbf{x}') \prod_i p(y_i|x'_i)^\beta}. \quad (54)$$

To quantify the information loss Φ_{MD} , the KL-divergence needs to be minimized with respect to the one-dimensional parameter β . We call $q(\mathbf{x}, \mathbf{y}; \beta)$ a “mismatched decoding model” \mathcal{M}_{MD} . The mismatched decoding model \mathcal{M}_{MD} forms a one-dimensional submanifold. As can be seen in Eq. 54, no interaction terms are included between x_i and y_j ($i \neq j$). Thus, \mathcal{M}_{MD} is included in the diagonally split graphical model \mathcal{M}_{DS} .

The optimal β^* , which minimizes the KL-divergence, is given by projecting $p(\mathbf{x}, \mathbf{y})$ to \mathcal{M}_{MD} . Since \mathcal{M}_{MD} is not an e -flat submanifold, it is difficult to obtain the analytical expression of $q^*(\mathbf{x}, \mathbf{y})$. However, the minimization of KL-divergence is a convex problem and thus, the optimal β^* can be easily found by numerical calculations such as gradient descent [6, 20].

4 Comparison of Various Measures of Integrated Information

We have derived four measures of integrated information from four different definitions of the split model. We elucidate their relations in this section.

First, \mathcal{M}_{FS} and \mathcal{M}_{DS} are e -flat submanifolds, forming exponential families. Therefore, we can directly apply the Pythagorean projection theorem and the projected $q^*(\mathbf{x}, \mathbf{y})$ is explicitly obtained by using the mixed coordinates. However, \mathcal{M}_G and \mathcal{M}_{MD} are curved submanifolds and thus, it is difficult to analytically obtain the projected $q^*(\mathbf{x}, \mathbf{y})$ in general.

The natural requirements for integrated information,

$$0 \leq \Phi \leq I(X, Y), \quad (55)$$

are satisfied for all the measures of integrated information except for \mathcal{M}_{FS} . This is because \mathcal{M}_{FS} does not include \mathcal{M}_I Eq. 22 while the other split models include \mathcal{M}_I .

In general, when $\mathcal{M}_1 \supset \mathcal{M}_2$,

$$\min_{q \in \mathcal{M}_1} D_{KL}[p : q] \leq \min_{q \in \mathcal{M}_2} D_{KL}[p : q] \quad (56)$$

and therefore,

$$\Phi_2 \geq \Phi_1. \quad (57)$$

We have proved

$$\mathcal{M}_{DS} \supset \mathcal{M}_{FS}, \quad \mathcal{M}_{DS} \supset \mathcal{M}_{MD}, \quad (58)$$

$$\mathcal{M}_G \supset \mathcal{M}_{FS}. \quad (59)$$

From these relations between the split models, we have the relations between the corresponding measures of integrated information,

$$\Phi_{FS} \geq \Phi_{DS}, \quad \Phi_{MD} \geq \Phi_{DS}, \quad \Phi_{FS} \geq \Phi_G. \quad (60)$$

\mathcal{M}_{FS} is included in the intersection of \mathcal{M}_{DS} and \mathcal{M}_G . \mathcal{M}_I is included in \mathcal{M}_{DS} , \mathcal{M}_G , and \mathcal{M}_{MD} .

The relations among four different measures of integrated information are schematically summarized in Fig. 4.

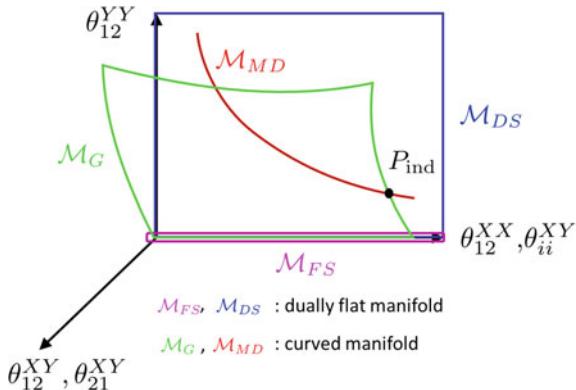


Fig. 4 Relations among four different split models. Fully split model \mathcal{M}_{FS} and diagonally split graphical model \mathcal{M}_{DS} are dually flat manifolds. \mathcal{M}_{FS} is represented by a magenta line on the axis of $\theta_{12}^{XX}, \theta_{ii}^{XY}$. \mathcal{M}_{DS} is represented by a blue square spanned by the two axes $\theta_{12}^{XX}, \theta_{ii}^{XY}$ and θ_{12}^{YY} . Causally split model (geometric model) \mathcal{M}_G and mismatched decoding model \mathcal{M}_{MD} are curved manifolds. \mathcal{M}_G is represented by a curved green surface. \mathcal{M}_{MD} is represented by a curved red line inside the surface of \mathcal{M}_{DS} . P_{ind} is an independent distribution of x and y , $P_{\text{ind}} = p(x)p(y)$, which is represented by a black point. P_{ind} is included in \mathcal{M}_{DS} , \mathcal{M}_{MD} , and \mathcal{M}_G but is not included in \mathcal{M}_{FS}

5 Conclusions

We studied four different measures of integrated information in a causal stochastic dynamical system from the unified viewpoint of information geometry. The four measures have their own meanings and characteristics. We elucidated their relations and a hierarchical structure of the measures (Fig. 4). We can define a measure of information transfer for each branch, but their effects are not additive but subadditive. Therefore, we need to study further collective behaviors of deleting branches [21]. This remains a future problem to be studied.

References

1. Tononi, G.: An information integration theory of consciousness. *BMC Neurosci.* **5**, 42 (2004). <https://doi.org/10.1186/1471-2202-5-42>
2. Balduzzi, D., Tononi, G.: Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Comput. Biol.* **4**(6), e1000091 (2008). <https://doi.org/10.1371/journal.pcbi.1000091>
3. Oizumi, M., Albantakis, L., Tononi, G.: From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput. Biol.* **10**(5), e1003588 (2014). <https://doi.org/10.1371/journal.pcbi.1003588>
4. Barrett, A.B., Barnett, L., Seth, A.K.: Multivariate Granger causality and generalized variance. *Phys. Rev. E* **81**(4), 041907 (2010). <https://doi.org/10.1103/PhysRevE.81.041907>
5. Tegmark, M.: Improved measures of integrated information. *PLoS Comput. Biol.* **12**(11), e1005123 (2016)
6. Oizumi, M., Amari, S., Yanagawa, T., Fujii, N., Tsuchiya, N.: Measuring integrated information from the decoding perspective. *PLoS Comput. Biol.* **12**(1), e1004654 (2016). <https://doi.org/10.1371/journal.pcbi.1004654>
7. Oizumi, M., Tsuchiya, N., Amari, S.: Unified framework for information integration based on information geometry. *Proc. Natl. Acad. Sci.* **113**(51), 14817–14822 (2016)
8. Ay, N.: Information geometry on complexity and stochastic interaction. *MPI MIS PREPRINT* 95 (2001)
9. Ay, N.: Information geometry on complexity and stochastic interaction. *Entropy* **17**(4), 2432–2458 (2015). <https://doi.org/10.3390/e17042432>
10. Amari, S.: *Information Geometry and Its Applications*. Springer, Berlin (2016)
11. Kanwal, M.S., Grochow, J.A., Ay, N.: Comparing information-theoretic measures of complexity in Boltzmann machines. *Entropy* **19**(7), 310–325 (2017)
12. Ay, N., Jost, J., Ván Lê, H., Schwachhöfer, L.: *Information Geometry*, vol. 64. Springer, Berlin (2017)
13. Pearl, J.: *Causality*. Cambridge University Press, Cambridge (2009)
14. Ay, N., Polani, D.: Information flows in causal networks. *Adv. Complex Syst.* **11**(01), 17–41 (2008)
15. Merhav, N., Kaplan, G., Lapidoth, A., Shitz, S.S.: On information rates for mismatched decoders. *IEEE Trans. Inf. Theory* **40**(6), 1953–1967 (1994)
16. Oizumi, M., Ishii, T., Ishibashi, K., Hosoya, T., Okada, M.: Mismatched decoding in the brain. *J. Neurosci.* **30**(13), 4815–4826 (2010)
17. Oizumi, M., Okada, M., Amari, S.: Information loss associated with imperfect observation and mismatched decoding. *Front. Comput. Neurosci.* **5** (2011)
18. Boly, M., Sasai, S., Gosseries, O., Oizumi, M., Casali, A., Massimini, M., et al.: Stimulus set meaningfulness and neurophysiological differentiation: a functional magnetic resonance imaging study. *PLoS One* **10**(5), e0125337 (2015). <https://doi.org/10.1371/journal.pone.0125337>

19. Haun, A.M., Oizumi, M., Kovach, C.K., Kawasaki, H., Oya, H., Howard, M.A., et al.: Conscious perception as integrated information patterns in human electrocorticography. *eNeuro* **4**(5), (2017). ENEURO-0085
20. Latham, P.E., Nirenberg, S.: Synergy, redundancy, and independence in population codes, revisited. *J. Neurosci.* **25**(21), 5195–5206 (2005)
21. Jost, J., Bertschinger, N., Olbrich, E., Ay, N., Frankel, S.: An information theoretic approach to system differentiation on the basis of statistical dependencies between subsystems. *Phys. A: Stat. Mech. Appl.* **378**(1), 1–10 (2007)

Information Geometry and Game Theory



Jürgen Jost, Nils Bertschinger, Eckehard Olbrich and David Wolpert

Abstract When the strict rationality underlying the Nash equilibria in game theory is relaxed, one arrives at the quantal response equilibria introduced by McKelvey and Palfrey. Here, the players are assigned parameters measuring their degree of rationality, and the resulting equilibria are Gibbs type distribution. This brings us into the realm of the exponential families studied in information geometry, with an additional structure arising from the relations between the players. Tuning these rationality parameters leads to a simple geometric proof of the Nash existence theorem that only employs intersection properties of submanifolds of Euclidean spaces and dispenses with the Brouwer fixed point theorem on which the classical proofs depend. Also, in this geometric framework, we can develop very efficient computational tools for studying examples. The method can also be applied when additional parameters are involved, like the capacity of an information channel.

Keywords Nash equilibrium · Quantal response equilibrium · Continuity method

1 Introduction

Game theory models agents that can choose between different options or moves by taking into account the effects of the moves of the other players in a fully rational manner. A Nash equilibrium (NE) in such a game is a selection of players' moves

J. Jost (✉) · E. Olbrich
Max Planck Institute for Mathematics in the Sciences, Inselstr. 22,
04103 Leipzig, Germany
e-mail: jjost@mis.mpg.de

N. Bertschinger
Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany

N. Bertschinger
Goethe University, Frankfurt am Main, Germany

D. Wolpert
Santa Fe Institute, Santa Fe, NM, USA

so that neither of them can increase their pay-offs by unilateral deviation from that choice of moves. In a pure NE, each player plays a single move, whereas in a mixed NE, players can play random mixtures of moves with fixed probabilities. Each game possesses at least one NE, and generically only finitely many. The outcomes of different NEs, that is, the pay-offs awarded to the individual players, are typically different in different NEs, and some can be better than others for some or all players. Also, they need not be optimal in the sense that the pay-offs could not be increased by *coordinated* deviations of the players. In particular, they need not be Pareto optimal. This then leads to the questions whether or by which mechanisms it is possible to induce transitions from NE to one that is superior for some or all players, or how to coordinate players to access Pareto optimal solutions.

As mentioned, players are assumed fully rational in the sense that they can determine their best moves in mutual anticipation of their opponents' actions. But this is neither always empirically observed, nor does it always achieve the best possible results, as the players may get stuck in a suboptimal equilibrium.

McKelvey and Palfrey [7] then introduced the concept of a quantal response equilibrium (QRE) in order to address those issues, that is, to achieve an extension of classical game theory that allows for transitions between different equilibria by embedding the discrete structure of a game into a differentiable framework and that parametrizes deviations from rationality.

Here, player i selects her move probabilities by a Gibbs distribution on the basis of the expected pay-offs given the move probabilities of her opponents. Her opponents do the same, and when the resulting probabilities match, in the sense that her move probabilities induce precisely those move probabilities for her opponents that enter into the Gibbs distribution that determines hers, and this is true for all players, then they are at a QRE. And since Gibbs distributions enter here, this brings us into the realm of exponential families, one of the core structures that are explored in information geometry (see [1, 2]). Beautiful relations obtain and await to be further explored.

Of course, the concept of a QRE is analogous to that of a NE which requires that a player's move leads to precisely those opponent moves against which it is a best response. Therefore, when the parameters of the Gibbs distributions (an inverse temperature in the jargon of statistical mechanics) that can be interpreted as expressing the degrees of rationality of the players tend to infinity the situation approaches the one of the original rational game, and the QREs converge to NEs. Naturally, this leads to the question whether all NE can occur as such limits, and if not, what are the criteria for such approximability by equilibria for not completely rational players. Conversely, when the rationality coefficient of a player goes to 0, in the limit, he will take all of his moves randomly with equal probability, independently of what the other players do.

It was already observed by McKelvey and Palfrey [7] that by tuning the parameters in quantal response games, that is, the degrees of rationality of the individual players, we can modify the equilibria and achieve transitions between different QREs or NEs. In particular, this also suggests to obtain NEs as limits of QEs by a path following technique. A corresponding algorithm was developed and applied by Turocy [14]. Here, we shall follow that approach and present a simple proof of the Nash

existence theorem that only needs the concept of an algebraic intersection number from homology theory, but not the Brouwer fixed point theorem or one of its variants that are the fundamental ingredients of Nash's original proof and its variants.

In [16], the bifurcation pattern in parameter space when varying those parameters was then analyzed. One can either try to follow solution curves for varying parameters continuously, or alternatively, when solution branches disappear for instance by saddle-node bifurcations, jump to a different solution branch. Here, we shall compute the bifurcation condition and discuss some examples in detail.

One can then naturally combine this with variations of other parameters; in particular, one can parametrize the players' pay-offs. For instance, as in [16], we can introduce an external controller that can modify the utilities via tax rates, perhaps in an attempt to steer the players towards a Pareto superior or for some other reason preferable equilibrium.

2 Preliminaries

We consider two (or more) players i with utility functions U_i with values

$$U_i(x_i^\alpha, x_{-i}^\gamma) \quad (1)$$

when player i plays x_i^α and the other player $-i$ plays x_{-i}^γ where α, γ run through index sets parametrizing the possible moves of the players. For abbreviation, we shall say that i plays α in place of x_i^α . Sometimes, we shall also use the pronoun "she" to refer to i , and "he" for $-i$. We let

$$p_i^\alpha := \text{probability that player } i \text{ plays } \alpha. \quad (2)$$

At a quantal response equilibrium (QRE), this probability has to satisfy

$$p_i^\alpha = \frac{1}{Z_i} \exp(\beta_i \sum_\gamma U_i(x_i^\alpha, x_{-i}^\gamma) p_{-i}^\gamma) \quad (3)$$

for each player i , where $0 \leq \beta_i \leq \infty$ is a coefficient indicating i 's degree of rationality, and

$$Z_i := \sum_\delta \exp(\beta_i \sum_\gamma U_i(x_i^\delta, x_{-i}^\gamma) p_{-i}^\gamma) \quad (4)$$

is the normalization factor that ensures $\sum_\alpha p_i^\alpha = 1$ for each i .

Even though the corresponding formula for $-i$ is simply obtained by exchanging i and $-i$ in (3), (4), we write it down explicitly as we shall frequently use it.

$$p_{-i}^\gamma = \frac{1}{Z_{-i}} \exp(\beta_{-i} \sum_\alpha U_{-i}(x_{-i}^\gamma, x_i^\alpha) p_i^\alpha) \quad (5)$$

$$\text{with } Z_{-i} = \sum_\eta \exp(\beta_{-i} \sum_\alpha U_{-i}(x_{-i}^\eta, x_i^\alpha) p_i^\alpha). \quad (6)$$

We note that for finite β s, $p_i^\alpha \neq 0$ in (3) and $p_{-i}^\gamma \neq 0$ in (5). For $\beta_i \rightarrow 0$, the p_i^α converge to constants, the same for all α , and analogously for p_{-i}^γ when $\beta_{-i} \rightarrow 0$. When we write

$$\pi_i^\alpha := \langle U_i(x_i^\alpha, \cdot) \rangle := \sum_\gamma U_i(x_i^\alpha, x_{-i}^\gamma) p_{-i}^\gamma \quad (7)$$

for the expected value of i 's utility when she plays α , at a QRE we have

$$p_i^\alpha = \frac{1}{Z_i} \exp(\beta_i \pi_i^\alpha) \quad (8)$$

and

$$\pi_i^\alpha = \sum_\gamma U_i(x_i^\alpha, x_{-i}^\gamma) \frac{1}{Z_{-i}} \exp(\beta_{-i} \pi_{-i}^\gamma). \quad (9)$$

Thus, the relation (7) between the p s can be replaced by the relation (9) between the π s.

The key point is, of course, that these relations are reciprocal, that is, they also hold for the opposite player $-i$. We also note

$$Z_i = \sum_\delta \exp(\beta_i \pi_i^\delta). \quad (10)$$

3 Information Geometry

We simplify our notation and put

$$V_{i,\gamma}^\alpha := \beta_i U_i(x_i^\alpha, x_{-i}^\gamma). \quad (11)$$

Then at a QRE,

$$p_i^\alpha = \frac{1}{Z_i} \exp\left(\sum_\gamma V_{i,\gamma}^\alpha p_{-i}^\gamma\right) \text{ and } p_{-i}^\gamma = \frac{1}{Z_{-i}} \exp\left(\sum_\delta V_{-i,\delta}^\gamma p_i^\delta\right). \quad (12)$$

This brings us into the setting of information geometry (see [1, 2]) where the $V_{i,\gamma}$ are observables with corresponding Lagrange multipliers p_{-i}^γ that are dual to the expectation values

$$\eta_{i,\gamma} := \sum_\delta V_{i,\gamma}^\delta p_i^\delta = \frac{\partial}{\partial p_{-i}^\gamma} \log Z_i. \quad (13)$$

We then obtain a Fisher metric, given by the covariance matrix of the observables and computed from second derivatives of $\log Z_i$,

$$g_{i;\mu,\nu} := \sum_{\alpha} V_{i,\mu}^{\alpha} V_{i,\nu}^{\alpha} p_i^{\alpha} - \sum_{\delta} V_{i,\mu}^{\delta} p_i^{\delta} \sum_{\epsilon} V_{i,\nu}^{\epsilon} p_i^{\epsilon} = \frac{\partial^2}{\partial p_{-i}^{\mu} \partial p_{-i}^{\nu}} \log Z_i. \quad (14)$$

By Legendre duality, with

$$\Xi_i := \sum_{\delta} p_i^{\delta} \log p_i^{\delta} \quad (15)$$

being the negative of the entropy of p_i^{δ} , we also have the relation

$$p_{-i}^{\gamma} = \frac{\partial}{\partial \eta_{i,\gamma}} \Xi_i. \quad (16)$$

$\log Z_i$ and Ξ_i are related by

$$\Xi_i = \sum_{\gamma} \eta_{i,\gamma} p_{-i}^{\gamma} - \log Z_i. \quad (17)$$

(See formulas (4.71) and (6.195) in [2] for an information geometric interpretation.)

When we also use the corresponding identities for $-i$, we obtain such relations as

$$\log Z_i + \log Z_{-i} + \sum_{\alpha} p_i^{\alpha} \log p_i^{\alpha} + \sum_{\gamma} p_{-i}^{\gamma} \log p_{-i}^{\gamma} = \sum_{\alpha, \gamma} p_i^{\alpha} p_{-i}^{\gamma} (V_{i,\gamma}^{\alpha} + V_{-i,\alpha}^{\gamma}). \quad (18)$$

On the right hand side, we have the sum of the expected payoffs (weighted with the rationality coefficients) averaged over their moves for the two players.

Anticipating some considerations in Sect. 4, we consider the situation where $\beta_i \rightarrow \infty$. There are two possibilities. Either there is a unique α_0 for which the expected pay-off $\sum_{\gamma} U_{i,\gamma}^{\alpha_0} p_{-i}^{\gamma}$ is maximal, in which case $p_i^{\alpha_0} \rightarrow 1$ and $p_i^{\alpha} \rightarrow 0$ for $\alpha \neq \alpha_0$, or there are several such values, in which case i is indifferent between the corresponding actions. The first case will lead us to a pure Nash equilibrium, the second to a mixed one, where the pay-off of i does not depend on her choice of actions among those best ones. See also Theorem 5.1 below.

We can consider the distributions $p_i^{\alpha} = \frac{1}{Z_i} \exp(\beta_i \sum_{\gamma} U_{i,\gamma}^{\alpha} p_{-i}^{\gamma})$ (12) as an exponential family with parameters p_{-i}^{γ} that satisfy the additional linear constraint $\sum_{\gamma} p_{-i}^{\gamma} = 1$. (We could formally eliminate that constraint by also varying β_i , that is, considering $\beta_i p_{-i}^{\gamma}$ as the exponential parameters. Those parameters then only have to be nonnegative.) Suppose that $-i$ has fewer action choices than i . Then the parameters dual to the p_{-i}^{γ} , the expectation values

$$\mathbb{E}_{p_i}(U_i(., x_{-i}^{\gamma})) = \sum_{\alpha} p_i^{\alpha} U_i(x_i^{\alpha}, x_{-i}^{\gamma}) \quad (19)$$

of the pay-offs (see e.g. (4.74) in [2]), do not determine the probabilities p_i^{α} completely, that is, there exist other values of those probabilities than given by (12) with the same expected pay-offs (19). By Theorem 2.8 in [2], the distribution

$p_i^\alpha = \frac{1}{Z_i} \exp(\beta_i \sum_\gamma U_{i,\gamma}^\alpha p_{-i}^\gamma)$ (12) is distinguished as that of highest entropy among all those with the same expected pay-offs $\mathbb{E}_{p_i}(U_i(., x_{-i}^\gamma)) = \sum_\alpha p_i^\alpha U_i(x_i^\alpha, x_{-i}^\gamma)$ for every action x_{-i}^γ of $-i$, and therefore also with the same expected payoff for that choice of his probabilities,

$$\mathbb{E}_{p_i} \left(\sum_\gamma U_i(., x_{-i}^\gamma) p_{-i}^\gamma \right) = \sum_\alpha p_i^\alpha \sum_\gamma U_i(x_i^\alpha, x_{-i}^\gamma) p_{-i}^\gamma. \quad (20)$$

Conversely, the expectation values (19) determine the parameters p_{-i}^γ of the exponential family (12). That is, there is a unique choice of the probabilities p_{-i}^γ so that the distribution is both a member of an exponential family of the form (12) and has the expectation values (19) prescribed. Of course, not all values of (19) can be realized. In order to get the maximal expected pay-off, $\max_\gamma p_i^\alpha U_i(x_i^\alpha, x_{-i}^\gamma)$, we need to have $\beta_i p_{-i}^{\gamma^*} \rightarrow \infty$ for the best values γ^* (cf. [10]), whereas the other probabilities p_{-i}^γ should be 0. We therefore move to the boundary of the exponential family of (12), see Thm. 2.8 in [2]. Of course, for the Nash equilibrium, a player i should rather optimize w.r.t. her own probabilities p_i^α . Again, this will lead to $\beta_i \rightarrow \infty$, and we shall explore that limit in Sect. 4.

Playing with the highest possible entropy while preserving the expected pay-off makes the player i least predictable for her opponent. Thus, information geometry naturally provides insight into the strategic choices of players.

4 The Nash Equilibrium Theorem

In this section, we provide a rather simple proof of the Nash equilibrium theorem. Our approach uses a path following argument, using quantal response equilibria as suggested in [7, 14], to establish a homotopy between the Nash situation and an intersection pattern of linear subspaces of Euclidean space. Topologically speaking, we exploit the invariance of a mod 2 intersection number between subspaces of Euclidean space under continuous deformation, that is, some standard textbook algebraic topology (see e.g. [13]). The details are quite elementary and geometrically intuitive.

The equilibrium theorem of Nash is the centerpiece of mathematical game theory, and thereby of much of modern microeconomic theory. Following the original work of Nash [9], this theorem is usually proven by some variant of the Brouwer fixed point theorem, like those of Kakutani or Ky Fan. Given that fixed point theorem, the proof of Nash is relatively easy and consists essentially in an appropriate reformulation of the statement. The Brouwer fixed point theorem itself is a deep and important result in topology. Its standard proof is of a combinatorial nature, most elegantly via Sperner's lemma, see e.g. [17]. It can also be proved analytically, with the help of the Weierstrass approximation theorem, see e.g. [4]. Kohlberg and Mertens [5] used a more abstract degree-theoretic argument to prove Nash. Another minimax argument based on the concept of exchangeable equilibria has recently been proposed in [12],

but in a beautiful piece of scholarship, Stein [11] pointed out a problem with that approach. Finally, there exists a path following approach to Nash [6, 15], different from the QRE based path following approach utilized here.

Let us now formulate the theorem. We consider a finite, strategic form, non-cooperative game with n players $i = 1, \dots, n$ that can choose between actions $\alpha = \alpha_i = 1, \dots, m_i$.¹ We shall also write $-i$ for the collection of opponents of player i , that is, $-i$ denotes the set $\{1, 2, \dots, i-1, i+1, \dots, n\}$. The choice of an action is probabilistic, that is, the player i chooses the action α with probability p_i^α . Thus,

$$\sum_{\alpha} p_i^\alpha = 1 \text{ for all } i. \quad (21)$$

In other words, the strategy space of player i is the $(m_i - 1)$ -dimensional simplex

$$\Sigma_i := \{(p_i^1, \dots, p_i^{m_i}) : p_i^\alpha \geq 0 \text{ for all } \alpha, \sum_{\beta} p_i^\beta = 1\}. \quad (22)$$

The collective strategy space then is the product simplex

$$\Sigma := \prod_j \Sigma_j. \quad (23)$$

The choice of probabilities of each player will depend on what the others do or what she expects them to do. That is, if she knows all the probabilities p_{-i}^γ , she determines her own probabilities p_i^α by some fixed rule. This rule will involve her utility function

$$U_i^*(\alpha_1, \dots, \alpha_n) =: U_i(\alpha, \gamma), \quad (24)$$

where now the first argument α stands for the action of player i herself whereas the second argument, the multiindex γ , stands for the collection of the actions $(\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$ of her opponents. i 's utility thus depends on her own action choice as well as on the ones of her opponents.

In game theory, players are assumed to be rational in the sense that they select actions that maximize their utility under the assumption that their opponents are rational as well. Thus, given all the other player's moves, player j plays his best response, that is, that move that maximizes his utility given his opponents' moves.

The theorem of Nash [9] then says that there exists an equilibrium, that is, a choice of action probabilities of the players that no player can gain by deviating from her action probabilities without triggering a response of at least one opponent that is better for him, but worse for her. Formally

¹All essentials features of our setting and reasoning will show themselves already for the case where $n = 2$ and $m_i = 2$ for all players, that is, for the simplest game where we have only two players each of which can choose between two possible actions or moves. Nevertheless, for the sake of generality, we shall consider arbitrary values of n and the m_i s.

Theorem *There exists a Nash equilibrium, that is, some $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i \in \Sigma_i$ for every i that maximizes the expected utility*

$$\sum_{\alpha_j=1,\dots,m_j} U_i^*(\alpha_1, \dots, \alpha_n) \prod_{j=1,\dots,n} \pi_j^{\alpha_j} = \sum_{\alpha, \gamma} U_i(\alpha, \gamma) \pi_i^\alpha \pi_{-i}^\gamma \quad (25)$$

for each player i w.r.t. his action p_i when the other players all play their π_{-i} .

We note that this equilibrium may be mixed, that is, some of the probabilities π_i^α at an equilibrium might be different from 0 and 1.

Here, we shall develop a proof of this theorem that only utilizes elementary properties of algebraic intersection numbers. The intersection numbers occurring will be ones between graphs of functions with values in Σ_i over the sets

$$\Sigma_{-i} := \prod_{j \neq i} \Sigma_j, \quad i = 1, \dots, n, \quad (26)$$

that provide some response, not necessarily a best one, of player i to the actions of her opponents. The collective state space of the opponents of i is Σ_{-i} , and the function for player i takes values in her own state space Σ_i . If we have n such graphs, they should generically intersect in finitely many points, when considered as submanifolds of Σ . We shall construct these graphs in such a way that they depend on some parameter β and converge for $\beta \rightarrow \infty$ towards the best response sets in the sense of Nash. Limits of intersections of those graphs will then be Nash equilibria.

While we shall use QREs, the discerning reader will realize that only certain continuity properties and limiting behaviors of this rule will be needed in the sequel. For the present purpose, it suffices that all player have the same rationality coefficient, that is,

$$\beta_i =: \beta \text{ for all } i, \quad (27)$$

and we shall play with that parameter β . The QRE rule (3) then becomes

$$p_i^\alpha = P_i^\alpha(p_{-i}^\gamma; \beta) := \frac{1}{Z_i(\beta)} \exp(\beta \sum_\gamma U_i(\alpha, \gamma) p_{-i}^\gamma) \quad (28)$$

for each player i . As before

$$Z_i(\beta) := \sum_\delta \exp(\beta \sum_\gamma U_i(\delta, \gamma) p_{-i}^\gamma) \quad (29)$$

is the normalization factor that ensures $\sum_\alpha p_i^\alpha = 1$ for each i . Because of this normalization, the range from which player i can choose her probabilities is the $(m_i - 1)$ -dimensional simplex Σ_i of (22).

An equilibrium is achieved when (28) holds for all players i simultaneously. Geometrically, this means that we look for a point of intersection of the graphs of the

functions $P_i^\alpha(p_{-i}^\gamma)$, $i = 1, \dots, n$. Since each such graph is of codimension $m_i - 1$ in the space Σ of (23) of dimension $\sum_j (m_j - 1)$, these graphs generically intersect in a finite collection of points.

At this point, this collection of intersection points is possibly empty. We shall show that the algebraic intersection number is always 1, for each value of β , so that there always has to be at least one intersection point. Moreover, we shall show that for $\beta \rightarrow \infty$, such intersection points converge to Nash equilibria of the original game, thereby showing the existence of such Nash equilibria.

For elementary properties of algebraic intersection numbers, we refer to [13]. A general theory of intersections of graphs is developed in [3].

We make the following observations.

1. For $\beta = 0$, (28) becomes

$$p_i^\alpha = \frac{1}{m_i} \text{ for all } i, \alpha. \quad (30)$$

Therefore, with orientations appropriately chosen, the algebraic intersection number for the graphs at $\beta = 0$ is 1.

2. For $0 \leq \beta < \infty$, an intersection cannot take place at any boundary point of Σ , as, from the definition (28),

$$P_i^\alpha(p_{-i}^\gamma; \beta) > 0 \text{ for all } i, \alpha, \gamma. \quad (31)$$

3. Since the graphs of $P_i^\alpha(p_{-i}^\gamma; \beta)$ depend continuously on β and since by 2., no intersection point can disappear at the boundary, the algebraic intersection number of the graphs is 1 for all $0 \leq \beta < \infty$. In particular, there is always at least one intersection point of those graphs in the interior of Σ .

We now consider the limits of the QRE graphs (28) and their intersection points for $\beta \rightarrow \infty$.

4. If for a given collection p_i^γ , there is a unique α_0 with

$$\alpha_0 = \operatorname{argmax}_\alpha \sum_\gamma U_i(\alpha, \gamma) p_{-i}^\gamma, \quad (32)$$

then

$$\lim_{\beta \rightarrow \infty} P_i^{\alpha_0}(p_{-i}^\gamma; \beta) = 1 \text{ and } \lim_{\beta \rightarrow \infty} P_i^\beta(p_{-i}^\gamma; \beta) = 0 \text{ for } \beta \neq \alpha_0. \quad (33)$$

This follows directly from (28), (29).

5. The set of tuples (p_i^γ) for which there exists more than one α' with

$$\alpha' = \operatorname{argmax}_\alpha \sum_\gamma U_i(\alpha, \gamma) p_{-i}^\gamma \quad (34)$$

is a union of hyperplanes in Σ_{-i} because the functions $\sum_\gamma U_i(\alpha, \gamma) p_{-i}^\gamma$ are linear w.r.t. the p_{-i}^γ . We call these hyperplanes singular. When crossing such a singular hyperplane, the maximizing α' in (34) changes. In particular, on the two sides of such a hyperplane, the maximizing α_0 in (32) is different. Thus, also the α_0 with (33) changes.

On such a singular hyperplane, that is when the opponents play a singular p_{-i}^γ , player i is indifferent between the different α' satisfying (34) as they all yield the same utility.

6. In particular, when we restrict the functions P_i^α to a line L (a one-dimensional subspace) in Σ_{-i} that is transversal to all the singular hyperplanes, we find some α_0 for which $\lim_{\beta \rightarrow \infty} P_i^{\alpha_0}(p_{-i}^\gamma; \beta)$ jumps from 0 to 1 at the intersection π of L with such a singular hyperplane.
7. Therefore, the entire line $(\pi, 0 \leq p_i^{\alpha_0} \leq 1)$ is contained in the pointwise limit set of the graphs of the functions $P_i^{\alpha_0}(p_{-i}^\gamma; \beta)$ for $\beta \rightarrow \infty$.
8. Thus, the limit set of the graphs of the $P_i^\alpha(p_{-i}^\gamma; \beta)$ for $\beta \rightarrow \infty$ consists either solely of the set $\Sigma_{-i} \times (p_i^\alpha = 0)$ or of $\Sigma_{-i} \times (p_i^\alpha = 1)$ or of some regions where $p_i^\alpha = 0$ and some where $p_i^\alpha = 1$ connected by pieces of hyperplanes in Σ_{-i} times sets $0 \leq p_i^\alpha \leq 1$ for those α for which the probabilities jump across the hyperplane.
9. Intersection points of the graphs of the functions $P_i^{\alpha_0}(p_{-i}^\gamma; \beta)$ for the different players i converge to intersection points of their pointwise limit sets. Some or all of these limit intersection points may lie in the boundary of Σ . It is also possible, however, that some of them lie in the connected pieces described in 8. where some or all of the p_i^α are undetermined.

We now see the *proof* of the Nash theorem: By 8., the limits of the QRE graphs yield the best response sets for the players. By 9., the intersection of these best response sets is not empty. Since by definition, an intersection point of the best response sets of all players is a Nash equilibrium, the existence of such an equilibrium follows. \square

5 Variations of QREs

We now want to utilize the QRE framework and its geometric interpretation to set up a computational scheme for investigating examples. We shall first compute the effect of parameter variations on the QRE values of the move probabilities. We shall obtain implicit systems of equations for the variations of the p_i^α and the p_{-i}^γ . We point out that since the β s and the probabilities of the QRE are not independent of each other, as the latter change when the former are varied, the standard information geometric formulas for parameter variations, like (4.72) in [2], need to be augmented by also differentiating those dependencies.

We return to the general case where each agent i has her own rationality coefficient β_i , and we shall consider a variation of that rationality coefficient. We have

$$\frac{1}{Z_i} \frac{d}{d\beta_i} \exp(\beta_i \sum_{\gamma} U_i(x_i^{\alpha}, x_{-i}^{\gamma}) p_{-i}^{\gamma}) = p_i^{\alpha} \sum_{\gamma} (\beta_i \frac{dp_{-i}^{\gamma}}{d\beta_i} + p_{-i}^{\gamma}) U_i(x_i^{\alpha}, x_{-i}^{\gamma}) \quad (35)$$

and

$$\frac{1}{Z_i} \frac{dZ_i}{d\beta_i} = \sum_{\gamma} (\beta_i \frac{dp_{-i}^{\gamma}}{d\beta_i} + p_{-i}^{\gamma}) \sum_{\delta} U_i(x_i^{\delta}, x_{-i}^{\gamma}) p_i^{\delta}. \quad (36)$$

From (3), (35), (36), we obtain

$$\frac{dp_i^{\alpha}}{d\beta_i} = p_i^{\alpha} \sum_{\gamma} (\beta_i \frac{dp_{-i}^{\gamma}}{d\beta_i} + p_{-i}^{\gamma}) (U_i(x_i^{\alpha}, x_{-i}^{\gamma}) - \sum_{\delta} U_i(x_i^{\delta}, x_{-i}^{\gamma}) p_i^{\delta}). \quad (37)$$

For the variation of p_{-i}^{γ} , we have the symmetric equation, except that we do not get the analogue of the term $+p_{-i}^{\gamma}$ because we vary only β_i , but not β_{-i} at this point. That is,

$$\frac{dp_{-i}^{\gamma}}{d\beta_i} = p_{-i}^{\gamma} \sum_{\alpha} \beta_{-i} \frac{dp_i^{\alpha}}{d\beta_i} (U_{-i}(x_{-i}^{\gamma}, x_i^{\alpha}) - \sum_{\eta} U_{-i}(x_{-i}^{\eta}, x_i^{\alpha}) p_{-i}^{\eta}). \quad (38)$$

We can insert (38) into (37) to obtain

$$\begin{aligned} \frac{dp_i^{\alpha}}{d\beta_i} &= p_i^{\alpha} \sum_{\gamma} p_{-i}^{\gamma} (\beta_i \sum_{\epsilon} \beta_{-i} \frac{dp_i^{\epsilon}}{d\beta_i} (U_{-i}(x_{-i}^{\gamma}, x_i^{\epsilon}) - \sum_{\eta} U_{-i}(x_{-i}^{\eta}, x_i^{\epsilon}) p_{-i}^{\eta}) + 1) \\ &\quad (U_i(x_i^{\alpha}, x_{-i}^{\gamma}) - \sum_{\delta} U_i(x_i^{\delta}, x_{-i}^{\gamma}) p_i^{\delta}). \end{aligned} \quad (39)$$

Let us discuss (37). The terms without the factor β_i are

$$p_i^{\alpha} (\pi_i^{\alpha} - \sum_{\delta} \pi_i^{\delta} p_i^{\delta}). \quad (40)$$

Thus, this yields a positive contribution when the expected utility for move α is larger than the average of the expected utilities over all moves. For given γ , this then is only overcompensated by the reaction of the opponent $-i$ when

$$\beta_i \frac{dp_{-i}^{\gamma}}{d\beta_i} < -p_{-i}^{\gamma}. \quad (41)$$

Of course, as we see from (38), $\frac{dp_{-i}^{\gamma}}{d\beta_i}$ becomes negative when $\frac{dp_i^{\alpha}}{d\beta_i}$ is positive for those α for which the return $U_{-i}(x_{-i}^{\gamma}, x_i^{\alpha})$ for the action γ is smaller than the average return over all actions of $-i$. Also, from (39) we see that this effect can be expected to be small when the product $\beta_i \beta_{-i}$ of the rationality coefficients is small.

Let us take a look at the special case when i has only two possible moves, $\alpha = +, -$. Let the value of β_i be such that $\frac{dp_i^\alpha}{d\beta_i} = 0$ for one of those α s and hence also for the other one, as $\sum_\alpha p_i^\alpha = 1$. In that case, from (38) then $\frac{dp_{-i}^\gamma}{d\beta_i} = 0$ for all γ (when i does not change her probabilities, $-i$ has no reason for a change either as his own rationality coefficient β_{-i} is not varying). Inserting this in (37) yields (cf. (40))

$$\pi_i^\alpha = \sum_\delta \pi_i^\delta p_i^\delta \quad (42)$$

for both values of α , that is,

$$\pi_i^+ = \pi_i^- . \quad (43)$$

Thus, the expected utility is the same for both moves. In other words, whenever the expected utilities are different for the two possible moves of i , a change in her rationality will also lead to a change of her QRE probabilities.

When, however, (43) holds, then the probabilities p_i^α of i are undetermined because it makes no difference for her which move she chooses. In this case, for an equilibrium, it should then not matter for $-i$ which move i plays, that is, his utility at his equilibrium should not depend on x_i^α , i.e. $\sum_\gamma U_{-i}(x_{-i}^\gamma, x_i^\alpha) p_{-i}^\gamma$ is the same for all α .

We continue to analyze the case where i has only two possible moves $+, -$. We now assume that condition (42) holds. When we use (42) in (39), we obtain

$$\frac{dp_i^\alpha}{d\beta_i} = p_i^\alpha \sum_\gamma p_{-i}^\gamma \beta_i \sum_\epsilon \beta_{-i} \frac{dp_i^\epsilon}{d\beta_i} U_{-i}(x_{-i}^\gamma, x_i^\epsilon) (U_i(x_i^\alpha, x_{-i}^\gamma) - \sum_\delta U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta) . \quad (44)$$

Using the fact that $\frac{dp_i^-}{d\beta_i} = -\frac{dp_i^+}{d\beta_i}$ because of $p_i^+ + p_i^- \equiv 1$, we obtain from (44)

$$\begin{aligned} \frac{dp_i^+}{d\beta_i} & \left(1 - p_i^+ \sum_\gamma p_{-i}^\gamma \beta_i \beta_{-i} (U_{-i}(x_{-i}^\gamma, x_i^+) - U_{-i}(x_{-i}^\gamma, x_i^-)) (U_i(x_i^+, x_{-i}^\gamma) \right. \\ & \left. - \sum_\delta U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta) \right) = 0 . \end{aligned} \quad (45)$$

We can also rewrite the last factor in (45) as

$$U_i(x_i^+, x_{-i}^\gamma) - \sum_\delta U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta = (U_i(x_i^+, x_{-i}^\gamma) - U_i(x_i^-, x_{-i}^\gamma)) p_i^- . \quad (46)$$

Thus, (45) becomes

$$\frac{dp_i^+}{d\beta_i} \left(1 - p_i^+ p_i^- \beta_i \beta_{-i} \sum_{\gamma} p_{-i}^\gamma (U_{-i}(x_{-i}^\gamma, x_i^+) - U_{-i}(x_{-i}^\gamma, x_i^-)) (U_i(x_i^+, x_{-i}^\gamma) - U_i(x_i^-, x_{-i}^\gamma)) \right) = 0. \quad (47)$$

Thus, unless the coefficient in (47) vanishes, the derivatives $\frac{dp_i^+}{d\beta_i}, \frac{dp_i^-}{d\beta_i}$ both vanish.

We note that this coefficient is symmetric in p_i^+ and p_i^- , that is, the same for the derivative of either of these two probabilities. Of course, vanishing of this coefficient is a very special and nongeneric condition. For instance, when we have a zero-sum game, that is

$$U_{-i}(x_{-i}^\gamma, x_i^\alpha) = -U_i(x_i^\alpha, x_{-i}^\gamma) \quad (48)$$

for all α, γ , then the product of the two differences in (47) is always negative which makes our coefficient always ≤ 1 , hence non-zero.

In any case, when all $\frac{dp_i^\alpha}{d\beta_i} = 0$, then also all $\frac{dp_{-i}^\gamma}{d\beta_i} = 0$. (If i does not change her probabilities, then $-i$ has no reason to change his probabilities either as his rationality coefficient is not varied.) In that case, that is, for such a specific value of β_i , we also obtain simplified relations for the second derivatives,

$$\frac{d^2 p_i^\alpha}{(d\beta_i)^2} = p_i^\alpha \sum_{\gamma} \beta_i \frac{d^2 p_{-i}^\gamma}{(d\beta_i)^2} (U_i(x_i^\alpha, x_{-i}^\gamma) - \sum_{\delta} U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta) \quad (49)$$

and

$$\frac{d^2 p_{-i}^\gamma}{(d\beta_i)^2} = p_{-i}^\gamma \sum_{\alpha} \beta_{-i} \frac{d^2 p_i^\alpha}{(d\beta_i)^2} (U_{-i}(x_{-i}^\gamma, x_i^\alpha) - \sum_{\eta} U_{-i}(x_{-i}^\eta, x_i^\alpha) p_{-i}^\eta), \quad (50)$$

and in combination

$$\begin{aligned} \frac{d^2 p_i^\alpha}{(d\beta_i)^2} &= p_i^\alpha \sum_{\gamma} p_{-i}^\gamma \beta_i \sum_{\epsilon} \beta_{-i} \frac{d^2 p_i^\epsilon}{(d\beta_i)^2} (U_{-i}(x_{-i}^\gamma, x_i^\epsilon) - \sum_{\eta} U_{-i}(x_{-i}^\eta, x_i^\epsilon) p_{-i}^\eta) \\ &\quad (U_i(x_i^\alpha, x_{-i}^\gamma) - \sum_{\delta} U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta), \end{aligned} \quad (51)$$

We return to the case where i has only two possible moves $+, -$. Analogously to (47), we obtain

$$\frac{d^2 p_i^+}{(d\beta_i)^2} \left(1 - p_i^+ p_i^- \beta_i \beta_{-i} \sum_{\gamma} p_{-i}^\gamma (U_{-i}(x_{-i}^\gamma, x_i^+) - U_{-i}(x_{-i}^\gamma, x_i^-)) (U_i(x_i^+, x_{-i}^\gamma) - U_i(x_i^-, x_{-i}^\gamma)) \right) = 0. \quad (52)$$

As before, unless the coefficient in (52) vanishes, the second derivatives $\frac{d^2 p_i^+}{(d\beta_i)^2}, \frac{d^2 p_i^-}{(d\beta_i)^2}$ both vanish. The same scheme then inductively applies to all higher derivatives. We conclude

Theorem 5.1 *For a QRE in a game where player i has only two possible moves, generically, her probabilities do not depend on her rationality coefficient β_i if and only if she is at an equilibrium where her expected utility is the same for both moves.*

This equilibrium of i can be a nontrivial function of β_i only if

$$1 - p_i^+ p_i^- \beta_i \beta_{-i} \sum_{\gamma} p_{-i}^\gamma (U_{-i}(x_{-i}^\gamma, x_i^+) - U_{-i}(x_{-i}^\gamma, x_i^-)) (U_i(x_i^+, x_{-i}^\gamma) - U_i(x_i^-, x_{-i}^\gamma)) = 0. \quad (53)$$

Proof We have seen that when the derivatives of her move probabilities w.r.t. β_i vanishes, then (43), that is, indifference, holds. Also, in that situation, generically, that is, unless (53) holds, all higher derivatives of her move probabilities vanish as well which implies that these probabilities are independent of β_i as they are analytic functions of β_i . Conversely, when (43) holds, then generically, the derivatives vanish. If the derivatives vanish generically, then they have to vanish always by continuity. \square

We now look at the bifurcation behavior. We consider (3), (5) as the following functional relationships

$$p_i = f_i(p_{-i}, \beta_i), \quad p_{-i} = f_{-i}(p_i, \beta_{-i}). \quad (54)$$

The bifurcation question then is whether locally (54) allows us to determine p_i as a function of β_i . That is, we ask whether the equation

$$\frac{\partial f_i}{\partial p_{-i}} \frac{\partial f_{-i}}{\partial p_i} \frac{\partial p_i}{\partial \beta_i} + \frac{\partial f_i}{\partial \beta_i} - \frac{\partial p_i}{\partial \beta_i} = 0 \quad (55)$$

can be solved for $\frac{\partial p_i}{\partial \beta_i}$ (here, $\frac{\partial p_i}{\partial \beta_i}$ is a vector, and its factors in (55) are matrices). Equation (55) cannot be solved in general if

$$\det \left(\frac{\partial f_i}{\partial p_{-i}} \frac{\partial f_{-i}}{\partial p_i} - \text{Id} \right) = 0. \quad (56)$$

Thus, in order to identify the bifurcation condition, we need to compute this determinant.

For abbreviation, we shall write $U_i(\alpha, \gamma) := U_i(x_i^\alpha, x_{-i}^\gamma)$ etc.
The components of the matrix in (56) are (using the standard Kronecker symbol)

$$\begin{aligned} & \sum_\gamma \frac{\partial p_i^\alpha}{\partial p_{-i}^\gamma} \frac{\partial p_{-i}^\gamma}{\partial p_i^\beta} - \delta_\beta^\alpha \\ = & \sum_\gamma \beta_i p_i^\alpha (U_i(\alpha, \gamma) - \sum_\delta U_i(\delta, \gamma) p_i^\delta) \beta_{-i} p_{-i}^\gamma (U_{-i}(\gamma, \beta) - \sum_\eta U_{-i}(\eta, \beta) p_{-i}^\eta) - \delta_\beta^\alpha. \end{aligned} \quad (57)$$

In order to evaluate the determinant of this matrix, we again specialize to the case where each player has only two moves, $+, -$. Then we can compute the determinant as

$$\begin{aligned} & [\beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(+, +) - U_i(-, +))(U_{-i}(+, +) - U_{-i}(-, +)) \\ & \quad + \beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(+, -) - U_i(-, -))(U_{-i}(-, +) - U_{-i}(+, +)) - 1] \\ & [\beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(-, +) - U_i(+, +))(U_{-i}(+, -) - U_{-i}(-, -)) \\ & \quad + \beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(-, -) - U_i(+, -))(U_{-i}(-, -) - U_{-i}(+, -)) - 1] \\ = & [\beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(+, +) - U_i(-, +))(U_{-i}(+, -) - U_{-i}(-, -)) \\ & \quad + \beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(+, -) - U_i(-, -))(U_{-i}(-, -) - U_{-i}(+, -))] \\ & [\beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(-, +) - U_i(+, +))(U_{-i}(+, +) - U_{-i}(-, +)) \\ & \quad + \beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(-, -) - U_i(+, -))(U_{-i}(-, +) - U_{-i}(+, +))] \\ = & 1 - \beta_i \beta_{-i} p_i^+ p_i^- p_{-i}^+ p_{-i}^- (U_i(+, +) - U_i(-, +) - U_i(+, -) + U_i(-, -)) \\ & (U_{-i}(+, +) - U_{-i}(-, +) - U_{-i}(+, -) + U_{-i}(-, -)). \end{aligned} \quad (58)$$

As before, we have $p_i^- = 1 - p_i^+$ and $p_{-i}^- = 1 - p_{-i}^+$.

The *bifurcation condition* then is that the r.h.s. of (58) vanishes for a pair p_i, p_{-i} that solves (3), (5).

We now look at a slightly different situation. We assume that both players have the same rationality coefficient $\beta = \beta_i = \beta_{-i}$, and we want to compute the derivatives of the move probabilities when this β varies for both of them simultaneously. For i , we can use (37) to obtain

$$\frac{dp_i^\alpha}{d\beta} = p_i^\alpha \sum_\gamma (\beta \frac{dp_{-i}^\gamma}{d\beta} + p_{-i}^\gamma) (U_i(x_i^\alpha, x_{-i}^\gamma) - \sum_\delta U_i(x_i^\delta, x_{-i}^\gamma) p_i^\delta), \quad (59)$$

and for $-i$, we obtain analogously

$$\frac{dp_{-i}^\gamma}{d\beta} = p_{-i}^\gamma \sum_\alpha (\beta \frac{dp_i^\alpha}{d\beta} + p_i^\alpha) (U_{-i}(x_{-i}^\gamma, x_i^\alpha) - \sum_\eta U_{-i}(x_{-i}^\eta, x_i^\alpha) p_{-i}^\eta). \quad (60)$$

In this case, in the 2-move situation, when for instance the derivatives w.r.t. β of the probabilities of $-i$ vanish, then this holds for i as well only if either (43) or if one

of the p_i^α vanishes, that is, when i only applies one of her moves. And the situation is now symmetric between i and $-i$.

Using that $\frac{dp_{-i}^-}{d\beta} = -\frac{dp_{-i}^+}{d\beta}$, we compute

$$\begin{aligned} \frac{dp_i^+}{d\beta} &= p_i^+ p_i^- (\beta \frac{dp_i^+}{d\beta} (U_i(x_i^+, x_{-i}^+) - U_i(x_i^-, x_{-i}^+) - (U_i(x_i^+, x_{-i}^-) - U_i(x_i^-, x_{-i}^-))) \\ &\quad + p_{-i}^+ (U_i(x_i^+, x_{-i}^+) - U_i(x_i^-, x_{-i}^+) + p_{-i}^- (U_i(x_i^+, x_{-i}^-) - U_i(x_i^-, x_{-i}^-))). \end{aligned} \quad (61)$$

If $\frac{dp_i^+}{d\beta} = 0$, and hence also $\frac{dp_i^-}{d\beta} = 0$, we obtain from (60)

$$\frac{dp_{-i}^+}{d\beta} = p_{-i}^+ p_{-i}^- \sum_{\alpha} p_i^\alpha (U_{-i}(x_{-i}^+, x_i^\alpha) - U_{-i}(x_{-i}^-, x_i^\alpha)). \quad (62)$$

We can insert this then into (61) and, using that $p_{-i}^- = 1 - p_{-i}^+$, obtain a quadratic equation for p_{-i}^+ as the condition for the vanishing of $\frac{dp_i^+}{d\beta}$ and $\frac{dp_i^-}{d\beta}$.

When, instead, the derivatives for both players vanish, we then obtain

$$\begin{aligned} (U_{-i}(x_{-i}^+, x_i^+) - U_{-i}(x_{-i}^-, x_i^+)) p_i^+ &= (U_{-i}(x_{-i}^-, x_i^-) - U_{-i}(x_{-i}^+, x_i^-)) p_i^- \text{ and} \\ (U_i(x_i^+, x_{-i}^+) - U_i(x_i^-, x_{-i}^+)) p_{-i}^+ &= (U_i(x_i^-, x_{-i}^-) - U_i(x_i^+, x_{-i}^-)) p_{-i}^-, \end{aligned} \quad (63)$$

that is, unless $U_{-i}(x_{-i}^+, x_i^+) = U_{-i}(x_{-i}^-, x_i^+)$ and so on,

$$\begin{aligned} p_i^+ &= \frac{U_{-i}(x_{-i}^-, x_i^-) - U_{-i}(x_{-i}^+, x_i^-)}{U_{-i}(x_{-i}^+, x_i^+) - U_{-i}(x_{-i}^-, x_i^+) + U_{-i}(x_{-i}^-, x_i^-) - U_{-i}(x_{-i}^+, x_i^-)} \quad \text{and} \\ p_{-i}^+ &= \frac{U_i(x_i^-, x_{-i}^-) - U_i(x_i^+, x_{-i}^-)}{U_i(x_i^+, x_{-i}^+) - U_i(x_i^-, x_{-i}^+) + U_i(x_i^-, x_{-i}^-) - U_i(x_i^+, x_{-i}^-)} \end{aligned} \quad (64)$$

and analogous equations for p_i^- , p_{-i}^- .

Again, when the first derivatives of the probabilities of the two players vanish, then also all higher derivatives vanish. Equation (64) are conditions for the probabilities that do not depend on the rationality coefficient β , and they provide an equilibrium that does not depend on β .

Equation (63) is also equivalent to

$$\begin{aligned} U_{-i}(x_{-i}^+, x_i^+) p_i^+ + U_{-i}(x_{-i}^-, x_i^-) p_i^- &= U_{-i}(x_{-i}^-, x_i^+) p_i^+ + U_{-i}(x_{-i}^+, x_i^-) p_i^- \text{ and} \\ U_i(x_i^+, x_{-i}^+) p_{-i}^+ + U_i(x_i^-, x_{-i}^-) p_{-i}^- &= U_i(x_i^-, x_{-i}^+) p_{-i}^+ + U_i(x_i^+, x_{-i}^-) p_{-i}^-. \end{aligned} \quad (65)$$

This simply means that both players expect the same pay-off for either move, because the other player plays with the right probabilities. By the QRE condition (3), all probabilities then are $= 1/2$. From (64), we then see that this can hold only when

the utility functions satisfy appropriate relationships, $U_i(x_i^-, x_{-i}^-) - U_i(x_i^+, x_{-i}^-) = U_i(x_i^+, x_{-i}^+) - U_i(x_i^-, x_{-i}^+)$ and $U_{-i}(x_{-i}^-, x_i^-) - U_{-i}(x_{-i}^+, x_i^-) = U_{-i}(x_{-i}^+, x_i^+) - U_{-i}(x_{-i}^-, x_i^+)$. In other words, this is a rather special case.

We now look again at the bifurcation condition. In fact, this condition is the same as in the previous case. We recall the functional relationships (54)

$$p_i = f_i(p_{-i}, \beta_i), \quad p_{-i} = f_{-i}(p_i, \beta_{-i}). \quad (66)$$

Above, we have analyzed whether locally (66) allows us to determine p_i as a function of β_i . Equivalently, we can ask whether we can solve for both p_i and p_{-i} as a function of β_i . Thus, in place of (55), we write

$$\frac{\partial f_i}{\partial p_{-i}} \frac{\partial p_{-i}}{\partial \beta_i} + \frac{\partial f_i}{\partial \beta_i} - \frac{\partial p_i}{\partial \beta_i} = 0 \quad (67)$$

$$\frac{\partial f_{-i}}{\partial p_i} \frac{\partial p_i}{\partial \beta_i} + \frac{\partial f_{-i}}{\partial \beta_i} - \frac{\partial p_{-i}}{\partial \beta_i} = 0. \quad (68)$$

The only difference to (55) is that we now also include a term $\frac{\partial f_{-i}}{\partial \beta_i}$, but this is irrelevant for the bifurcation condition which now is written as

$$\det \begin{pmatrix} \text{Id} & \frac{\partial f_i}{\partial p_{-i}} \\ \frac{\partial f_{-i}}{\partial p_i} & \text{Id} \end{pmatrix} = 0 \quad (69)$$

which is equivalent to (56). In particular, in the case of two players with two moves $+, -$, we arrive, as in (58), at the bifurcation condition for $\beta = \beta_i = \beta_{-i}$

$$1 = \beta^2 p_i^+(1 - p_i^+) p_{-i}^+(1 - p_{-i}^+) (U_i(+, +) - U_i(-, +) - U_i(+, -) + U_i(-, -)) \\ (U_{-i}(+, +) - U_{-i}(-, +) - U_{-i}(+, -) + U_{-i}(-, -)). \quad (70)$$

In particular, a necessary condition for a bifurcation to occur at some positive value of β is that

$$(U_i(+, +) - U_i(-, +) - U_i(+, -) + U_i(-, -))(U_{-i}(+, +) \\ - U_{-i}(-, +) - U_{-i}(+, -) + U_{-i}(-, -)) > 0. \quad (71)$$

This condition excludes many of the standard games. More quantitatively, since $p_i^+(1 - p_i^+) p_{-i}^+(1 - p_{-i}^+) \leq \frac{1}{16}$, a necessary condition for a bifurcation is

$$\beta^2 (U_i(+, +) - U_i(-, +) - U_i(+, -) + U_i(-, -))(U_{-i}(+, +) \\ - U_{-i}(-, +) - U_{-i}(+, -) + U_{-i}(-, -)) \geq \frac{1}{16}, \quad (72)$$

that is, β cannot be too small.

For small enough values of β_i, β_{-i} , there is a single QRE near $(p_i^+, p_{-i}^+) = (\frac{1}{2}, \frac{1}{2})$. We can then follow that solution branch. As McKelvey–Palfrey [7, 8] showed, for a generic game, and when we play only with a single parameter, in our case $\beta = \beta_i = \beta_{-i}$, there are only saddle-node type bifurcations, and this branch then extends to the limit where β goes to infinity and converges to a Nash equilibrium of the original game. In general, we can follow this branch and check whether a bifurcation occurs along it. Assuming the necessary condition (71), there are two possibilities. Either for the solutions (p_i^+, p_{-i}^+) , at least one of them converges to 0 or 1, in which case the right hand side of (70) may always stay smaller than 1 so that the bifurcation condition is never satisfied. Or both p_i^+ and p_{-i}^+ converge to values strictly between 0 and 1, that is, to some mixed Nash equilibrium. In that case, when (71) holds, the right hand side of (70) eventually gets larger than 1 when β increases, and therefore, by continuity, there must be some value where the necessary bifurcation condition (70) is satisfied. For a generic game, this condition then is also sufficient, and we find a bifurcation. Below, we shall present a careful analysis of some examples.

It should be clear how to also compute derivatives w.r.t. other parameters than β_i .

6 An Example

We consider a simple example. The game has the following pay-offs, i being the row and $-i$ being the column player:

$$\begin{array}{cc|cc} & 2 & 1 & 0 \\ & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array} \quad (73)$$

The first move of each player, that is, up for i and left for $-i$, will be denoted by +, the second on, down for i and right for $-i$, by -. An appropriately indexed p will again stand for the corresponding probabilities.

We note that this game is symmetric in the sense that it remains invariant if we simultaneously interchange the two players and the labels of the move, as the pay-off of i for some combination of moves equals the pay-off of $-i$ when the opposite moves are selected. This symmetry will then extend to the QRE situation when we also exchange the two rationality coefficients. This symmetry, however, will not affect our qualitative conclusions; it will only simplify the understanding of the plots presented.

The Nash equilibria can be easily understood in the following geometric manner. In general, the expected pay-offs of i and $-i$ are given by

$$(U_i(x_i^+, x_{-i}^+)p_{-i}^+ + U_i(x_i^+, x_{-i}^-)p_{-i}^-)p_i^+ + (U_i(x_i^-, x_{-i}^+)p_{-i}^+ + U_i(x_i^-, x_{-i}^-)p_{-i}^-)p_i^- \quad (74)$$

and

$$(U_{-i}(x_{-i}^+, x_i^+)p_i^+ + U_{-i}(x_{-i}^+, x_i^-)p_i^-)p_{-i}^+ + (U_{-i}(x_{-i}^-, x_i^+)p_i^+ + U_{-i}(x_{-i}^-, x_i^-)p_i^-)p_{-i}^- \quad (75)$$

Since $p_i^- = 1 - p_i^+$, for each fixed probabilities of her opponent, the expected payoff of i is a linear function of her p_i^+ which is constrained to the unit interval $[0, 1]$. When this linear function has a negative slope, the maximum is achieved for $p_i^+ = 0$, when the slope is positive, at $p_i^+ = 1$, and when the slope is 0, she is indifferent. In the present example, the slope is positive for $p_{-i}^+ > 1/3$, negative for $p_{-i}^+ < 1/3$, and 0 for $p_{-i}^+ = 1/3$. Since the game is symmetric, the corresponding holds for $-i$. Therefore, the Nash equilibria are given by

$$\begin{aligned} p_i^+ = 1 = p_{-i}^+, & \text{ the right endpoints of two lines with positive slope} \\ p_i^+ = 0 = p_{-i}^+, & \text{ the left endpoints of two lines with negative slope} \\ p_i^+ = 2/3, p_{-i}^+ = 1/3, & \text{ the intersection of the two lines with 0 slope.} \end{aligned} \quad (76)$$

At the mixed equilibrium, the expected pay-off for each player is $2/3$, according to (74), which is smaller than the pay-offs at the pure equilibria which are $2/1$ and $1/2$.

We can put this example into a more general perspective. We consider p_i^+ and p_{-i}^+ as coordinates, ranging from 0 to 1, of course. By (74), noting $p_i^- = 1 - p_i^+$ and $p_{-i}^- = 1 - p_{-i}^+$, for each value of p_{-i}^+ , we find either a single value of p_i^+ or the line $0 \leq p_i^+ \leq 1$ of those values that maximize the pay-off of i . The collection of these maximizing sets when p_{-i}^+ ranges from 0 to 1 is a connected set that connects the line $p_{-i}^+ = 0$ with the line $p_{-i}^+ = 1$. Similarly, the collection of the maximizing values of p_{-i}^+ when p_i^+ ranges from 0 to 1 connects the line $p_i^+ = 0$ with the line $p_i^+ = 1$. Therefore these two maximizing sets need to meet at a boundary point or intersect at least once. Since the intersections correspond to the Nash equilibria, this geometrically demonstrates the existence of a Nash equilibrium. This simple topological argument naturally extends to the case of more than two players with more than two possible actions for each player.

The same kind of reasoning also works for QRE equilibria. For each value of p_{-i}^+ , we find the optimal p_i^+ , and therefore, when p_{-i}^+ varies, this collection of optimal values of p_i^+ connects the line $p_{-i}^+ = 0$ with the line $p_{-i}^+ = 1$, and analogously for $-i$, and the intersections of these curves (or their higher dimensional analogues in the general case) then yield the QREs. In fact, when the rationality parameters of the players go to 0, the limiting response curves are the lines $p_i^+ = \frac{1}{2}$ and $p_{-i}^+ = \frac{1}{2}$, resp., which intersect at the common value $+ = \frac{1}{2}$. From this, the existence of intersections for positive parameter values can then also be deduced from a homotopy argument.

We can also combine this kind of reasoning in a single simple scheme. We consider a general game with finitely many players j each of which has n_j (finitely many) possible actions. For $\beta = 0$, the players are completely irrational in the sense that they play all their actions with the same probability $\frac{1}{n_j}$. Thus, the corresponding graphs of the action probabilities of one player as a function of the probabilities of the other players intersect in a single point, the point where each plays all actions with the same probability. In terms of algebraic topology, the intersection number mod 2 (that is, we consider homology with coefficients in \mathbb{Z}_2) is 1 *cite some book on algebraic topology*. This algebraic intersection number is a homotopy invariant, and it is therefore the same for any combination of the rationality coefficients β_j . Note

that, by definition of an algebraic intersection number, this does not mean that the corresponding graphs for the action probabilities of each player as a function of the probabilities of the other players always have to intersect in a single point. However, it does imply that the intersection is nonempty, that is, they need to intersect at least once. Thus, there always exists at least one QRE. By continuity, this extends to the limit where the β_j go to infinity. And since a limit of QREs is a Nash equilibrium, we have deduced the existence of the latter for any game, without having to invoke the Brouwer fixed point theorem.

We return to our specific example. According to (3), (4), the QRE is given by

$$p_i^+ = \frac{\exp(\beta_i 2p_{-i}^+)}{\exp(\beta_i 2p_i^+) + \exp(\beta_i(1-p_{-i}^+))} =: f_i(p_{-i}^+) \quad (77)$$

$$p_{-i}^+ = \frac{\exp(\beta_{-i} p_i^+)}{\exp(\beta_{-i} 2(1-p_i^+)) + \exp(\beta_{-i} p_i^+)} =: f_{-i}(p_i^+). \quad (78)$$

We note that both these functions are strictly increasing. We have the symmetry

$$f_i(p) + f_{-i}(1-p) = 1 \text{ for all } p. \quad (79)$$

Inserting one of the equations of (77), (78) into the other then yields a fixed point equation for p_i^+ or p_{-i}^+ . Equivalently, we look for the intersections of the two graphs resulting from (77), (78), that is for p_i^+ as a function of p_{-i}^+ and for p_{-i}^+ as a function of p_i^+ .

We shall now assume for a moment that

$$\beta_i = \beta_{-i} =: \beta. \quad (80)$$

For large enough β , the two graphs then intersect thrice. One intersection point is symmetric, that is, $p_{-i}^+ = 1 - p_i^+$ and unstable, whereas the two other ones are nonsymmetric and stable, but symmetric to each other; one of them is near $(1, 1)$, the other near $(0, 0)$ for large β . In order to analyze the bifurcation that takes place when β becomes smaller, we observe that

$$f_i\left(\frac{1}{3}\right) = \frac{1}{2} \text{ and } f_{-i}\left(\frac{2}{3}\right) = \frac{1}{2} \quad (81)$$

and

$$f_i\left(\frac{1}{2}\right) >, =, < \frac{2}{3} \text{ and } f_{-i}\left(\frac{1}{2}\right) <, =, > \frac{1}{3} \text{ for } \beta >, =, < 2 \log_e 2, \text{ resp.} \quad (82)$$

Therefore, for $\beta > 2 \log_e 2$, we have an unstable fixed point with $1/2 < p_i^+ < 2/3$, because, iterating (77) and using (81), (82), when i plays $p_i^+ = 1/2$, then $-i$ will play $p_{-i}^+ < 1/3$ which then induces i to play $p_i^+ < 1/2$, but when she plays $p_i^+ = 2/3$, then he will play $p_{-i}^+ = 1/2$, which then induces her to play $p_i^+ > 2/3$. Thus, for every $\beta > 2 \log_e 2$, there must be such an unstable fixed point with $1/2 < p_i^+ < 2/3$.

Consequently, because the first graph (77) starts above and ends below the second one (when $p_i^+ = 0$, then from (78) $p_{-i}^+ > 0$ for finite β , and conversely, and analogously for $p_i^+ = 1$), but at the unstable fixed point, the first graph crosses the second one from below, there must also exist two stable fixed points. Thus, the qualitative situation is the same for all such β .

In fact, we have an implicit equation for the unstable fixed point. Because of the symmetry (78), when $p_i^+ = f_i(p_{-i}^+) = 1 - p_{-i}^+$, then also $f_{-i}(p_i^+) = p_{-i}^+$, which leads to the equation

$$\exp(3\beta p_{-i}^+ - \beta) = \frac{1 - p_{-i}^+}{p_{-i}^+}. \quad (83)$$

For $\beta \rightarrow \infty$, we obtain the solution $p_{-i}^+ = \frac{1}{3}$, whereas for $\beta \rightarrow 0$, we obtain $p_{-i}^+ = \frac{1}{2}$.

For $\beta = 2 \log_e 2$, from (81), (82), we also obtain explicitly the two fixed points

$$\left(\frac{1}{2}, \frac{1}{3}\right) \text{ and } \left(\frac{2}{3}, \frac{1}{2}\right). \quad (84)$$

For small β , both functions f_i, f_{-i} are close to the constant function $\equiv \frac{1}{2}$ which intersect only once in a stable fixed point. Therefore, because of the above symmetry, there must be some critical bifurcation value β^* where the two stable fixed points for $\beta > \beta^*$ merge with the unstable one in a pitchfork bifurcation to leave only a single stable fixed point for $\beta < \beta^*$. According to (70), the bifurcation condition is

$$(\beta^*)^2 p_i^+ (1 - p_i^+) p_{-i}^+ (1 - p_{-i}^+) = \frac{1}{9} \quad (85)$$

where p_i^+ and p_{-i}^+ here are the values at the fixed point. Since these values lie between $\frac{1}{2}$ and $\frac{2}{3}$, and between $\frac{1}{3}$ and $\frac{1}{2}$, resp., we obtain bounds for the possible range of β^* ; in fact, $1 < \beta^* < 2$.

We now give up (80). This will then also destroy the symmetry between the fixed points. We keep β_i fixed, but let β_{-i} go to 0. Then the response curve (78) of $-i$ becomes very flat, the value of p_{-i}^+ being close to 1/2 for every p_i^+ . It then intersects the response curve (77) of i only once. At this intersection, the value of p_{-i}^+ is still close to 1/2. Therefore, it is larger than the value at the lower equilibrium in the symmetric case, but smaller than the one at the upper equilibrium in that case. In particular, when $-i$ becomes less rational, that is, his β_{-i} becomes smaller, than the resulting equilibrium is better for him, and worse for i , than that equilibrium in the symmetric case that is more favorable to i . Thus, when $-i$ becomes less rational, i can no longer reach as good a situation as against a more rational opponent. However, the other stable equilibrium which is good for $-i$ and bad for i has disappeared by a bifurcation according to (58) when β_{-i} moves from larger to smaller values while β_i is kept fixed. The bifurcation here is a saddle-node bifurcation.

We can also view this as follows. When $-i$ stays rational, that is, β_{-i} is kept large while i becomes less rational, that is, β_i is lowered, then the two equilibria starting

from $(1, 1)$ and $\left(\frac{2}{3}, \frac{1}{3}\right)$ eventually merge and disappear by a saddle-node bifurcation, and only the one coming from $(0, 0)$, that is, the one coming from the NE less favorable for i , remains. When we reverse the role of the players, that is, keep i rational and let $-i$ become less rational, the opposite happens. When we move between these two saddle-node bifurcations in parameter space, there must be a transition through another bifurcation, generically a pitchfork bifurcation. And because of the symmetry, this pitchfork bifurcation will occur at a point in parameter space where $\beta_i = \beta_{-i}$.

The preceding qualitative analysis does not depend on the symmetry of the particular game (73), except for the location of the pitchfork bifurcation on the diagonal in parameter space. For instance, if we change it to

$$\begin{array}{c|cc} 3 & 1 & 0 \\ \hline 0 & | & 0 \\ 0 & | & 2 \end{array}, \quad (86)$$

we still obtain the same qualitative behavior, except that the pitchfork bifurcation no longer occurs at equal values of β_i and β_{-i} , but at certain values $\beta_i \neq \beta_{-i}$.

We now change the game (73) by taking the negative utility functions, i.e.,

$$\begin{array}{c|cc} 2 & 1 & 0 \\ \hline 0 & | & 0 \\ 0 & | & 2 \end{array}. \quad (87)$$

(Formally, this is equivalent to looking at negative values of the β s in the original game. Therefore, we can combine the QRE analysis for both games in a single plot.)

Analogously to (77), (78), the QREs are given by

$$p_i^+ = \frac{\exp(-\beta_i 2 p_{-i}^+)}{\exp(-\beta_i 2 p_{-i}^+) + \exp(-\beta_{-i}(1-p_{-i}^+))} =: f_i(p_{-i}^+) \quad (88)$$

$$p_{-i}^+ = \frac{\exp(-\beta_{-i} p_i^+)}{\exp(-\beta_{-i} 2(1-p_i^+)) + \exp(-\beta_{-i} p_i^+)} =: f_{-i}(p_i^+). \quad (89)$$

(I.e., we have redefined the functions f_i and f_{-i} according to the present game.)

As before, for large enough values of β_i and β_{-i} , we have 3 QREs, close the 3 Nash equilibria $(p_i^+, p_{-i}^+) = (0, 1), (1, 0), (\frac{2}{3}, \frac{1}{3})$. We observe that the first of the two pure NEs, $(0, 1)$ is preferred by the two players when they take the risk of the loss resulting from a switch of the opponent into account. (81) continues to hold, i.e.,

$$f_i\left(\frac{1}{3}\right) = \frac{1}{2} \text{ and } f_{-i}\left(\frac{2}{3}\right) = \frac{1}{2} \quad (90)$$

Since $f_i(p)$ and $f_{-i}(p)$ are now both monotonically decreasing functions that converge to the constant function $\equiv \frac{1}{2}$ when the β s go to 0, we see that for sufficiently small β s, the two intersection points starting from the Nash equilibria $(1, 0), (\frac{2}{3}, \frac{1}{3})$ will disappear by a saddle-node bifurcation whereas the one starting from the preferred NE $(0, 1)$ will continue into the final intersection $(\frac{1}{2}, \frac{1}{2})$ when the β s become

0. This implies that there will be no QRE branch in the β_i, β_{-i} space that connects the NE with either of the other ones. Those two other ones, are connected by such a branch. In particular, one cannot move from the more risky NE $(1, 0)$ to the less risky one $(0, 1)$ by tuning the β s in this example. (The details of the bifurcation analysis are not hard to verify, and the behavior can also be seen in the plots.)

7 Another Example – Not so Nice

Turocy [14] showed that the principal branch of any QRE correspondence satisfying a monotonicity property converges to the risk-dominant equilibrium in 2×2 games. In particular, not every NE is reached as a limit of a QR. With the tools that we have developed above, we can investigate this in detail at a concrete example. With notation as in (73), we consider the following pay-off matrix

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1 & 0 \\ 2 & 0 & 0 \end{array} \quad (91)$$

Here, i prefers the first move, +, except when $-i$ plays +, in which case she is indifferent. Likewise, $-i$ prefers +, except if i plays +, in which case she is indifferent. Thus, since both of them prefer their first move, there is a tendency to end up at $(1, 1)$ even though this is not Pareto optimal. This effect will now become clearer when we analyze the QREs.

With finite β s, analogously as in (77), (78), we put

$$p_i^+ = \frac{\exp(\beta_i(2-p_{-i}^+))}{\exp(\beta_i(2-p_{-i}^+))+\exp(\beta_{-i}p_{-i}^+)} =: f_i(p_{-i}^+) \quad (92)$$

$$p_{-i}^+ = \frac{\exp(\beta_{-i}(2-p_i^+))}{\exp(\beta_{-i}(2-p_i^+))+\exp(\beta_i p_i^+)} =: f_{-i}(p_i^+). \quad (93)$$

Then $f_i(p)$ is a decreasing function with

$$f_i(1) = \frac{1}{2}, \quad f_i(0) > \frac{1}{2}, \quad (94)$$

and the same holds for f_{-i} . Therefore, the intersection of their graphs, that is, the QRE, has to be contained in the upper quadrant, that is, it has to occur for

$$\frac{1}{2} < p_i^+ < 1, \quad \frac{1}{2} < p_{-i}^+ < 1. \quad (95)$$

This then also constrains the possible limits of QREs for $\beta_i, \beta_{-i} \rightarrow \infty$ to that region. In particular, those two Nash equilibria that are strict for at least one of the players, that is +, - and -, + are not limits of QREs. In particular, when we look at the

symmetric situation $\beta_i = \beta_{-i} \rightarrow \infty$, the limit is $+, +$ which is not strict as either player could increase the other's pay-off while keeping her/his own.

8 Controlling the Game

After having analyzed the preceding examples, we can now address the issue who can control the behavior of the players in a quantal response game in which way, and what effects can be achieved by such control.

The first possibility is that the individual players can set their β s, at least within a certain range. That is, players can for instance determine the value of their β s so as to maximize their expected pay-offs. They can either do this without regard for similar actions or counteractions of the other players, but rather assuming that the others keep their values constant. One might label such a scenario as "anarchy". Or they could try to anticipate the reactions of their opponents, as in classical game theory. When the ranges of the β s are constrained, the players thus play some game with a continuous move space. Again, they will find an equilibrium, but this will not be explored here.

A different possibility is that the β s are set by some external regulator, for instance with the aim of inducing the players to a collectively superior, perhaps Pareto optimal, equilibrium, different from the one that they may achieve when left to their own devices. Of course, it does not make much sense to assume that an external regulator can set the degrees of rationality of the players. But we can easily find an interpretation that avoids that problem, on the basis of the following simple observation. The QRE probabilities in (3) depend only on the products $\beta_i U_i$. Therefore, changing β_i by some factor λ has the same effect as, that is, is not distinguishable from multiplying the U_i s by that factor λ . But then, for instance, the factor λ could be 1 – the tax rate, which, of course, can be set in many circumstances by an external regulator, and perhaps even differently for different players. In other words, increasing taxes in quantal response games has the same effect as making players less rational. Thus, an external regulator could set up some tax rates for the players so as to achieve some collective aim.

More generally, in all three scenarios (anarchy, game, or regulator), the coefficients could be functions of time, that is, they could take different values at different times. For instance, in the anarchy scenario, after players have individually adapted their β s without anticipating the actions of other players, they may find themselves in a situation where they again prefer to change their β s, and perhaps they might even have a different range of options at that time. Also, the regulator could try to steer the participants along some path leading from one equilibrium to a better one. For that purpose, the regulator will have to understand the bifurcation behavior of the QREs as analyzed in our examples.

9 Channel Dependence

We now look at an asymmetric situation where player i has some information about the move of player $-i$ and can react accordingly. We observe that a pure Nash equilibrium can persist under this condition when $-i$ knows that i can react to his moves, as he can then anticipate that reaction. When the game possesses an equilibrium where $-i$ plays a mixed strategy, then he will be put at a disadvantage, as i has the possibility to react differently to the different moves in $-i$'s mixture. Consequently, he may change his strategy, and this may then turn out to be disadvantageous for i as well. For instance, let us consider a game with the following pay-offs, i being the row and $-i$ being the column player:

$$\begin{array}{cc|c} & & 10 | -2 & 0 | 2 \\ & 9 | 1 & 1 | -1 \end{array} \quad (96)$$

Without information about each other, each player can play at the Nash equilibrium, that is, at $p_i^+ = \frac{1}{3}$, $p_{-i}^+ = \frac{1}{2}$ in which case the expected pay-off for i is 5, the one for $-i$ 0. When i now knows which move $-i$ is playing, and $-i$ knows that i knows that, then $-i$ will always play his second move. In that case, his pay-off is reduced to -1 while the one of i goes down to 1. In particular, the pay-off of the player that is provided with the additional information about the other's move is more drastically reduced than that of the other one who only knows that his opponent has that information, but does not have the reciprocal information about her move.

Here, we are interested in situations where the information that i obtains is incomplete, that is, where it may get distorted by some channel noise. The channel puts out symbols $d \in D$ in response to the action γ of $-i$. i can then observe these symbols. We let $p_\gamma^d = p(d|\gamma)$ denote the probability that the symbol d appears when $-i$ chooses γ . i can thus select a mapping

$$a : D \rightarrow \{\alpha\}, \quad (97)$$

that is, select her actual move depending on the symbol d received. In principle, the choice of a symbol for a given a can also be probabilistic, but here we put the probabilities into the choice of the map a rather than into the operation of a . That is, the map a is chosen with a certain probability, for instance as given by (98) below, but when it is chosen, the reaction of i to the symbol d is determined. Of course, the reaction could ignore the symbol, that is, each map could select a single move α regardless of the symbol d . This then would simply amount to the choice of a move α , as considered in the previous sections. Thus, the present generalization consists in allowing for moves depending on some information about the opponent as contained in the symbol d . To repeat it once more: The choice of the map a does not depend on the actual symbol d received, only the actual move $\alpha = a(d)$ does.

We then have the following condition for a QRE: i selects a map a with probability

$$p_i^a = \frac{1}{Z_i} \exp(\beta_i \sum_{\gamma,d} U_i(x_i^{a(d)}, x_{-i}^\gamma) p_{-i}^\gamma p_\gamma^d), \quad (98)$$

while $-i$ plays as before with probabilities

$$p_{-i}^\gamma = \frac{1}{Z_{-i}} \exp(\beta_{-i} \sum_\alpha U_{-i}(x_{-i}^\gamma, x_i^\alpha) p_i^\alpha), \quad (99)$$

where now

$$p_i^\alpha = \sum_{a,d} \delta(\alpha = a(d)) p_i^a p(d) \quad (100)$$

with $p(d) = \sum_\gamma p_{-i}^\gamma p_\gamma^d$, since the map a is assumed deterministic; when the maps are not deterministic, this is simply replaced by $p_i^\alpha = \sum_{a,d} p(\alpha|a(d)) p_i^a p(d)$. Here, the map a is chosen with probability given by (98), and when the symbol d occurs, it has to be checked whether (or with which probability in the non-deterministic case) a takes the value α for that symbol.

We then also have

$$p_i(\alpha|\gamma) = \sum_{a,d} \delta(\alpha = a(d)) p_i^a p_\gamma^d. \quad (101)$$

According to the scheme developed in Sect. 5, we can compute the effect of a variation of a probability p_γ^d on the move probabilities of the players (in the sums below, η will run over the move set of $-i$, e over the channel symbol set D , and b over the set of maps of i from D to her move set):

$$\frac{dp_i^a}{dp_\gamma^d} = p_i^a \beta_i \sum_{\eta,e} \left(\frac{dp_{-i}^\eta}{dp_\gamma^d} p_\eta^e + p_{-i}^\eta \frac{dp_i^e}{dp_\gamma^d} \right) (U_i(x_i^{a(e)}, x_i^\eta) - \sum_b U_i(x_i^{b(e)}, x_i^\eta) p_i^b). \quad (102)$$

Here, we assume that $\frac{dp_\eta^e}{dp_\gamma^d} = 0$ for $\eta \neq \gamma$, that is, only the probability of the symbol d in response to the specific action γ of $-i$ changes, but not the response to other actions. Also, $\frac{dp_\gamma^d}{dp_\gamma^d} = 1$. When we only have two possible symbols, denoted by d and d^* , we also have $\frac{dp_{\gamma}^{d^*}}{dp_\gamma^d} = -1$.

When we neglect the variation of the probabilities of the opponent $-i$, that is, suppress $\frac{dp_{-i}^\eta}{dp_\gamma^d}$, which is an indirect effect, and write \sim in place of $=$ to indicate this, we get for the two symbol case

$$\begin{aligned} \frac{dp_i^a}{dp_\gamma^d} &\sim p_i^a \beta_i p_{-i}^\gamma (U_i(x_i^{a(d)}, x_i^\gamma) - \sum_b U_i(x_i^{b(d)}, x_i^\gamma) p_i^b \\ &\quad - (U_i(x_i^{a(d^*)}, x_i^\gamma) - \sum_b U_i(x_i^{b(d^*)}, x_i^\gamma) p_i^b)). \end{aligned} \quad (103)$$

We obtain the following – rather obvious – conclusion: When we neglect the reaction of the opponent, the probability of the strategy map a will increase when the symbol d occurs more frequently in response to γ if upon receipt of d it plays a better move against γ compared to the average over all strategy maps than upon receipt of the other symbol d^* .

We now consider

$$\frac{dp_{-i}^\eta}{dp_\gamma^d} = p_{-i}^\eta \beta_{-i} \sum_\alpha \frac{dp_i^\alpha}{dp_\gamma^d} (U_{-i}(x_{-i}^\eta, x_i^\alpha) - \sum_\lambda U_{-i}(x_{-i}^\lambda, x_i^\alpha) p_{-i}^\lambda), \quad (104)$$

cf. (38). Since $p_i^\alpha = \sum_{a,d} \delta(\alpha = a(d)) p_i^a \sum_\gamma p_{-i}^\gamma p_\gamma^d$ (cf. (100)), we have, again for the case of only two symbols,

$$\frac{dp_i^\alpha}{dp_\gamma^d} = \sum_{a,e} \frac{dp_i^a}{dp_\gamma^d} \delta(\alpha = a(e)) \sum_\lambda p_{-i}^\lambda p_\lambda^e + \sum_a (\delta(\alpha = a(d)) - \delta(\alpha = a(d^*))) p_i^a p_{-i}^\gamma. \quad (105)$$

Here, the first part comes from the variation of i 's map, whereas the second term arises from the change of the frequency of the symbol d .

From (104), we see that the probability of the move η will decrease when on average the probability of those moves α of i will increase against which η is not a good response. Again, this is obvious.

In any case, the change of equilibrium caused by the difference in expected payoffs for $-i$ could well counteract the gain of i from more precise information about the behavior of her opponent. In particular, when a symbol d becomes more frequent upon the move γ , and that symbol favors a move of i that is disadvantageous for $-i$, then $-i$ may decrease the probability of that specific move γ . This is the same mechanism as in the deterministic game discussed in the beginning of this section. Of course, if $-i$ is ignorant about the fact that i possesses information about his moves, then he will not change his strategy, and i can then benefit from the information according to (103).

References

1. Amari, S.: Information Geometry and Its Applications. Applied Mathematical Sciences, vol. 194. Springer, Berlin (2016)
2. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. Ergebnisse der Mathematik. Springer, Berlin (2017)
3. Giacinta, M., Modica, G., Souček, J.: Cartesian Currents in the Calculus of Variations I. Ergebnisse, vol. 37. Springer, Berlin (1998)
4. Heuser, H.: Analysis II, Teubner (1980)
5. Kohlberg, E., Mertens, J.-F.: On the strategic stability of equilibria. *Econometrica* **54**, 1003–1037 (1986)
6. Lemke, C.E., Howson Jr., J.T.: Equilibrium points in bimatrix games. *SIAM J. Appl. Math.* **12**, 413–423 (1964)

7. McKelvey, R., Palfrey, T.: Quantal response equilibria for normal form games. *Games Econ. Behav.* **10**, 6–38 (1995)
8. McKelvey, R., Palfrey, T.: A statistical theory of equilibrium in games. *Jpn. Econ. Rev.* **47**, 186–209 (1996)
9. Nash, J.: Equilibrium points in n -person games. *Proc. Natl. Acad. Sci.* **36**, 48–49 (1950)
10. Oikonomou, V., Jost, J.: Periodic Strategies. A New Solution Concept and an Algorithm for Non-trivial Strategic Form Games (to appear)
11. Stein, N.: Error in Nash existence proof, 13 Sep 2010. [arXiv:1005.3045v3](https://arxiv.org/abs/1005.3045v3) [cs.GT]
12. Stein, N., Parillo, P., Ozdaglar, A.: A new proof of Nash's theorem via exchangeable equilibria, 13 Sep 2010. [arXiv:1005.3045v3](https://arxiv.org/abs/1005.3045v3) [cs.GT]
13. Stöcker, R., Zieschang, H.: *Algebraische Topologie*, Teubner (1994)
14. Turocy, T.: A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games Econ. Behav.* **51**, 243–263 (2005)
15. Wilson, R.: Computing equilibria of N -person games. *SIAM J. Appl. Math.* **21**, 80–87 (1971)
16. Wolpert, D.H., Harre, M., Olbrich, E., Bertschinger, N., Jost, J.: Hysteresis effects of changing parameters in noncooperative games. *Phys. Rev. E* **85**, 036102 (2012). <https://doi.org/10.1103/PhysRevE.85.036102>
17. Zeidler, E.: *Nonlinear Functional Analysis and Its Applications. IV: Applications to Mathematical Physics*. Springer, Berlin (1985)

Higher Order Equivalence of Bayes Cross Validation and WAIC



Sumio Watanabe

Abstract It was proved in the previous paper (Watanabe, J Mach Learn Res, 11:3571–3591, (2010), [16]) that Bayes cross validation is asymptotically equivalent to the widely applicable information criterion (WAIC), even if the posterior distribution can not be approximated by any normal distribution. In the present paper, we prove that they are equivalent to each other according to the second order asymptotics, if the posterior distribution can be approximated by some normal distribution. Based on this equivalence, it is also shown that the Bayes cross validation and WAIC are asymptotically equivalent as functions of a hyperparameter of a prior.

Keywords Cross validation · WAIC · Asymptotic equivalence · Second order

1 Introduction

In this section, we introduce the background of the Bayesian statistical inference [5] and the main result of this paper.

Let $q(x)$ be a probability density function on the N dimensional real Euclidean space \mathbb{R}^N , and X_1, X_2, \dots, X_n be independent random variables on \mathbb{R}^N which are distributed according to $q(x)$. We assume that the set of all zero points of $q(x)$ has the zero measure on \mathbb{R}^N . A sample is denoted by $X^n = (X_1, X_2, \dots, X_n)$, where n is the number of independent random variables or the sample size. The average $\mathbb{E}[\cdot]$ shows the expectation value over all training sets X^n .

A statistical model or a learning machine is defined by $p(x|w)$ which is a probability density function of $x \in \mathbb{R}^N$ for a given parameter $w \in W \subset \mathbb{R}^d$. Also we assume that the set of all zero points of $p(x|w)$ has the zero measure on \mathbb{R}^N for an arbitrary w . The true distribution $q(x)$ is said to be realizable by $p(x|w)$ if there exists a parameter w such that $q(x) = p(x|w)$. In this paper, we do not assume that $q(x)$

S. Watanabe (✉)

Tokyo Institute of Technology, Tokyo, Japan
e-mail: swatanab@c.titech.ac.jp

is realizable by $p(x|w)$ in general. A nonnegative function $\varphi(w)$ on the parameter set W is called an improper prior. In other words, we study general cases

$$\int \varphi(w) dw \neq 1. \quad (1)$$

The posterior distribution of w for a given training sample X^n is denoted by

$$p(w|X^n) = \frac{1}{Z(\varphi)} \varphi(w) \prod_{i=1}^n p(X_i|w), \quad (2)$$

where $Z(\varphi)$ is a normalizing constant.

$$Z(\varphi) = \int \varphi(w) \prod_{i=1}^n p(X_i|w) dw.$$

We assume that $Z(\varphi)$ is finite with probability one for $n \geq n_0$ for some $n_0 > 0$ and the sample size n is assumed to be $n > n_0 + 1$. The expectation value of a given function $f(w)$ over the posterior distribution is defined by

$$\mathbb{E}_\varphi[f(w)] = \int f(w) p(w|X^n) dw, \quad (3)$$

where the used prior φ is explicitly shown as \mathbb{E}_φ . The predictive distribution is defined by the conditional density of x for a given training set X^n which is the average of a statistical model over the posterior distribution,

$$p(x|X^n) = \mathbb{E}_\varphi[p(x|w)]. \quad (4)$$

Note that $p(w|X^n)$ and $p(x|X^n)$ are density functions of w and x respectively, which are different functions. The generalization loss is defined by

$$G(\varphi) = - \int q(x) \log p(x|X^n) dx, \quad (5)$$

where $q(x)$ is the true distribution. Note that $G(\varphi)$ is not a constant but a random variable. The average generalization loss is defined by $\mathbb{E}[G(\varphi)]$. The Bayesian leave-one-out cross validation (CV) [4, 12, 13] is defined by

$$\text{CV}(\varphi) = - \frac{1}{n} \sum_{i=1}^n \log p(X_i|X^n \setminus X_i) \quad (6)$$

$$= \frac{1}{n} \sum_{i=1}^n \log \mathbb{E}_\varphi \left[\frac{1}{p(X_i|w)} \right], \quad (7)$$

where $X^n \setminus X_i$ is a sample leaving X_i out. The equivalence between Eqs. (6) and (7) are shown in [16]. A numerical calculation method of CV by Eq. (7) using the posterior distribution by the Markov chain Monte Carlo method is sometimes called the importance sampling cross validation, whose divergence phenomenon and its improvement have been studied [3, 6, 9, 12, 13]. The training loss $T(\varphi)$ and the functional variance $V(\varphi)$ are respectively defined by

$$T(\varphi) = -\frac{1}{n} \sum_{i=1}^n \log \mathbb{E}_\varphi[p(X_i|w)], \quad (8)$$

$$V(\varphi) = \frac{1}{n} \sum_{i=1}^n \{\mathbb{E}_\varphi[(\log p(X_i|w))^2] - \mathbb{E}_\varphi[\log p(X_i|w)]^2\}. \quad (9)$$

Note that $T(\varphi) \leq \text{CV}(\varphi)$ holds in general by Cauchy–Schwarz inequality. The widely applicable information criterion (WAIC) [15–17] is defined by

$$\text{WAIC}(\varphi) = T(\varphi) + V(\varphi). \quad (10)$$

For a real number α , the functional cumulant function is defined by

$$F_{\text{cum}}(\alpha) = \frac{1}{n} \sum_{i=1}^n \log \mathbb{E}_\varphi[p(X_i|w)^\alpha]. \quad (11)$$

Then, as is shown in [16],

$$\text{CV}(\varphi) = F_{\text{cum}}(-1), \quad (12)$$

$$T(\varphi) = -F_{\text{cum}}(1), \quad (13)$$

$$V(\varphi) = F_{\text{cum}}''(0), \quad (14)$$

$$\text{WAIC}(\varphi) = -F_{\text{cum}}(1) + F_{\text{cum}}''(0). \quad (15)$$

In the previous papers [16, 17], we proved by singular learning theory that, even if $q(x)$ is unrealizable by a statistical model or even if the posterior distribution can not be asymptotically approximated by any normal distribution,

$$\begin{aligned} \mathbb{E}[\text{CV}(\varphi)] &= \mathbb{E}[G(\varphi)] + O\left(\frac{1}{n^2}\right), \\ \mathbb{E}[\text{WAIC}(\varphi)] &= \mathbb{E}[G(\varphi)] + O\left(\frac{1}{n^2}\right), \\ \text{WAIC}(\varphi) &= \text{CV}(\varphi) + O_p\left(\frac{1}{n^2}\right). \end{aligned}$$

However, it has been left unknown the effect by the choice of the prior to CV, WAIC, and the generalization loss.

In this paper, we prove that, if the posterior distribution can be approximated by some normal distribution, then for arbitrary priors $\varphi(w)$ and $\varphi_0(w)$,

$$\text{CV}(\varphi) = \text{CV}(\varphi_0) + O_p\left(\frac{1}{n^2}\right), \quad (16)$$

$$\text{WAIC}(\varphi) = \text{WAIC}(\varphi_0) + O_p\left(\frac{1}{n^2}\right), \quad (17)$$

$$\text{WAIC}(\varphi) = \text{CV}(\varphi) + O_p\left(\frac{1}{n^3}\right), \quad (18)$$

resulting that WAIC and CV are asymptotically equivalent according to prior evaluation. Also we clarify the concrete functions $O_p(1/n^2)$ in Eqs. (16), (17).

Remark In the hyperparameter optimization problem, the minimization of the minus log marginal likelihood or the free energy

$$F_{\text{free}}(\varphi) = -\log \int \varphi(w) \prod_{i=1}^n p(X_i|w) dw \quad (19)$$

is sometimes studied [1, 7]. In order to employ this method, a prior should be proper, because, if it is not proper, minimization of $F_{\text{free}}(\varphi)$ has no meaning, since $F_{\text{free}}(\varphi) \rightarrow -\infty$ by $\varphi(w) \rightarrow \infty$. In this paper, we show that the hyperparameter which minimizes $F_{\text{free}}(\varphi)$ among proper priors does not minimize either $\mathbb{E}[G(\varphi)]$ or $G(\varphi)$ even asymptotically. The optimal prior for a minimax problem and the higher order asymptotics of the Bayes generalization loss were investigated in [2, 8], respectively. This paper firstly clarifies the higher order asymptotics of the Bayes cross validation and WAIC.

2 Main Results

2.1 Definitions and Conditions

In this section, we introduce several notations, regularity conditions, and definitions of mathematical relations between priors.

The set of parameters W is assumed to be an open subset of \mathbb{R}^d . In this paper, $\varphi_0(w)$ and $\varphi(w)$ are arbitrary fixed and alternative priors, respectively. We assume that, for an arbitrary $w \in W$, $\varphi_0(w) > 0$ and $\varphi(w) > 0$. They are improper in general. The prior ratio function $\phi(w)$ is denoted by

$$\phi(w) = \frac{\varphi(w)}{\varphi_0(w)}.$$

If $\varphi_0(w) \equiv 1$, then $\phi(w) = \varphi(w)$. The empirical log loss function and the maximum a posteriori (MAP) estimator \hat{w} are respectively defined by

$$L(w) = -\frac{1}{n} \sum_{i=1}^n \log p(X_i|w) - \frac{1}{n} \log \varphi_0(w), \quad (20)$$

$$\hat{w} = \arg \min_{w \in W} L(w), \quad (21)$$

where either $L(w)$ or \hat{w} does not depend on $\varphi(w)$. If $\varphi_0(w) \equiv 1$, then \hat{w} is equal to the maximum likelihood estimator (MLE). The average log loss function and the parameter that minimizes it are respectively defined by

$$\mathcal{L}(w) = - \int q(x) \log p(x|w) dx, \quad (22)$$

$$w_0 = \arg \min_{w \in W} \mathcal{L}(w). \quad (23)$$

In this paper we use the following notations for simple description.

Notations

(1) A parameter is denoted by $w = (w^1, w^2, \dots, w^k, \dots, w^d) \in \mathbb{R}^d$. Remark that w^k means the k th element of w , which does not mean w to the power of k .

(2) For a given real function $f(w)$ and nonnegative integers k_1, k_2, \dots, k_m , we define

$$f_{k_1 k_2 \dots k_m} = f_{k_1 k_2 \dots k_m}(w) = \frac{\partial^m f}{\partial w^{k_1} \partial w^{k_2} \dots \partial w^{k_m}}(w). \quad (24)$$

(3) We adopt Einstein's summation convention and k_1, k_2, k_3, \dots are used for such suffices. For example,

$$X_{k_1 k_2} Y^{k_2 k_3} = \sum_{k_2=1}^d X_{k_1 k_2} Y^{k_2 k_3}.$$

In other words, if a suffix k_i appears both upper and lower, it means automatic summation over $k_i = 1, 2, \dots, d$. In this paper, for each k_1, k_2 , $X^{k_1 k_2} = X_{k_2}^{k_1} = X_{k_1 k_2}$.

In order to prove the main theorem, we need the regularity conditions. In this paper, we do not study singular learning theory.

Regularity Conditions

- (1) **(Parameter Set)** The parameter set W is an open set in \mathbb{R}^d .
- (2) **(Smoothness of Models)** The functions $\log \varphi(w)$, $\log \varphi_0(w)$, and $\log p(x|w)$ are C^∞ -class functions of $w \in W$, in other words, they are infinitely many times differentiable.
- (3) **(Identifiability of Parameter)** There exists a unique $w_0 \in W$ which minimizes the average log loss function $\mathcal{L}(w)$. There exists a unique $\hat{w} \in W$ which minimizes

$L(w)$ with probability one. It is assumed that the convergence in probability $\hat{w} \rightarrow w_0$ ($n \rightarrow \infty$) holds.

(4) **(Regularity Condition)** The matrix $\mathcal{L}_{k_1 k_2}(w_0)$ is invertible. Also the matrix $L_{k_1 k_2}(w)$ is invertible for almost all w in a neighborhood of w_0 with probability one. Let $J^{k_1 k_2}(w)$ be the inverse matrix of $L_{k_1 k_2}(w)$.

(5) **(Well-Definedness and Concentration of Posterior)** We assume that, for an arbitrary $|\alpha| \leq 1$ and $j = 1, 2, \dots, n+1$,

$$\mathbb{E}_{X_{n+1}} \mathbb{E} \left[\left| \log \mathbb{E}_\varphi [p(X_j | w)^\alpha] \right| \right] < \infty. \quad (25)$$

The same inequality as Eq.(25) holds for $\varphi_0(w)$ instead of $\varphi(w)$. Let $Q(X^n, w)$ be an arbitrary finite times product of

$$\begin{aligned} & (\log \varphi(w))_{k_1 k_2 \dots k_p}, \\ & (\log \varphi_0(w))_{k_1 k_2 \dots k_q}, \\ & \frac{1}{n} \sum_{i=1}^n \prod (\log p(X_i | w))_{k_1 k_2 \dots k_r}, \\ & (J^{k_1 k_2}(w))_{k_1 k_2 \dots k_s}, \\ & w^{k_1}, \end{aligned}$$

where $|\alpha| \leq 1$, $p, q, r, s \geq 0$ and \prod shows a finite product of a combination (k_1, k_2, \dots, k_r) . Let

$$W(\varepsilon) = \{w \in W; |w - \hat{w}| < n^{\varepsilon-1/2}\}. \quad (26)$$

It is assumed that there exists $\varepsilon > 0$, for an arbitrary such product $Q(X^n, w)$,

$$\mathbb{E}[\sup_{W(\varepsilon)} |Q(X^n, w)|] < \infty, \quad (27)$$

$$\mathbb{E}[Q(X^n, \hat{w})] \rightarrow \mathbb{E}[Q(X^n, w_0)], \quad (28)$$

and that, for arbitrary $|\alpha| \leq 1$ and $\beta > 0$,

$$\begin{aligned} & \frac{\mathbb{E}_\varphi [Q(X^n, w) p(X_j | w)^\alpha]}{\mathbb{E}_\varphi [p(X_j | w)^\alpha]} \\ &= \left(1 + o_p\left(\frac{1}{n^\beta}\right)\right) \frac{\int_{W(\varepsilon)} Q(X^n, w) p(X_j | w)^\alpha \prod_{i=1}^n p(X_i | w) \varphi(w) dw}{\int_{W(\varepsilon)} p(X_j | w)^\alpha \prod_{i=1}^n p(X_i | w) \varphi(w) dw}, \end{aligned} \quad (29)$$

where $o_p(1/n^\beta)$ satisfies $n^\beta \mathbb{E}[|o_p(1/n^\beta)|] \rightarrow 0$. Also we assume that the same equation as Eq. (29) holds for $\varphi_0(w)$ instead of $\varphi(w)$.

Explanation of Regularity Condition. (1) In this paper, we assume that $p(x|w)$ is regular at w_0 , that is to say, the second order matrix $\mathcal{L}_{k_1 k_2}(w_0)$ is positive definite. If this condition is not satisfied, then such $(q(x), p(x|w))$ is called singular. The results of this paper do not hold for singular learning machines.

(2) Conditions Eqs. (27) and (28) ensure the finiteness of the expectation values and concentration of the posterior distribution. The condition of the concentration, Eq. (29), is set by the following mathematical reason. Let $S(w)$ be a function which takes the minimum value $S(\hat{w}) = 0$ at $w = \hat{w}$. If $S_{k_1 k_2}(\hat{w})$ is positive definite, then by using the saddle point approximation in the neighborhood of \hat{w} ,

$$\exp(-nS(w)) \approx \exp\left(-\frac{n}{2} S_{k_1 k_2}(\hat{w})(w - \hat{w})^{k_1}(w - \hat{w})^{k_2}\right),$$

hence the orders of integrations inside and outside of $W(\epsilon)$ are respectively given by

$$\begin{aligned} \int_{W(\epsilon)} \exp(-nS(w)) dw &= O(1/n^{d/2}), \\ \int_{W \setminus W(\epsilon)} \exp(-nS(w)) dw &= O(\exp(-n^\varepsilon)). \end{aligned}$$

Therefore the integration over $W \setminus W(\varepsilon)$ converges to zero faster than that over $W(\epsilon)$ as $n \rightarrow \infty$.

Definition (Empirical Mathematical Relations between Priors) The prior ratio is defined by $\phi(w) = \varphi(w)/\varphi_0(w)$. Then the empirical mathematical relation between two priors $\varphi(w)$ and $\varphi_0(w)$ at a parameter w is defined by

$$M(\phi, w) = A^{k_1 k_2} (\log \phi)_{k_1} (\log \phi)_{k_2} + B^{k_1 k_2} (\log \phi)_{k_1 k_2} + C^{k_1} (\log \phi)_{k_1}, \quad (30)$$

where

$$\begin{aligned} J^{k_1 k_2}(w) &= \text{Inverse matrix of } L_{k_1 k_2}(w), \\ A^{k_1 k_2}(w) &= \frac{1}{2} J^{k_1 k_2}(w), \\ B^{k_1 k_2}(w) &= \frac{1}{2} (J^{k_1 k_2}(w) + J^{k_1 k_3}(w) J^{k_2 k_4}(w) F_{k_3, k_4}(w)), \\ C^{k_1}(w) &= J^{k_1 k_2}(w) J^{k_3 k_4}(w) F_{k_2 k_4, k_3}(w) - \frac{1}{2} J^{k_1 k_2}(w) J^{k_3 k_4}(w) L_{k_2 k_3 k_4}(w), \\ &\quad - \frac{1}{2} J^{k_1 k_2}(w) J^{k_3 k_4}(w) J^{k_5 k_6}(w) L_{k_2 k_3 k_5}(w) F_{k_4, k_6}(w), \end{aligned}$$

where $L_{k_1 k_2}(w)$ and $L_{k_1 k_2 k_3}(w)$ are the second and third derivatives of $L(w)$ respectively as defined by Eq. (24) and

$$F_{k_1, k_2}(w) = \frac{1}{n} \sum_{i=1}^n (\log p(X_i|w))_{k_1} (\log p(X_i|w))_{k_2},$$

$$F_{k_1 k_2, k_3}(w) = \frac{1}{n} \sum_{i=1}^n (\log p(X_i|w))_{k_1 k_2} (\log p(X_i|w))_{k_3}.$$

Remark By the definition, $M(1, w) = 0$. Note that neither $A^{k_1 k_2}(w)$, $B^{k_1 k_2}(w)$, nor $C^{k_1}(w)$ depends on any prior $\varphi(w)$ or $\varphi_0(w)$.

Definition (*Average Mathematical Relations between Priors*) The average mathematical relation $\mathcal{M}(\phi, w)$ is defined by the same manner as Eq. (30) by using

$$\mathcal{J}^{k_1 k_2}(w) = \text{Inverse matrix of } \mathcal{L}_{k_1 k_2}(w), \quad (31)$$

$$\mathcal{L}_{k_1 k_2}(w) = \int (-\log p(x|w))_{k_1 k_2} q(x) dx, \quad (32)$$

$$\mathcal{L}_{k_1 k_2 k_3}(w) = \int (-\log p(x|w))_{k_1 k_2 k_3} q(x) dx, \quad (33)$$

$$\mathcal{F}_{k_1, k_2}(w) = \int (\log p(x|w))_{k_1} (\log p(x|w))_{k_2} q(x) dx, \quad (34)$$

$$\mathcal{F}_{k_1 k_2, k_3}(w) = \int (\log p(x|w))_{k_1 k_2} (\log p(x|w))_{k_3} q(x) dx, \quad (35)$$

instead of $J^{k_1 k_2}(w)$, $L_{k_1 k_2}(w)$, $L_{k_1 k_2 k_3}(w)$, $F_{k_1, k_3}(w)$, and $F_{k_1 k_2, k_3}(w)$ respectively. The self-average mathematical relation $\langle M \rangle(\phi, w)$ is defined by the same manner as $M(\phi, w)$ by using

$$\langle J^{k_1 k_2} \rangle(w) = \text{Inverse matrix of } \langle L_{k_1 k_2} \rangle(w), \quad (36)$$

$$\langle L_{k_1 k_2} \rangle(w) = \int (-\log p(x|w))_{k_1 k_2} p(x|w) dx, \quad (37)$$

$$\langle L_{k_1 k_2 k_3} \rangle(w) = \int (-\log p(x|w))_{k_1 k_2 k_3} p(x|w) dx, \quad (38)$$

$$\langle F_{k_1, k_2} \rangle(w) = \int (\log p(x|w))_{k_1} (\log p(x|w))_{k_2} p(x|w) dx, \quad (39)$$

$$\langle F_{k_1 k_2, k_3} \rangle(w) = \int (\log p(x|w))_{k_1 k_2} (\log p(x|w))_{k_3} p(x|w) dx, \quad (40)$$

instead of $J^{k_1 k_2}(w)$, $L_{k_1 k_2}(w)$, $L_{k_1 k_2 k_3}(w)$, $F_{k_1, k_3}(w)$, and $F_{k_1 k_2, k_3}(w)$ respectively.

Remark In the self-average case, it holds that $\langle L_{k_1 k_2} \rangle(w) = \langle F_{k_1, k_2} \rangle(w)$, hence $\langle M \rangle(\phi, w)$ can be calculated by the same manner as Eq. (30) by using

$$\begin{aligned}\langle A^{k_1 k_2} \rangle(w) &= \frac{1}{2} \langle J^{k_1 k_2} \rangle(w), \\ \langle B^{k_1 k_2} \rangle(w) &= \langle J^{k_1 k_2} \rangle(w), \\ \langle C^{k_1} \rangle(w) &= \langle J^{k_1 k_2} \rangle(w) \langle J^{k_3 k_4} \rangle(w) \langle F_{k_2 k_4, k_3} \rangle(w) \\ &\quad - \langle J^{k_1 k_2} \rangle(w) \langle J^{k_3 k_4} \rangle(w) \langle L_{k_2 k_3 k_4} \rangle(w).\end{aligned}$$

instead of $A^{k_1 k_2}(w)$, $B^{k_1 k_2}(w)$ and $C^{k_1}(w)$.

2.2 Main Theorem

The following is the main result of this paper.

Theorem 1 Assume the regularity conditions (1), (2), ..., and (5). Let $M(\phi, w)$ and $\mathcal{M}(\phi, w)$ be the empirical and average mathematical relations between $\varphi(w)$ and $\varphi_0(w)$, where $\phi(w) = \varphi(w)/\varphi_0(w)$. Then

$$\text{CV}(\varphi) = \text{CV}(\varphi_0) + \frac{M(\phi, \hat{w})}{n^2} + O_p\left(\frac{1}{n^3}\right), \quad (41)$$

$$\mathbb{E}[\text{CV}(\varphi)] = \mathbb{E}[\text{CV}(\varphi_0)] + \frac{\mathcal{M}(\phi, w_0)}{n^2} + O\left(\frac{1}{n^3}\right), \quad (42)$$

$$\text{WAIC}(\varphi) = \text{WAIC}(\varphi_0) + \frac{M(\phi, \hat{w})}{n^2} + O_p\left(\frac{1}{n^3}\right), \quad (43)$$

$$\mathbb{E}[\text{WAIC}(\varphi)] = \mathbb{E}[\text{WAIC}(\varphi_0)] + \frac{\mathcal{M}(\phi, w_0)}{n^2} + O\left(\frac{1}{n^3}\right), \quad (44)$$

$$\text{CV}(\varphi) = \text{WAIC}(\varphi) + O_p\left(\frac{1}{n^3}\right), \quad (45)$$

and

$$M(\phi, \hat{w}) = \mathcal{M}(\phi, w_0) + O_p\left(\frac{1}{n^{1/2}}\right), \quad (46)$$

$$M(\phi, \mathbb{E}_w[w]) = M(\phi, \hat{w}) + O_p\left(\frac{1}{n}\right), \quad (47)$$

$$\mathbb{E}[M(\phi, \hat{w})] = \mathcal{M}(\phi, w_0) + O\left(\frac{1}{n}\right). \quad (48)$$

On the other hand,

$$\begin{aligned}G(\varphi) &= G(\varphi_0) + \frac{1}{n} (\hat{w}^{k_1} - (w_0)^{k_1}) (\log \phi)_{k_1}(\hat{w}) + O_p\left(\frac{1}{n^2}\right) \\ &= G(\varphi_0) + O_p\left(\frac{1}{n^{3/2}}\right),\end{aligned} \quad (49)$$

$$\mathbb{E}[G(\varphi)] = \mathbb{E}[G(\varphi_0)] + \frac{\mathcal{M}(\phi, w_0)}{n^2} + o\left(\frac{1}{n^3}\right). \quad (50)$$

From Theorem 1, the five mathematical facts are derived.

- (1) Assume that a prior $\varphi(w)$ has a hyperparameter. Let $h(f)$ be the hyperparameter that minimizes a given function $f(\varphi)$. By Eqs.(41) and (43), $h(\text{CV})$ and $h(\text{WAIC})$ can be directly found by minimizing the empirical mathematical relation $M(\phi, \hat{w})$ asymptotically. By Eqs.(46) and (50), $h(\text{CV})$ and $h(\text{WAIC})$ is asymptotically equal to $h(\mathbb{E}[G])$.
- (2) The variance of $h(\text{CV})$ is asymptotically equal to that of $h(\text{WAIC})$, however, the former is larger than the latter when the sample size n is finite, in experiments.
- (3) In calculation of the mathematical relation $M(\phi, \hat{w})$, the MAP estimator \hat{w} can be replaced by the posterior average parameter $\mathbb{E}_w[w]$ asymptotically.
- (4) By Eq.(49), the variance of the random generalization loss $G(\varphi) - G(\varphi_0)$ is larger than those of $\text{CV}(\varphi) - \text{CV}(\varphi_0)$ and $\text{WAIC}(\varphi) - \text{WAIC}(\varphi_0)$.
- (5) It was proved in [14, 15] that

$$\mathbb{E}[G(\varphi_0)] = \text{constant} + d/(2n) + o(1/n),$$

where d is the dimension of the parameter set, even if $q(x)$ is not realizable by $p(x|w)$. Assume that there exist finite sets of real values $\{d_k\}$ and $\{\gamma_k\}$, where $\gamma_k > 1$, such that

$$\mathbb{E}[G(\varphi_0)] = \frac{d}{2n} + \sum_k \frac{d_k}{n^{\gamma_k}} + o\left(\frac{1}{n^2}\right).$$

Since $\mathbb{E}[\text{CV}(\varphi_0)]$ of X^n is equal to $\mathbb{E}[G(\varphi_0)]$ of X^{n-1} and

$$\frac{1}{n-1} - \frac{1}{n} = \frac{1}{n^2} + o\left(\frac{1}{n^2}\right),$$

it immediately follows from Theorem 1 that

$$\mathbb{E}[G(\varphi)] = \mathbb{E}[G(\varphi_0)] + \frac{\mathcal{M}(\phi, w_0)}{n^2} + o\left(\frac{1}{n^2}\right), \quad (51)$$

$$\mathbb{E}[\text{CV}(\varphi)] = \mathbb{E}[G(\varphi_0)] + \frac{d/2 + \mathcal{M}(\phi, w_0)}{n^2} + o\left(\frac{1}{n^2}\right), \quad (52)$$

$$\mathbb{E}[\text{WAIC}(\varphi)] = \mathbb{E}[G(\varphi_0)] + \frac{d/2 + \mathcal{M}(\phi, w_0)}{n^2} + o\left(\frac{1}{n^2}\right). \quad (53)$$

Theorem 2 *Assume the regularity conditions (1), (2),..., and (5). If there exists a parameter w_0 such that $q(x) = p(x|w_0)$, then*

$$\langle M \rangle(\phi, \hat{w}) = M(\phi, \hat{w}) + O_p\left(\frac{1}{\sqrt{n}}\right), \quad (54)$$

$$\langle M \rangle(\phi, \hat{w}) = \mathcal{M}(\phi, w_0) + O_p\left(\frac{1}{\sqrt{n}}\right). \quad (55)$$

By Theorem 2, if the true distribution is realizable by a statistical model or a learning machine, then the empirical mathematical relation can be replaced by its self-average. The variance of the self-average mathematical relation is often smaller than the original one, hence the variance of the estimated hyperparameter by using the self-average is made smaller.

Based on Theorems 1 and 2, we define new information criteria for hyperparameter optimization, the widely applicable information criterion for a regular case and a regular case using self-average,

$$\text{WAICR} = \frac{M(\phi, \hat{w})}{n^2}, \quad (56)$$

$$\text{WAICRS} = \frac{\langle M \rangle(\phi, \hat{w})}{n^2}, \quad (57)$$

where \hat{w} can be replaced by $\mathbb{E}_{\varphi_0}[w]$. If $q(x)$ is realizable by $p(x|w)$, then the optimal hyperparameter can be chosen by minimization of WAICRS.

3 Example

A simple but nontrivial example is a normal distribution whose mean and standard deviation are $(m, 1/s)$,

$$p(x|m, s) = \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{s}{2}(x - m)^2\right). \quad (58)$$

For a prior distribution, we study

$$\varphi(m, s|\lambda, \mu, \varepsilon) = s^\mu \exp\left(-\frac{\lambda sm^2 + \varepsilon s}{2}\right), \quad (59)$$

where $(\lambda, \mu, \varepsilon)$ is a set of hyperparameters. Note that the prior is improper in general. If $\lambda > 0$, $\mu > -1/2$ and $\varepsilon > 0$, the prior can be made proper by

$$\Phi(m, s|\lambda, \mu, \varepsilon) = \frac{1}{C} \varphi(m, s|\lambda, \mu, \varepsilon),$$

where

$$C = \sqrt{\frac{2\pi}{\lambda}} (\varepsilon/2)^{-\mu-1/2} \Gamma(\mu + 1/2).$$

We use a fixed prior as $\varphi_0(m, s) \equiv 1$, then the empirical log loss function is given by

$$L(m, s) = -\frac{1}{2} \log \frac{s}{2\pi} + \frac{s}{2n} \sum_{i=1}^n (X_i - m)^2. \quad (60)$$

Let $M_j = (1/n) \sum_{i=1}^n (X_i - \hat{m})^j$ ($j = 2, 3, 4$). The MAP estimator is equal to the MLE $\hat{w} = (\hat{m}, \hat{s})$, where $\hat{s}^2 = 1/M_2$, resulting that

$$A^{k_1 k_2}(\hat{w}) = \begin{pmatrix} 1/(2\hat{s}) & 0 \\ 0 & \hat{s}^2 \end{pmatrix}, \quad (61)$$

$$B^{k_1 k_2}(\hat{w}) = \begin{pmatrix} 1/\hat{s} & -\hat{s}^2 M_3/2 \\ -\hat{s}^2 M_3/2 & (\hat{s}^2 + \hat{s}^4 M_4)/2 \end{pmatrix}, \quad (62)$$

$$C^{k_1}(\hat{w}) = (0, \hat{s} + \hat{s}^3 M_3). \quad (63)$$

Also the self-average mathematical relation is given by

$$\langle A^{k_1 k_2} \rangle(\hat{w}) = \begin{pmatrix} 1/(2\hat{s}) & 0 \\ 0 & \hat{s}^2 \end{pmatrix}, \quad (64)$$

$$\langle B^{k_1 k_2} \rangle(\hat{w}) = 2 \langle A^{k_1 k_2} \rangle(\hat{w}), \quad (65)$$

$$\langle C^{k_1} \rangle(\hat{w}) = (0, \hat{s}). \quad (66)$$

The prior ratio function is $\phi(w) = \varphi(w)$, hence the derivatives of the log prior ratio are

$$(\log \phi)_1(\hat{w}) = -\lambda \hat{s} \hat{m},$$

$$(\log \phi)_2(\hat{w}) = -\frac{\lambda \hat{m}^2}{2} + \mu/\hat{s} - \frac{\varepsilon}{2},$$

$$(\log \phi)_{11}(\hat{w}) = -\lambda \hat{s},$$

$$(\log \phi)_{12}(\hat{w}) = -\lambda \hat{m},$$

$$(\log \phi)_{22}(\hat{w}) = -\mu/\hat{s}^2.$$

Therefore, the empirical and self-average mathematical relations are respectively

$$\begin{aligned}
M(\phi, \hat{m}, \hat{s}) &= \frac{1}{2} \lambda^2 \hat{s} \hat{m}^2 + (-\lambda \hat{s} \hat{m}^2 / 2 + \mu - \varepsilon \hat{s} / 2)^2 \\
&\quad + (-\lambda \hat{s} \hat{m}^2 / 2 + \mu / 2 - \varepsilon \hat{s} / 2)(1 + \hat{s}^2 M_4) - \lambda + \lambda \hat{m} \hat{s}^2 M_3, \\
\langle M \rangle(\phi, \hat{m}, \hat{s}) &= \frac{1}{2} \lambda^2 \hat{s} \hat{m}^2 + (-\lambda \hat{s} \hat{m}^2 / 2 + \mu - \varepsilon \hat{s} / 2)^2 \\
&\quad + 4(-\lambda \hat{s} \hat{m}^2 / 2 + \mu / 2 - \varepsilon \hat{s} / 2) - \lambda.
\end{aligned}$$

When $\lambda = \varepsilon = 0$, $M(\varphi, \hat{m}, \hat{s})$ is minimized at $\mu = -(1 + \hat{s}^2 M_4)/4$, whereas $\langle M \rangle(\phi, \hat{m}, \hat{s})$ at $\mu = -1$.

In this model, we can derive the exact forms of CV, WAIC, DIC [10, 11], and the free energy, hence we can compare the optimal hyperparameters for these criteria. Let

$$Z_n(X, \alpha) = \int p(X|w)^\alpha \prod_{i=1}^n p(X_i|w) \varphi(w) dw.$$

Then

$$Z_n(X, \alpha) = \frac{1}{(2\pi)^{(n+\alpha-1)/2}} \exp\left(-\frac{\log a(\alpha)}{2} - c(\alpha) \log d(\alpha)\right) \Gamma(c(\alpha)),$$

where $\Gamma(\)$ is the gamma function and

$$\begin{aligned}
a(\alpha) &= \alpha + \lambda + n, \\
b_i(\alpha) &= \alpha X + \sum_{j=1}^n X_j^2, \\
c(\alpha) &= \mu + (\alpha + n + 1)/2, \\
d_i(\alpha) &= (1/2)(\alpha X + \sum_{j=1}^n X_j^2 - b_i(\alpha)^2/a(\alpha) + \varepsilon).
\end{aligned}$$

All criteria can be calculated by using $Z(X, \alpha)$ by their definitions,

$$\begin{aligned}
\text{CV}(\varphi) &= -\frac{1}{n} \sum_{i=1}^n \{\log Z_n(0, 0) - \log Z_n(X_i, -1)\}, \\
\text{WAIC}(\varphi) &= -\frac{1}{n} \sum_{i=1}^n \{\log Z_n(X_i, 1) - \log Z_n(0, 0) - \frac{\partial^2}{\partial \alpha^2}(\log Z_n(X_i, 0))\}, \\
\text{DIC}(\varphi) &= -\frac{1}{n} \sum_{i=1}^n \{2 \frac{\partial}{\partial \alpha}(\log Z_n(X_i, 0)) - \log p(X_i, \bar{m}, \bar{s})\}, \\
F_{\text{free}}(\varphi) &= -\log Z_n(0, 0) + \log Z_0(0, 0),
\end{aligned}$$

Table 1 Averages and standard errors of criteria

	μ	ΔC	ΔW	WAICR	WAICRS	ΔD	ΔG
Average	-1	-0.00194	-0.00175	-0.00147	-0.00165	0.00332	-0.00156
STD	-1	0.00101	0.00080	0.00062	0.00001	0.00001	0.01292
Average	1	0.00506	0.00489	0.00450	0.00467	0.00006	0.00445
STD	1	0.00095	0.00076	0.00059	0.00004	0.00002	0.01250

where $\bar{m} = b(0)/a(0)$ and $\bar{s} = (2\mu + n + 1)/(\sum_i X_i^2 - b(0)^2/a(0) + \varepsilon)$.

A numerical experiment was conducted. A true distribution $q(x)$ was set as $\mathcal{N}(1, 1^2)$. We study a case $n = 25$. Ten thousands independent training sets were collected. A statistical model and a prior were defined by Eqs. (58) and (59) respectively. The fixed prior was $\varphi_0(w) \equiv 1$. We set $\lambda = \varepsilon = 0.01$, and studied the optimization problem of the hyperparameter μ . Firstly, we compared averages and standard deviations of criteria. In Table 1, for the two cases $\mu = \pm 1$, averages and standard deviations of

$$\begin{aligned}\Delta C &= \text{CV}(\varphi) - \text{CV}(\varphi_0), \\ \Delta W &= \text{WAIC}(\varphi) - \text{WAIC}(\varphi_0), \\ \text{WAICR} &= M(\phi, \hat{w})/n^2, \\ \text{WAICRS} &= \langle M \rangle(\phi, \hat{w})/n^2, \\ \Delta D &= \text{DIC}(\varphi) - \text{DIC}(\varphi_0), \\ \Delta G &= G(\varphi) - G(\varphi_0),\end{aligned}$$

are shown. In this experiment, averages of ΔC , ΔW , WAICR, and WAICRS were almost equal to that of ΔG , however that of ΔD was not. The standard deviations were

$$\sigma(\Delta G) >> \sigma(\Delta C) > \sigma(\Delta W) > \sigma(\text{WAICR}) > \sigma(\text{WAICRS}) \cong \sigma(\Delta \text{DIC}).$$

The standard deviation of ΔG was largest which is consistent to Theorem 1. Note that CV had the larger variance than WAIC. WAICRS gave the most precise result.

Secondly, we compared the distributions of the chosen hyperparameters by criteria. One hundred candidate hyper parameters in the interval $(-2.5, 2.5]$ were compared and the optimal hyperparameter for each criterion was chosen by minimization. Remark that the interval for the free energy was set as $(-0.5, 2.5]$ because the prior is proper if and only if $\mu > -0.5$. In Table 2, averages (A), standard deviations (STD), and $A \pm 2STD$ of optimal hyperparameters are shown.

In this case, the optimal hyperparameter for the minimum generalization loss is almost equal to (-1) , whose prior is improper. By CV, WAIC, WAICR, WAICRS, the optimal hyperparameter was almost chosen, whereas by DIC or the free energy, it was not. The standard deviations of chosen hyperparameters were

Table 2 Chosen hyperparameters in normal distribution

	h(CV)	h(WAIC)	h(WAICR)	h(WAICRS)	h(DIC)	h(F)
Average	-0.9863	-0.9416	-0.9329	-0.9993	0.4512	-0.2977
STD	0.2297	0.19231	0.1885	0.0059	0.0077	0.0106
A-2STD	-1.4456	-1.3262	-1.3100	-1.0112	0.4358	-0.3188
A+2STD	-0.5269	-0.5569	-0.5559	-0.9874	0.4667	-0.2766

$$\begin{aligned} \sigma(h(\text{CV})) &> \sigma(h(\text{WAIC})) > \sigma(h(\text{WAICR})) \\ &> \sigma(h(F)) > \sigma(h(\text{DIC})) > \sigma(h(\text{WAICRS})). \end{aligned}$$

In this experiment, neither the marginal likelihood nor DIC was appropriate for the predictive prior design.

4 Basic Lemmas

The main purpose of this paper is to prove Theorems 1 and 2. In this section we prepare several lemmas which are used in the proof of the main theorem. The proofs of these lemmas consists of exhaustive calculations, hence they are given in [18].

For arbitrary function $f(w)$, we define the expectation values by

$$\mathbb{E}_\varphi^{(\pm j)}[f(w)] = \frac{\int f(w)\varphi(w)p(X_j|w)^{\pm 1} \prod_{i=1}^n p(X_i|w)dw}{\int \varphi(w)p(X_j|w)^{\pm 1} \prod_{i=1}^n p(X_i|w)dw}.$$

Then the predictive distribution of x using $X^n \setminus X_j$ is $\mathbb{E}_\varphi^{(-j)}[p(x|w)]$. Thus its log loss for the test sample X_j is

$$-\log \mathbb{E}_\varphi^{(-j)}[p(X_j|w)].$$

The log loss of the leave-one-out cross validation is then given by its summation,

$$\text{CV}(\varphi) = -\frac{1}{n} \sum_{j=1}^n \log \mathbb{E}_\varphi^{(-j)}[p(X_j|w)]. \quad (67)$$

Lemma 1 Let $\phi(w) = \varphi(w)/\varphi_0(w)$. The cross validation and the generalization loss satisfy the following equations.

$$\text{CV}(\varphi) = \text{CV}(\varphi_0) + \frac{1}{n} \sum_{j=1}^n \{\log \mathbb{E}_{\varphi_0}^{(-j)}[\phi(w)] - \log \mathbb{E}_{\varphi_0}[\phi(w)]\}, \quad (68)$$

$$G(\varphi) = G(\varphi_0) - \mathbb{E}_{X_{n+1}} \left[\log \mathbb{E}_{\varphi_0}^{(+n+1)}[\phi(w)] - \log \mathbb{E}_{\varphi_0}[\phi(w)] \right]. \quad (69)$$

Definition The log loss function for $X^n \setminus X_j$ is defined by

$$L(w, -j) = -\frac{1}{n} \sum_{i \neq j}^n \log p(X_i|w) - \frac{1}{n} \log \varphi_0(w).$$

The MAP estimator for $X^n \setminus X_j$ is denoted by

$$\check{w}_j = \arg \min L(w, -j).$$

Lemma 2 Let $f(w)$ be a function $Q(X^n, w)$ which satisfies the regularity conditions (1), (2), ..., (5). Then there exist functions $R_1(f, w)$ and $R_2(f, w)$ which satisfy

$$\mathbb{E}_{\varphi_0}[f(w)] = f(\hat{w}) + \frac{R_1(f, \hat{w})}{n} + \frac{R_2(f, \hat{w})}{n^2} + O_p\left(\frac{1}{n^3}\right), \quad (70)$$

$$\mathbb{E}_{\varphi_0}^{(-j)}[f(w)] = f(\check{w}_j) + \frac{R_1(f, \check{w}_j)}{n-1} + \frac{R_2(f, \check{w}_j)}{(n-1)^2} + O_p\left(\frac{1}{n^3}\right), \quad (71)$$

where $R_1(f, w)$ is given by

$$R_1(f, w) = \frac{1}{2} f_{k_1 k_2}(w) J^{k_1 k_2}(w) - \frac{1}{2} f_{k_1}(w) V^{k_1}(w). \quad (72)$$

and $V^{k_1}(w) = J^{k_1 k_2}(w) J^{k_3 k_4}(w) L_{k_2 k_3 k_4}(w)$.

We do not need the concrete form of $R_2(f, w)$ in the proof of the main theorem.

Lemma 3 Let $|\alpha| \leq 1$. If m is a positive odd number,

$$\frac{\mathbb{E}_{\varphi_0}[p(X_k|w)^\alpha \prod_{j=1}^m (w^{k_j} - \hat{w}^{k_j})]}{\mathbb{E}_{\varphi_0}[p(X_k|w)^\alpha]} = O_p\left(\frac{1}{n^{(m+1)/2}}\right). \quad (73)$$

If m is a positive even number,

$$\frac{\mathbb{E}_{\varphi_0}[p(X_k|w)^\alpha \prod_{j=1}^m (w^{k_j} - \hat{w}^{k_j})]}{\mathbb{E}_{\varphi_0}[p(X_k|w)^\alpha]} = O_p\left(\frac{1}{n^{m/2}}\right). \quad (74)$$

For $m = 2, 4$,

$$\mathbb{E}_{\varphi_0} \left[\prod_{j=1}^2 (w^{k_j} - \hat{w}^{k_j}) \right] = \frac{1}{n} J^{k_1 k_2} + O_p\left(\frac{1}{n^2}\right), \quad (75)$$

$$\begin{aligned} \mathbb{E}_{\varphi_0} \left[\prod_{j=1}^4 (w^{k_j} - \hat{w}^{k_j}) \right] &= \frac{1}{n^2} (J^{k_1 k_2} J^{k_3 k_4} + J^{k_1 k_3} J^{k_2 k_4} + J^{k_1 k_4} J^{k_2 k_3}) \\ &\quad + O_p\left(\frac{1}{n^3}\right). \end{aligned} \quad (76)$$

Definition We use several functions of w in the proof.

$$S^{k_1}(w) = J^{k_1 k_2}(w) J^{k_3 k_4}(w) F_{k_2 k_3, k_4}(w), \quad (77)$$

$$T^{k_1}(w) = J^{k_1 k_2}(w) J^{k_3 k_4}(w) J^{k_5 k_6}(w) L_{k_2 k_3 k_5}(w) F_{k_4, k_6}(w), \quad (78)$$

$$U^{k_1 k_2}(w) = J^{k_1 k_3}(w) J^{k_2 k_4}(w) F_{k_3, j_4}(w). \quad (79)$$

Lemma 4 Let \hat{w} and \check{w}_j be the MAP estimators for X^n and $X^n \setminus X_j$, respectively. Then

$$\frac{1}{n} \sum_{j=1}^n \{(\check{w}_j)^{k_1} - \hat{w}^{k_1}\} = \frac{1}{n^2} S_{k_1}(\hat{w}) - \frac{1}{2n^2} T_{k_1}(\hat{w}) + O_p\left(\frac{1}{n^3}\right), \quad (80)$$

$$\frac{1}{n} \sum_{j=1}^n \{(\check{w}_j)^{k_1} - \hat{w}^{k_1}\} \{(\check{w}_j)^{k_2} - \hat{w}^{k_2}\} = \frac{1}{n^2} U^{k_1 k_2}(\hat{w}) + O_p\left(\frac{1}{n^3}\right). \quad (81)$$

For an arbitrary C^∞ -class function $f(w)$

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \{f(\check{w}_j) - f(\hat{w})\} &= \frac{1}{n^2} \{f_{k_1}(\hat{w})(S^{k_1}(\hat{w}) - \frac{1}{2} T^{k_1}(\hat{w})) \\ &\quad + \frac{1}{2} f_{k_1 k_2}(\hat{w}) U^{k_1 k_2}(\hat{w})\} + O_p\left(\frac{1}{n^3}\right). \end{aligned} \quad (82)$$

5 Proof of Theorem 1

In this section, we prove the main theorems. The proof of Theorem 1 consists of the five parts, Cross validation, WAIC, mathematical relations, averages, and random generalization loss.

5.1 Proof of Theorem 1, Cross Validation

In this subsection, we prove Eq.(41) in Theorem 1. By Lemma 2,

$$\begin{aligned}\mathbb{E}_{\varphi_0}[\phi(w)] &= \phi(\hat{w}) \left(1 + \frac{R_1(\phi, \hat{w})}{\phi(\hat{w})n} + \frac{R_2(\phi, \hat{w})}{\phi(\hat{w})n^2} \right) + O_p\left(\frac{1}{n^3}\right), \\ \mathbb{E}_{\varphi_0}^{(-j)}[\phi(w)] &= \phi(\check{w}_j) \left(1 + \frac{R_1(\phi, \check{w}_j)}{\phi(\check{w}_j)(n-1)} + \frac{R_2(\phi, \check{w}_j)}{\phi(\check{w}_j)(n-1)^2} \right) + O_p\left(\frac{1}{n^3}\right),\end{aligned}\quad (84)$$

For an arbitrary C^∞ -class function $f(w)$, by Lemma 4,

$$\begin{aligned}\frac{f(\check{w}_j)}{n-1} - \frac{f(\hat{w})}{n} &= \frac{f(\check{w}_j) - f(\hat{w})}{n-1} + \frac{f(\hat{w})}{n(n-1)} \\ &= \frac{f(\hat{w})}{n^2} + O_p\left(\frac{1}{n^3}\right),\end{aligned}\quad (85)$$

$$\begin{aligned}\frac{f(\check{w}_j)}{(n-1)^2} - \frac{f(\hat{w})}{n^2} &= \frac{f(\check{w}_j) - f(\hat{w})}{(n-1)^2} + \frac{(2n-1)f(\hat{w})}{n^2(n-1)^2} \\ &= O_p\left(\frac{1}{n^3}\right).\end{aligned}\quad (86)$$

By using Lemma 1 and by applying these equations for $f(w) = R_1(w)/\phi(w)$, $R_2(w)/\phi(w)$,

$$\begin{aligned}\text{CV}(\varphi) - \text{CV}(\varphi_0) &= \frac{1}{n} \sum_{j=1}^n \{\log \mathbb{E}_{\varphi_0}^{(-j)}[\phi(w)] - \log \mathbb{E}_{\varphi_0}[\phi(w)]\} \\ &= \frac{1}{n} \sum_{j=1}^n \{\log \phi(\check{w}_j) - \log \phi(\hat{w})\} + \frac{R_1(\hat{w})}{\phi(\hat{w})n^2} + O_p\left(\frac{1}{n^3}\right).\end{aligned}\quad (87)$$

By using Lemmas 2 and 4,

$$\begin{aligned}\text{CV}(\varphi) - \text{CV}(\varphi_0) &= \frac{1}{n^2} (\log \phi)_{k_1} (S^{k_1} - \frac{1}{2} T^{k_1}) + \frac{1}{2n^2} (\log \phi)_{k_1 k_2} U^{k_1 k_2} \\ &\quad + \frac{1}{2\phi n^2} (\phi_{k_1 k_2} J^{k_1 k_2} - \phi_{k_1} V^{k_1}) + O_p\left(\frac{1}{n^3}\right).\end{aligned}\quad (88)$$

Then by using

$$\begin{aligned}\phi_{k_1}/\phi &= (\log \phi)_{k_1}, \\ \phi_{k_1 k_2}/\phi &= (\log \phi)_{k_1 k_2} + (\log \phi)_{k_1} (\log \phi)_{k_2},\end{aligned}$$

it follows that

$$\begin{aligned} \text{CV}(\varphi) - \text{CV}(\varphi_0) &= \frac{1}{n^2} (\log \phi)_{k_1} (S^{k_1} - \frac{1}{2} T^{k_1} - \frac{1}{2} V^{k_1}) \\ &\quad + \frac{1}{2n^2} (\log \phi)_{k_1 k_2} (U^{k_1 k_2} + J^{k_1 k_2}) \\ &\quad + \frac{1}{2n^2} (\log \phi)_{k_1} (\log \phi)_{k_2} J^{k_1 k_2} + O_p\left(\frac{1}{n^3}\right), \end{aligned} \quad (89)$$

which completes Eq.(41). \square

5.2 Proof of Theorem 1, WAIC

In this subsection we prove Eqs.(43) and (45) in Theorem 1. In the following, we prove Eq.(45). In order to prove Eq.(45), it is sufficient to prove Eq.(45) in the case $\varphi(w)0 = \varphi_0(w)$ for an arbitrary $\varphi_0(w)$. Let the functional cumulant generating function for $\varphi_0(w)$ be

$$F_{\text{cum}}^0(\alpha) = \frac{1}{n} \sum_{i=1}^n \log \mathbb{E}_{\varphi_0}[p(X_i|w)^\alpha].$$

For a natural number j , we define the j th functional cumulant by

$$C_j(\alpha) \equiv \frac{\partial^j}{\partial \alpha^j} F_{\text{cum}}^0(\alpha).$$

Then by definition, $F_{\text{cum}}^0(0) = 0$ and

$$\text{CV}(\varphi_0) = F_{\text{cum}}^0(-1), \quad (90)$$

$$\text{WAIC}(\varphi_0) = T(\varphi_0) + V(\varphi_0)/n = -F_{\text{cum}}^0(1) + C_2(0). \quad (91)$$

For a natural number j , let $m_j(X_i, \alpha)$ be

$$m_j(X_i, \alpha) = \frac{\mathbb{E}_{\varphi_0}[\eta(X_i, w)^j \exp(-\alpha \eta(X_i, w))]}{\mathbb{E}_{\varphi}[\exp(-\alpha \eta(X_i, w))]},$$

where $\eta(X_i, w) = \log p(X_i|w) - \log p(X_i|\hat{w})$. Note that

$$\eta(X_i, w) = (w^{k_1} - \hat{w}^{k_1}) \ell_{k_1}(X_i, \hat{w}) + O_p((w - \hat{w})^2). \quad (92)$$

Therefore, if j is an odd number, by using Lemma 3,

$$m_j(X_i, \alpha) = O_p\left(\frac{1}{n^{(j+1)/2}}\right), \quad (93)$$

or if j is an even number

$$m_j(X_i, \alpha) = O_p\left(\frac{1}{n^{j/2}}\right). \quad (94)$$

Since $p(X_i|\hat{w})$ is a constant function of w ,

$$\begin{aligned} C_6(\alpha) &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^6}{\partial \alpha^6} \log \mathbb{E}_{\varphi_0}[p(X_i|w)^\alpha] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^6}{\partial \alpha^6} \log \mathbb{E}_{\varphi_0}[(p(X_i|w)/p(X_i|\hat{w}))^\alpha] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^6}{\partial \alpha^6} \log \mathbb{E}_{\varphi_0}[\exp(-\alpha \eta(X_i|w))] \\ &= \frac{1}{n} \sum_{i=1}^n \{m_6 - 6m_5m_1 - 15m_4m_2 + 30m_4m_1^2 - 10m_3^2 + 120m_3m_2m_1 \\ &\quad - 120m_3m_1^3 + 30m_2^3 - 270m_2^2m_1^2 + 360m_2m_1^4 - 120m_1^6\} = O_p\left(\frac{1}{n^3}\right), \end{aligned} \quad (95)$$

where $m_k = m_k(X_i, \alpha)$ in Eq. (95). Hence by Eq. (90),

$$\text{CV}(\varphi_0) = \sum_{j=1}^5 \frac{(-1)^j}{j!} C_j(0) + O_p\left(\frac{1}{n^3}\right). \quad (96)$$

On the other hand, by Eq. (91),

$$\text{WAIC}(\varphi_0) = \sum_{j=1}^5 \frac{-1}{j!} C_j(0) + C_2(0) + O_p\left(\frac{1}{n^3}\right). \quad (97)$$

It follows that

$$\text{WAIC}(\varphi_0) = \text{CV}(\varphi_0) - \frac{1}{12} C_4(0) + O_p\left(\frac{1}{n^3}\right).$$

Hence the main difference between CV and WAIC is $C_4(0)/12$. In order to prove Eq. (45), it is sufficient to prove $C_4(0) = O_p(1/n^3)$.

$$\begin{aligned}
C_4(0) &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^4}{\partial \alpha^4} \log \mathbb{E}_\varphi [p(X_i|w)^\alpha] \Big|_{\alpha=0} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\partial^4}{\partial \alpha^4} \log \mathbb{E}_\varphi [(p(X_i|w)/p(X_i|\hat{w}))^\alpha] \Big|_{\alpha=0} \\
&= \frac{1}{n} \sum_{i=1}^n \{m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4\},
\end{aligned} \tag{98}$$

where $m_k = m_k(X_i, 0)$ in Eq. (98). By Eqs. (93) and (94),

$$C_4(0) = \frac{1}{n} \sum_{i=1}^n \{m_4(X_i, 0) - 3m_2(X_i, 0)^2\} + O_p\left(\frac{1}{n^3}\right).$$

By using a notation

$$F_{k_1, k_2, k_3, k_4} \equiv \frac{1}{n} \sum_{i=1}^n \ell_{k_1}(X_i) \ell_{k_2}(X_i) \ell_{k_3}(X_i) \ell_{k_4}(X_i),$$

it follows that by Eq. (92) and Lemma 3,

$$\frac{1}{n} \sum_{i=1}^n m_4(X_i) = \frac{3}{n^2} J^{k_1 k_2} J^{k_3 k_4} F_{k_1, k_2, k_3, k_4} + O_p(1/n^3), \tag{99}$$

$$\frac{1}{n} \sum_{i=1}^n m_2(X_i)^2 = \frac{1}{n^2} J^{k_1 k_2} J^{k_3 k_4} F_{k_1, k_2, k_3, k_4} + O_p(1/n^3), \tag{100}$$

resulting that $C_4(0) = O_p(1/n^3)$, which completes Eq. (45). Then, Eq. (43) is immediately derived using Eqs. (41) and (45). \square

5.3 Mathematical Relations Between Priors

In this subsection, we prove Eqs. (46), (47), and (48). Since \hat{w} minimizes $L(w)$,

$$L_{k_1}(\hat{w}) = 0.$$

There exists w^* such that $\|w^* - w_0\| \leq \|\hat{w} - w_0\|$ and that

$$L_{k_1}(w_0) + L_{k_1 k_2}(w^*)(\hat{w} - w_0) = 0.$$

By the regularity condition (3), $\hat{w} \rightarrow w_0$ resulting that $w^* \rightarrow w_0$. By using the central limit theorem,

$$\hat{w} - w_0 = (L_{k_1 k_2}(w^*))^{-1} L_{k_1}(w_0) = O_p\left(\frac{1}{\sqrt{n}}\right). \quad (101)$$

Also by the central limit theorem, for an arbitrary $w \in W$,

$$L_{k_1 k_2}(w) = \mathbb{E}[L_{k_1 k_2}(w)] + \frac{\beta_{k_1 k_2}}{n^{1/2}}, \quad (102)$$

$$L_{k_1 k_2 k_3}(w) = \mathbb{E}[L_{k_1 k_2 k_3}(w)] + \frac{\beta_{k_1 k_2 k_3}}{n^{1/2}}, \quad (103)$$

$$F_{k_1, k_2}(w) = \mathbb{E}[F_{k_1, k_2}(w)] + \frac{\gamma_{k_1 k_2}}{n^{1/2}}, \quad (104)$$

$$F_{k_1 k_2, k_3}(w) = \mathbb{E}[F_{k_1 k_2, k_3}(w)] + \frac{\gamma_{k_1 k_2 k_3}}{n^{1/2}}, \quad (105)$$

where $\beta_{k_1 k_2}$, $\beta_{k_1 k_2 k_3}$, $\gamma_{k_1 k_2}$ and $\gamma_{k_1 k_2 k_3}$ are constant order random variables, whose expectation values are equal to zero. By the definitions,

$$\mathcal{L}_{k_1 k_2}(w) = \mathbb{E}[L_{k_1 k_2}(w)] + O\left(\frac{1}{n}\right), \quad (106)$$

$$\mathcal{L}_{k_1 k_2 k_3}(w) = \mathbb{E}[L_{k_1 k_2 k_3}(w)] + O\left(\frac{1}{n}\right), \quad (107)$$

$$\mathcal{F}_{k_1, k_2}(w) = \mathbb{E}[F_{k_1, k_2}(w)] + O\left(\frac{1}{n}\right), \quad (108)$$

$$\mathcal{F}_{k_1 k_2, k_3}(w) = \mathbb{E}[F_{k_1 k_2, k_3}(w)] + O\left(\frac{1}{n}\right), \quad (109)$$

Let $\beta \equiv \{\beta_{k_1 k_2}\}$ and $\Lambda \equiv \{L_{k_1 k_2}(w)\}$. Then by Eqs.(102) and (106),

$$\begin{aligned} \mathcal{J}(w) &= \mathbb{E}[\Lambda]^{-1} + O\left(\frac{1}{n}\right) \\ &= (\Lambda - \beta/\sqrt{n})^{-1} + O\left(\frac{1}{n}\right) \\ &= (\Lambda(1 - \Lambda^{-1}\beta/\sqrt{n}))^{-1} + O\left(\frac{1}{n}\right) \\ &= (1 + \Lambda^{-1}\beta/\sqrt{n})\Lambda^{-1} + O_p\left(\frac{1}{n}\right) \\ &= J(w) + \Lambda^{-1}\beta\Lambda^{-1}/\sqrt{n} + O_p\left(\frac{1}{n}\right). \end{aligned}$$

Hence

$$\begin{aligned} J^{k_1 k_2}(w) &= \mathcal{J}^{k_1 k_2}(w) + O_p(1/\sqrt{n}), \\ \mathbb{E}[J^{k_1 k_2}(w)] &= \mathcal{J}^{k_1 k_2}(w) + O(1/n). \end{aligned}$$

It follows that

$$\begin{aligned} M(\phi, w_0) &= \mathcal{M}(\phi, w_0) + O_p(1/\sqrt{n}), \\ \mathbb{E}[M(\phi, w_0)] &= \mathcal{M}(\phi, w_0) + O(1/n). \end{aligned}$$

Hence

$$\begin{aligned} M(\phi, \hat{w}) &= M(\phi, w_0) + (\hat{w} - w_0)^{k_1} (M(\phi, w_0))_{k_1} = \mathcal{M}(\phi, w_0) + O_p\left(\frac{1}{\sqrt{n}}\right), \\ \mathbb{E}[M(\phi, \hat{w})] &= \mathcal{M}(\phi, w_0) + O\left(\frac{1}{n}\right), \end{aligned}$$

which shows Eqs. (46) and (48). Then Eq. (47) is immediately derived by the fact $\hat{w} - \mathbb{E}_\varphi[w] = O_p(1/n)$ by Lemma 3. \square

5.4 Proof of Theorem 1, Averages

In this subsection, we prove we show Eqs. (42), (44), and (50).

Firstly, Eq. (42) is derived from Eqs. (41) and (48). Secondly, Eq. (44) is derived from Eqs. (43) and (48). Lastly, let us prove Eq. (50). Let $\text{CV}_n(\varphi)$ and $G_n(\varphi)$ be the cross validation and the generalization losses for X^n , respectively. Then by the definition, for an arbitrary φ ,

$$\begin{aligned} \mathbb{E}[G_n(\varphi)] &= \mathbb{E}[\text{CV}_{n+1}(\varphi)] \\ &= \mathbb{E}[\text{CV}_{n+1}(\varphi_0)] + \frac{\mathbb{E}[M(\varphi, \hat{w})]}{(n+1)^2} + O\left(\frac{1}{n^3}\right) \\ &= \mathbb{E}[G_n(\varphi_0)] + \frac{\mathbb{E}[M(\varphi, \hat{w})]}{n^2} + O\left(\frac{1}{n^3}\right), \end{aligned} \quad (110)$$

where we used $1/n^2 - 1/(n+1)^2 = O(1/n^3)$, which completes Eq. (50). \square

5.5 Proof of Theorem 1, Random Generalization Loss

In this subsection, we prove Eq. (49) in Theorem 1. We use a notation $\ell(n+1, w) = \log p(X_{n+1}|w)$. Let \bar{w} be the parameter that minimizes

$$-\frac{1}{n+1} \sum_{i=1}^{n+1} \log p(X_i|w) - \frac{1}{n+1} \log \varphi(w) = \frac{n}{n+1} \left\{ L(w) - \frac{1}{n} \ell(n+1, w) \right\}.$$

Since \bar{w} minimizes $L(w) - \ell(n+1, w)/n$,

$$L_{k_1}(\bar{w}) - \frac{1}{n} \ell_{k_1}(n+1, \bar{w}) = 0. \quad (111)$$

By applying the mean value theorem to Eq.(111), there exists \bar{w}^* such that

$$L_{k_1}(\hat{w}) + (\bar{w}^{k_2} - \hat{w}^{k_2}) L_{k_1 k_2}(\bar{w}^*) - \frac{1}{n} \ell_{k_1}(n+1, \bar{w}) = 0.$$

By using $L_{k_1}(\hat{w}) = 0$ and positive definiteness of $L_{k_1 k_2}(\hat{w})$,

$$\bar{w}^{k_1} - \hat{w}^{k_1} = O_p \left(\frac{1}{n} \right). \quad (112)$$

By applying the higher order mean value theorem to Eq.(111), there exists w^{**} such that

$$(\bar{w}^{k_2} - \hat{w}^{k_2}) L_{k_1 k_2}(\hat{w}) + \frac{1}{2} (\bar{w}^{k_2} - \hat{w}^{k_2})(\bar{w}^{k_3} - \hat{w}^{k_3}) L_{k_1 k_2 k_3}(\hat{w}^{**}) - \frac{1}{n} \ell_{k_1}(n+1, \bar{w}) = 0.$$

By Eq.(112), the second term of this equation is $O_p(1/n^2)$. The inverse matrix of $L_{k_1 k_2}(\hat{w})$ is $J^{k_1 k_2}(\hat{w})$,

$$\bar{w}^{k_1} - \hat{w}^{k_1} = \frac{1}{n} J^{k_1 k_2}(\hat{w}) \ell_{k_2}(n+1, \hat{w}) + O_p \left(\frac{1}{n^2} \right). \quad (113)$$

By Eq.(101), $\hat{w} - w_0 = O_p(1/\sqrt{n})$. Hence by the expansion of $(\hat{w} - w_0)$,

$$\begin{aligned} \mathbb{E}_{X_{n+1}} [\ell_{k_2}(n+1, \hat{w})] &= \mathbb{E}_{X_{n+1}} [(\log p(X_{n+1}|\hat{w}))_{k_2}] \\ &= (\hat{w}^{k_3} - (w_0)^{k_3}) \mathbb{E}_{X_{n+1}} [(\log p(X_{n+1}|w_0))_{k_2 k_3}] + O_p \left(\frac{1}{n} \right), \end{aligned} \quad (114)$$

where we used $\mathbb{E}_{X_{n+1}} [(\log p(X_{n+1}|w_0))_{k_2}] = 0$. By Eqs.(113) and (114),

$$\begin{aligned} \mathbb{E}_{X_{n+1}} [\bar{w}^{k_1} - \hat{w}^{k_1}] &= -\frac{1}{n} J^{k_1 k_2}(\hat{w}) (\hat{w}^{k_3} - (w_0)^{k_3}) \mathbb{E}[L_{k_2 k_3}(w_0)] + O_p \left(\frac{1}{n^2} \right) \\ &= -\frac{1}{n} (\hat{w}^{k_1} - (w_0)^{k_1}) + O_p \left(\frac{1}{n^2} \right), \end{aligned}$$

where we used $J^{k_2 k_3}(\hat{w}) = (\mathbb{E}[L_{k_2 k_3}(w_0)])^{-1} + O_p(1/n^{1/2})$. Therefore, by Lemmas 1 and 2

$$\begin{aligned}
G(\varphi) - G(\varphi_0) &= -\mathbb{E}_{X_{n+1}}[\log \mathbb{E}_{\varphi_0}^{(+n+1)}[\phi(w)] - \log \mathbb{E}_{\varphi_0}[\phi(w)]] \\
&= -\mathbb{E}_{X_{n+1}}[-\log(\phi(\bar{w}) + \frac{R_1(\phi, \bar{w})}{n+1}) + \log(\phi(\hat{w}) + \frac{R_1(\phi, \hat{w})}{n})] + O_p\left(\frac{1}{n^2}\right) \\
&= -\mathbb{E}_{X_{n+1}}[(\bar{w}^{k_1} - \hat{w}^{k_1})(\log \phi)_{k_1}(\hat{w})] + O_p\left(\frac{1}{n^2}\right) \\
&= \frac{1}{n}(\hat{w}^{k_1} - (w_0)^{k_1})(\log \phi)_{k_1}(\hat{w}) + O_p\left(\frac{1}{n^2}\right).
\end{aligned} \tag{115}$$

By Eq.(101), $\hat{w} - w_0 = O_p(1/\sqrt{n})$, we obtain Eq.(49). \square

5.6 Proof of Theorem 2

If there exists a parameter which satisfies $q(x) = p(x|w_0)$, then $\hat{w} - w_0 = O_p(1/\sqrt{n})$, $\langle L_{k_1 k_2} \rangle(w) = \mathcal{L}_{k_1 k_2}(w) + O_p(1/\sqrt{n})$, $\langle L_{k_1 k_2 k_3} \rangle(w) = \mathcal{L}_{k_1 k_2}(w) + O_p(1/\sqrt{n})$, $\langle F_{k_1, k_2} \rangle(w) = \mathcal{F}_{k_1, k_2}(w) + O_p(1/\sqrt{n})$, and $\langle F_{k_1 k_2, k_3} \rangle(w) = \mathcal{F}_{k_1 k_2, k_3}(w) + O_p(1/\sqrt{n})$. Hence Theorem 2 is obtained. \square

6 Discussions

In this section, we discuss several points about predictive prior design.

6.1 Summary of Results

In this paper, we have shown the mathematical properties of Bayes CV, WAIC, and the generalization loss. Let us summarize the results of this paper.

- (1) CV and WAIC are applicable to predictive prior design. Theoretically CV and WAIC are asymptotically equivalent, whereas experimentally the variance of WAIC is smaller than CV. If the regularity conditions are satisfied, then CV and WAIC can be approximated by WAICR. The variance of WAICR is smaller than CV and WAIC.
- (2) If the true distribution is realizable by a statistical model, then CV and WAIC can be estimated by WAICRS. The variance of WAICRS is smaller than WAICR.
- (3) The marginal likelihood is not appropriate for predictive prior design.

6.2 Divergence Phenomenon of CV and WAIC

In this subsection we study a divergence phenomenon of CV, WAIC, and the marginal likelihood. Let the maximum likelihood estimator be w_{mle} and $\delta_{mle}(w) =$

$\delta(w - w_{mle})$. Then for an arbitrary proper prior $\varphi(w)$, $\text{CV}(\varphi) \geq \text{CV}(\delta_{mle})$, $\text{WAIC}(\varphi) \geq \text{WAIC}(\delta_{mle})$, and $F_{\text{free}}(\varphi) \geq F_{\text{free}}(\delta_{mle})$. Hence, if a candidate prior can be made to converge to $\delta_{mle}(w)$, then minimizing these criteria results in the maximum likelihood method, where Theorem 1 does not hold. Therefore, in the optimization problem of a hyperparameter, then the prior should be design so that such a phenomenon can not occur.

6.3 Training and Testing Sets

In practical applications of machine learning, we often prepare both a set of training samples X^n and a set of test samples Y^m , where X^n and Y^m are independent. This method is sometimes called the holdout cross validation. Then we have a basic question, “Does the optimal hyperparameter chosen by CV or WAIC using a training set X^n minimize the generalization loss estimated using a test set Y^m ?”. The theoretical answer to this question is No, because, as is shown in Theorem 1, the optimal hyperparameter for X^n asymptotically minimizes $\mathbb{E}[G(X^n)]$ but not $G(X^n)$. The hyperparameter which minimizes $G(X^n)$ can not be found by any of CV, WAIC, DIC, or the marginal likelihood. On the other hand, the hyperparameter which minimizes $\mathbb{E}[G(X^n)]$, then CV or WAIC can be applied.

7 Conclusion

In this paper, we proved the higher order asymptotic equivalence of Bayes cross validation and WAIC. By using these equivalence, we can choose the optimal hyper parameter which minimizes the average generalization loss.

Acknowledgements This research was partially supported by the Ministry of Education, Science, Sports and Culture in Japan, Grant-in-Aid for Scientific Research 23500172, 25120013, and 15K00331.

References

1. Akaike, H.: Likelihood and Bayes procedure. In: Bernald, J.M. (ed.) Bayesian Statistics, pp. 143–166 (1980)
2. Clarke, B.S., Barron, A.R.: Jeffreys’ prior is asymptotically least favorable under entropy risk. *J. Stat. Plan. Inference* **41**, 36–60 (1994)
3. Epifani, I., MacEachern, S.N., Peruggia, M.: Case-deletion importance sampling estimators: central limit theorems and related results. *Electron. J. Stat.* **2**, 774–806 (2008)
4. Gelfand, A.E., Dey, D.K., Chang, H.: Model determination using predictive distributions with implementation via sampling-based method. *Bayesian Stat.* **4**, 147–167 (1992)

5. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis, 3rd edn. Chapman and Hall/CRC, New York (2013)
6. Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for bayesian models. *Stat. Comput.* **24**, 997–1016 (2014). <https://doi.org/10.1007/s11222-013-9416-2>
7. Good, I.J.: Rational decisions. *J. R. Stat. Soc. Ser. B* **14**, 107–114 (1952)
8. Komaki, F.: On asymptotic properties of predictive distributions. *Biometrika* **83**(2), 299–313 (1996)
9. Peruggia, M.: On the variability of case-detection importance sampling weights in the bayesian linear model. *J. Am. Stat. Assoc.* **92**, 199–207 (1997)
10. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A.: Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **64**(4), 583–639 (2002)
11. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A.: The deviance information criterion: 12 years on. *J. R. Stat. Soc. Ser. B* (2014). <https://doi.org/10.1111/rssb.12062>
12. Vehtari, A., Lampinen, J.: Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* **14**(10), 2439–2468 (2002)
13. Vehtari, A., Ojanen, J.: A survey of Bayesian predictive methods for model assessment, selection and comparison. *Stat. Surv.* **6**, 142–228 (2012)
14. Watanabe, S.: Algebraic analysis for nonidentifiable learning machines. *Neural Comput.* **13**(4), 899–933 (2001)
15. Watanabe, S.: Algebraic Geometry and Statistical Learning Theory. Cambridge University Press, Cambridge (2009)
16. Watanabe, S.: Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3591 (2010)
17. Watanabe, S.: Equations of states in singular statistical estimation. *Neural Netw.* **23**(1), 20–34 (2010)
18. Watanabe, S.: Bayesian cross validation and WAIC for predictive prior design in regular asymptotic theory (2015). [arXiv:1503.07970](https://arxiv.org/abs/1503.07970)

Restricted Boltzmann Machines: Introduction and Review



Guido Montúfar

Abstract The restricted Boltzmann machine is a network of stochastic units with undirected interactions between pairs of visible and hidden units. This model was popularized as a building block of deep learning architectures and has continued to play an important role in applied and theoretical machine learning. Restricted Boltzmann machines carry a rich structure, with connections to geometry, applied algebra, probability, statistics, machine learning, and other areas. The analysis of these models is attractive in its own right and also as a platform to combine and generalize mathematical tools for graphical models with hidden variables. This article gives an introduction to the mathematical analysis of restricted Boltzmann machines, reviews recent results on the geometry of the sets of probability distributions representable by these models, and suggests a few directions for further investigation.

Keywords Hierarchical model · Latent variable model · Exponential family · Mixture model · Hadamard product · Non-negative tensor rank · Expected dimension · Universal approximation · Kullback–Leibler divergence · Divergence maximization

1 Introduction

This article is intended as an introduction to the mathematical analysis of the restricted Boltzmann machine. Complementary to other existing and excellent introductions, we emphasize mathematical structures in relation to the geometry of the set of distributions that can be represented by this model. There is a large number of works on theory and applications of restricted Boltzmann machines. We review a selection

G. Montúfar (✉)

Department of Mathematics and Department of Statistics,
University of California, Los Angeles, Los Angeles, USA
e-mail: montufar@math.ucla.edu

G. Montúfar

Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

of recent results in a way that, we hope, can serve as a guide to this rich subject, and lets us advertise some of the interesting and challenging problems that still remain to be addressed.

Brief Overview

A Boltzmann machine is a model of pairwise interacting units that update their states over time in a probabilistic way depending on the states of the adjacent units. Boltzmann machines have been motivated as models for parallel distributed computing [1–3]. They can be regarded as stochastic versions of Hopfield networks [4], which serve as associative memories. They are closely related to mathematical models of interacting particles studied in statistical physics, especially the Ising model [5, Chapter 14]. For each fixed choice of interaction strengths and biases in the network, the collective of units assumes different states at relative frequencies that depend on their associated energy, in what is known as a Gibbs-Boltzmann probability distribution [6]. As pair interaction models, Boltzmann machines define special types of hierarchical log-linear models, which are special types of exponential family models [7] closely related to undirected graphical models [8, 9]. In contrast to the standard discussion of exponential families, Boltzmann machines usually involve hidden variables. Hierarchical log-linear models are widely used in statistics. Their geometric properties are studied especially in information geometry [10–13] and algebraic statistics [14, 15]. The information geometry of the Boltzmann machine was first studied by Amari, Kurata, and Nagaoka [16].

A restricted Boltzmann machine (RBM) is a special type of a Boltzmann machine where the pair interactions are restricted to be between an observed set of units and an unobserved set of units. These models were introduced in the context of harmony theory [17] and unsupervised two layer networks [18]. RBMs played a key role in the development of greedy layer-wise learning algorithms for deep layered architectures [19, 20]. A recommended introduction to RBMs is [21]. RBMs have been studied intensively, with tools from optimization, algebraic geometry, combinatorics, coding theory, polyhedral geometry, and information geometry among others. Some of the advances over the past few years include results in relation to their approximation properties [22–25], dimension [26–28], semi-algebraic description [29, 30], efficiency of representation [31, 32], sequential optimization [33, 34], statistical complexity [35], sampling and training [33, 34, 36, 37], information geometry [11, 16, 38].

Organization

This article is organized as follows. In Sect. 2 we introduce Boltzmann machines, Gibbs sampling, and the associated probability models. In Sect. 3 we introduce restricted Boltzmann machines and discuss various perspectives, viewing the probability models as marginals of exponential families with Kronecker factoring sufficient

statistics, as products of mixtures of product distributions, and as feedforward networks with soft-plus activations. We also discuss a piecewise linear approximation called the tropical RBM model, which corresponds to a feedforward network with rectified linear units. In Sect. 4 we give a brief introduction to training by maximizing the likelihood of a given data set. We comment on gradient, contrastive divergence, natural gradient, and EM methods. Thereafter, in Sect. 5 we discuss the Jacobian of the model parametrization and the model dimension. In Sect. 6 we discuss the representational power, covering two hierarchies of representable distributions, namely mixtures of product distributions and hierarchical log-linear models, depending on the number of hidden units of the RBM. In Sect. 7 we use the representation results to obtain bounds on the approximation errors of RBMs. In Sect. 8 we discuss semi-algebraic descriptions and a recent result for a small RBM. Finally, in Sect. 9 we collect a few open questions and possible research directions.

2 Boltzmann Machines

A Boltzmann machine is a network of stochastic units. Each unit, or neuron, can take one of two states. A joint state of all units has an associated energy value which is determined by pair interactions and biases. The states of the units are updated in a stochastic manner at discrete time steps, whereby lower energy states are preferred over higher energy ones. In the limit of infinite time, the relative number of visits of each state, or the relative probability of observing each state, converges to a fixed value that is exponential in the energy differences. The set of probability distributions that result from all possible values of the pair interactions and biases, forms a manifold of probability distributions called the Boltzmann machine probability model. The probability distributions for a subset of visible units are obtained via marginalization, adding the probabilities of all joint states that are compatible with the visible states. We make these notions more specific in the following.

Pairwise Interacting Units

We consider a network defined by a finite set of nodes N and a set of edges $I \subseteq \binom{N}{2}$ connecting pairs of nodes. Each node $i \in N$ corresponds to a random variable, or unit, with states $x_i \in \{0, 1\}$. The joint states of all units are vectors $x = (x_i)_{i \in N} \in \{0, 1\}^N$. Each unit $i \in N$ has an associated bias $\theta_i \in \mathbb{R}$, and each edge $\{i, j\} \in I$ has an associated interaction weight $\theta_{\{i, j\}} \in \mathbb{R}$. For any given value of the parameter $\theta = ((\theta_i)_{i \in N}, (\theta_{\{i, j\}})_{\{i, j\} \in I})$, the energy of the joint states x is given by

$$E(x; \theta) = - \sum_{i \in N} \theta_i x_i - \sum_{\{i, j\} \in I} \theta_{\{i, j\}} x_i x_j, \quad x \in \{0, 1\}^N. \quad (1)$$

In particular, the negative energy function $-E(\cdot; \theta)$ is a linear combination of the functions $x \mapsto x_i$, $i \in N$, $x \mapsto x_i x_j$, $\{i, j\} \in I$, with coefficients θ . It takes lower values when pairs of units with positive interaction take the same states, or also when units with positive bias take state one.

State Updates, Gibbs Sampling

The Boltzmann machine updates the states of its units at discrete time steps, in a process known as Gibbs sampling. Given a state $x^{(t)} \in \{0, 1\}^N$ at time t , the state $x^{(t+1)}$ at the next time step is created by selecting a unit $i \in N$, and then setting $x_i^{(t+1)} = 1$ with probability

$$\Pr(x_i^{(t+1)} = 1 | x^{(t)}) = \sigma\left(\sum_{\{i,j\} \in I} \theta_{\{i,j\}} x_j^{(t)} + \theta_i\right), \quad (2)$$

or $x_i^{(t+1)} = 0$ with complementary probability $\Pr(x_i^{(t+1)} = 0 | x^{(t)}) = 1 - \Pr(x_i^{(t+1)} = 1 | x^{(t)})$. Here $\sigma : s \mapsto 1/(1 + \exp(-s))$ is the standard logistic function. In particular, the quotient of the probabilities of setting either $x_i = 1$ or $x_i = 0$ is the exponential energy difference $\sum_{\{i,j\} \in I} \theta_{\{i,j\}} x_j + \theta_i$ between the two resulting joint states. The activation probability (2) can be regarded as the output value of a deterministic neuron with inputs x_j weighted by $\theta_{\{i,j\}}$ for all adjacent j s, bias θ_i , and activation function σ .

If the unit i to be updated at time t is selected according to a probability distribution r over N , and $T_i(x^{(t+1)} | x^{(t)})$ denotes the Markov transition kernel when choosing unit i , then the total transition kernel is

$$T = \sum_{i \in N} r(i) T_i.$$

In other words, if the state at time t is $x^{(t)}$, then the state $x^{(t+1)}$ at the next time step is drawn from the probability distribution $T(\cdot | x^{(t)})$. More generally, if $p^{(t)}$ is a probability distribution over joint states $x^{(t)} \in \{0, 1\}^N$ at time t , then at time $t+1$ we have the probability distribution

$$p^{(t+1)} = p^{(t)} \cdot T.$$

The one step transition kernel T is non zero only between state vectors $x^{(t)}$ and $x^{(t+1)}$ that differ at most in one entry. However, if r is strictly positive, then there is a

positive probability of transitioning from any state to any other state in N time steps, so that the N -th power T^N is strictly positive, implying that T is a primitive kernel.

Stationary Limit Distributions

If T is a primitive kernel, then there is a unique distribution p with $\lim_{t \rightarrow \infty} p^0 \cdot T^t = p$, for all start state distributions p^0 . This follows from a theorem by Geman and Geman, which also shows that p is the Gibbs-Boltzmann distribution

$$p(x; \theta) = \frac{1}{Z(\theta)} \exp(-E(x; \theta)), \quad x \in \{0, 1\}^N, \quad (3)$$

with the energy function $E(\cdot; \theta)$ given in (1) and normalizing partition function $Z(\theta) = \sum_{x'} \exp(-E(x'; \theta))$.

The set of stationary distributions (3), for all $\theta \in \mathbb{R}^{|N|+|I|}$, is the Boltzmann machine probability model with interaction structure $G = (N, I)$. This is an exponential family with sufficient statistics $x_i, i \in N, x_i x_j, \{i, j\} \in I$ and canonical or exponential parameter θ . It is a smooth manifold of dimension $|N| + |I|$, contained in the $2^N - 1$ dimensional simplex of probability distributions on $\{0, 1\}^N$,

$$\Delta_{\{0,1\}^N} = \left\{ p \in \mathbb{R}^{\{0,1\}^N} : p(x) \geq 0 \text{ for all } x \in \{0, 1\}^N, \text{ and } \sum_{x \in \{0, 1\}^N} p(x) = 1 \right\}.$$

Hidden Units, Visible Marginal Distributions

We will be interested in a situation where only a subset $V \subseteq N$ of all units can be observed, while the other units $H = N \setminus V$ are unobserved or hidden. Given the probability distribution $p(x; \theta)$ over the states $x = (x_V, x_H) \in \{0, 1\}^V \times \{0, 1\}^H$ of all units, the marginal probability distribution over the visible states x_V is given by

$$p(x_V; \theta) = \sum_{x_H \in \{0, 1\}^H} p(x; \theta), \quad x_V \in \{0, 1\}^V. \quad (4)$$

The set of marginal probability distributions, for all choices of θ , is a subset of the $2^V - 1$ dimensional simplex $\Delta_{\{0,1\}^V}$. It is the image of the fully observable Boltzmann machine probability manifold by the linear map that computes marginal distributions. In general this set is no longer a manifold. It may have a rather complex shape with self intersections and dimension strictly smaller than that of the manifold of distributions of all units. We will be concerned with the properties of this set in the special case where interaction edges are only allowed between visible and hidden units.

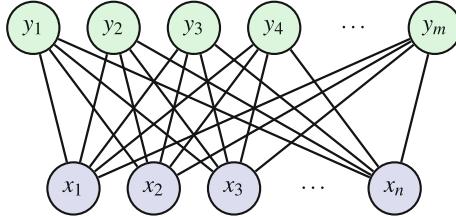


Fig. 1 RBM as a graphical model with visible units x_1, \dots, x_n and hidden units y_1, \dots, y_m . Each edge has an associated interaction weight w_{ji} , each visible node has an associated bias weight b_i , and each hidden node an associated bias weight c_j

3 Restricted Boltzmann Machines

The restricted Boltzmann machine (RBM) is a special type of Boltzmann machine where the interactions are restricted to be between visible and hidden units, such that $I = \{(i, j) : i \in V, j \in H\}$. This is illustrated in Fig. 1. The corresponding probability distributions take the form

$$p(x; \theta) = \frac{1}{Z(\theta)} \sum_{y \in \{0, 1\}^H} \exp(y^\top Wx + c^\top y + b^\top x), \quad x \in \{0, 1\}^V. \quad (5)$$

Here x is the state of the visible units, y is the state of the hidden units, Z is the partition function, and $\theta = (W, b, c)$ denotes the parameters, composed of the interaction weights $W = (w_{j,i})_{j \in H, i \in V}$, the biases of the visible units $b = (b_i)_{i \in V}$, and the biases of the hidden units $c = (c_j)_{j \in H}$. The RBM probability model with n visible and m hidden units is the set of probability distributions of the form (5), for all possible choices of θ . We denote this set by $\text{RBM}_{n,m}$. We will write $[n] = \{1, \dots, n\}$ and $[m] = \{1, \dots, m\}$ to enumerate the visible and hidden units, respectively. We write $\mathcal{X} = \{0, 1\}^V$ for the state space of the visible units, and $\mathcal{Y} = \{0, 1\}^H$ for that of the hidden units.

An RBM probability model can be interpreted in various interesting and useful ways, as we discuss in the following. These are views of the same object and are equivalent in that sense, but they highlight different aspects.

Product of Mixtures

One interpretation the RBM is as a *product of experts* model, meaning that it consists of probability distributions which are normalized entrywise products with factors coming from some fixed models. Factorized descriptions are familiar from graphical models, where one considers probability distributions that factorize into potential functions, which are arbitrary positive valued functions that depend only on certain fixed subsets of all variables. We discuss graphical models in more depth in Sect. 6. In

the case of RBMs, each factor model is given by mixtures of product distributions. A *product distribution* is a distribution of multiple variables which factorizes as an outer product $q(x_1, \dots, x_n) = \prod_{i \in [n]} q_i(x_i)$ of distributions q_i of the individual variables. A *mixture distribution* is a convex combination $q(x) = \sum_k \lambda_k q_k(x)$, where the λ_k are non-negative weights adding to one, and the q_k are probability distributions from some given set. Indeed, the RBM distributions can be written as

$$\begin{aligned} p(x; \theta) &= \frac{1}{Z(\theta)} \sum_{y \in \{0,1\}^m} \exp(y^\top Wx + c^\top y + b^\top x) \\ &= \frac{1}{Z(\theta)} \exp(b^\top x) \prod_{j \in [m]} (1 + \exp(W_{j:}x + c_j)) \\ &= \frac{1}{Z(\theta)} \prod_{j \in [m]} (\exp(W'_{j:}x) + \exp(c_j) \exp(W''_{j:}x)). \end{aligned} \quad (6)$$

Here $W'_{j:}$ and $W''_{j:} = W_{j:} + W'_{j:}$ can be chosen arbitrarily in \mathbb{R}^n for all $j \in [m]$, with $b = \sum_{j \in [m]} W'_{j:}$. In turn, for any mixture weights $\lambda_j \in (0, 1)$ we can find suitable $c_j \in \mathbb{R}$, and for any distributions $p'_{j,i}$ and $p''_{j,i}$ on $\mathcal{X}_i = \{0, 1\}$ suitable $W'_{j,i}$ and $W''_{j,i}$, such that

$$p(x; \theta) = \frac{1}{Z(\theta)} \prod_{j \in [m]} \left(\lambda_j \prod_{i \in [n]} p'_{j,i}(x_i) + (1 - \lambda_j) \prod_{i \in [n]} p''_{j,i}(x_i) \right). \quad (7)$$

This shows that the RBM model can be regarded as the set distributions that are entrywise products of m terms, with each term being a mixture of two product distributions over the visible states.

Products of experts can be trained in an efficient way, with methods such as contrastive divergence, which we will outline in Sect. 4. Products of experts also relate to the notion of distributed representations, where each observation is explained by multiple latent causes. This allows RBMs to create exponentially many inference regions, or possible categorizations of input examples, on the basis of only a polynomial number of parameters. This sets RBMs apart from mixture models, and provides one way of breaking the curse of dimensionality, which is one motivation for choosing one network architecture over another in the first place. We discuss more about this further below and in Sect. 6.

Tensors and Polynomial Parametrization

A probability distribution on $\{0, 1\}^n$ can be regarded as an n -way table or tensor with entries indexed by $x_i \in \{0, 1\}$, $i \in [n]$. A tensor p is said to have rank one if it can be factorized as $p = p_1 \otimes \cdots \otimes p_n$, where each p_i is a vector. Thus, non-negative rank one tensors correspond to product distributions. A tensor is said to have non-negative rank k if it can be written as the sum of k non-negative tensors of rank 1, and k is the

smallest number for which this is possible. Tensors of non-negative rank at most k correspond to mixtures of k product distributions. The RBM distributions are, up to normalization, the tensors that can be written as Hadamard (i.e., entrywise) products of m factor tensors of non-negative rank at most two. The representable tensors have the form

$$p = \prod_{j \in [m]} (q'_{j,1} \otimes \cdots \otimes q'_{j,n} + q''_{j,1} \otimes \cdots \otimes q''_{j,n}), \quad (8)$$

where the $q'_{j,i}$ and $q''_{j,i}$ are non-negative vectors of length two.

In particular, we note that, up to normalization, the RBM distributions have a polynomial parametrization

$$p = \left(\prod_{i \in [n]} \omega_{0,i}^{x_i} \right) \prod_{j \in [m]} \left(1 + \omega_{j,0} \prod_{i \in [n]} \omega_{j,i}^{x_i} \right), \quad (9)$$

with parameters $\omega_{0,i} = \exp(b_i) \in \mathbb{R}_+$, $\omega_{j,0} = \exp(c_j) \in \mathbb{R}_+$, $j \in [m]$, $\omega_{j,i} = \exp(W_{j,i}) \in \mathbb{R}_+$, $(i, j) \in [n] \times [m]$. The fact that RBMs have a polynomial parametrization makes them, like many other probability models, amenable to be studied with tools from algebra. This is the realm of *algebraic statistics*. Introductions to this area at the intersection of mathematics and statistics are [14, 15]. In algebraic geometry one studies questions such as the dimension and degree of solution sets of polynomial equations. When translated to statistics, these questions relate to parameter identifiability, the number of maximizers of the likelihood function, and other important properties of statistical models.

Kronecker Products, Harmonium Models

As we have seen, the joint distributions of a Boltzmann machine form an exponential family over the states of all units. That is, the joint distributions are given by exponentiating and normalizing vectors from an affine space,

$$p(x, y; \theta) = \frac{1}{Z(\theta)} \exp(\theta^\top F(x, y)), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (10)$$

Here the sufficient statistics $F_1, \dots, F_d: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ span the affine space in question. For an RBM, the sufficient statistics F have a special structure. Recall that the Kronecker product of two matrices is defined by $(a_{i,j})_{i,j} \otimes (b_{k,l})_{k,l} = (a_{i,j} b_{k,l})_{i,j}$. The sufficient statistics for the exponential family of the RBM can be written as a Kronecker product

$$F(x, y) = F^V(x) \otimes F^H(y), \quad (x, y) \in \mathcal{X} \times \mathcal{Y}, \quad (11)$$

where $F^V(x) = (1, x_1, \dots, x_n)^\top$ and $F^H(y) = (1, y_1, \dots, y_m)^\top$ are sufficient statistics of the independence models of the n visible binary units and the m hidden binary

units. The independence model is the exponential family of product distributions, $\frac{1}{Z} \exp(\sum_i \theta_i F_i^V(x)) = \frac{1}{Z} \exp(w^\top x + c) = \frac{1}{Z} \prod_{i \in [n]} \exp(w_i x_i)$.

The Kronecker product structure allows us to express the conditional distribution of hidden units given visible units, and vice versa, in the following simple way. Given two vectors a, b , write $\langle a, b \rangle$ for their inner product $a^\top b = \sum_i a_i b_i$. Take any parameter vector $\theta \in \mathbb{R}^{(n+1)(m+1)}$ and arrange its entries into a matrix $\Theta \in \mathbb{R}^{(m+1) \times (n+1)}$, going column by column. Then

$$\begin{aligned}\langle \theta, F(x, y) \rangle &= \langle \theta, F^V(x) \otimes F^H(y) \rangle \\ &= \langle \Theta^\top F^H(y), F^V(x) \rangle \\ &= \langle \Theta F^V(x), F^H(y) \rangle.\end{aligned}$$

These expression describe following probability distributions:

$$\begin{aligned}p(x, y; \theta) &= \frac{1}{Z(\theta)} \exp(\langle \theta, F(x, y) \rangle) \\ p(x|y; \theta) &= \frac{1}{Z(\Theta^\top F^H(y))} \exp(\langle \Theta^\top F^H(y), F^V(x) \rangle) \\ p(y|x; \theta) &= \frac{1}{Z(\Theta F^V(x))} \exp(\langle \Theta F^V(x), F^H(y) \rangle).\end{aligned}$$

Geometrically, ΘF^V is a linear projection of F^V into the parameter space of the exponential family with sufficient statistics F^H and, similarly, $\Theta^\top F^H$ is a linear projection of F^H into the parameter space of an exponential family for the visible variables. This is illustrated in Fig. 2.

Restricted Mixtures of Products

The marginal distributions can always be written as

$$p(x; \theta) = \sum_y p(x, y; \theta) = \sum_y p(y; \theta) p(x|y; \theta), \quad x \in \mathcal{X}.$$

In the case of an RBM, the conditional distributions are product distributions $p(x|y; \theta) = \prod_{i \in [n]} p(x_i|y; \theta)$. In turn, the RBM model consists of mixtures of product distributions, with mixture weights $p(y; \theta)$. However, the marginal $p(y; \theta)$ and the tuple of conditionals $p(x|y; \theta)$ have a specific and constrained structure. For instance, as can be seen in Fig. 2 for the model $\text{RBM}_{3,2}$, the mixture components have parameter vectors that are affinely dependent. One implication is that $\text{RBM}_{3,2}$ cannot represent any distribution with large values on the even parity strings 000, 011, 101, 110 and small values on the odd parity strings 001, 010, 100, 111. This kind of constraint, coming from constraints on the mixture components, have

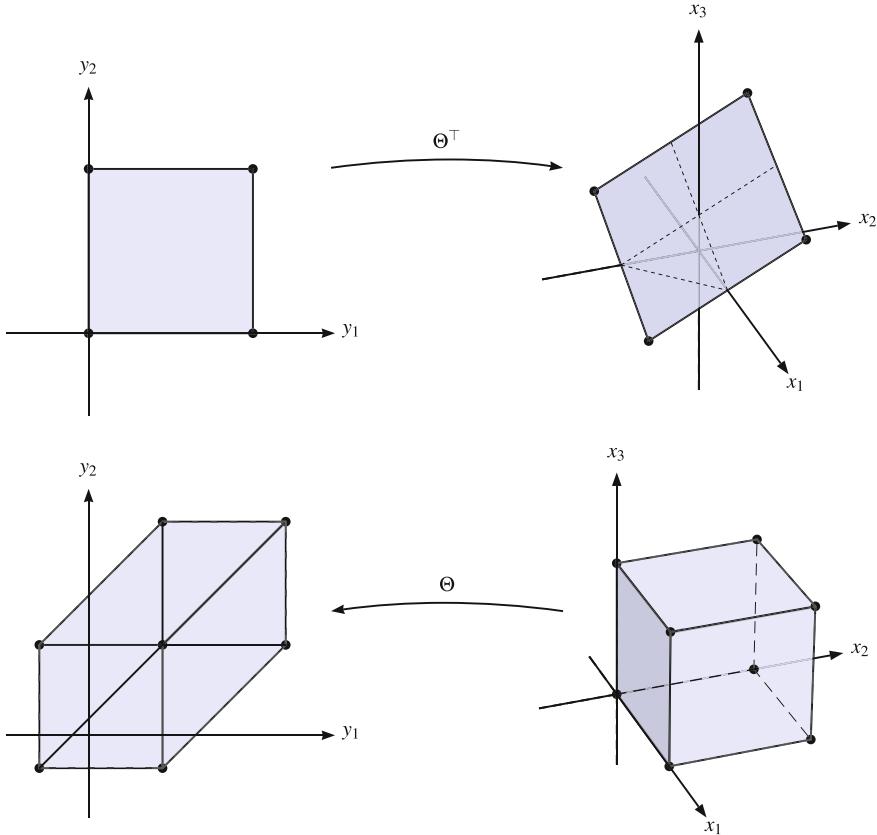


Fig. 2 For an RBM, the conditional distributions $p(X|y; \theta)$ of the visible variables given the hidden variables, are the elements of an exponential family with sufficient statistics F^V and parameters given by projections $\Theta^\top F^H(y)$ of the sufficient statistics F^H of the hidden variables. Similarly, $p(Y|x; \theta)$ are exponential family distributions with sufficient statistics F^H and parameters $\Theta F^V(x)$. The figure illustrates these vectors for $\text{RBM}_{3,2}$ and a choice of θ

been studied in [32]. An exact description of the constraints that apply to the probability distributions within $\text{RBM}_{3,2}$ was obtained recently in [30]. We comment on this later in Sect. 8.

Superposition of Soft-Plus Units

Another useful way of viewing RBMs is as follows. The description as products of mixtures shows that in RBMs the log-probabilities are sums of independent terms. More precisely, they are superpositions of m soft-plus units and one linear unit:

$$\log(p(x; \theta)) = \sum_{j \in [m]} \log(1 + \exp(W_{j:}x + c_j)) + b^T x - \log(Z(\theta)). \quad (12)$$

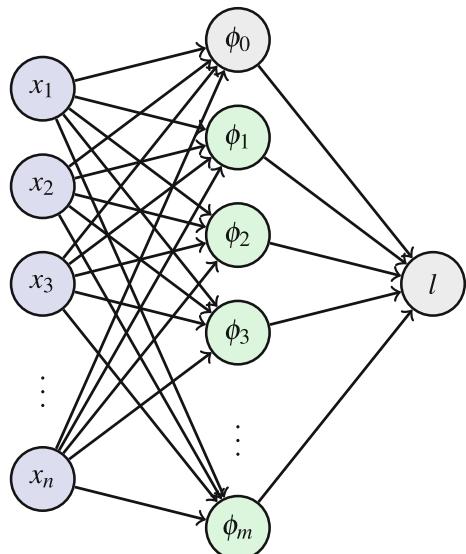
A *soft-plus unit* computes a real valued affine function of its arguments, $x \mapsto w^T x + c$, and then applies the soft-plus non linearity $s \mapsto \log(1 + \exp(s))$. A linear unit simply computes $x \mapsto b^T x + c$.

Log-probabilities correspond uniquely to probability distributions. When studying the space of representable log-probabilities, it is helpful to allow ourselves to add or disregard additive constants, since they correspond to scaling factors that cancel out with the normalization of the probability distributions.

The RBM model can be regarded as the set of negative energy functions (log-probabilities modulo additive constants) that can be computed by a feedforward network with one hidden layer of m soft-plus units and one linear unit, and a single output unit adding the outputs of the hidden units. The situation is illustrated in Fig. 3. Feedforward networks are often conceptually easier than stochastic networks or probabilistic graphical models. One point to note is that the output unit of the RBM energy network only computes unweighted sums.

A type of computational unit that is closely related to the soft-plus unit is the *rectified linear unit* (ReLU). A ReLU computes a real valued affine function of its arguments, $x \mapsto w^T x + c$, followed by rectification $s \mapsto [s]_+ = \max\{0, s\}$. As it turns out, if we replace the soft-plus units by ReLUs in Eq. (12), we obtain the so-called tropical RBM model, which is a piecewise linear version of the original model that facilitates a number of computations. We discuss more details of this relationship in the next paragraph.

Fig. 3 An RBM model can be regarded as the set of log-probabilities which are computable as the sum of a linear unit ϕ_0 and m soft-plus units $\phi_j, j = 1, \dots, m$



Tropical RBM, Superposition of ReLUs

The tropical RBM model is the set of vectors that we obtain when evaluating log-probabilities of the RBM model using the max-plus algebra and disregarding additive constants. We replace sums by maximum, so that a log-probability vector $l(x; \theta) = \sum_y \exp(y^\top Wx + b^\top x + c^\top y)$, $x \in \mathcal{X}$, becomes $\Phi(x; \theta) = \max_y \{y^\top Wx + b^\top x + c^\top y\}$, $x \in \mathcal{X}$. We can write this more compactly as

$$\Phi(x; \theta) = \theta^\top F(x, h(x; \theta)), \quad x \in \mathcal{X}, \quad (13)$$

where $F(x, y) = (1, x_1, \dots, x_n)^\top \otimes (1, y_1, \dots, y_m)^\top$ is the vector of sufficient statistics, and $h(x; \theta) = \operatorname{argmax}_y \theta^\top F(x, y) = \operatorname{argmax}_y p(y|x; \theta)$ is the *inference function* that returns the most probable y given x . In particular, the tropical RBM model is the image of a piecewise linear map.

We note the following decomposition, which expresses the tropical RBM model as a superposition of one linear unit and m ReLUs. We have

$$\begin{aligned} \Phi(x; \theta) &= \max_y \{y^\top Wx + b^\top x + c^\top y\} \\ &= b^\top x + \sum_{j \in [m]} \max_{y_j} \{y_j W_{j,:} x + c_j y_j\} \\ &= b^\top x + \sum_{j \in [m]} [W_{j,:} x + c_j]_+. \end{aligned}$$

In turn, the tropical RBM is the set of vectors computable by a sum of one linear unit $x \mapsto b^\top x$ and m ReLUs $x \mapsto [w^\top x + c]_+ = \max\{0, w^\top x + c\}$.

The set of functions that can be represented by a ReLU is closed under multiplication by non-negative scalars. Hence the unweighted sums of m ReLUs, $\sum_{j \in [m]} [w_j^\top x + c_j]_+$, express the same set of functions as the conic combinations of m ReLUs, $\sum_{j \in [m]} \alpha_j [\bar{w}_j^\top x + \bar{c}_j]_+$, where $\alpha_j \geq 0$, $j \in [m]$. For analysis and visualization, we can disregard positive multiplicative factors, and consider convex combinations of m normalized ReLUs. We can normalize each function such that its entry sum equals one. Zero functions cannot be normalized in this way, but they are equivalent to constant functions. The set of normalized functions expressible by a ReLU with two binary inputs is shown in Fig. 4. A sum of m ReLUs can realize any convex combinations of m points from this set. Affine functions with positive values correspond to the horizontal square in the middle of the figure, and constant functions to the point at the center of the square. Adding positive / negative constants to a given point corresponds to moving from it towards / away from the center.

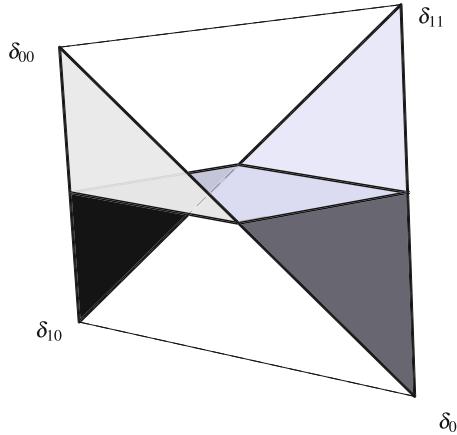


Fig. 4 Illustration of the set of functions $([w^\top x + c]_+, x \in \{0, 1\}^2)$, that can be represented by a ReLU with two binary inputs. This corresponds to the tropical RBM model with zero biases on the visible units. For the visualization of this 3 dimensional set in $\mathbb{R}_{\geq 0}^4$, we scaled the vectors to have entry sum 1 (the zero function is identified with the one function), which results in the shown subset of the simplex with vertices δ_x the indicators of individual inputs $x \in \{0, 1\}^2$

Other Generalizations

There are numerous generalizations of the regular RBM model.

- A Boltzmann machine can be defined with discrete non-binary states, real valued Gaussian units, or any other type of probability model for each unit. If the hidden variables are defined to take k possible values each, then the RBM defines a Hadamard product of tensors of non-negative rank at most k [27]. In particular, this is a generalization of mixtures of products models. Visible units with more than two states have been used, for example, in collaborative filtering [39].
- Viewed as Kronecker product models, with distributions $\frac{1}{Z(\theta)} \sum_y \exp(\theta^\top F^V(x) \otimes F^H(y))$, RBMs can be generalized to have arbitrary factors F^V and F^H , rather than just sufficient statistics of independence models. In this case, the conditional distributions of the visible variables, given the hidden variables, are distributions from the exponential family specified by F^V . This setting has been discussed in [28] and in [40] by the name *exponential family harmonium*.
- We can extend the setting of pair interactions to models with higher order interactions, called higher order Boltzmann machines [41].
- Other generalizations include deep architectures, such as deep belief networks [20] and deep Boltzmann machines [42]. Here one considers a stack of layers of units, with interactions restricted to pairs of units at adjacent layers. The representational power of deep belief networks has been studied in [22, 43–45] and that of deep Boltzmann machines in [46].

- For some applications, such as discriminative tasks, structured output prediction, stochastic control, one splits the visible units into a set of inputs and a set of outputs. The representational power of *conditional RBMs* has been studied in [47].
- Another line of generalizations are quantum models [48].
- A recent overview on RBM variants for diverse applications was given in [49].

4 Basics of Training

We give a short introduction to training. The general idea of training is to adjust the parameters of the Boltzmann machine such that it behaves in a desirable way. To do this, we first decide on a function to measure the desirability of the different possible behaviors, and then maximize that function over the model parameters. The first explicit motivation and derivation of a learning algorithm for Boltzmann machines is by Ackley, Hinton, and Sejnowski [1], based on statistical mechanics. Given a set of examples, the algorithm modifies the interaction weights and biases of the network so as to construct a generative model that produces examples with the same probability distribution of the provided examples.

Maximizing the Likelihood of a Data Set

Based on a set of examples, we aim at generating examples with the same probability distribution. To this end, we can maximize the log-likelihood of the provided examples with respect to the Boltzmann machine model parameters. For a set of examples $x^1, \dots, x^N \in \{0, 1\}^n$, the log-likelihood is

$$L(\theta) = \sum_{i=1}^N \log p(x^i; \theta) = \sum_x p_{\text{data}}(x) \log p(x; \theta), \quad (14)$$

where p_{data} is the empirical data distribution $p_{\text{data}}(x) = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}(x)$, $x \in \mathcal{X}$, and $p(x; \theta)$, $x \in \mathcal{X}$, is the model distribution with parameter $\theta \in \mathbb{R}^d$. Maximizing (14) with respect to θ is equivalent to minimizing the Kullback–Leibler divergence $D(p_{\text{data}} \| p_\theta)$ from p_{data} to the model distribution $p_\theta \equiv p(\cdot; \theta)$, again with respect to θ . The divergence is defined as

$$D(p_{\text{data}} \| p_\theta) = \sum_x p_{\text{data}}(x) \log \frac{p_{\text{data}}(x)}{p(x; \theta)}. \quad (15)$$

In some cases the minimum might not be attained by any value of the parameter θ . However, it is attained as $D(p_{\text{data}} \| p)$ for some distribution p in the closure of $\{p_\theta : \theta \in \mathbb{R}^d\} \subseteq \Delta_{\mathcal{X}}$.

Likelihood Gradient

In most cases, we do not know how to maximize the log-likelihood in closed form (we discuss a recent exception to this in Sect. 8). We can search for a maximizer by initializing the parameters at some random value $\theta^{(0)}$ and iteratively adjusting them in the direction of the gradient, as

$$\theta^{(t+1)} = \theta^{(t)} + \alpha_t \nabla L(\theta^{(t)}), \quad (16)$$

until some convergence criterion is met. Here the *learning rate* $\alpha_t > 0$ is a hyper-parameter of the learning criterion that needs to be specified. Typically the user tries a range of values. Often in practice, the parameter updates are computed based only on subsets of the data at the time, in what is known as on-line, mini-batch, or stochastic gradient.

Writing $F: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^d$ for the sufficient statistics of an exponential family of joint distributions of visible and hidden variables, we have

$$\nabla L(\theta) = \langle F \rangle_{\text{data}} - \langle F \rangle_\theta. \quad (17)$$

Here $\nabla = (\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_d})^\top$ is the column vector of partial derivatives with respect to the model parameters, $\langle \cdot \rangle_{\text{data}}$ stands for the expectation value with respect to the joint probability distribution $p_{\text{data}}(x)p_\theta(y|x)$, and $\langle \cdot \rangle_\theta$ stands for the expectation with respect to the joint distribution $p_\theta(x, y)$.

The computation of the gradient can be implemented as follows. We focus on the binary RBM, for which the sufficient statistics take the form

$$F(x, y) = (F_I, F_V, F_H)(x, y) = ((y_j x_i)_{j \in H, i \in V}, (x_i)_{i \in V}, (y_j)_{j \in H}), \quad (x, y) \in \{0, 1\}^V \times \{0, 1\}^H.$$

For the expectation value in (17) involving the data distribution:

- Write a data matrix $\tilde{X} = (x^1, \dots, x^N)$.
- Collect the activation probabilities of the individual hidden units, in response to each visible data vector, into a matrix $\tilde{Y} = \sigma(c \cdot \mathbf{1}_{1 \times N} + W \cdot \tilde{X})$. Here σ is the logistic function $s \mapsto 1/(1 + \exp(-s))$ applied entrywise to the argument, and $\mathbf{1}_{1 \times N}$ is the $1 \times N$ matrix of ones.
- Then

$$\begin{aligned} \langle F_I \rangle_{\text{data}} &= \tilde{Y} \cdot \tilde{X}^\top / N, \\ \langle F_V \rangle_{\text{data}} &= \tilde{X} \cdot \mathbf{1}_{N \times 1} / N, \\ \langle F_H \rangle_{\text{data}} &= \tilde{Y} \cdot \mathbf{1}_{N \times 1} / N. \end{aligned} \quad (18)$$

This calculation is relatively tractable, with order Nnm operations.

For the expectation in (17) with respect to the model distribution:

- Write X for the matrix with columns all vectors in $\{0, 1\}^n$ and Y for the matrix with columns all vectors in $\{0, 1\}^m$.
- Let $P_{Y \times X}$ be the matrix with entries $p_\theta(x, y)$, with rows and columns indexed by y and x .
- Then

$$\begin{aligned}\langle F_I \rangle_\theta &= Y \cdot P_{Y \times X} \cdot X^\top, \\ \langle F_V \rangle_\theta &= \mathbb{1}_{1 \times 2^m} \cdot P_{Y \times X} \cdot X^\top, \\ \langle F_H \rangle_\theta &= Y \cdot P_{Y \times X} \cdot \mathbb{1}_{2^n \times 1}.\end{aligned}\tag{19}$$

This calculation is possible for small models, but it can quickly become intractable. Since $P_{Y \times X}$ has 2^m rows and 2^n columns, computing its partition function and the expectations requires exponentially many operations in the number of units. In applications n and m may be in the order of hundreds or thousands. In order to overcome the intractability of this computation, a natural approach is to approximate the expectation values by sample averages. We discuss this next.

Contrastive Divergence

The expectations $\langle F \rangle_\theta$ with respect to the model distribution can be approximated in terms of sample averages obtained by Gibbs sampling the RBM. One method based on this idea is *contrastive divergence* (CD) [50]. This method has been enormously valuable in practical applications and is the standard learning algorithm for RBMs. The CD algorithm can be implemented as follows.

- As before, write a data matrix $\tilde{X} = (x^1, \dots, x^N)$.
- Then update the state of the hidden units of the RBM by

$$\tilde{Y} = (\sigma(c \cdot \mathbb{1}_{1 \times N} + W \cdot \tilde{X}) \geq \text{rand}_{m \times N}).$$

- Update the state of the visible units by

$$\hat{X} = (\sigma(b \cdot \mathbb{1}_{1 \times N} + W^\top \tilde{Y}) \geq \text{rand}_{n \times N}).$$

These updates are the Gibbs sampling state updates described in Eq. (2), computed in parallel for all hidden and visible units. Here $\text{rand}_{n \times N}$ is an $n \times N$ array of independent variables uniformly distributed in $[0, 1]$, and \geq is evaluated entrywise as a logic gate with binary outputs.

- Now use the reconstructed data \hat{X} to compute $\langle F \rangle_{\text{recon}}$ in the same way as \tilde{X} was used to compute $\langle F \rangle_{\text{data}}$ in Eq. (18). The approximate model sample average $\langle F \rangle_{\text{recon}}$ is then used as an approximation of $\langle F \rangle_\theta$.

This calculation involves only order Nnm operations, and remains tractable even for relatively large n and m in the order of thousands.

CD is an approximation to the maximum likelihood gradient. The bias of this method with respect to the actual gradient has been studied theoretically in [33]. There are a number of useful variants of the basic CD method. One can use k Gibbs updates, instead of just one, in what is known as the CD_k method. The larger k , the more one can expect the samples to follow the model distribution. In this spirit, there is also the persistent CD method (PCD) [51], where each sampling chain is initialized at previous samples, rather than at examples from the data set. Another useful technique in this context is parallel tempering [21, 52]. Moreover, basic gradient methods are often combined with other strategies, such as momentum, weight decay, preconditioners, second order methods. For more details see the introduction to training RBMs [53] and the useful practical guide [54].

Natural Gradient

A natural modification of the standard gradient method is the *natural gradient*, which is based on the notion that the parameter space has an underlying geometric structure. This is the point of view of *information geometry* [11, 12, 55]. A recent mathematical account on this topic is given in the book [13]. The natural gradient method was popularized with Amari's paper [56], which discusses how this method is efficient in learning. In this setting, the ordinary gradient is replaced by a Riemannian gradient, which leads to a parameter update rule of the form

$$\theta^{(t+1)} = \theta^{(t)} + \alpha_t G^{-1}(\theta^{(t)}) \nabla L(\theta^{(t)}), \quad (20)$$

where G is the Fisher information [57]. For a given parametric model $\{p_\theta : \theta \in \mathbb{R}^d\}$, the Fisher information is defined as

$$G(\theta) = \mathbb{E}_\theta [\nabla \log p(X; \theta) \cdot \nabla^\top \log p(X; \theta)].$$

Here $\mathbb{E}_\theta[\cdot]$ denotes expectation with respect to the model distribution $p(X; \theta) \equiv p_\theta$. Amari, Kurata, and Nagaoka [16] discuss the statistical meaning of the Fisher metric. The inverse Fisher matrix divided by the number of observations describes the behavior of the expected square error (covariance matrix) of the maximum likelihood estimator.

For an exponential family model with sufficient statistics $F: \mathcal{X} \rightarrow \mathbb{R}^d$ and log-partition function $\psi(\theta) = \log Z(\theta)$, the Fisher matrix can be given as the Hessian of the log-partition function, as

$$G(\theta) = \nabla \nabla^\top \psi(\theta) = \mathbb{E}_\theta[F \cdot F^\top] - \mathbb{E}_\theta[F] \cdot \mathbb{E}_\theta[F]^\top = \text{Cov}_\theta[F],$$

which is the covariance of F with respect to the exponential family distribution. This matrix is full rank iff the exponential family parametrization is minimal, meaning

that the functions $F_1, \dots, F_d: \mathcal{X} \rightarrow \mathbb{R}$ are linearly independent and do not contain the constant function 1 in their linear span.

Consider now the RBM model as the set of visible marginals of an exponential family with sufficient statistics $F: \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}^d$. The gradient of the visible log-probabilities is

$$\nabla \log p(x; \theta) = \mathbb{E}_\theta[F|x] - \mathbb{E}_\theta[F], \quad (21)$$

where $\mathbb{E}_\theta[F|x] = \sum_y F(x, y)p(y|x; \theta)$ is the conditional expectation of F , given the visible state x , and $\mathbb{E}_\theta[F] = \sum_{x,y} F(x, y)p(x, y; \theta)$ is the expectation with respect to the joint distribution over visible and hidden states. The Fisher matrix takes the form

$$\begin{aligned} G(\theta) &= \mathbb{E}_\theta[\mathbb{E}_\theta[F|X] \cdot \mathbb{E}_\theta[F|X]^\top] - \mathbb{E}_\theta[F] \cdot \mathbb{E}_\theta[F]^\top \\ &= \text{Cov}_\theta[\mathbb{E}_\theta[F|X]]. \end{aligned}$$

The rank of this matrix is equal to the rank of the Jacobian $J(\theta) = [\nabla p(x; \theta)]_x$ of the parametrization of the visible marginal distributions. Verifying whether and when the Fisher matrix of the RBM has full rank, is a non-trivial problem that we will discuss further in Sect. 5.

In models with hidden variables, the Fisher matrix is not always full rank. An area that studies the statistical effects of this is *singular learning theory*; see [35, 58]. In practice, for the purpose of parameter optimization, the natural gradient works well even when the model involves singularities, at least so long as the parameter updates don't step into the singular set. The advantages of the natural gradient over the regular gradient have been demonstrated in numerous applications. It tends to be better at handling plateaus, thus reducing the number of required parameter updates, and also to find better local optimizers. On the other hand, computing the Fisher matrix and its inverse is challenging for large systems. Approximations of the relevant expectation values still require a computational overhead over the regular gradient, and in some cases, it is not clear how to balance optimization with other statistical considerations. Approximating the Fisher matrix in an efficient and effective way is an active topic of research. RBMs have been discussed specifically in [59, 60]. Following the notions of the natural gradient, recent works also investigate alternatives and variants of the Fisher metric, for instance based on the Wasserstein metric [61, 62].

Doubly Minimization, EM Algorithm

Amari [11, Section 8.1.3] discusses an alternative view on the maximum likelihood estimation problem in probability models with hidden variables. See also [16, 63]. The idea is to regard this as an optimization problem over the model of joint distributions of both visible and hidden variables. Given an empirical data distribution p_V over visible states $x \in \mathcal{X}$, consider the set of joint distributions over $(x, y) \in \mathcal{X} \times \mathcal{Y}$ that are compatible with p_V :

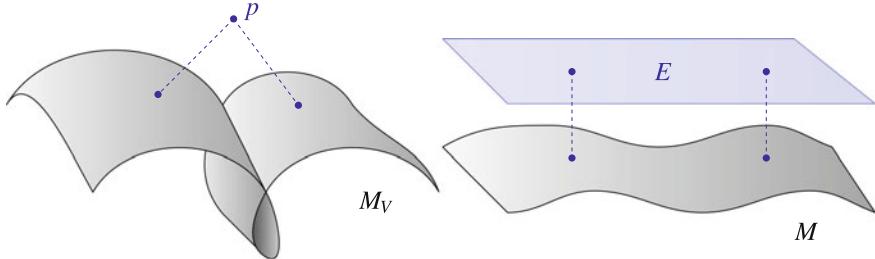


Fig. 5 Schematic illustration of the maximum likelihood estimation problem over the set of visible marginal distributions M_V , and over the set of joint distributions M prior to marginalization

$$E = \left\{ p(x, y) : \sum_{y \in \mathcal{Y}} p(x, y) = p_V(x) \right\}.$$

This *data manifold* E , being defined by linear equality constraints, is a special type of linear model. Note that it can be written as $E = \{p(x, y) = p_V(x)p(y|x)\}$, where we fix the marginal distribution $p_V(x)$ and are free to choose arbitrary conditional distributions $p(y|x)$ of hidden states given the visible states.

Taking this view, we no longer minimize the divergence from p_V to our model M_V of visible marginal distributions $q_V(x; \theta) = \sum_y q(x, y; \theta)$, but rather we seek for the distributions $q(x, y; \theta)$ in the model M of joint distributions, with the smallest divergence from the data manifold E . The situation is illustrated schematically in Fig. 5.

When working with the data manifold E and the joint model M , the maximum likelihood estimation problem becomes a double minimization problem

$$\min_{p \in E, q \in M} D(p \| q). \quad (22)$$

The minimum of this problem equals the minimum of the original problem

$$\min_{q_V \in M_V} D(p_V \| q_V).$$

To see this, use the chain rule for probability, $P(x, y) = P(x)P(y|x)$, to write

$$\begin{aligned} \min_{p \in E, q \in M} D(p \| q) &= \min_{p \in E, q \in M} \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{q(x, y)} \\ &= \min_{q \in M} \sum_x p_V(x) \log \frac{p_V(x)}{q_V(x)} + \min_{p(y|x)} \sum_x p_V(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)} \\ &= \min_{q_V \in M_V} D(p_V \| q_V). \end{aligned}$$

For simplicity of exposition, we are assuming that the sets E and M are so that the minimum can be attained, e.g., they are closed.

The expression (22) hints at an approach to computing the minimizers. Namely, we can iteratively minimize with respect to each of the two arguments.

- For any fixed value of the second argument, $q \in M$, minimization of the divergence over the first argument $p \in E$ is a convex problem, because E is a linear model. This is solved by the *e-projection* of q onto E , which is given simply by setting $p(y|x) = q(y|x)$.
- For any fixed value of the first argument, $p \in E$, the minimization over the second argument $q \in M$ is also a convex problem, because M is an exponential family. It is solved by the *m-projection* of p onto M , which is given by the unique distribution q in M for which $\sum_{x,y} F(x, y)q(x, y) = \sum_{x,y} F(x, y)p(x, y)$.

This procedure corresponds to the expectation maximization (EM) algorithm [64].

Optimization Landscape

In general, for a model with hidden variables, we must assume that the log-likelihood function $L(\theta)$ is non-concave. Gradient methods and other local techniques, such as contrastive divergence and EM, may only allow us to reach critical points or locally optimal solutions. The structure of the optimization landscape and critical points of these methods is the subject of current studies. In Sect. 8 we discuss results from [30] showing that an RBM model can indeed have several local optimizers with different values of the likelihood function, but also that in some cases, the optimization problem may be solvable in closed form.

5 Dimension

From a geometric standpoint, a basic question we are interested in, is the dimension of the set of distributions that can be represented by our probability model. The dimension is useful when comparing a model against other models, or when testing hypotheses expressed in terms of equality constraints. Under mild conditions, if the dimension is equal to the number of parameters, then the Fisher matrix is regular almost everywhere and the model is generically locally identifiable.

A Boltzmann machine with all units observed is an exponential family, and its dimension can be calculated simply as the dimension of the linear space spanned by the sufficient statistics, disregarding constant functions. This is precisely equal to the number of parameters of the model, since the statistics associated with each of the parameters, bias and interaction weights, are linearly independent.

When some of the units of the Boltzmann machine are hidden, as is usually the case, the set of observable distributions is no longer an exponential family, but rather a linear projection of an exponential family. The marginalization map takes

the high dimensional simplex $\Delta_{\mathcal{X} \times \mathcal{Y}}$ to the low dimensional simplex $\Delta_{\mathcal{X}}$. Such a projection can in principle collapse the dimension of the set that is being projected. A simple example where this happens is the set of product distributions. The visible marginals of an independence model are simply the independent distributions of the observed variables, meaning that the hidden variables and their parameters do not contribute to the dimension of the observable model. Another well-known example is the set of mixtures of three product distributions of four binary variables. This model has dimension 13, instead of 14 that one would expect from the number of model parameters. Computing the dimension of probability models with hidden variables often corresponds to challenging problems in algebraic geometry, most prominently the dimension of secant varieties, which correspond to mixture models.

Tropical Approach

The first investigation of the dimension of the RBM model was by Cueto, Morton, and Ottaviani [26], using tools from tropical geometry and secant varieties. The tropical approach to the dimension of secant varieties was proposed by Ottaviani [65]. It can be used in great generality, and it was also used to study non-binary versions of the RBM [27].

As mentioned in Sect. 3, the tropical RBM consists of piecewise linear approximation of the log-probability vectors of the RBM. The dimension of the tropical RBM is often easy to estimate by combinatorial arguments. A theorem by Bieri and Ottaviani [65, 66] implies that the dimension of the tropical RBM model is a lower bound on the dimension of the original RBM model. Using this method, [26] proved that the RBM model has the expected dimension for most combinations of n and m . However, a number of cases were left open. In fact, for the tropical RBM those cases are still open. A different approach to the dimension of RBMs was proposed in [28], which allowed verifying the conjecture that it always has the expected dimension. In the following we discuss this approach and how it compares to the tropical approach.

Jacobian Rank of RBMs and Mixtures of Products

The dimension of a smoothly parametrized model can be computed as the maximum rank of the Jacobian of the parametrization. For a parametrization $p(x; \theta) = \sum_y p(x, y; \theta)$, with $p(x, y; \theta) = \frac{1}{Z(\theta)} \exp(\sum_i \theta^\top F(x, y))$, the columns of the Jacobian matrix are

$$J_{:x}(\theta) = \sum_y p(x, y; \theta)(F(x, y) - \sum_{x', y'} p(x', y'; \theta)F(x', y')), \quad x \in \mathcal{X}. \quad (23)$$

Now we need to consider the specific F and evaluate the maximum rank of the matrix J over the parameter space. In order to simplify this, one possibility is to consider the limit of large parameters θ . The corresponding limit distributions usually

have a reduced support and the sum in (23) has fewer nonzero terms. As shown in [28], the dimension bounds from the tropical approach can be obtained in this manner. On the other hand, it is clear that after taking such limits, it is only possible to lower bound the maximum rank. Another problem is that, when the number of parameters is close to the cardinality of \mathcal{X} , the rank of the limit matrices is not always easy to compute, with block structure arguments leading to challenging combinatorial problems, such as accurately estimating the maximum cardinality of error correcting codes.

For the analysis it is convenient to work with the denormalized model, which includes all positive scalar multiples of the probability distributions. The dimension of the original model is simply one less. Following (23), and as discussed in [28], the Jacobian for the denormalized RBM is equivalent to the matrix with columns

$$\sum_y p(y|x; \theta) F(x, y) = \sum_y p(y|x; \theta) \hat{x} \otimes \hat{y} = \hat{x} \otimes \hat{\sigma}(Wx + c), \quad x \in \mathcal{X}, \quad (24)$$

where we write $\hat{v} = (1, v^\top)^\top$ for the vector v with an additional 1. Here $\sigma(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$ can be regarded as the derivative of the soft-plus function $\log(1 + \exp(\cdot))$. The j th coordinate of $\sigma(Wx + c)$ ranges between 0 and 1, taking larger values the farther x lies in the positive side of the hyperplane $H_j = \{r \in \mathbb{R}^V : W_j \cdot r + c_j = 0\}$. In the case of the tropical RBM, the Jacobian is equivalent to the matrix with columns

$$\hat{x} \otimes \hat{1}_{[Wx+c]_+}, \quad x \in \mathcal{X},$$

where now $\hat{1}_{[\cdot]_+}$ corresponds to the derivative of the rectification non-linearity $[\cdot]_+$. The j th coordinate indicates whether the point x lies on the positive side of the hyperplane H_j . The matrices for the RBM and the tropical RBM are illustrated in Fig. 6.

In [28] it is shown that (24) can approximate the following matrix, equivalent to the Jacobian of a mixture of $m + 1$ product distributions model, arbitrarily well at generic parameters:

$$\hat{x} \otimes \hat{\sigma}'(\tilde{W}x + \tilde{c}), \quad x \in \mathcal{X}.$$

Here $\sigma'(\tilde{W}x + \tilde{c}) = \frac{\exp(\tilde{W}x + \tilde{c})}{\sum_j \exp(\tilde{W}_j \cdot x + \tilde{c}_j)}$ is a soft-max unit. In turn, the dimension of the RBM model is bounded below by the dimension of the mixture model. But the results from [67] imply that mixture models of binary product distributions have the expected dimension (except in one case, which for the RBM can be verified by other means). This implies that the RBM model always has the expected dimension:

Theorem 1 ([28, Corollary 26]). *For any $n, m \in \mathbb{N}$ the model RBM _{n,m} , with n visible and m hidden binary units, has dimension $\min\{2^n - 1, (n + 1)(m + 1) - 1\}$.*

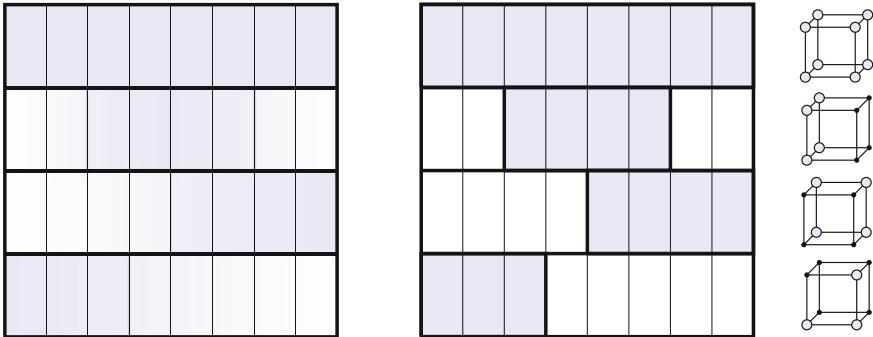


Fig. 6 Illustration of the Jacobian matrix for an RBM with three visible and three hidden units, and its tropical counterpart, together with the corresponding slicings of the visible sufficient statistics. Rows correspond to model parameters and columns to visible states

This result implies that, unless the number of parameters exceeds $2^n - 1$, almost every probability distribution in the RBM model can be represented by only finitely many different choices of the parameters. One trivial way in which the parameters are not unique, is that we can permute the hidden units without changing the represented distributions, $\sum_{j \in H} \log(1 + \exp(w_j x + c_j)) = \sum_{j \in H} \log(1 + \exp(w_{\pi(j)} x + c_{\pi(j)}))$ for all $\pi \in H!$. On the other hand, there are also a few probability distributions that can be represented by infinitely many different choices of the parameters. For instance, if $w_j = 0$, then the choice of c_j is immaterial.

The characterization of the parameter fibers $\{\theta \in \mathbb{R}^d : p_\theta = p\}$ of the distributions p that can be represented by an RBM model is an important problem, with implications on the parameter optimization problem, which still requires more investigation. We can ask in the first place whether a given distribution p can be represented by an RBM model. We discuss this in the next section.

6 Representational Power

The representational power of a probability model can be studied from various angles. An idea is that each parameter allows us to model certain features or properties of the probability distributions. The question then is how to describe and interpret these features. As we have seen, each hidden unit of an RBM can be interpreted as contributing entrywise multiplicative factors which are arbitrary mixtures of two product distributions. Alternatively, each hidden unit can be interpreted as adding a soft-plus unit to the negative energy function of the visible distributions.

Now we want to relate these degrees of freedom with the degrees of freedom of other families of distributions for which we have a good intuition, or for which we can maximize the likelihood function in closed form and compute metrics of the representational power, such as the maximum divergence. The natural approach to

this problem is by showing that there exist choices of parameters for which the model realizes a given distribution of interest, or, more generally, a class of distributions of interest. We note that another approach, which we will discuss in Sect. 8, is by showing that any constraints that apply on the set of distributions from the RBM are less stringent than the constraints that apply on the distributions of interest.

Overview

The representational power of RBMs has been studied in many works. Le Roux and Bengio [22] showed that each hidden unit of an RBM can model the probability of one elementary event. Freund and Haussler [68] used similar arguments to discuss universal approximation. In [44] it was shown that each hidden unit can model the probability of two elementary events of Hamming distance one, which implied improved bounds on the minimal number of hidden units that is sufficient for universal approximation. Generalizing this, [24] showed that each hidden unit can model a block of elementary events with a weighted product distribution, provided certain conditions on the support sets are satisfied. Another line of ideas was due to [25], showing that each hidden unit can model the coefficient of a monomial in a polynomial representation of the energy function. This analysis was refined in [23], showing that each hidden unit can model the coefficients of as many as n monomials in the energy function.

Mixtures of Products and Partition Models

We discuss a result from [24] showing that an RBM with m hidden units can represent mixtures of $m + 1$ product distributions, provided the support sets of m of the mixture components are disjoint. The support of a distribution p on \mathcal{X} is $\text{supp}(p) := \{x \in \mathcal{X} : p(x) > 0\}$. The idea is as follows. Consider an entrywise product of the form

$$p_0(x)(1 + \lambda p_1(x)) = p_0(x) + \lambda p_0(x)p_1(x), \quad x \in \mathcal{X}. \quad (25)$$

If p_0 and p_1 are product distributions, then so is $p_2 = p_0 p_1$. This is a direct consequence of the fact that the set of product distributions has an affine set of exponential parameters, $\exp(w_0^\top x)\exp(w_1^\top x) = \exp((w_0 + w_1)^\top x) = \exp(w_2^\top x)$. In turn, an entrywise product of the form (25) expresses a linear combination of product distributions, provided that p_0 and p_1 are product distributions. The last requirement can be relaxed to hold only over the intersection of the support sets of p_0 and p_1 , since the entrywise product will vanish on the other entries either way. When we renormalize, the linear combination becomes a mixture of product distributions, whereby the relative mixture weights are controlled by λ .

Now recall from Sect. 3 that the RBM distributions can be written as

$$p(x; \theta) = \frac{1}{Z(\theta)} \exp(b^\top x) \prod_{j \in H} (1 + \exp(c_j) \exp(W_{j,:}x)). \quad (26)$$

By the previous discussion, we can interpret each factor in (26) as adding a mixture component $p_j(x) = \frac{1}{Z} \exp(W_{j,:}x)$, which is a product distribution, so long as the distribution obtained from the preceding factors is a product distribution over the support of p_j . Being an exponential family distribution, p_j has full support, but it can approximate product distributions with restricted support arbitrarily well.

A similar discussion applies to non-binary variables, as shown in [27]. We denote by $\text{RBM}_{\mathcal{X}, \mathcal{Y}}$ the RBM with visible states $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and hidden states $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$. This is the set of marginals of the exponential family with sufficient statistics given by the Kronecker product of the statistics of the independence models on \mathcal{X} and \mathcal{Y} , respectively.

Theorem 2 ([43, Theorem 3]) *Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ be finite sets. The model $\text{RBM}_{\mathcal{X}, \mathcal{Y}}$ can approximate any mixture distribution $p(x) = \sum_{i=0}^m \lambda_i p_i(x)$, $x \in \mathcal{X}$, arbitrarily well, where p_0 is any product distribution, and p_i are respectively for all $i \in [m]$, any mixtures of $(|\mathcal{Y}_i| - 1)$ product distributions, with support sets satisfying $\text{supp}(p_i) \cap \text{supp}(p_j) = \emptyset$ for all $1 \leq i < j \leq m$.*

In particular, the binary $\text{RBM}_{n,m}$ can approximate, to within any desired degree of accuracy, any mixture of $m + 1$ product distributions with disjoint supports. Given a collection of disjoint sets $A_1, \dots, A_{m+1} \subseteq \mathcal{X}$, the set of mixtures $p = \sum_j \lambda_j p_j$, where each p_j is a product distribution with support set A_j , is an exponential family on $\cup_j A_j$. More precisely, its topological closure coincides with that of an exponential family with sufficient statistics $\mathbb{1}_{A_j}, \mathbb{1}_{A_j} x_i, i = 1, \dots, n, j = 1, \dots, m + 1$. Theorem 2 shows that an RBM can represent all such exponential families, for all choices of disjoint sets A_1, \dots, A_{m+1} .

A *partition model* is a special type of mixture model, consisting of all mixtures of a fixed set of uniform distributions on disjoint support sets. Partition models are interesting not only because of their simplicity, but also because they are optimally approximating exponential families of a given dimension. If all support sets of the components, or blocks, have the same size, then the partition model attains the smallest uniform approximation error, measured in terms of the Kullback–Leibler divergence, among all exponential families that have the same dimension [69]. The previous theorem shows that RBMs can approximate certain partition models arbitrarily well. In particular we have:

Corollary 3 *Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ and $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$ be finite sets. Let \mathcal{P} be the partition model with partition blocks $\{x_1\} \times \cdots \times \{x_k\} \times \mathcal{X}_{k+1} \times \cdots \times \mathcal{X}_n$ for all $(x_1, \dots, x_k) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_k$. If $1 + \sum_{j \in [m]} (|\mathcal{Y}_j| - 1) \geq (\prod_{i \in [k]} |\mathcal{X}_i|) / \max_{j \in [k]} |\mathcal{X}_j|$, then each distribution contained in \mathcal{P} can be approximated arbitrarily well by distributions from $\text{RBM}_{\mathcal{X}, \mathcal{Y}}$.*

Hierarchical Models

Intuitively, each hidden unit of an RBM should be able to mediate certain interactions between the visible units. To make this more concrete, we may ask which distributions from a hierarchical model can be expressed in terms of an RBM, or which parameters of a hierarchical model can be modeled in terms of the hidden units of an RBM. Younes [25] showed that a binary hierarchical model with a total of K pure higher order interactions can be modeled by an RBM with K hidden units. Later, [23] showed that each hidden unit of an RBM can model several parameters of a hierarchical model simultaneously.

Consider a set $S \subseteq 2^V$ of subsets of V . A hierarchical model with interactions S is defined as the set of probability distributions p that can be factorized as

$$p(x) = \prod_{\lambda \in S} \psi_\lambda(x), \quad x \in \mathcal{X}, \quad (27)$$

where each $\psi_\lambda : \mathcal{X} \rightarrow \mathbb{R}_+$ is a positive valued function that only depends on the coordinates λ , i.e., satisfies $\psi_\lambda(x) = \psi_\lambda(x')$ whenever $x_i = x'_i$ for all $i \in \lambda$. In practice, we choose a basis to express the potentials as parametrized functions. The set S is conveniently defined as the set of cliques of a graph $G = (V, E)$, and hence these models are also known as hierarchical graphical models. These models are very intuitive and have been studied in great detail. Each factor ψ_λ is interpreted as allowing us to model arbitrary interactions between the variables $x_i, i \in \lambda$, independently of the variables $x_j, j \in V \setminus \lambda$. Hence, they are a good reference to compare the representational power other models, which is what we want to do for RBMs in the following.

At a high level, the difficulty of comparing RBMs and hierarchical models stems from the fact that their parameters contribute different types of degrees of freedom. While a hidden unit can implement interactions among all visible units it is connected to, certain constraints apply on the values of these interactions. For example, the set of interaction coefficients among two visible variables that can be modeled by one hidden unit is shown in Fig. 7.

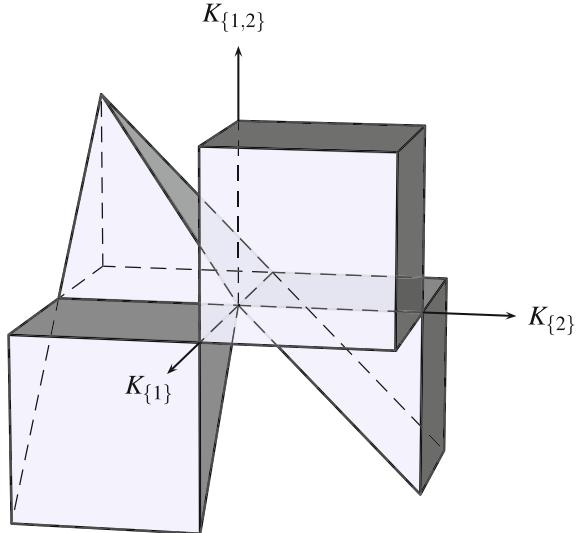
To proceed with more details, we first fix a coordinate system. Hierarchical models are conveniently expressed in terms of a basis of orthogonal functions known as characters. For each $\lambda \subseteq V$ we have a function

$$\sigma_\lambda(x) = \prod_{i \in \lambda} (-1)^{x_i}, \quad x \in \{0, 1\}^V.$$

The functions $\sigma_\lambda, \lambda \subseteq V$, are orthogonal, with $\sum_x \sigma_\lambda(x) \sigma_\mu(x) = 2^n \delta_{\lambda, \mu}$. In turn, we can express any given vector $l \in \mathbb{R}^{\{0, 1\}^V}$ as

$$l(x) = \sum_{\lambda \subseteq V} J_\lambda \sigma_\lambda(x), \quad x \in \{0, 1\}^V,$$

Fig. 7 Interaction coefficients expressible by one RBM hidden unit. Shown is the set of coefficients $(K_{\{1\}}, K_{\{2\}}, K_{\{1,2\}}) \in \mathbb{R}^3$, clipped to a cube centered at the origin, of the polynomials $K_\emptyset + K_{\{1\}}x_1 + K_{\{2\}}x_2 + K_{\{1,2\}}x_1x_2$ expressible in terms of a soft-plus unit on binary inputs. Figure adapted from [23]



where the coefficients are given by

$$J_\lambda = \frac{1}{2^n} \sum_{x \in \{0,1\}^V} \sigma_\lambda(x) l(x), \quad \lambda \subseteq V.$$

The change of coordinates from the standard basis δ_x , $x \in \{0, 1\}^V$, to the basis of characters σ_λ , $\lambda \subseteq V$, can be interpreted as a Möbius inversion, or also as a Fourier transform.

If we replaced the states $\{0, 1\}$ with $\{+1, -1\}$, we could write each σ_λ as a monomial $\prod_{i \in \lambda} x_i$. But we can also use a basis of monomials without changing the states. For each $\lambda \subseteq V$, let

$$\pi_\lambda(x) = \prod_{i \in \lambda} x_i, \quad x \in \{0, 1\}^V. \quad (28)$$

Although this is no longer an orthogonal basis, it is conceptually simple and very frequently used in practice. Moreover, for an inclusion closed set $S \subseteq 2^V$, the span of π_λ , $\lambda \in S$, equals that of σ_λ , $\lambda \in S$, such that both bases have the same hierarchical coordinate sub-spaces.

For an inclusion closed set $S \subseteq 2^V$, the binary hierarchical model with interactions S can be parametrized as the exponential family \mathcal{E}_S of distributions of the form

$$p(x) = \frac{1}{Z} \exp \left(\sum_{\lambda \in S} J_\lambda \prod_{i \in \lambda} x_i \right), \quad x \in \{0, 1\}^V, \quad (29)$$

with parameters $J_\lambda \in \mathbb{R}$, $\lambda \in S$.

Now we proceed with the representation of the parameters of a hierarchical model in terms of an RBM. Recall that the log-probabilities $l = \log(p)$ in the model $\text{RBM}_{n,m}$ are sums of a linear unit and m soft-plus units. For a linear unit $w^\top x + c$, the polynomial coefficients are simply $K_\emptyset = c$, $K_{\{i\}} = w_i$, $i \in V$, and $K_\lambda = 0$ for all $\lambda \subseteq V$ with $|\lambda| \geq 2$. For a soft-plus unit, [23] obtains a partial characterization of the possible polynomial coefficients. In particular, it shows the following.

Lemma 4 ([23, Lemma 5]) *Consider a subset $B \subseteq V$, and let $J_{B \cup \{j\}} \in \mathbb{R}$, $j \in V \setminus B$, and $\varepsilon > 0$. Then there are $w \in \mathbb{R}^V$ and $c \in \mathbb{R}$ such that the soft-plus unit $\log(1 + \exp(w^\top x + c))$ is equal to a polynomial $\sum_\lambda K_\lambda \prod_{i \in \lambda} x_i$ with coefficients satisfying $|K_{B \cup \{j\}} - J_{B \cup \{j\}}| \leq \varepsilon$ for all $j \in V \setminus B$, and $|K_C| \leq \varepsilon$ for all $C \neq B$, $B \cup \{j\}$, $j \in V \setminus B$.*

This says that each hidden unit of an RBM can model arbitrarily the parameters of a hierarchical model corresponding to the monomials that cover $\prod_{i \in B} x_i$, for any fixed choice of $B \subseteq V$, while at the same time setting all other parameters arbitrarily close to zero, except for the parameter associated with $\prod_{i \in B} x_i$, whose value may be coupled to the values of the other parameters.

We can use this result to describe hierarchical models that can be represented by an RBM. Since each hidden unit of the RBM can model certain subsets of parameters of hierarchical models, we just need to find a sufficiently large number of hidden units which together can model all the required parameters. For example:

- $\text{RBM}_{3,1}$ contains the hierarchical models \mathcal{E}_S with $S = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, S = \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}\}$, $S = \{\{1\}, \{2\}, \{3\}, \{1, 3\}, \{2, 3\}\}$. It does not contain the *no-three-way interaction model*, with $S = S_2 = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$.
- The model $\text{RBM}_{3,2}$ contains the no-three-way interaction model \mathcal{E}_S with $S = S_2$. It does not contain the full interaction model, with $S = S_3$. In particular, this model is not a universal approximator.

In general, finding a minimal cover of the relevant set of parameters of hierarchical models in terms of subsets of parameters of the form described in Lemma 4 relates to well-known problems in the theory of combinatorial designs. For S consisting of all sets up to a given cardinality, we can obtain the following bounds.

Theorem 5 ([23, Theorem 11]) *Let $1 \leq k \leq n$ and $\mathcal{X} = \{0, 1\}^V$. Every distribution from the hierarchical model \mathcal{E}_{S_k} , with $S_k = \{\lambda \subseteq V : |\lambda| \leq k\}$, can be approximated arbitrarily well by distributions from $\text{RBM}_{n,m}$ whenever*

$$m \geq \min \left\{ \sum_{j=2}^k \binom{n-1}{j-1}, \frac{\log(n-1) + 1}{n+1} \sum_{j=2}^k \binom{n+1}{j} \right\}.$$

We note that in specific cases there are sharper bounds available, listed in [23].

The hidden units and parameters of an RBM can be employed to model different kinds of hierarchical models. For instance, a limited number of hidden units could

model the set of full interactions among a small subset of visible variables, or, alternatively, to model all k -wise interactions among a large set of visible units. Exactly characterizing the largest hierarchical models that can be represented by an RBM is still an open problem for $n \geq 4$.

Universal Approximation

The universal approximation question asks for the smallest model within a class of models, which is able to approximate any given probability distribution on its domain to within any desired degree of accuracy. This is a special case of the problems discussed in the previous paragraphs. A direct consequence of Theorem 11 is

Corollary 6 *Let $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ and $\mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$ be finite sets. The model $\text{RBM}_{\mathcal{X}, \mathcal{Y}}$ is a universal approximator whenever*

$$1 + \sum_{j \in [m]} (|\mathcal{Y}_j| - 1) \geq |\mathcal{X}| / \max_{i \in [n]} |\mathcal{X}_i|.$$

When all units are binary, this implies that an RBM with $2^{n-1} - 1$ hidden units is a universal approximator of distributions on $\{0, 1\}^n$. Theorem 5 improves this bound as follows:

Corollary 7 ([23, Corollary 12]). *Every distribution on $\{0, 1\}^n$ can be approximated arbitrarily well by distributions from $\text{RBM}_{n,m}$ whenever*

$$m \geq \min \left\{ 2^{n-1} - 1, \frac{2(\log(n-1) + 1)}{n+1} (2^n - (n+1) - 1) + 1 \right\}.$$

This is the sharpest general upper bound that is available at the moment. A slightly looser but simpler bound is $\frac{2(\log(n)+1)}{n+1} 2^n - 1$. Again, in specific cases there are sharper bounds available, listed in [23].

In terms of the necessary number of hidden units for universal approximation, bounds have been harder to obtain. In the general case, we only have lower bounds coming from parameter counting arguments:

Proposition 8 *Let \mathcal{M} be an exponential family over $\mathcal{X} \times \mathcal{Y}$ and \mathcal{M}_V the set of marginals on \mathcal{X} . If \mathcal{M}_V is a universal approximator, then \mathcal{M}_V has dimension $|\mathcal{X}| - 1$ and \mathcal{M} has dimension at least $|\mathcal{X}| - 1$.*

This implies that for $\text{RBM}_{n,m}$ to be a universal approximator, necessarily $m \geq 2^n/(n+1) - 1$. There is still a logarithmic gap between the upper and lower bounds. Further closing this gap is an important theoretical problem, which could help us obtain a more complete understanding of the representational power question. In a few small cases we can obtain the precise numbers. For instance, for $n = 2$, the minimal size of a universal approximator is $m = 1$. For $n = 3$ it is $m = 3$. But already for $n = 4$ we can only bound the exact value between 3 and 6.

Relative Representational Power

As we have seen, RBMs can represent certain mixtures of product distributions. Complementary to this, it is natural to ask how large a mixture of products is needed in order to represent an RBM. Following Sect. 3, an RBM model consists of tensors which are entrywise products of tensors of with non-negative rank at most two. For many combinations of n and m it turns out that the RBM model represents tensors of the maximum possible rank, 2^m , which implies that the smallest mixture of products that contain the RBM model is as large as one could possibly expect, having 2^m components:

Theorem 9 ([32, Theorem 1.2]) *The smallest k for which the model $\mathcal{M}_{n,k}$, consisting of arbitrary mixtures of k product distributions of n binary variables, contains the model RBM $_{n,m}$, is bounded by $\frac{3}{4}n \leq \log_2(k) \leq n - 1$ when $m \geq n$, by $\frac{3}{4}n \leq \log_2(k) \leq m$ when $\frac{3}{4}n \leq m \leq n$, and satisfies $\log_2(k) = m$ when $m \leq \frac{3}{4}n$.*

As shown in [32] RBMs can express distributions with many more strong modes than mixtures of products with the same number of parameters. A strong mode is a local maximum of the probability distribution, with value larger than the sum of all its neighbors, whereby the vicinity structure is defined by the Hamming distance over the set of elementary events. Distributions with many strong modes have a large non-negative tensor rank. At the same time, [32] shows that an RBM does not always contain a mixture of products model with the same number of parameters. The size of the largest mixture of products that is contained in an RBM is still an open problem.

For hierarchical models, Lemma 4 allows us to formulate an analogous result. The lemma implies that a hidden unit can create non-zero values of any parameter of any arbitrary hierarchical model. In turn, the smallest hierarchical model that contains an RBM must have all possible interactions and hence it is as large as one could possibly expect:

Proposition 10 *Let $n, m \in \mathbb{N}$. The smallest $S \subseteq 2^V$ for which the hierarchical model \mathcal{E}_S on $\{0, 1\}^V$ contains RBM $_{n,m}$ is $S = 2^V$.*

7 Divergence Bounds

Instead of asking for the sets of distributions that can be approximated arbitrarily well by an RBM, we can take a more refined standpoint and ask for the error in the approximation of a given target distribution. The best possible uniform upper bound on the divergence to a model \mathcal{M} is $D_{\mathcal{M}} = \max_p D(p \parallel \mathcal{M}) = \max_p \inf_{q \in \mathcal{M}} D(p \parallel q)$.

Maximizing the divergence to a model, over the set of all possible targets, is an interesting problem in its own right. For instance, the divergence to an independence model is called multi-information and can be regarded as a measure of complexity. The multi-information can be used as an objective function in certain learning problems, as a way to encourage behaviors that are both predictable and diverse. The

divergence maximization problem is challenging, even in the case of exponential families with closed formulas for the maximum likelihood estimators. For exponential families models the divergence maximization problem has been studied in particular by Matúš [70], Ay [71], and Rauh [72].

In the case of RBMs, as with most machine learning models used in practice, the situation is further complicated, since we do not have closed formulas for the error minimizers of a given target. The approximation errors of RBMs were studied in [24] by showing that RBMs contain a number of exponential families and providing upper bounds on the divergence to such families. The approach was formulated more generally in [73]. In [74] it was shown how to obtain upper bounds on the expected value of the approximation error, when the target distributions are sampled from a given prior. In the following we discuss some of these bounds and also a divergence bound derived from the hierarchical models presented in Sect. 6.

Upper Bounds from Unions of Mixtures of Products and Hierarchical Models

The Kullback–Leibler divergence from a distribution q to another distribution p is

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

Given some p , we are interested in the best approximation within a given model \mathcal{M} . We consider the function that maps each possible target distribution p to

$$D(p\|\mathcal{M}) = \inf_{q \in \mathcal{M}} D(p\|q).$$

The divergence to a partition model \mathcal{P}_A with blocks A_k , $k = 1, \dots, K$, is bounded above by $D(\cdot\|\mathcal{P}_A) \leq \max_k \log |A_k|$. This bound is in fact tight. Corollary 3 shows that RBMs can represent certain partition models. This implies the following bound.

Theorem 11 ([43, Theorem 5]) *Let $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ and $\mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$ be finite sets. If $1 + \sum_{j \in [m]} (|\mathcal{Y}_j| - 1) \geq |\mathcal{X}_{\Lambda \setminus \{k\}}$ for some $\Lambda \subseteq [n]$ and $k \in \Lambda$, then*

$$D(\cdot\|\text{RBM}_{\mathcal{X}, \mathcal{Y}}) \leq \log |\mathcal{X}_{[n] \setminus \Lambda}|.$$

Instead of partition models, we can also consider mixtures of product distributions with disjoint supports, as described in Theorem 2. As discussed in [24] the divergence to a mixture of models with disjoint supports can be bounded tightly from above by the maximum divergence to one of the component models over targets with the same support. Consider a model \mathcal{M} consisting of mixtures $\sum_j \lambda_j p_j$ of distributions $p_j \in \mathcal{M}_j$, where \mathcal{M}_j consists of distributions supported on A_j , and $A_i \cap A_j = \emptyset$ whenever $i \neq j$. Then

$$\max_p D(p\|\mathcal{M}) = \max_j \max_{p: \text{ supp}(p) \subseteq A_j} D(p\|\mathcal{M}_j).$$

We know that the RBM contains several mixtures of products with disjoint supports. Hence we can further improve the divergence upper bounds by considering the divergence to the union of all the models that are contained in the RBM model. This gives the following bound.

Theorem 12 ([73, Theorem 2]) *If $m \leq 2^{n-1} - 1$,*

$$D(\cdot\|\text{RBM}_{n,m}) \leq \left(n - \lfloor \log_2(m+1) \rfloor - \frac{m+1}{2^{\lfloor \log_2(m+1) \rfloor}} \right) \log(2).$$

A corresponding analysis for RBMs with non-binary units still needs to be worked out.

We can also bound the divergence in terms of the hierarchical models described in Theorem 5, instead of the partition models and mixtures of products mentioned above. Matúš [70] studies the divergence to hierarchical models, and proves, in particular, the following bound.

Lemma 13 ([70, Corollary 3]). *Consider an inclusion closed set $S \subseteq 2^V$ and the hierarchical model \mathcal{E}_S on $\{0, 1\}^V$. Then $D(\cdot\|\mathcal{E}_S) \leq \min_{A \in S} \log |\mathcal{X}_{V \setminus A}|$.*

In conjunction with Theorem 7, this directly implies the following bound.

Corollary 14 *Let $n, m \in \mathbb{N}$, and let k be the largest integer with $m \geq \frac{\log(k)+1}{k+1} 2^{k+1} - 1$. Then $D(\cdot\|\text{RBM}_{n,m}) \leq (n-k) \log(2)$.*

A version of this result for non-binary variables and bounding the divergence to unions of hierarchical models still need to be worked out.

Divergence to Polyhedral Exponential Families

The previous results estimate the divergence to an RBM model by looking at the divergence to exponential families or unions of exponential families that are contained within the RBM model (or within its closure, to be more precise). More generally, we might be interested in estimating the divergence to models whose set of log-probabilities forms a polyhedral shape, as the one shown in Fig. 7. Each face of a polyhedron can be extended to an affine space, and hence corresponds to a piece of an exponential family. This allows us to compute the maximum likelihood estimators of a polyhedral family in the following way. A related discussion was conducted recently in [75] in the context of mixtures of products, and in [30] in the context of RBMs.

Given a target distribution p and a model with log-probabilities from a polyhedron \mathcal{M} we proceed as follows.

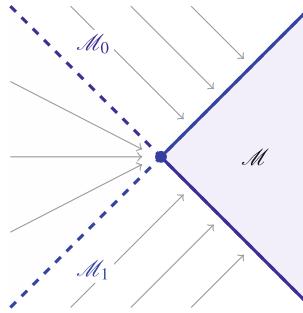


Fig. 8 Illustration of the maximum likelihood projections onto a model whose log-probabilities form a polyhedron. Here the polyhedron \mathcal{M} consists of the points on the positive side of two hyperplanes, \mathcal{M}_0 and \mathcal{M}_1 . Each face of the polyhedron extends to an affine space that corresponds to an exponential family. For each possible target, each exponential family has a unique maximum likelihood projection point. Arrows indicate how targets project to the different faces of \mathcal{M}

- For each face \mathcal{M}_i of \mathcal{M} , we define a corresponding exponential family \mathcal{E}_i . Any basis of the affine hull of \mathcal{M}_i forms a sufficient statistics, and we can take any point in \mathcal{M}_i as a reference measure.
- Then we compute the maximum likelihood estimator $q_i = \operatorname{arginf}_{q \in \mathcal{E}_i} D(p \| q)$ for each individual exponential family \mathcal{E}_i . For exponential families the maximum likelihood estimation problem is concave and has a unique solution (possibly on the closure of the exponential family).
- Then we verify which of the projections q_i are feasible, meaning that they satisfy the constraints of the corresponding face \mathcal{M}_i .
- Finally, we select among the feasible projections, the one with the smallest divergence to the target distribution p . This is illustrated in Fig. 8.

Tightness of the Bounds

In the previous paragraphs we provided upper bounds on the divergence from arbitrary target distributions to an RBM model. One may wonder about the tightness of these bounds. For the special case of independence models, which are RBMs with no hidden units, the bounds are tight, provided all visible variables have state spaces of equal cardinality. However, already in the case of one single hidden unit, the exact value of the maximum divergence is not known in general.

Experiments on small RBMs [24, 43] seem to indicate that the bounds provided in the previous paragraphs are in good agreement with the actual values. Empirical studies are difficult because of two opposing effects. On the one hand, sequential optimization methods may only lead to sub-optimal approximations of a given target. In fact, part of the motivation for deriving theoretical upper bounds is to monitor the quality of our sequential optimization methods. On the other hand, finding a target distribution with maximum divergence to the model may be a difficult problem itself.

It may be that the vast majority of possible targets are not as far to the model as the divergence maximizer. In turn, the theoretical upper bounds could appear pessimistic for most of the targets. In [74] it is shown how to estimate the expected value of the divergence when the target distributions are sampled from a Dirichlet distribution. The average values tend to be indeed much lower than the maximum values.

A recent work [30] shows that the model $\text{RBM}_{3,2}$ has a boundary described in terms of a union of exponential families, and uses this description to obtain the divergence maximizers to the model. It shows that the divergence bounds obtained in Theorem 12 are tight for this particular model.

Theorem 15 ([30, Theorem 3]). *The maximum divergence to $\text{RBM}_{3,2}$ is $\frac{1}{2} \log 2$. The maximizers are $\frac{1}{4}(\delta_{000} + \delta_{011} + \delta_{101} + \delta_{110})$ and $\frac{1}{4}(\delta_{001} + \delta_{010} + \delta_{100} + \delta_{111})$. For each of these targets, there is one distinct projection point on each of the six boundary pieces of $\text{RBM}_{3,2}$.*

8 Implicit Description

So far we have discussed probability models presented explicitly, as parametric families of distributions. RBMs can also be expressed implicitly, in terms of constraints that apply to the distributions within the model, and only to the distributions within the model. Indeed, since RBMs have a polynomial parametrization, they can be described semi-algebraically as the set of real solutions to a collection of polynomial equations and polynomial inequalities. The *implicitization problem* consists of replacing a parametric description with a description as the solution set of a collection of equations and inequalities. Finding implicit characterizations for graphical models with hidden variables is a significant challenge and a central topic within algebraic statistics [14, 15]. In principle both, explicit and implicit presentations, can be challenging to interpret in general, for instance when the parametrization is convoluted, or when the constraints correspond to complicated properties of the distributions. However, in some cases the implicit descriptions have a very intuitive statistical interpretation and can allow us to make significant advances over what is possible with a parametric description alone. Implicit descriptions can be extremely useful for hypothesis testing, membership testing, and other related problems. So far there are not many results on the implicit description of RBMs. The following discussion is intended as a motivation.

Markov Properties

A fully observable undirected graphical model can be defined in terms of the factorization property (27). Each of the factors can be considered as a parameter, or can be easily parametrized, as shown in (29). Graphical models are usually also motivated and defined in terms of so-called Markov properties, or conditional independence

statements. These are constraints that characterize the probability distributions in the model. Undirected graphical models encode conditional independence relations in terms of the structure of the graph. Specifically, a probability distribution is contained in an undirected graphical model with graph G if and only if it satisfies all conditional independence statements encoded by the graph G , namely

$$X_A \perp\!\!\!\perp X_B \mid X_C, \quad (30)$$

whenever A, B, C are disjoint subsets of V for which any path connecting a point in A and a point in B , passes through C . Equation (30) means that p satisfies the equations $p(x_A, x_B | x_C) = p(x_A | x_C)p(x_B | x_C)$, or, equivalently,

$$p(x_A, x_B, x_C) \sum_{x'_A, x'_B} p(x'_A, x'_B, x_C) - \sum_{x'_B} p(x_A, x'_B, x_C) \sum_{x'_A} p(x'_A, x_B, x_C) = 0,$$

for all $x_A \in \mathcal{X}_A, x_B \in \mathcal{X}_B, x_C \in \mathcal{X}_C$. These are quadratic binomial equations in the indeterminates $p(x) \in \mathbb{R}, x \in \mathcal{X}$. A famous theorem by Hammersley and Clifford [76] gives the correspondence between the conditional independence constraints and the factorization property of the joint distributions in a fully observable graphical model. This correspondence is usually limited to strictly positive probability distributions. For distributions that are not strictly positive, which lie at the boundary of the probability simplex, the correspondence is more subtle in general and has been investigated in [77]. The main point here is that we can formulate a parametric set of functions in terms of constraints, or properties of distributions. Moreover, at least in the case of fully observable undirected graphical models, the constraints have an intuitive statistical interpretation.

Constraints in a Small RBM

A natural question is what are the constraints that define the visible distributions in a an RBM, and more generally, in a hierarchical model with hidden variables. Aside from RBMs with one single hidden unit, which correspond to mixtures of two product distributions, the RBM with 4 visible and 2 hidden variables has been studied, which turns out to be a hyper-surface defined as the zero set of a polynomial with over a trillion monomials [29].

The constraints that apply to $\text{RBM}_{3,2}$ were studied in [32], obtaining a coarse description of the model. The full semi-algebraic description of this model was then obtained in [30]. The characterization is as follows.

Theorem 16 ([30, Theorem 1]). *The model $\text{RBM}_{3,2}$ is the union of six basic semi-algebraic sets, each described by two inequalities, namely:*

$$\begin{aligned}
& \{p_{000}p_{011} \geq p_{001}p_{010}, \quad p_{100}p_{111} \geq p_{101}p_{110}\} \\
& \{p_{000}p_{011} \leq p_{001}p_{010}, \quad p_{100}p_{111} \leq p_{101}p_{110}\} \\
& \{p_{000}p_{101} \geq p_{001}p_{100}, \quad p_{010}p_{111} \geq p_{011}p_{110}\} \\
& \{p_{000}p_{101} \leq p_{001}p_{100}, \quad p_{010}p_{111} \leq p_{011}p_{110}\} \\
& \{p_{000}p_{110} \geq p_{100}p_{010}, \quad p_{001}p_{111} \geq p_{101}p_{011}\} \\
& \{p_{000}p_{110} \leq p_{100}p_{010}, \quad p_{001}p_{111} \leq p_{101}p_{011}\}.
\end{aligned}$$

Each pair of inequalities represents the non-negativity or non-positivity of two determinants. These determinants capture the conditional correlations of two of the variables, given the value of the third variable. The conditional correlation is either non-negative or non-positive for both possible values of the third variable.

This theorem gives a precise description of the geometry of the model. The model is full dimensional in the ambient probability simplex. Hence the description involves only inequalities and no equations (aside from the normalization constraint $\sum_x p_x = 1$). Setting either of the inequalities to an equation gives a piece of the boundary of the model. Each boundary piece is an exponential family which can be interpreted as the set of mixtures of one arbitrary product distribution and one product distribution with support on the states with fixed value of one of the variables, similar to the distributions described in Theorem 2. For these exponential families we can compute the maximum likelihood estimators in closed form, as described in the previous paragraph, and also obtain the exact maximizers of the divergence, given in Theorem 15. With the implicit description at hand [30] also shows that the model $\text{RBM}_{3,2}$ is equal to the mixture model of three product distributions, and that it does not contain any distributions with 4 modes, both statements that had been conjectured in [32].

Coarse Necessary Constraints

Obtaining the exact constraints that define an RBM model can be difficult in general. In Sect. 6 we described submodels of the RBM, which can be interpreted as constraints that are sufficient for probability distributions to be contained in the model, but not necessary. A complementary alternative is to look for constraints that are necessary for distributions to be in the model, but not sufficient. These sometimes are easier to obtain and interpret. An example are strong mode inequalities in mixtures of product distributions [32], and information theoretic inequalities in Bayesian networks [78]. Mode inequality constraints for RBMs have been studied in [32]. Another possible direction was suggested in [30], namely to consider the inequality constraints that apply to mixtures of two product distributions and how they combine when building Hadamard products.

9 Open Problems

The theory of RBMs is by no means a finished subject. In the following, I collect a selection of problems, as a sort of work program, addressing which I think is important towards obtaining a more complete picture of RBMs and advancing the theory of graphical models with hidden variables in general.

1. Can we find non-trivial constraints on the sets of representable probability distributions? A related type of questions has been investigated in [32], with focus on the approximation of distributions with many modes, or mixtures of product distributions.
2. Closely related to the previous item, given the number n of visible units, what is the smallest number m of hidden units for which $\text{RBM}_{n,m}$ is a universal approximator? Alternatively, can we obtain lower bounds on the number of hidden units of an RBM that is a universal approximator? Here, of course, we are interested in lower bounds that do not readily follow from parameter counting arguments.
The first open case is $n = 4$, for which we have bounds $3 \leq m \leq 6$.
3. What is the smallest tropical RBM that is a universal approximator? Equivalently, what is the smallest m for which a sum of one affine function and m ReLUs can express any function of n binary variables?
4. Characterize the support sets of the distributions in the closure of an RBM. We note that characterizing the support sets of distributions in the closure of an exponential family corresponds to describing the faces its convex support polytope.
5. Also in relation to the first item, obtain an implicit description of the RBM model. The work [30] gives the description of $\text{RBM}_{3,2}$ and ideas for the inequality constraints of larger models. Interesting cases to consider are $\text{RBM}_{4,3}$ (this might be the full probability simplex), $\text{RBM}_{5,2}$, $\text{RBM}_{6,5}$. For the latter [32] obtained some linear inequality constraints.
6. Can we produce explicit descriptions of the maximum likelihood estimators? Here [30] indicates possible avenues.
7. Describe the structure of the likelihood function of an RBM. In particular, what is the number of local and global optimizers? How does this number depend on the empirical data distribution?
8. Describe the critical points of the EM algorithm for an RBM model or for its Zariski closure.
9. Characterize the sets of parameters that give rise to the different distributions expressible by an RBM. When this is finite, are there parameter symmetries other than those coming from relabeling units and states?
10. What is the maximum possible value of the divergence to an RBM model, $D_{n,m} = \max_{p \in \Delta_{\{0,1\}^n}} \inf_{q \in \text{RBM}_{n,m}} D(p \| q)$, and what are the divergence maximizers? We know $D_{3,0} = 2 \log 2$ from results for independence models (see, e.g., [73]), and $D_{3,2} = \frac{1}{2} \log 2$ (see Theorem 16 and [30]). The first open case is $D_{3,1}$. Discussions with Johannes Rauh suggest $-\frac{3}{4} \log_2(2\sqrt{3} - 3)$.

11. In relation to the previous item, can we provide lower bounds on the maximum divergence from a given union of exponential families?
12. Does the tropical RBM model have the expected dimension? In [26] it was conjectured that it does. The problem remains open, even though [28] gave a proof for the RBM. The description of the tropical RBM as a superposition of ReLUs could be useful here.
13. What is the largest mixture of product distributions that is contained in the RBM model? A result from [32] shows that RBMs do not always contain mixtures of products of the same dimension.
14. What are the largest hierarchical models that are contained in the closure of an RBM model? A partial characterization of the polynomials that are expressible in terms of soft-plus and rectified linear units on binary inputs was obtained in [23]. A full characterization is still missing.
15. Generalize the analysis of hierarchical models contained in RBM models to the case of non-binary variables (both visible and hidden).

Acknowledgements I thank Shun-ichi Amari for inspiring discussions over the years. This review article originated at the IGAIA IV conference in 2016 dedicated to his 80th birthday. I am grateful to Nihat Ay, Johannes Rauh, Jason Morton, and more recently Anna Seigal for our collaborations. I thank Fero Matúš for discussions on the divergence maximization for hierarchical models, lastly at the MFO Algebraic Statistics meeting in 2017. I thank Bernd Sturmfels for many fruitful discussions, and Dave Ackley for insightful discussions at the Santa Fe Institute in 2016. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no 757983).

References

1. Ackley, D.H., Hinton, G.E., Sejnowski, T.J.: A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**(1), 147–169 (1985)
2. Hinton, G.E., Sejnowski, T.J.: Analyzing cooperative computation. In: Proceedings of the Fifth Annual Conference of the Cognitive Science Society. Rochester, NY (1983)
3. Hinton, G.E., Sejnowski, T.J.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Chapter learning and relearning in boltzmann machines, pp. 282–317. MIT Press, USA (1986)
4. Hopfield, J.J.: Neurocomputing: Foundations of Research. Chapter neural networks and physical systems with emergent collective computational abilities, pp. 457–464. MIT Press, USA (1988)
5. Huang, K.: Statistical Mechanics. Wiley, New York (2000)
6. Gibbs, J.: Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundations of Thermodynamics. C. Scribner’s sons (1902)
7. Brown, L.: Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory. Institute of Mathematical Statistics, USA (1986)
8. Jordan, M.I.: Graphical models. *Stat. Sci.* **19**(1), 140–155 (2004)
9. Lauritzen, S.L.: Graphical Models. Oxford University Press, USA (1996)
10. Amari, S.: Information geometry on hierarchical decomposition of stochastic interactions. *IEEE Trans. Inf. Theory* **47**, 1701–1711 (1999)
11. Amari, S.: Information Geometry and its Applications. Applied mathematical sciences, vol. 194. Springer, Japan (2016)

12. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of mathematical monographs. American Mathematical Society (2007)
13. Ay, N., Jost, J., Lê, H., Schwachhöfer, L.: Information Geometry. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 64. Springer, Berlin (2017)
14. Drton, M., Sturmfels, B., Sullivant, S.: *Lectures on Algebraic Statistics*. Springer, Oberwolfach Seminars (2009)
15. Sullivant, S.: *Algebraic Statistics* (2018)
16. Amari, S., Kurata, K., Nagaoka, H.: Information geometry of Boltzmann machines. *IEEE Trans. Neural Netw.* **3**(2), 260–271 (1992)
17. Smolensky, P.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Chapter information processing in dynamical systems: foundations of harmony theory, vol. 1, pp. 194–281. MIT Press, USA (1986)
18. Freund, Y., Haussler, D.: Unsupervised learning of distributions on binary vectors using two layer networks. In: Moody, J.E., Hanson, S.J., Lippmann, R.P. (eds.) *Advances in Neural Information Processing Systems 4*, pp. 912–919. Morgan-Kaufmann (1992)
19. Bengio, Y.: Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2**(1), 1–127 (2009). Also published as a book. Now Publishers
20. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006)
21. Fischer, A., Igel, C.: An introduction to restricted Boltzmann machines. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 14–36. Springer, Heidelberg (2012)
22. Le Roux, N., Bengio, Y.: Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput.* **20**(6), 1631–1649 (2008)
23. Montúfar, G., Rauh, J.: Hierarchical models as marginals of hierarchical models. *Int. J. Approx. Reason.* **88**, 531–546 (2017). (Supplement C)
24. Montúfar, G., Rauh, J., Ay, N.: Expressive power and approximation errors of restricted Boltzmann machines. *Adv. Neural Inf. Process. Syst.* **24**, 415–423 (2011)
25. Younes, L.: Synchronous Boltzmann machines can be universal approximators. *Appl. Math. Lett.* **9**(3), 109–113 (1996)
26. Cueto, M.A., Morton, J., Sturmels, B.: Geometry of the restricted Boltzmann machine. In: Viana, M.A.G., Wynn, H.P. (eds.) *Algebraic methods in statistics and probability II*, AMS Special Session, vol. 2. American Mathematical Society (2010)
27. Montúfar, G., Morton, J.: Discrete restricted Boltzmann machines. In: *Proceedings of the 1-st International Conference on Learning Representations (ICLR2013)* (2013)
28. Montúfar, G., Morton, J.: Dimension of marginals of Kronecker product models. *SIAM J. Appl. Algebra Geom.* **1**(1), 126–151 (2017)
29. Cueto, M.A., Tobis, E.A., Yu, J.: An implicitization challenge for binary factor analysis. *J. Symb. Comput.* **45**(12), 1296–1315 (2010)
30. Seigal, A., Montúfar, G.: Mixtures and products in two graphical models. To Appear *J. Algebraic Stat.* (2018). [arXiv:1709.05276](https://arxiv.org/abs/1709.05276)
31. Martens, J., Chattopadhyay, A., Pitassi, T., Zemel, R.: On the representational efficiency of restricted Boltzmann machines. In: *Advances in Neural Information Processing Systems 26*, pp. 2877–2885. Curran Associates, Inc., USA (2013)
32. Montúfar, G., Morton, J.: When does a mixture of products contain a product of mixtures? *SIAM J. Discret. Math.* **29**(1), 321–347 (2015)
33. Fischer, A., Igel, C.: Bounding the bias of contrastive divergence learning. *Neural Comput.* **23**(3), 664–673 (2010)
34. Fischer, A., Igel, C.: A bound for the convergence rate of parallel tempering for sampling restricted Boltzmann machines. *Theor. Comput. Sci.* **598**, 102–117 (2015)
35. Aoyagi, M.: Stochastic complexity and generalization error of a restricted Boltzmann machine in Bayesian estimation. *J. Mach. Learn. Res.* **99**, 1243–1272 (2010)
36. Fischer, A., Igel, C.: Contrastive divergence learning may diverge when training restricted Boltzmann machines. In: *Frontiers in Computational Neuroscience*. Bernstein Conference on Computational Neuroscience (BCCN 2009) (2009)

37. Salakhutdinov, R.: Learning and evaluating Boltzmann machines. Technical report, 2008
38. Karakida, R., Okada, M., Amari, S.: Dynamical analysis of contrastive divergence learning: restricted Boltzmann machines with Gaussian visible units. *Neural Netw.* **79**, 78–87 (2016)
39. Salakhutdinov, R., Mnih, A., Hinton, G.E.: Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th international conference on Machine learning, ICML '07, pp. 791–798. ACM, NY (2007)
40. Welling, M., Rosen-Zvi, M., Hinton, G.E.: Exponential family harmoniums with an application to information retrieval. *Adv. Neural Inf. Process. Syst.* **17**, 1481–1488 (2005)
41. Sejnowski, T.J.: Higher-order Boltzmann machines. *Neural Networks for Computing*, pp. 398–403. American Institute of Physics (1986)
42. Salakhutdinov, R., Hinton, G.E.: Deep Boltzmann machines. In: Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS 09), pp. 448–455 (2009)
43. Montúfar, G.: Universal approximation depth and errors of narrow belief networks with discrete units. *Neural Comput.* **26**(7), 1386–1407 (2014)
44. Montúfar, G., Ay, N.: Refinements of universal approximation results for deep belief networks and restricted Boltzmann machines. *Neural Comput.* **23**(5), 1306–1319 (2011)
45. Sutskever, I., Hinton, G.E.: Deep, narrow sigmoid belief networks are universal approximators. *Neural Comput.* **20**(11), 2629–2636 (2008)
46. Montúfar, G.: Deep narrow Boltzmann machines are universal approximators. International Conference on Learning Representations (ICLR 15) (2015). [arXiv:1411.3784](https://arxiv.org/abs/1411.3784)
47. Montúfar, G., Ay, N., Ghazi-Zahedi, K.: Geometry and expressive power of conditional restricted Boltzmann machines. *J. Mach. Learn. Res.* **16**, 2405–2436 (2015)
48. Amin, M.H., Andriyash, E., Rolfe, J., Kulchytskyy, B., Melko, R.: Quantum Boltzmann machine. *Phys. Rev. X* **8**, 021050 (2018)
49. Zhang, N., Ding, S., Zhang, J., Xue, Y.: An overview on restricted Boltzmann machines. *Neurocomputing* **275**, 1186–1199 (2018)
50. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14**, 1771–1800 (2002)
51. Tieleman, T.: Training restricted Boltzmann machines using approximations to the likelihood gradient. In: Proceedings of the 25th International Conference on Machine Learning, ICML '08, pp. 1064–1071. ACM, USA (2008)
52. Salakhutdinov, R.: Learning in Markov random fields using tempered transitions. In: Bengio, Y., Schuurmans, D., Lafferty, J.D., Williams, C.K.I., Culotta, A. (eds.) *Advances in Neural Information Processing Systems 22*, pp. 1598–1606. Curran Associates, Inc., (2009)
53. Fischer, A., Igel, C.: Training restricted Boltzmann machines: an introduction. *Pattern Recognit.* **47**(1), 25–39 (2014)
54. Hinton, G.E.: A practical guide to training restricted Boltzmann machines, version 1. Technical report, UTML2010-003, University of Toronto, 2010
55. Amari, S.: Differential-geometrical Methods in Statistics. Lecture notes in statistics. Springer, Berlin (1985)
56. Amari, S.: Natural gradient works efficiently in learning. *Neural Comput.* **10**(2), 251–276 (1998)
57. Rao, R.C.: Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91 (1945)
58. Watanabe, S.: Algebraic Geometry and Statistical Learning Theory. Cambridge University Press, USA (2009)
59. Grosse, R., Salakhutdinov, R.: Scaling up natural gradient by sparsely factorizing the inverse fisher matrix. In: Bach, F., Blei, D. (eds.) *Proceedings of the 32nd International Conference on Machine Learning of Research*, vol. 37, pp. 2304–2313. PMLR, France, 07–09 Jul (2015)
60. Pascanu, R., Bengio, Y.: Revisiting natural gradient for deep networks. In: International Conference on Learning Representations 2014 (Conference Track) (2014)
61. Li, W., Montúfar, G.: Natural gradient via optimal transport I (2018). [arXiv:1803.07033](https://arxiv.org/abs/1803.07033)
62. Montavon, G., Müller, K.-R., Cuturi, M.: Wasserstein training of restricted boltzmann machines. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 3718–3726. Curran Associates Inc., USA (2016)

63. Csiszár, I., Tusnády, G.: Information Geometry and Alternating minimization procedures. *Statistics and decisions* (1984). Supplement Issue 1
64. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B (Methodological)* **39**(1), 1–38 (1977)
65. Draisma, J.: A tropical approach to secant dimensions. *J. Pure Appl. Algebra* **212**(2), 349–363 (2008)
66. Bieri, R., Groves, J.: The geometry of the set of characters induced by valuations. *Journal für die reine und angewandte Mathematik* **347**, 168–195 (1984)
67. Catalisano, M., Geramita, A., Gimigliano, A.: Secant varieties of $\mathbb{P}^1 \times \cdots \times \mathbb{P}^1$ (n -times) are not defective for $n \geq 5$. *J. Algebraic Geom.* **20**, 295–327 (2011)
68. Freund, Y., Haussler, D.: Unsupervised learning of distributions on binary vectors using two layer networks. Technical report, Santa Cruz, CA, USA 1994
69. Rauh, J.: Optimally approximating exponential families. *Kybernetika* **49**(2), 199–215 (2013)
70. Matúš, F.: Divergence from factorizable distributions and matroid representations by partitions. *IEEE Trans. Inf. Theory* **55**(12), 5375–5381 (2009)
71. Matúš, F., Ay, N.: On maximization of the information divergence from an exponential family. In: Proceedings of the WUPES'03, pp. 199–204 (2003)
72. Rauh, J.: Finding the maximizers of the information divergence from an exponential family. *IEEE Trans. Inf. Theory* **57**(6), 3236–3247 (2011)
73. Montúfar, G., Rauh, J., Ay, N.: Geometric Science of Information: First International Conference, GSI 2013, Paris, France, August 28–30, 2013. Proceedings. Chapter maximal information divergence from statistical models defined by neural networks, pp. 759–766. Springer, Heidelberg (2013)
74. Montúfar, G., Rauh, J.: Scaling of model approximation errors and expected entropy distances. *Kybernetika* **50**(2), 234–245 (2014)
75. Allman, E., Cervantes, H.B., Evans, R., Hoşten, S., Kubjas, K., Lemke, D., Rhodes, J., Zwiernik, P.: Maximum likelihood estimation of the latent class model through model boundary decomposition (2017)
76. Hammersley, J.M., Clifford, P.E.: Markov Random Fields on Finite Graphs and Lattices (1971). Unpublished manuscript
77. Geiger, D., Meek, C., Sturmfels, B.: On the toric algebra of graphical models. *Ann. Stat.* **34**(3), 1463–1492 (2006)
78. Steudel, B., Ay, N.: Information-theoretic inference of common ancestors. *Entropy* **17**(4), 2304 (2015)

Part II

**Infinite-Dimensional Information
Geometry**

Information Geometry of the Gaussian Space



Giovanni Pistone

Abstract We discuss the Pistone-Sempi exponential manifold on the finite-dimensional Gaussian space. We consider the role of the entropy, the continuity of translations, Poincaré-type inequalities, the generalized differentiability of probability densities of the Gaussian space.

Keywords Information geometry · Pistone-Sempi exponential manifold
Gaussian Orlicz space · Gaussian Orlicz-Sobolev space

1 Introduction

The Information Geometry (IG) set-up based on exponential Orlicz spaces [19], as further developed in [6, 8, 15, 16, 18, 20], has reached a satisfying consistency, but has a basic defect. In fact, it is unable to deal with the structure of the measure space on which probability densities are defined. When the basic space is \mathbb{R}^n one would like to discuss for example transformation models as sub-manifold of the exponential manifold, which is impossible without some theory about the effect of transformation of the state space on the relevant Orlicz spaces. Another example of interest are evolution equations for densities, such as the Fokker–Planck equation, which are difficult to discuss in this set-up without considering Gaussian Orlicz–Sobolev spaces. See an example of such type of applications in [3–5].

In [10] the idea of an exponential manifold in a Gaussian space has been introduced and the idea is applied to the study of the spatially homogeneous Boltzmann equation. In the second part of that paper, it is suggested that the Gaussian space allows to consider Orlicz–Sobolev spaces with Gaussian weight of [12, Ch. II] as a set-up for exponential manifolds.

G. Pistone (✉)

de Castro Statistics, Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8,
10122 Turin, Italy

e-mail: giovanni.pistone@carloalberto.org

URL: <http://www.giannidiorestino.it>

In Sect. 2 we discuss some properties of the Gauss-Orlicz spaces. Most results are quite standard, but are developed in some detail because to the best of our knowledge the case of interest is not treated in standard treatises. Notable examples and Poincaré-type inequalities are considered in Sect. 3.

The properties of the exponential manifold in the Gaussian case that are related with the smoothness of translation and the existence of mollifiers are presented in Sect. 4. A short part of this section is based on the conference paper [17]. Gaussian Orlicz-Sobolev space are presented in Sect. 5. Only basic notions on Sobolev's spaces are used here, mainly using the presentation by Haim Brezis [2, Ch. 8–9].

Part of the results presented here were announced in an invited talk at the conference IGAIA IV Information Geometry and its Applications IV, June 12–17, 2016, Liblice, Czech Republic.

2 Orlicz Spaces with Gaussian Weight

All along this paper, the sample space is the real Borel space $(\mathbb{R}^n, \mathcal{B})$ and M denotes the standard n -dimensional Gaussian density (M because of J.C. Maxwell!),

$$M(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}|x|^2\right), \quad x \in \mathbb{R}^n.$$

2.1 Generalities

First, we review basic facts about Orlicz spaces. Our reference on Orlicz space is J. Musielak monograph [12, Ch. II].

On the probability space $(\mathbb{R}^n, \mathcal{B}, M)$, called here the *Gaussian space*, the couple of Young functions $(\cosh - 1)$ and its conjugate $(\cosh - 1)_*$ are associated with the Orlicz space $L^{(\cosh - 1)}(M)$ and $L^{(\cosh - 1)_*}(M)$, respectively.

The space $L^{(\cosh - 1)}(M)$ is called *exponential space* and is the vector space of all functions such that $\int (\cosh - 1)(\alpha f(x))M(x) dx < \infty$ for some $\alpha > 0$. This is the same as saying that the moment generating function $t \mapsto \int e^{tf(x)}M(x) dx$ is finite on a open interval containing 0.

If $x, y \geq 0$, we have $(\cosh - 1)'(x) = \sinh(x)$, $(\cosh - 1)'_*(y) = \sinh^{-1}(y) = \log(y + \sqrt{1 + y^2})$, $(\cosh - 1)_*(y) = \int_0^y \sinh^{-1}(t) dt$. The Fenchel-Young inequality is

$$xy \leq (\cosh - 1)(x) + (\cosh - 1)_*(y) = \int_0^x \sinh(s) ds + \int_0^y \sinh^{-1}(t) dt$$

and

$$\begin{aligned} (\cosh - 1)(x) &= x \sinh(x) - (\cosh - 1)_*(\sinh(x)) ; \\ (\cosh - 1)_*(y) &= y \sinh^{-1}(y) - (\cosh - 1)(\sinh^{-1}(y)) \\ &= y \log(y + (1 + y^2)^{1/2}) - (1 + y^2)^{-1/2} . \end{aligned}$$

The conjugate Young function $(\cosh - 1)_*$ is associated with the *mixture space* $L^{(\cosh - 1)_*}(M)$. In this case, we have the inequality

$$(\cosh - 1)_*(ay) \leq C(a)(\cosh - 1)_*(y), \quad C(a) = \max(|a|, a^2). \quad (1)$$

In fact

$$\begin{aligned} (\cosh - 1)_*(ay) &= \int_0^{ay} \frac{ay - t}{\sqrt{1+t^2}} dt \\ &= a^2 \int_0^y \frac{y - s}{\sqrt{1+a^2s^2}} ds = a \int_0^y \frac{y - s}{\sqrt{\frac{1}{a^2} + s^2}} ds . \end{aligned}$$

The inequality (1) follows easily by considering the two cases $a > 1$ and $a < 1$. As a consequence, $g \in L^{(\cosh - 1)_*}(M)$ if, and only if, $\int (\cosh - 1)_*(g(y))M(y) dy < \infty$.

In the theory of Orlicz spaces, the existence of a bound of the type (1) is called Δ_2 -property, and it is quite relevant. In our case, it implies that the mixture space $L^{(\cosh - 1)_*}(M)$ is the dual space of its conjugate, the exponential space $L^{(\cosh - 1)}(M)$. Moreover, a separating sub-vector space e.g., $C_0^\infty(\mathbb{R}^n)$ is norm-dense.

In the definition of the associated spaces, the couple $(\cosh - 1)$ and $(\cosh - 1)_*$ is equivalent to the couple defined for $x, y > 0$ by $\Phi(x) = e^x - 1 - x$ and $\Psi(y) = (1 + y) \log(1 + y) - y$. In fact, for $t > 0$ we have $\log(1 + t) \leq \log(y + \sqrt{1 + t^2})$ and

$$\log(t + \sqrt{1 + t^2}) \leq \log(t + \sqrt{1 + 2t + t^2}) = \log(1 + 2t) ,$$

so that we derive by integration the inequality

$$\Psi(y) \leq (\cosh - 1)_*(y) \leq \frac{1}{2}\Psi(2y) .$$

In turn, conjugation gives

$$\frac{1}{2}\Phi(x) \leq (\cosh - 1)(x) \leq \Phi(x) .$$

The *exponential space* $L^{(\cosh - 1)}(M)$ and the *mixture space* $L^{(\cosh - 1)_*}(M)$ are the spaces of real functions on \mathbb{R}^n respectively defined using the conjugate Young

functions $\cosh - 1$ and $(\cosh - 1)_*$. The exponential space and the mixture space are given norms by defining the closed unit balls of $L^{(\cosh - 1)}(M)$ and $L^{(\cosh - 1)_*}(M)$, respectively, by

$$\left\{ f \left| \int (\cosh - 1)(f(x)) M(x) dx \leq 1 \right. \right\}, \quad \left\{ g \left| \int (\cosh - 1)_*(g(x)) M(x) dx \leq 1 \right. \right\}.$$

Such a norm is called Luxemburg norm.

The Fenchel-Young inequality

$$xy \leq (\cosh - 1)(x) + (\cosh - 1)_*(y)$$

implies that $(f, g) \mapsto \mathbb{E}_M [fg]$ is a separating duality, precisely

$$\left| \int f(x)g(x)M(x) dx \right| \leq 2 \|f\|_{L^{(\cosh - 1)}(M)} \|g\|_{L^{(\cosh - 1)_*}(M)}.$$

A random variable g has norm $\|g\|_{L^{(\cosh - 1)_*}(M)}$ bounded by ρ if, and only if, $\|g/\rho\|_{L^{(\cosh - 1)_*}(M)} \leq 1$, that is $\mathbb{E}_M [(\cosh - 1)_*(g/\rho)] \leq 1$, which in turn implies $\mathbb{E}_M [(\cosh - 1)_*(\alpha g)] = \mathbb{E}_M [(\cosh - 1)_*(\alpha \rho(g/\rho))] \leq \rho\alpha$ for all $\alpha \geq 0$. This is not true for the exponential space $L^{(\cosh - 1)}(M)$.

It is possible to define a dual norm, called Orlicz norm, on the exponential space, as follows. We have $\|f\|_{(L^{(\cosh - 1)_*}(M))^*} \leq 1$ if, and only if $\left| \int f(x)g(x)M(x) dx \right| \leq 1$ for all g such that $\int (\cosh - 1)_*(g(x))M(x) dx \leq 1$. With this norm, we have

$$\left| \int f(x)g(x)M(x) dx \right| \leq \|f\|_{(L^{(\cosh - 1)_*}(M))^*} \|g\|_{L^{(\cosh - 1)_*}(M)} \quad (2)$$

The Orlicz norm and the Luxemburg norm are equivalent, precisely,

$$\|f\|_{L^{(\cosh - 1)}(M)} \leq \|f\|_{L^{(\cosh - 1)_*}(M)^*} \leq 2 \|f\|_{L^{(\cosh - 1)}(M)}.$$

2.2 Entropy

The use of the exponential space is justified by the fact that for every 1-dimensional exponential family $I \ni \theta \mapsto p(\theta) \propto e^{\theta V}$, I neighborhood of 0, the sufficient statistics V belongs to the exponential space. The statistical interest of the mixture space resides in its relation with entropy.

If f is a positive density of the Gaussian space, $\int f(x)M(x) dx = 1$, we define its entropy to be $\text{Ent}(f) = - \int f(x) \log f(x)M(x) dx$. As $x \log x \geq x - 1$, the integral is well defined. It holds

$$-\int f(x) \log^+ f(x) M(x) dx \leq \text{Ent}(f) \leq e^{-1} - \int f(x) \log^+ f(x) M(x) dx , \quad (3)$$

where \log^+ is the positive part of \log .

Proposition 1 A positive density f of the Gaussian space has finite entropy if, and only if, f belongs to the mixture space $L^{(\cosh -1)_*}(M)$.

Proof We use Eq.(3) in order to show the equivalence. For $x \geq 1$ it holds

$$2x \leq x + \sqrt{1+x^2} \leq (1+\sqrt{2})x .$$

It follows

$$\log 2 + \log x \leq \log \left(x + \sqrt{1+x^2} \right) = \sinh^{-1}(x) \leq \log \left(1 + \sqrt{2} \right) + \log x ,$$

and, taking the integral \int_1^y with $y \geq 1$, we get

$$\begin{aligned} & \log 2(y-1) + y \log y - y + 1 \\ & \leq (\cosh -1)_*(y) - (\cosh -1)_*(1) \\ & \leq \log \left(1 + \sqrt{2} \right) (y-1) + y \log y - y + 1 , \end{aligned}$$

then, substituting $y > 1$ with $\max(1, f(x))$, $f(x) > 0$,

$$\begin{aligned} & (\log 2 - 1)(f(x) - 1)^+ + f(x) \log^+ f(x) \\ & \leq (\cosh -1)_*(\max(1, f(x))) - (\cosh -1)_*(1) \\ & \leq (\log \left(1 + \sqrt{2} \right) - 1)(f(x) - 1)^+ + f(x) \log^+ f(x) . \end{aligned}$$

By taking the Gaussian integral, we have

$$\begin{aligned} & (\log 2 - 1) \int (f(x) - 1)^+ M(x) dx + \int f(x) \log^+ f(x) M(x) dx \\ & \leq \int (\cosh -1)_*(\max(1, f(x))) M(x) dx - (\cosh -1)_*(1) \\ & \leq (\log \left(1 + \sqrt{2} \right) - 1) \int (f(x) - 1)^+ M(x) dx + \int f(x) \log^+ f(x) M(x) dx , \end{aligned}$$

which in turn implies the statement because $f \in L^1(M)$ and

$$\begin{aligned}
& \int (\cosh -1)_*(f(x))M(x) dx + (\cosh -1)_*(1) \\
&= \int (\cosh -1)_*(\max(1, f(x)))M(x) dx \\
&\quad + \int (\cosh -1)_*(\min(1, f(x)))M(x) dx .
\end{aligned}$$

□

Of course, this proof does not depend on the Gaussian assumption.

2.3 Orlicz and Lebesgue Spaces

We discuss now the relations between the exponential space, the mixture space, and the Lebesgue spaces. This provides a first list of classes of functions that belong to the exponential space or to the mixture space. The first item in the proposition holds for a general base probability measure, while the other is proved in the Gaussian case.

Proposition 2 *Let $1 < a < \infty$.*

1.

$$L^\infty(M) \hookrightarrow L^{(\cosh -1)}(M) \hookrightarrow L^a(M) \hookrightarrow L^{(\cosh -1)*}(M) \hookrightarrow L^1(M).$$

2. *If $\Omega_R = \{x \in \mathbb{R}^n \mid |x| < R\}$, the restriction operator is defined and continuous in the cases*

$$L^{(\cosh -1)}(M) \rightarrow L^a(\Omega_R), \quad L^{(\cosh -1)*}(M) \rightarrow L^1(\Omega_R)$$

Proof 1. See [12, Ch. II].

2. For all integers $n \geq 1$,

$$\begin{aligned}
1 &\geq \int (\cosh -1) \left(\frac{f(x)}{\|f\|_{L^{(\cosh -1)}(M)}} \right) M(x) dx \\
&\geq \int_{\Omega_R} \frac{1}{(2n)!} \left(\frac{f(x)}{\|f\|_{L^{(\cosh -1)}(M)}} \right)^{2n} M(x) dx \\
&\geq \frac{(2\pi)^{-n/2} e^{-R^2/2}}{(2n)! \|f\|_{L^{(\cosh -1)}(M)}} \int_{\Omega_R} (f(x))^{2n} dx.
\end{aligned}$$

□

3 Notable Bounds and Examples

There is a large body of literature about the analysis of the Gaussian space $L^2(M)$. In order to motivate our own construction and to connect it up, in this section we have collected some results about notable classes of functions that belongs to the exponential space $L^{(\cosh^{-1})}(M)$ or to the mixture space $L^{(\cosh^{-1})_*}(M)$. Some of the examples will be used in the applications of Orlicz-Sobolev spaces in the Information Geometry of the Gaussian space. Basic references on the analysis of the Gaussian space are [11, V.1.5], [21, 4.2.1], and [13, Ch. 1].

3.1 Polynomial Bounds

The exponential space $L^{(\cosh^{-1})}(M)$ contains all functions $f \in C^2(\mathbb{R}^n; \mathbb{R})$ whose Hessian is uniformly dominated by a constant symmetric matrix. In such a case, $f(x) = f(0) + \nabla f(0)x + \frac{1}{2}x^* \text{Hess } f(\bar{x})x$, with $x^* \text{Hess } f(y)x \leq \lambda |x|^2$, $y \in \mathbb{R}^n$, and $\lambda \geq 0$ being the largest non-negative eigen-value of the dominating matrix. Then for all real α ,

$$\int_{\mathbb{R}^n} e^{\alpha f(x)} M(x) dx < \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{\alpha f(0) + \nabla f(0)x + \frac{1}{2}(\alpha\lambda - 1)|x|^2} dx$$

and the RHS is finite for $\alpha < \lambda^{-1}$. In particular, $L^{(\cosh^{-1})}(M)$ contains all polynomials with degree up to 2.

An interesting simple application of the same argument is the following. Assume $p = e^v$ is a positive density on the Gaussian space such that

$$e^{A_1(x)} \leq e^{v(x)} \leq e^{A_2(x)}, \quad x \in \mathbb{R}^n,$$

for suitable second order polynomials A_1, A_2 . Then $v \in L^{(\cosh^{-1})}(M)$. Inequalities of this type appear in the theory of parabolic equations e.g., see [21, Ch. 4].

The mixture space $L^{(\cosh^{-1})_*}(M)$ contains all random variables $f: \mathbb{R}^d \rightarrow \mathbb{R}$ which are bounded by a polynomial, in particular, all polynomials. In fact, all polynomials belong to $L^2(M) \subset L^{(\cosh^{-1})_*}(M)$.

3.2 Densities of Exponential Form

In this paper, we are specially interested in densities of the Gaussian space of the form $f = e^v$, that is $\int e^{v(x)} M(x) dx = 1$. Let us now consider simple properties of the mappings $f \mapsto v = \log f$ and $v \mapsto f = e^v$.

We have seen in Proposition 1 that $f = e^v \in L^{(\cosh^{-1})_*}(M)$ if, and only if,

$$-\text{Ent}(\mathrm{e}^v) = \int \mathrm{e}^{v(x)} v(x) M(x) dx < \infty.$$

As $\lim_{x \rightarrow +\infty} \frac{\cosh(x)}{x \mathrm{e}^x} = 0$, we do not expect $v \in L^{(\cosh^{-1})}(M)$ to imply $f = \mathrm{e}^v \in L^{(\cosh^{-1})^*}(M)$.

As $(\cosh - 1)(\alpha \log y) = (y^\alpha + y^{-\alpha})/2 - 1$, $\alpha > 0$, then $v = \log f \in L^{(\cosh^{-1})}(M)$ if, and only if, both f^α and $f^{-\alpha}$ both belong to $L^1(M)$ for some $\alpha > 0$. In the case $\|v\|_{L^{(\cosh^{-1})}(M)} < 1$, then we can take $\alpha > 1$ and $f \in L^\alpha(M) \subset L^{(\cosh^{-1})^*}(M)$. In conclusion, $\exp: v \mapsto \mathrm{e}^v$ maps the open unit ball of $L^{(\cosh^{-1})}(M)$ into $\cup_{\alpha > 1} L^\alpha(M) \subset L^{(\cosh^{-1})^*}(M)$.

This issue is discussed in the next Sect. 4.

3.3 Poincaré-Type Inequalities

Let us denote by $C_b^k(\mathbb{R}^n)$ the space of functions with derivatives up to order k , each bounded by a constant. We write $C_p^k(\mathbb{R}^n)$ if all the derivative are bounded by a polynomial. We discuss below inequalities related to the classical Gaussian Poincaré inequality, which reads, in the 1-dimensional case,

$$\int \left(f(x) - \int f(y) M(y) dy \right)^2 M(x) dx \leq \int |f'(x)|^2 M(x) dx , \quad (4)$$

for all $f \in C_p^1(\mathbb{R}^n)$. We are going to use the same techniques used in the classical proof of (4) e.g., see [13].

If X, Y are independent standard Gaussian variables, then

$$X' = \mathrm{e}^{-t} + \sqrt{1 - \mathrm{e}^{-2t}} Y, \quad Y' = \sqrt{1 - \mathrm{e}^{-2t}} X - \mathrm{e}^{-t} Y$$

are independent standard Gaussian random variables for all $t \geq 0$. Because of that, it is useful to define Ornstein–Uhlenbeck semi-group by the Mehler formula

$$P_t f(x) = \int f(\mathrm{e}^{-t} x + \sqrt{1 - \mathrm{e}^{-2t}} y) M(y) dy , \quad t \geq 0, \quad f \in C_p(\mathbb{R}^n) . \quad (5)$$

For any convex function Φ , Jensen's inequality gives

$$\begin{aligned} & \int \Phi(P_t f(x)) M(x) dx \\ & \leq \int \int \Phi(f(\mathrm{e}^{-t} x + \sqrt{1 - \mathrm{e}^{-2t}} y)) M(y) dy M(x) dx \\ & = \int \Phi(f(x)) M(x) dx . \end{aligned}$$

In particular, this shows that, for all $t \geq 0$, $f \mapsto P_t f$ is a contraction for the norm of both the mixture space $L^{(\cosh -1)_*}(M)$ and the exponential space $L^{(\cosh -1)}(M)$.

Moreover, if $f \in C_p^1(\mathbb{R}^n)$, we have

$$\begin{aligned} f(x) - \int f(y)M(y) dy \\ = P_0(x) - P_\infty f(x) \\ = - \int_0^\infty \frac{d}{dt} P_t f(x) dt \\ = \int_0^\infty \int \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}y) \cdot \left(e^{-t}x - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}y \right) M(y) dy dt \quad (6) \\ \leq \int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt \times \\ \int \left| \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}y) \right| \left| \sqrt{1-e^{-2t}}x - e^{-t}y \right| M(y) dy . \quad (7) \end{aligned}$$

Note that

$$\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt = \int_0^1 \frac{ds}{\sqrt{1-s^2}} = \frac{\pi}{2} .$$

We use this remark and (7) to prove our first inequality.

Proposition 3 *If $f \in C_p^1(\mathbb{R}^n)$ and $\lambda > 0$ is such that*

$$C\left(\lambda \frac{\pi}{2}\right) \int C(|y|)M(y) dy = 1 , \quad C(a) = \max(|a|, a^2) , \quad (8)$$

then

$$\begin{aligned} & \int (\cosh -1)_* \left(\lambda \left(f(x) - \int f(y)M(y) dy \right) \right) M(x) dx \\ & \leq \int (\cosh -1)_* (|\nabla f(x)|) M(x) dx , \end{aligned}$$

that is

$$\left\| f - \int f(y)M(y) dy \right\|_{L^{(\cosh -1)_*}(M)} \leq \lambda^{-1} \|\nabla f\|_{L^{(\cosh -1)_*}(M)} .$$

Proof Jensen's inequality applied to Eq.(7) gives

$$\begin{aligned} & (\cosh -1)_* \left(\lambda \left(f(x) - \int f(y)M(y) dy \right) \right) \leq \int_0^\infty \frac{2}{\pi} \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt \\ & \times \int (\cosh -1)_* \left(\lambda \frac{\pi}{2} \left| \nabla f(\sqrt{1-e^{-2t}}x + e^{-t}y) \right| \left| \sqrt{1-e^{-2t}}x - e^{-t}y \right| \right) M(y) dy \quad (9) \end{aligned}$$

Now we use of the bound in Eq. (1), namely $(\cosh -1)_*(ay) \leq C(a)(\cosh -1)_*(y)$ if $a > 0$, where $C(a) = \max(|a|, a^2)$, and further bound for $a, k > 0$

$$C(ka) = ka \vee k^2 a^2 \leq kC(a) \vee k^2 C(a) = C(k)C(a) ,$$

to get

$$\begin{aligned} & (\cosh -1)_* \left(\lambda \frac{\pi}{2} \left| \nabla f(e^{-t}x + \sqrt{1 - e^{-2t}}y) \right| \left| \sqrt{1 - e^{-2t}}x - e^{-t}y \right| \right) \\ & \leq C \left(\lambda \frac{\pi}{2} \right) C \left(\left| \sqrt{1 - e^{-2t}}x - e^{-t}y \right| \right) (\cosh -1)_* \left(\left| \nabla f(e^{-t}x + \sqrt{1 - e^{-2t}}y) \right| \right) . \end{aligned} \quad (10)$$

Taking the expected value of both sides of the inequality resulting from (9) and (10), we get

$$\begin{aligned} & \int (\cosh -1)_* \left(\lambda \left(f(y) - \int f(x)M(x) dx \right) \right) M(y) dy \\ & \leq C \left(\lambda \frac{\pi}{2} \right) \int C(|y|)M(y) dy \int (\cosh -1)_* (|\nabla f(x)|)M(x) dx , \end{aligned}$$

We conclude by choosing a proper value of λ . \square

The same argument does not work in the exponential space. We have assume the boundedness of derivatives i.e., a Lipschitz assumption.

Proposition 4 *If $f \in C_b^1(\mathbb{R}^n)$ with $\sup \{ |\nabla f(x)| \mid x \in \mathbb{R}^n \} = m$ then*

$$\left\| f - \int f(y)M(y) dy \right\|_{L^{(\cosh -1)}(M)} \leq \frac{\pi}{2\sqrt{2\log 2}} m .$$

Proof Jensen's inequality applied to Eq. (7) and the assumption give

$$\begin{aligned} & (\cosh -1) \left(\lambda \left(f(x) - \int f(y)M(y) dy \right) \right) \\ & \leq \int (\cosh -1) \left(\lambda \frac{\pi}{2} mx \right) M(x) dx = \exp \left(\frac{\lambda^2}{2} \frac{\pi^2}{4} m^2 \right) - 1 . \end{aligned}$$

To conclude, choose λ such that the the RHS equals 1. \square

Remark 1 Both Propositions 3 and 4 are related with interesting results on the Gaussian space other then bounds on norms. For example, if f is a density of the Gaussian space, then the first one is a bound on the lack of uniformity $f - 1$, which, in turn, is related with the entropy of f . As a further example, consider a case where $\int f(x)M(x) dx = 0$ and $\|\nabla f\|_\infty < \infty$. In such a case, we have a bound on the

Laplace transform of f , which in turn implies a bound on large deviations of the random variable f .

To prepare the proof of an inequality for the exponential space, we start from Eq.(6) and observe that for $f \in C_p^2(\mathbb{R}^n)$ we can write

$$\begin{aligned} f(x) - \int f(y)M(y) dy \\ = \int_0^\infty e^{-t} \left(\int \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}y)M(y) dy \right) \cdot x dt \\ - \int_0^\infty e^{-2t} \int \nabla \cdot \nabla f(e^{-t}x + \sqrt{1-e^{-2t}}y)M(y) dy dt , \end{aligned}$$

where integration by parts and $(\partial/\partial y_i)M(y) = -y_i M(y)$ have been used to get the last term.

If we write $f_i(z) = \frac{\partial}{\partial z_i}$ and $f_{ii}(z) = \frac{\partial^2}{\partial z_i^2}f(z)$ then

$$\frac{\partial}{\partial x_i} P_t f(x) = e^{-t} P_t f_i(x)$$

and

$$\frac{\partial^2}{\partial x_i^2} P_t f(x) = e^{-2t} P_t f_{ii}(x) ,$$

so that

$$f(x) - \int f(y)M(y) dy = \int_0^\infty (x \cdot \nabla P_t f(x) - \nabla \cdot \nabla P_t f(x)) dt .$$

If $g \in C_b^2(\mathbb{R}^n)$ we have

$$\begin{aligned} & \int g(x) \left(f(x) - \int f(y)M(y) dy \right) M(x) dx \\ &= \int_0^\infty \left(\int g(x)x \cdot \nabla P_t f(x) M(x) dx - \int g(x)\nabla \cdot \nabla P_t f(x) M(x) dx \right) dt \\ &= \int_0^\infty \left(\int g(x)x \cdot \nabla P_t f(x) M(x) dx + \int \nabla(g(x)M(x)) \cdot \nabla P_t f(x) dx \right) dt \\ &= \int_0^\infty \int \nabla g(x) \cdot \nabla P_t f(x) M(x) dx dt \\ &= \int_0^\infty e^{-t} \int \nabla g(x) \cdot P_t \nabla f(x) M(x) dx dt . \end{aligned} \tag{11}$$

Let $|\cdot|_1$ and $|\cdot|_2$ be two norms on \mathbb{R}^n such that $|x \cdot y| \leq |x|_1 |y|_2$. Define the covariance of $f, g \in C_p^2(\mathbb{R}^n)$ to be

$$\begin{aligned}
& \text{Cov}_M(f, g) \\
&= \int \left(f(x) - \int f(y) M(y) dy \right) g(x) M(x) dx \\
&= \int \left(f(x) - \int f(y) M(y) dy \right) \left(g(x) - \int g(y) M(y) dy \right) M(x) dx .
\end{aligned}$$

Proposition 5 If $f, g \in C_p^2(\mathbb{R}^n)$, then

$$|\text{Cov}_M(f, g)| \leq \|\nabla f\|_{L^{(\cosh-1)*}(M)} \Big|_1 \|\nabla g\|_{(L^{(\cosh-1)*}(M))^*} \Big|_2 .$$

Proof We use Eq.(11) and the inequality (2).

$$\begin{aligned}
& \left| \int \nabla g(x) \cdot P_t \nabla f(x) M(x) dx \right| \\
&\leq \sum_{i=1}^n \left| \int g_i(x) P_t f_i(x) M(x) dx \right| \\
&\leq \sum_{i=1}^n \|g_i\|_{L^{(\cosh-1)*}(M)^*} \|P_t f_i\|_{L^{(\cosh-1)*}(M)} \\
&\leq \sum_{i=1}^n \|g_i\|_{L^{(\cosh-1)*}(M)^*} \|f_i\|_{L^{(\cosh-1)*}(M)} \\
&\leq \|\nabla g\|_{L^{(\cosh-1)*}(M)} \Big|_1 \|\nabla f\|_{L^{(\cosh-1)*}(M)} \Big|_2 .
\end{aligned}$$

□

If g_n is a sequence such that $\nabla g_n \rightarrow 0$ in $L^{(\cosh-1)}(M)$, then the inequality above shows that $g_n - \int g_n(x) M(x) dx \rightarrow 0$.

4 Exponential Manifold on the Gaussian Space

In this section we first review the basic features of our construction of IG as it was discussed in the original paper [19]. Second, we see how the choice of the Gaussian space adds new features, see [10, 16]. We normally use capital letters to denote random variables and write $\mathbb{E}_M[U] = \int U(x) M(x) dx$.

We define $B_M = \{U \in L^{(\cosh-1)}(M) | \mathbb{E}_M[U] = 0\}$. The positive densities of the Gaussian space we consider are all of the exponential form $p = e^U / Z_M(U)$, with $U \in B_M \subset L^{(\cosh-1)}(M)$ and normalization (moment functional, partition functional) $Z_M(U) = \mathbb{E}_M[U] < \infty$.

We can also write $p = e^{U - K_M(U)}$, where $K_M(U) = \log Z_M(U)$ is called cumulant functional. Because of the assumption $\mathbb{E}_M[U] = 0$, the chart mapping

$$s_M : p \mapsto \log p - \mathbb{E}_M [\log p] = U$$

is well defined.

Both the extended real functions Z_M and K_M are convex on B_M . The common proper domain of Z_M and of K_M contains the open unit ball of B_M . In fact, if $\mathbb{E}_M [(\cosh - 1)(\alpha U)] \leq 1$, $\alpha > 1$, then, in particular, $Z_M(U) \leq 4$.

We denote \mathcal{S}_M the interior of the proper domain of the cumulant functional. The set \mathcal{S}_M is nonempty, convex, star-shaped, and solid i.e., the generated vector space is B_M itself.

We define the maximal exponential model to be the set of densities on the Gaussian space $\mathcal{E}(M) = \{e^{U-K_M(U)} | U \in \mathcal{S}_M\}$.

We prove below that the mapping $e_M = s_M^{-1} : \mathcal{S}_M \rightarrow \mathcal{E}(M)$ is smooth. The chart mapping itself s_M is not and induces on $\mathcal{E}(M)$ a topology that we do not discuss here.

Proposition 6 *The mapping $e_M : \mathcal{S}_M \ni U \mapsto e^{U-K_M(U)}$ is continuously differentiable in $L^{(\cosh - 1)_*}(M)$ with derivative $d_H e_M(U) = e_M(U)(U - \mathbb{E}_{e_M(U) \cdot M}[H])$.*

Proof We split the proof into numbered steps.

1. If $U \in \mathcal{S}_M$ then $\alpha U \in \mathcal{S}_M$ for some $\alpha > 1$. Moreover, $(\cosh - 1)_*(y) \leq C(\alpha) |y|^\alpha$. Then

$$\mathbb{E}_M [(\cosh - 1)_*(e^U)] \leq \text{const } \mathbb{E}_M [e^{\alpha U}] < \infty .$$

- so that $e^U \in L^{(\cosh - 1)_*}(M)$. It follows that $e_M(U) = e^{U-K_M(U)} \in L^{(\cosh - 1)_*}(M)$.
2. Given $U \in \mathcal{S}_M$, as \mathcal{S}_M is open in the exponential space $L^{(\cosh - 1)}(M)$, there exists a constant $\rho > 0$ such that $\|H\|_{L^{(\cosh - 1)}(M)} \leq \rho$ implies $U + H \in \mathcal{S}_M$. In particular,

$$U + \frac{\rho}{\|U\|_{L^{(\cosh - 1)}(M)}} U = \frac{\|U\|_{L^{(\cosh - 1)}(M)} + \rho}{\|U\|_{L^{(\cosh - 1)}(M)}} U \in \mathcal{S}_M .$$

We have, from the Hölder's inequality with conjugate exponents

$$\frac{2(\|U\|_{L^{(\cosh - 1)}(M)} + \rho)}{2\|U\|_{L^{(\cosh - 1)}(M)} + \rho}, \quad \frac{2(\|U\|_{L^{(\cosh - 1)}(M)} + \rho)}{\rho},$$

that

$$\begin{aligned} & \mathbb{E}_M \left[\exp \left(\frac{2\|U\|_{L^{(\cosh - 1)}(M)} + \rho}{2\|U\|_{L^{(\cosh - 1)}(M)}} (U + H) \right) \right] \\ & \leq \mathbb{E}_M \left[\exp \left(\frac{\|U\|_{L^{(\cosh - 1)}(M)} + \rho}{\|U\|_{L^{(\cosh - 1)}(M)}} U \right) \right]^{\frac{2\|U\|_{L^{(\cosh - 1)}(M)} + \rho}{2(\|U\|_{L^{(\cosh - 1)}(M)} + \rho)}} \\ & \quad \times \mathbb{E}_M \left[\exp \left(\frac{(2\|U\|_{L^{(\cosh - 1)}(M)} + \rho)(\|U\|_{L^{(\cosh - 1)}(M)} + \rho)}{\rho \|U\|_{L^{(\cosh - 1)}(M)}} H \right) \right]^{\frac{\rho}{2(\|U\|_{L^{(\cosh - 1)}(M)} + \rho)}} . \end{aligned}$$

In the RHS, the first factor is finite because the random variable under \exp belong to \mathcal{S}_M , while the second factor is bounded by a fixed constant for all H such that

$$\|H\|_{L^{(\cosh-1)}(M)} \leq \frac{\rho \|U\|_{L^{(\cosh-1)}(M)}}{(2\|U\|_{L^{(\cosh-1)}(M)} + \rho)(\|U\|_{L^{(\cosh-1)}(M)} + \rho)}.$$

This shows that e_M is locally bounded in $L^{(\cosh-1)}(M)$.

3. Let us now consider $\prod_{i=1}^m H_i e^U$ with $\|H_i\|_{L^{(\cosh-1)}(M)} \leq 1$ for $i = 1, \dots, m$ and $U \in \mathcal{S}_M$. Chose an $\alpha > 1$ such that $\alpha U \in \mathcal{S}_M$, and observe that, because of the previous item applied to αU , the mapping $U \mapsto \mathbb{E}_M[e^{\alpha U}]$ is uniformly bounded in a neighborhood of U by a constant $C(U)$. As $\alpha > (\alpha+1)/2 > 1$ and we have the inequality $(\cosh - a)_*(y) \leq \frac{C((1+\alpha)/2)}{(1+\alpha)/2} |y|^{(1+\alpha)/2}$. It follows, using the $(m+1)$ -terms Fenchel-Young inequality for conjugate exponents $2\alpha/(\alpha+1)$ and $2m\alpha/(\alpha-1)$ (m times), that

$$\begin{aligned} & \mathbb{E}_M \left[(\cosh - 1)_* \left(\prod_{i=1}^m H_i e^U \right) \right] \\ & \leq \frac{C((1+\alpha)/2)}{(1+\alpha)/2} \mathbb{E}_M \left[\prod_{i=1}^m |H_i|^{(1+\alpha)/2} e^{(1+\alpha)U/2} \right] \\ & \leq \frac{C((1+\alpha)/2)}{(1+\alpha)/2} \left(\mathbb{E}_M [e^{\alpha U}] + \sum_{i=1}^m \mathbb{E}_M [|H_i|^{m\alpha(1+\alpha)/(\alpha-1)}] \right) \\ & \leq \frac{C((1+\alpha)/2)}{(1+\alpha)/2} \left(C(U) + \sum_{i=1}^m \|H_i\|_{L^{m\alpha(1+\alpha)/(\alpha-1)}(M)}^{m\alpha(1+\alpha)/(\alpha-1)} \right), \end{aligned}$$

which is bounded by a constant depending on U and α . We have proved that the multi-linear mapping $(H_1, \dots, H_m) \mapsto \prod_{i=1}^m H_i e^U$ is continuous from $(L^{(\cosh-1)}(M))^m$ to $L^{(\cosh-1)*}(M)$, uniformly in a neighborhood of U .

4. Let us consider now the differentiability of $U \mapsto e^U$. For $U + H \in \mathcal{S}_M$, it holds

$$\begin{aligned} 0 \leq e(U + H) - e^U - e^U H &= \int_0^1 (1-s)e^{U+sH} H^2 ds \\ &= \int_0^1 (1-s)e^{(1-s)U+s(U+H)} H^2 ds \\ &\leq \int_0^1 (1-s)^2 e^U H^2 ds + \int_0^1 s(1-s)e^{U+H} H^2 ds \\ &= \left(\frac{1}{3}e^U + \frac{1}{6}e^{U+H} \right) H^2. \end{aligned}$$

Because of the previous item, the RHS is bounded by a constant times $\|H\|_{L^{(\cosh-1)}(M)}^2$ for $\|H\|_{L^{(\cosh-1)}(M)}$ small, which in turn implies the differentiability. Note that the bound is uniform in a neighborhood of U .

5. It follows that Z_M and K_M are differentiable and also e_M is differentiable with locally uniformly continuous derivative.

□

We turn to discuss the approximation with smooth random variables. We recall that $(\cosh - 1)_*$ satisfies the Δ_2 -bound

$$(\cosh - 1)_*(ay) \leq \max(|a|, a^2)(\cosh - 1)_*(y)$$

hence, bounded convergence holds for the mixture space $L^{(\cosh-1)_*}(M)$. That, in turn, implies separability. This is not true for the exponential space $L^{(\cosh-1)}(M)$. Consider for example $f(x) = |x|^2$. This function belongs in $L^{(\cosh-1)}(M)$, but, if $f_R(x) = f(x)(|x| \geq R)$, then

$$\int (\cosh - 1)(\epsilon^{-1} f_R(x)) M(x) dx \geq \frac{1}{2} \int_{|x| > R} e^{\epsilon^{-1}|x|^2} M(x) dx = +\infty, \quad \text{if } \epsilon \leq 2,$$

hence there is no convergence to 0. However, the truncation of $f(x) = |x|$ does converge.

While the exponential space $L^{(\cosh-1)}(M)$ is not separable nor reflexive, we have the following weak property. Let $C_0(\mathbb{R}^n)$ and $C_0^\infty(\mathbb{R}^n)$ respectively denote the space of continuous real functions with compact support and the space of infinitely-differentiable real functions on \mathbb{R}^n with compact support. The following proposition was stated in [17, Prop. 2].

Proposition 7 *For each $f \in L^{(\cosh-1)}(M)$ there exist a nonnegative function $h \in L^{(\cosh-1)}(M)$ and a sequence $g_n \in C_0^\infty(\mathbb{R}^n)$ with $|g_n| \leq h$, $n = 1, 2, \dots$, such that $\lim_{n \rightarrow \infty} g_n = f$ a.e. As a consequence, $C_0^\infty(\mathbb{R}^n)$ is weakly dense in $L^{(\cosh-1)}(M)$.*

Proof Our proof uses a monotone class argument [7, Ch. II]. Let \mathcal{H} be the set of all random variables $f \in L^{(\cosh-1)}(M)$ for which there exists a sequence $g_n \in C_0(\mathbb{R}^n)$ such that $g_n(x) \rightarrow f(x)$ a.s. and $|g_n(x)| \leq |f(x)|$. Let us show that \mathcal{H} is closed for monotone point-wise limits of positive random variables. Assume $f_n \uparrow f$ and $g_{n,k} \rightarrow f_n$ a.s. with $|g_{n,k}| \leq f_n \leq f$. Each sequence $(g_{nk})_k$ is convergent in $L^1(M)$ then, for each n we can choose a g_n in the sequence such that $\|f_n - g_n\|_{L^1(M)} \leq 2^{-n}$. It follows that $|f_n - g_n| \rightarrow 0$ a.s. and also $f - g_n = (f - f_n) + (f_n - g_n) \rightarrow 0$ a.s. Now we can apply the monotone class argument to $C_0(\mathbb{R}^n) \subset \mathcal{H}$. The conclusion follows from the uniform density of $C_0^\infty(\mathbb{R}^n)$ in $C_0(\mathbb{R}^n)$. □

The point-wise bounded convergence of the previous proposition implies a result of local approximation in variation of finite-dimensional exponential families.

Proposition 8 *Given $U_1, \dots, U_m \in B_M$, consider the exponential family*

$$p_\theta = \exp \left(\sum_{j=1}^m \theta_j U_j - \psi(\theta) \right), \quad \theta \in \Theta.$$

There exists a sequence $(U_1^k, \dots, U_m^k)_{k \in \mathbb{N}}$ in $C_0^\infty(\mathbb{R}^n)^m$ and an $\alpha > 0$ such that the sequence of exponential families

$$p_\theta^k = \exp \left(\sum_{j=1}^m \theta_j U_j^k - \psi_k(\theta) \right)$$

is convergent in variation to p_θ for all θ such that $\sum |\theta_j| < \alpha$.

Proof For each $j = 1, \dots, m$ there exists a point-wise converging sequence $(U_j^k)_{k \in \mathbb{N}}$ in $C_0^\infty(\mathbb{R}^n)$ and a bound $h_i \in L^{(\cosh-1)}(M)$. Define $h = \wedge_{j=1, \dots, m} h_j$. Let $\alpha > 0$ be such that $\mathbb{E}_M[(\cosh-1)(\alpha h)] \leq 1$, which in turn implies $\mathbb{E}_M[e^{\alpha h}] \leq 4$. Each $\sum_j \theta_j U_j^k$ is bounded in absolute value by αh if $\sum_{j=1}^m |\theta_j| < \alpha$.

As

$$\psi(\theta) = K_M \left(\sum_{j=1}^m \theta_j U_j \right) = \log \mathbb{E}_M \left[e^{\sum_{j=1}^m \theta_j U_j} \right]$$

and

$$\psi_k(\theta) = \log \mathbb{E}_M \left[e^{\sum_{j=1}^m \theta_j U_j^k} \right]$$

dominated convergence implies $\psi_k(\theta) \rightarrow \psi(\theta)$ and hence $p_k(x; \theta) \rightarrow p(x; \theta)$ for all x if $\sum_{j=1}^m |\theta_j| < \alpha$. Sheffé lemma concludes the proof. \square

4.1 Maximal Exponential Manifold as an Affine Manifold

The maximal exponential model $\mathcal{E}(M) = \{e^{U-K_M(U)} | U \in B_M\}$ is an elementary manifold embedded into $L^{(\cosh-1)_*}(M)$ by the smooth mapping $e_M: \mathcal{S}_M \rightarrow L^{(\cosh-1)_*}(M)$. There is actually an atlas of charts that makes it into an affine manifold, see [16]. We discuss here some preliminary results about this important topic.

An elementary computation shows that

$$(\cosh-1)^2(u) = \frac{1}{2}(\cosh-1)(2u) - 2(\cosh-1)(u) \leq \frac{1}{2}(\cosh-1)(2u) \quad (12)$$

and, iterating,

$$(\cosh-1)^{2k}(u) \leq \frac{1}{2^k}(\cosh-1)(2^k u).$$

Proposition 9 If $f, g \in \mathcal{E}(M)$, then $L^{(\cosh-1)}(f \cdot M) = L^{(\cosh-1)}(g \cdot M)$.

Proof Given any $f \in \mathcal{E}(M)$, with $f = e^{U - K_M(U)}$ and $U \in \mathcal{S}_M$, and any $V \in L^{(\cosh^{-1})}(M)$, we have from Fenchel-Young inequality and Eq.(12) that

$$\begin{aligned} & \int (\cosh - 1)(\alpha V(x)) f(x) M(x) dx \\ & \leq \frac{1}{2^{k+1} k} \int (\cosh - 1)(2k\alpha V(x)) M(x) dx \\ & \quad + \frac{2k - 1}{2k} Z_M(U)^{\frac{2k}{2k-1}} \int \exp\left(\frac{2k}{2k-1} U\right) M(x) dx . \end{aligned}$$

If k is such that $\frac{2k}{2k-1} U \in \mathcal{S}_M$, one sees that $V \in L^{(\cosh^{-1})}(f \cdot M)$. We have proved that $L^{(\cosh^{-1})}(M) \subset L^{(\cosh^{-1})}(f \cdot M)$.

Conversely,

$$\begin{aligned} & \int (\cosh - 1)(\alpha V(x)) M(x) dx = \int (\cosh - 1)(\alpha V(x)) f^{-1}(x) f(x) M(x) dx \\ & \leq \frac{1}{2^{k+1} k} \int (\cosh - 1)(2k\alpha V(x)) f(x) M(x) dx \\ & \quad + \frac{2k - 1}{2k} Z_M(U)^{\frac{1}{2k-1}} \int \exp\left(\frac{1}{2k-1} U\right) M(x) dx . \end{aligned}$$

If $\frac{1}{2k-1} U \in \mathcal{S}_M$, one sees that $V \in L^{(\cosh^{-1})}(f \cdot M)$ implies $V \in L^{(\cosh^{-1})}(M)$. \square

The affine manifold is defined as follows. For each $f \in \mathcal{E}(M)$, we define the Banach space

$$B_f = \left\{ U \in L^{(\cosh^{-1})}(f \cdot M) \mid \mathbb{E}_{f \cdot M}[U] = 0 \right\} = \left\{ U \in L^{(\cosh^{-1})}(M) \mid \mathbb{E}_M[Uf] = 0 \right\},$$

and the chart

$$s_f: \mathcal{E}(M) \ni g \mapsto \log \frac{g}{f} - \mathbb{E}_{f \cdot M} \left[\log \frac{g}{f} \right].$$

It is easy to verify the following statement, which defines the *exponential affine manifold*. Specific properties related with the Gaussian space are discussed in the next Sect. 4.2 and space derivatives in Sect. 5.

Proposition 10 *The set of charts $s_f: \mathcal{E}(M) \rightarrow B_f$ is an affine atlas of global charts on $\mathcal{E}(M)$.*

On each fiber $S_p \mathcal{E}(M) = B_p$ of the statistical bundle the covariance $(U, V) \mapsto \mathbb{E}_M[UV] = \langle U, V \rangle_p$ provides a natural metric. In that metric the natural gradient of a smooth function $F: \mathcal{E}(M) \rightarrow \mathbb{R}$ is defined by

$$\frac{d}{dt} F(p(t)) = \langle \text{grad } F(p(t)), Dp(t) \rangle_{p(t)},$$

where $t \mapsto p(t)$ is a smooth curve in $\mathcal{E}(M)$ and $Dp(t) = \frac{d}{dt} \log p(t)$ is the expression of the velocity.

4.2 Translations and Mollifiers

In this section, we start to discuss properties of the exponential affine manifold of Proposition 10 which depend on the choice of the Gaussian space as base probability space.

Because of the lack of norm density of the space of infinitely differentiable functions with compact support $C_0^\infty(\mathbb{R}^n)$ in the exponential space $L^{(\cosh^{-1})}(M)$, we introduce the following classical the definition of Orlicz class.

Definition 1 We define the *exponential class*, $C_0^{(\cosh^{-1})}(M)$, to be the closure of $C_0(\mathbb{R}^n)$ in the exponential space $L^{(\cosh^{-1})}(M)$.

We recall below the characterization of the exponential class.

Proposition 11 Assume $f \in L^{(\cosh^{-1})}(M)$ and write $f_R(x) = f(x)(|x| > R)$. The following conditions are equivalent:

1. The real function $\rho \mapsto \int (\cosh - 1)(\rho f(x)) M(x) dx$ is finite for all $\rho > 0$.
2. $f \in C^{(\cosh^{-1})}(M)$.
3. $\lim_{R \rightarrow \infty} \|f_R\|_{L^{(\cosh^{-1})}(M)} = 0$.

Proof This is well known e.g., see [12, Ch.II]. A short proof is given in our note [17, Prop. 3]. \square

Here we study of the action of translation operator on the exponential space $L^{(\cosh^{-1})}(M)$ and on the exponential class $C_0^{(\cosh^{-1})}(M)$. We consider both translation by a vector, $\tau_h f(x) = f(x - h)$, $h \in \mathbb{R}^n$, and translation by a probability measure, of convolution, μ , $\tau_\mu f(x) = \int f(x - y) \mu(dy) = f * \mu(x)$. A small part of this material was published in the conference paper [17, Prop. 4–5].

Proposition 12 (Translation by a vector)

1. For each $h \in \mathbb{R}^n$, the translation mapping $L^{(\cosh^{-1})}(M) \ni f \mapsto \tau_h f$ is linear and bounded from $L^{(\cosh^{-1})}(M)$ to itself. In particular,

$$\|\tau_h f\|_{L^{(\cosh^{-1})}(M)} \leq 2 \|f\|_{L^{(\cosh^{-1})}(M)} \quad \text{if } |h| \leq \sqrt{\log 2}.$$

2. For all $g \in L^{(\cosh^{-1})^*}(M)$ we have

$$\langle \tau_h f, g \rangle_M = \langle f, \tau_h^* g \rangle_M, \quad \tau_h^* g(x) = e^{-h \cdot x - \frac{1}{2}|h|^2} \tau_{-h} g(x),$$

and $|h| \leq \sqrt{\log 2}$ implies $\|\tau_h^* g\|_{L^{(\cosh^{-1})}(M)^*} \leq 2 \|g\|_{L^{(\cosh^{-1})}(M)^*}$. The translation mapping $h \mapsto \tau_h^* g$ is continuous in $L^{(\cosh^{-1})^*}(M)$.

3. If $f \in C_0^{(\cosh^{-1})}(M)$ then $\tau_h f \in C_0^{(\cosh^{-1})}(M)$, $h \in \mathbb{R}^n$, and the mapping $\mathbb{R}^n : h \mapsto \tau_h f$ is continuous in $L^{(\cosh^{-1})}(M)$.

Proof 1. Let us first prove that $\tau_h f \in L^{(\cosh^{-1})}(M)$. Assume $\|f\|_{L^{(\cosh^{-1})}(M)} \leq 1$. For each $\rho > 0$, writing $\Phi = \cosh - 1$,

$$\begin{aligned} \int \Phi(\rho \tau_h f(x)) M(x) dx &= \int \Phi(\rho f(x-h)) M(x) dx \\ &= \int \Phi(\rho f(z)) M(z+h) dz = e^{-\frac{1}{2}|h|^2} \int e^{-z \cdot h} \Phi(\rho f(z)) M(z) dz , \end{aligned}$$

hence, using Hölder inequality and the inequality in Eq.(12),

$$\begin{aligned} \int \Phi(\rho \tau_h f(x)) M(x) dx &\leq e^{-\frac{1}{2}|h|^2} \left(\int e^{-2z \cdot h} M(z) dz \right)^{\frac{1}{2}} \left(\int \Phi^2(\rho f(z)) M(z) dz \right)^{\frac{1}{2}} \\ &\leq \frac{1}{\sqrt{2}} e^{\frac{|h|^2}{2}} \left(\int \Phi(2\rho f(z)) M(z) dz \right)^{\frac{1}{2}} . \end{aligned} \quad (13)$$

Take $\rho = 1/2$, so that $\mathbb{E}_M [\Phi(\tau_h \frac{1}{2} f(x))] \leq \frac{1}{\sqrt{2}} e^{\frac{|h|^2}{2}}$, which implies $f \in L^{(\cosh^{-1})}(M)$. Moreover, $\|\tau_h f\|_{L^{(\cosh^{-1})}(M)} \leq 2$ if $\frac{1}{\sqrt{2}} e^{\frac{|h|^2}{2}} \leq 1$.

The semi-group property $\tau_{h_1+h_2} f = \tau_{h_1} \tau_{h_2} f$ implies the boundedness for all h .

2. The computation of τ_h^* is

$$\begin{aligned} \langle \tau_h f, g \rangle_M &= \int f(x-h) g(x) M(x) dx \\ &= \int f(x) g(x+h) M(x+h) dx \\ &= \int f(x) e^{-h \cdot x - \frac{1}{2}|h|^2} \tau_{-h} g(x) M(x) dx \\ &= \langle f, \tau_h^* g \rangle_M . \end{aligned}$$

Computing Orlicz norm of the mixture space, we find

$$\begin{aligned} \|\tau_h^* g\|_{(L^{(\cosh^{-1})}(M))^*} &= \sup \{ \langle f, \tau_h^* g \rangle_M \mid \|f\|_{L^{(\cosh^{-1})}(M)} \leq 1 \} \\ &= \sup \{ \langle \tau_h f, g \rangle_M \mid \|f\|_{L^{(\cosh^{-1})}(M)} \leq 1 \} . \end{aligned}$$

From the previous item we know that $|h| \leq \sqrt{\log 2}$ implies

$$\begin{aligned} \langle \tau_h f, g \rangle_M &\leq \|\tau_h f\|_{L^{(\cosh^{-1})}(M)} \|g\|_{(L^{(\cosh^{-1})}(M))^*} \\ &\leq 2 \|f\|_{L^{(\cosh^{-1})}(M)} \|g\|_{(L^{(\cosh^{-1})}(M))^*} , \end{aligned}$$

hence $\|\tau_h^* g\|_{(L^{(\cosh-1)}(M))^*} \leq 2 \|g\|_{(L^{(\cosh-1)}(M))^*}$.

Consider first the continuity at 0. We have for $|h| \leq \sqrt{\log 2}$ and any $\phi \in C_0^\infty(\mathbb{R}^n)$ that

$$\begin{aligned} & \|\tau_h g - g\|_{(L^{(\cosh-1)}(M))^*} \\ & \leq \|\tau_h(g - \phi)\|_{(L^{(\cosh-1)}(M))^*} + \|\tau_h\phi - \phi\|_{(L^{(\cosh-1)}(M))^*} + \|\phi - g\|_{(L^{(\cosh-1)}(M))^*} \\ & \leq 3 \|g - \phi\|_{(L^{(\cosh-1)}(M))^*} + \sqrt{2} \|\tau_h\phi - \phi\|_\infty . \end{aligned}$$

The first term in the RHS is arbitrary small because of the density of $C_0^\infty(\mathbb{R}^n)$ in $L^{(\cosh-1)_*}(M)$, while the second term goes to zero as $h \rightarrow 0$ for each ϕ .

The general case follows from the boundedness and the semi-group property.

3. If $f \in C_0^{(\cosh-1)}(M)$, then, by Proposition 11, the RHS of Eq. (13) is finite for all ρ , which in turn implies that $\tau_h f \in C_0^{(\cosh-1)}(M)$ because of Proposition 11(1). Other values of h are obtained by the semi-group property.

The continuity follows from the approximation argument, as in the previous item. \square

We denote by \mathcal{P} the convex set of probability measures on \mathbb{R}^n and call *weak convergence* the convergence of sequences in the duality with $C_b(\mathbb{R}^n)$. In the following proposition we denote by \mathcal{P}_e the set of probability measures μ such that $h \mapsto e^{\frac{1}{2}|h|^2}$ is integrable. For example, this is the case when μ is Gaussian with variance $\sigma^2 I$, $\sigma^2 < 1$, or when μ has a bounded support. Weak convergence in \mathcal{P}_e means $\mu_n \rightarrow \mu$ weakly and $\int e^{\frac{1}{2}|h|^2} \mu_n(dh) \rightarrow \int e^{\frac{1}{2}|h|^2} \mu(dh)$. Note that we study here convolutions for the limited purpose of deriving the existence of smooth approximations obtained by *mollifiers*, see [2, 108–109].

Proposition 13 (Translation by a probability) *Let $\mu \in \mathcal{P}_e$.*

1. *The mapping $f \mapsto \tau_\mu f$ is linear and bounded from $L^{(\cosh-1)}(M)$ to itself. If, moreover, $\int e^{\frac{1}{2}|h|^2} \mu(dh) \leq \sqrt{2}$, then $\|\tau_\mu f\|_{L^{(\cosh-1)}(M)} \leq 2 \|f\|_{L^{(\cosh-1)}(M)}$.*
2. *If $f \in C_0^{(\cosh-1)}(M)$ then $\tau_\mu f \in C_0^{(\cosh-1)}(M)$. The mapping $\mathcal{P}: \mu \mapsto \tau_\mu f$ is continuous at δ_0 from the weak convergence to the $L^{(\cosh-1)}(M)$ norm.*

Proof 1. Let us write $\Phi = \cosh - 1$ and note the Jensen's inequality

$$\begin{aligned} \Phi(\rho \tau_\mu f(x)) &= \Phi\left(\rho \int f(x-h) \mu(dh)\right) \\ &\leq \int \Phi(\rho f(x-h)) \mu(dh) = \int \Phi(\rho \tau_h f(x)) \mu(dh) . \end{aligned}$$

By taking the Gaussian expectation of the previous inequality we have, as in the previous item,

$$\begin{aligned}
\mathbb{E}_M [\Phi(\rho \tau_\mu f)] &\leq \int \int \Phi(\rho f(x-h)) M(x) dx \mu(dh) \\
&= \int e^{-\frac{1}{2}|h|^2} \int e^{-h \cdot z} \Phi(\rho f(z)) M(z) dz \mu(dh) \\
&\leq \frac{1}{\sqrt{2}} \int e^{\frac{1}{2}|h|^2} \mu(dh) \mathbb{E}_M [\Phi(2\rho f)].
\end{aligned} \tag{14}$$

If $\|f\|_{L^{(\cosh^{-1})}(M)} \leq 1$ and $\rho = 1/2$, the RHS is bounded, hence $\tau_\mu f \in L^{(\cosh^{-1})}(M)$. If, moreover, $\int e^{\frac{1}{2}|h|^2} \mu(dh) \leq \sqrt{2}$, then the RHS is bounded by 1, hence $\|\tau_\mu f\|_{L^{(\cosh^{-1})}(M)} \leq 2$.

2. We have found above that for each $\rho > 0$ it holds (14), where the right-end-side is finite for all ρ under the current assumption. It follows from Proposition 11 that $\tau_h f \in C_0^{(\cosh^{-1})}(M)$.

To prove the continuity at δ_0 , assume $\int e^{\frac{1}{2}|h|^2} \mu(dh) \leq \sqrt{2}$, which is always feasible if $\mu \rightarrow \delta_0$ in \mathcal{P}_e weakly. Given $\epsilon > 0$, choose $\phi \in C_0^\infty(\mathbb{R}^n)$ so that $\|f - \phi\|_{L^{(\cosh^{-1})}(M)} < \epsilon$. We have

$$\begin{aligned}
&\|\tau_\mu f - f\|_{L^{(\cosh^{-1})}(M)} \\
&\leq \|\tau_\mu(f - \phi)\|_{L^{(\cosh^{-1})}(M)} + \|\tau_\mu \phi - \phi\|_{L^{(\cosh^{-1})}(M)} + \|\phi - f\|_{L^{(\cosh^{-1})}(M)} \\
&\leq 3\epsilon + A^{-1} \|\tau_\mu \phi - \phi\|_\infty,
\end{aligned}$$

where $A = \|1\|_{L^{(\cosh^{-1})}(M)}$. As $\lim_{\mu \rightarrow \delta_0} \|\tau_\mu \phi - \phi\|_\infty = 0$, see e.g. [11, III-1.9], the conclusion follows. \square

We use the previous propositions to show the existence of smooth approximations through sequences of mollifiers. A *bump* function is a non-negative function ω in $C_0^\infty(\mathbb{R}^n)$ such that $\int \omega(x) dx = 1$. It follows that $\int \lambda^{-n} \omega(\lambda^{-1}x) dx = 1$, $\lambda > 0$ and the family of mollifiers $\omega_\lambda(dx) = \lambda^{-n} \omega(\lambda^{-1}x) dx$, $\lambda > 0$, converges weakly to the Dirac mass at 0 as $\lambda \downarrow 0$ in \mathcal{P}_e . Without restriction of generality, we shall assume that the support of ω is contained in $[-1, +1]^n$.

For each $f \in L^{(\cosh^{-1})}(M)$ we have

$$\tau_{\omega_\lambda}(x) = f * \omega_\lambda(x) = \int f(x-y) \lambda^{-n} \omega(\lambda^{-1}y) dy = \int_{[-1, +1]^n} f(x-\lambda z) \omega(z) dz.$$

For each Φ convex we have by Jensen's inequality that

$$\Phi(f * \omega_\lambda(x)) \leq (\Phi \circ f) * \omega_\lambda(x)$$

and also

$$\begin{aligned} \int \Phi(f * \omega_\lambda(x)) M(x) dx &\leq \int \int_{[-1,+1]^n} \Phi \circ f(x - \lambda z) \omega(z) dz M(x) dx \\ &= \int \Phi \circ f(y) \left(\int_{[-1,+1]^n} \exp\left(-\lambda \langle z, y \rangle - \frac{\lambda^2}{2} |z|^2\right) \omega(z) dz \right) M(y) dy \\ &\leq \int \Phi \circ f(y) M(y) dy . \end{aligned}$$

Proposition 14 (Mollifiers) *Let be given a family of mollifiers ω_λ , $\lambda > 0$. For each $f \in C_0^{(\cosh^{-1})}(M)$ and for each $\lambda > 0$ the function*

$$\tau_{\omega_\lambda} f(x) = \int f(x - y) \lambda^{-n} \omega(\lambda^{-1} y) dy = f * \omega_\lambda(x)$$

belongs to $C^\infty(\mathbb{R}^n)$. Moreover,

$$\lim_{\lambda \rightarrow 0} \|f * \omega_\lambda - f\|_{L^{(\cosh^{-1})}(M)} = 0 .$$

Proof Any function in $L^{(\cosh^{-1})}(M)$ belongs to $L_{\text{loc}}^1(\mathbb{R}^n)$, hence

$$x \mapsto \int f(x - y) \omega_\lambda(y) dy = \int f(z) \omega_\lambda(z - x) dz$$

belongs to $C^\infty(\mathbb{R}^n)$, see e.g. [2, Ch.4]. Note that $\int e^{|h|^2/2} \omega_\lambda(dh) < +\infty$ and then apply Proposition 13(2).

Remark 2 Properties of weighted Orlicz spaces with the Δ_2 -property can be sometimes deduced from the properties on the un-weighted spaces by suitable embeddings, but this is not the case for the exponential space. Here are two examples.

1. Let $1 \leq a < \infty$. The mapping $g \mapsto gM^{\frac{1}{a}}$ is an isometry of $L^a(M)$ onto $L^a(\mathbb{R}^n)$. As a consequence, for each $f \in L^1(\mathbb{R}^n)$ and each $g \in L^a(M)$ we have $\left\| \left[f * (gM^{\frac{1}{a}}) \right] M^{-\frac{1}{a}} \right\|_{L^a(M)} \leq \|f\|_{L^1(\mathbb{R}^n)} \|g\|_{L^a(M)}$.
2. The mapping

$$g \mapsto \text{sign}(g) (\cosh - 1)_*^{-1}(M(\cosh - 1)_*(g))$$

is a surjection of $L^{(\cosh^{-1})_*}(\mathbb{R}^n)$ onto $L^{(\cosh^{-1})_*}(M)$ with inverse

$$h \mapsto \text{sign}(h) (\cosh - 1)_*^{-1}(M^{-1}(\cosh - 1)_*(h)) .$$

It is surjective from unit vectors (for the Luxemburg norm) onto unit vectors.

We conclude this section by recalling the following tensor property of the exponential space and of the mixture space, see [10].

Proposition 15 Let us split the components $\mathbb{R}^n x \mapsto (x_1, x_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and denote by M_1, M_2 , respectively, the Maxwell densities on the factor spaces.

1. A function f belongs to $L^{(\cosh^{-1})}(M)$ if and only if for one $\alpha > 0$ the partial integral $x_1 \rightarrow \int (\cosh^{-1})(\alpha f(x_1, x_2)) M(x_2) dx_2$ is M_1 -integrable.
2. A function f belongs to $L^{(\cosh^{-1})_*}(M)$ if and only if the partial integral $x_1 \rightarrow \int (\cosh^{-1})_*(f(x_1, x_2)) M(x_2) dx_2$ is M_1 -integrable.

4.3 Gaussian Statistical Bundle

It is an essential feature of the exponential affine manifold on $\mathcal{E}(M)$ discussed in Sect. 4.1 that the exponential *statistical bundle*

$$S\mathcal{E}(M) = \{(p, U) \mid p \in \mathcal{E}(M), U \in B_p\},$$

with $B_p = \{U \in L^{(\cosh^{-1})}(p \cdot M) \mid \mathbb{E}_{p \cdot M}[U] = 0\}$ is an expression of the tangent bundle in the atlas $\{s_p \mid p \in \mathcal{E}(M)\}$. This depends on the fact that all fibers B_p are actually a closed subspace of the exponential space $L^{(\cosh^{-1})}(M)$. This has been proved in Proposition 9. The equality of the spaces $L^{(\cosh^{-1})}(p \cdot M)$ and $L^{(\cosh^{-1})}(M)$ is equivalent to $p \in \mathcal{E}(M)$, see the set of equivalent conditions called Portmanteau Theorem in [20].

We now investigate whether translation statistical models are sub-set of the maximal exponential model $\mathcal{E}(M)$ and whether they are sub-manifolds. Proper sub-manifolds of the exponential affine manifold should have a tangent bundle that splits the statistical bundle.

Let $p \in \mathcal{E}(M)$ and write $f = p \cdot M$. Then f is a positive probability density of the Lebesgue space and so are all its translations

$$\tau_h f(x) = p(x-h)M(x-h) = e^{h \cdot x - \frac{1}{2}|h|^2} \tau_h p(x) \cdot M(x) = \tau_{-h}^* p(x) \cdot M(x).$$

From Propositions 6 and 12(2) we know that the translated densities $\tau_{-h}^* p$, are in $L^{(\cosh^{-1})_*}(M)$ for all $h \in \mathbb{R}^n$ and the dependence on h is continuous.

Let us consider now the action of the translation on the values of the chart s_M . If $s_M(p) = U$, that is $p = e^{U - K_M(U)}$ with $U \in \mathcal{S}_M$, then

$$\begin{aligned} \tau_{-h}^* p(x) \\ = e^{h \cdot X - \frac{1}{2}|h|^2} e^{U(x-h) - K_M(U)} = \exp\left(h \cdot X - \frac{1}{2}|h|^2 + \tau_h U - K_M(U)\right) \\ = \exp\left((h \cdot X + \tau_h U - \mathbb{E}_M[\tau_h U]) - \left(K_M(U) + \frac{1}{2}|h|^2 - \mathbb{E}_M[\tau_h U]\right)\right). \end{aligned}$$

Here $\tau_h U \in L^{(\cosh^{-1})}(M)$ because of Proposition 12(1). If $\tau_{-h}^* p \in \mathcal{E}(M)$, then

$$s_M(\tau_{-h}^* p) = h \cdot X + \tau_h U - \mathbb{E}_M [\tau_h U] .$$

The expected value of the translated $\tau_h U$ is

$$\mathbb{E}_M [\tau_h U] = \int U(x - h) M(x) dx = e^{-\frac{1}{2}|h|^2} \int e^{-h \cdot x} U(x) M(x) dx .$$

We have found that the action of the translation on the affine coordinate $U = s_M(p)$ of a density $p \in \mathcal{E}(M)$ is

$$U \mapsto h \cdot X + \tau_h U - e^{-\frac{1}{2}|h|^2} \mathbb{E}_M [e^{-h \cdot X} U] , \quad (15)$$

and we want the resulting value belong to \mathcal{S}_M , i.e. we want to show that

$$\begin{aligned} & \mathbb{E}_M [\exp (\gamma (h \cdot X + \tau_h U - \mathbb{E}_M [\tau_h U]))] \\ &= e^{\gamma \mathbb{E}_M [\tau_h U]} \mathbb{E}_M [e^{\gamma h \cdot X}] \mathbb{E}_M [e^{\gamma \tau_h U}] \\ &= e^{\frac{\gamma^2}{2}|h|^2 + \gamma \mathbb{E}_M [\tau_h U]} \mathbb{E}_M [e^{\gamma \tau_h U}] . \end{aligned}$$

is finite for γ in a neighborhood of 0.

We have the following result.

- Proposition 16**
1. If $p \in \mathcal{E}(M)$, for all $h \in \mathbb{R}^n$ the translated density $\tau_{-h}^* p$ is in $\mathcal{E}(M)$.
 2. If $s_M(p) \in C_0^{(\cosh^{-1})}(M)$, then $s_M(\tau_{-h}^* p) \in C_0^{(\cosh^{-1})}(M) \cap \mathcal{S}_M$ for all $h \in \mathbb{R}^n$ and dependence in h is continuous.

Proof 1. For each γ and conjugate exponents α, β , we have

$$\begin{aligned} \mathbb{E}_M [e^{\gamma \tau_h U}] &= e^{-\frac{1}{2}|h|^2} \int e^{-h \cdot x} e^{\gamma U(x)} M(x) dx \\ &\leq e^{-\frac{1}{2}|h|^2} \left(\frac{1}{\alpha} e^{\frac{\alpha^2}{2}|h|^2} + \frac{1}{\beta} \mathbb{E}_M [e^{\beta \gamma U}] \right) . \end{aligned}$$

As $U \in \mathcal{S}_M$, then $\mathbb{E}_M [e^{\pm aU}] < \infty$ for some $a > 1$, and we can take $\beta = \sqrt{a}$ and $\gamma = \pm \sqrt{a}$.

2. Under the assumed conditions on U the mapping $h \mapsto \tau_h U$ is continuous in $C_0^{(\cosh^{-1})}(M)$ because of Proposition 12(3). So is $h \mapsto \mathbb{E}_M [\tau_h U]$. As $X_i \in C_0^{(\cosh^{-1})}(M)$, $i = 1, \dots, n$, the same is true for $h \mapsto h \cdot X$. In conclusion, the translated U of (15) belongs to $C_0^{(\cosh^{-1})}(M)$.

□

The proposition above shows that the translation statistical model $\tau_{-h}^* p$, $h \in \mathbb{R}^m$ is well defined as a subset of $\mathcal{E}(M)$. To check if it is a differentiable sub-manifold, we want to compute the velocity of a curve $t \mapsto \tau_{h(t)}^* p$, that is

$$\frac{d}{dt} \left(h(t) \cdot X + \tau_{h(t)} U - \mathbb{E}_M [\tau_{h(t)}] U \right).$$

That will require first of all the continuity in h , hence $U \in C_0^{(\cosh-1)}(M)$, and moreover we want to compute $\partial/\partial h_i U(x-h)$, that is the gradient of U . This task shall be the object of the next section.

Cases other than translations are of interest. Here are two sufficient conditions for a density to be in $\mathcal{E}(M)$.

Proposition 17 1. Assume $p > 0$ M -a.s., $\mathbb{E}_M [p] = 1$, and

$$\mathbb{E}_M [p^{n_1/(n_1-1)}] \leq 2^{n_1/(n_1-1)}, \quad \mathbb{E}_M [p^{-1/(n_2-1)}] \leq 2^{n_2/(n_2-1)} \quad (16)$$

for some natural $n_1, n_2 > 2$. Then $p \in \mathcal{E}(M)$, the exponential spaces are equal, $L^{(\cosh-1)}(M) = L^{(\cosh-1)}(p \cdot M)$, and for all random variable U

$$\|U\|_{L^{(\cosh-1)}(p \cdot M)} \leq 2^{n_1} \|U\|_{L^{(\cosh-1)}(M)}, \quad (17)$$

$$\|U\|_{L^{(\cosh-1)}(M)} \leq 2^{n_2} \|U\|_{L^{(\cosh-1)}(p \cdot M)}. \quad (18)$$

2. Condition (16) holds for $p = \sqrt{\pi/2} |X_i|$ and for $p = X_i^2$, $i = 1, \dots, n$.
3. Let χ be a diffeomorphism of \mathbb{R}^n and such that both the derivatives are uniformly bounded in norm. Then the density of the image under χ of the standard Gaussian measure belongs to $\mathcal{E}(M)$.

Proof 1. The bound on the moments in Eq.(16) is equivalent to the inclusion in $\mathcal{E}(M)$ because of the definition of \mathcal{S}_M , or see [20, Th.4.7(vi)]. Assume $\|U\|_{L^{(\cosh-1)}(M)} \leq 1$, that is $\mathbb{E}_M [(\cosh-1)(U)] \leq 1$. From Hölder inequality and the elementary inequality in Eq.(12), we have

$$\begin{aligned} \mathbb{E}_{f \cdot M} \left[(\cosh-1) \left(\frac{U}{2^{n_1}} \right) \right] &= \mathbb{E}_M \left[(\cosh-1) \left(\frac{U}{2^{n_1}} \right) f \right] \\ &\leq \mathbb{E}_M \left[(\cosh-1) \left(\frac{U}{2^{n_1}} \right)^{n_1} \right]^{1/n_1} \mathbb{E}_M [f^{n_1/(n_1-1)}]^{(n_1-1)/n_1} \leq \frac{1}{2} \cdot 2 = 1 \end{aligned}$$

For the other direction, assume $\|U\|_{L^{(\cosh-1)}(f \cdot M)} \leq 1$, that is $\mathbb{E}_M [\Phi(U)f] \leq 1$, so that

$$\begin{aligned} \mathbb{E}_M \left[(\cosh-1) \left(\frac{U}{2^{n_2}} \right) \right] &= \mathbb{E}_M \left[(\cosh-1) \left(\frac{U}{2^{n_2}} \right) f^{1/n_2} f^{-1/n_2} \right] \\ &\leq \mathbb{E}_M \left[(\cosh-1) \left(\frac{U}{2^{n_2}} \right)^{n_2} f \right]^{1/n_2} \mathbb{E}_M [f^{-1/(n_2-1)}]^{(n_2-1)/n_2} \leq \frac{1}{2} \cdot 2 = 1. \end{aligned}$$

2. Simple computations of moments.
3. We consider first the case where $\chi(0) = 0$, in which case we have the following inequalities. If we define $\alpha^{-1} = \sup \{ \|d\chi(x)\|^2 \mid x \in \mathbb{R}^n \}$, then $\alpha |\chi(x)|^2 \leq |x|^2$

for all $x \in \mathbb{R}^n$ and equivalently, $\alpha |x|^2 \leq |\chi^{-1}(x)|^2$. In a similar way, if we define $\beta^{-1} = \sup \left\{ \|d\chi^{-1}(y)\|^2 \mid y \in \mathbb{R}^n \right\}$, then $\beta |\chi^{-1}(y)|^2 \leq |y|^2$ and $\beta |x|^2 \leq |\chi(x)|^2$.

The density of the image probability is $M \circ \chi^{-1} |\det d\chi^{-1}|$ and we want to show that for some $\epsilon > 0$ the following inequalities both hold,

$$\mathbb{E}_M \left[\left(\frac{M \circ \chi^{-1} |\det d\chi^{-1}|}{M} \right)^{1+\epsilon} \right] < \infty$$

and

$$\mathbb{E}_{M \circ \chi^{-1} |\det d\chi^{-1}|} \left[\left(\frac{M}{M \circ \chi^{-1} |\det d\chi^{-1}|} \right)^{1+\epsilon} \right] < \infty.$$

The first condition is satisfied as

$$\begin{aligned} & \int |\det d\chi^{-1}(x)|^{1+\epsilon} \left(\frac{M(\chi^{-1}(x))}{M(x)} \right)^{1+\epsilon} M(x) dx \\ &= \int |\det d\chi^{-1}(x)|^{1+\epsilon} M(\chi^{-1}(x))^{1+\epsilon} M(x)^{-\epsilon} dx \\ &\leq (2\pi)^{-n/2} \beta^{-\frac{(1+\epsilon)n}{2}} \int \exp \left(-\frac{1}{2} \left((1+\epsilon) |\chi^{-1}(x)|^2 - \epsilon |x|^2 \right) \right) dx \\ &= (2\pi)^{-n/2} \beta^{-\frac{(1+\epsilon)n}{2}} \int \exp \left(-\frac{|x|^2}{2} \left((1+\epsilon) \frac{|\chi^{-1}(x)|^2}{|x|^2} - \epsilon \right) \right) dx \\ &\leq (2\pi)^{-n/2} \beta^{-\frac{(1+\epsilon)n}{2}} \int \exp \left(-\frac{|x|^2}{2} ((1+\epsilon)\alpha - \epsilon) \right) dx, \end{aligned}$$

where we have used the Hadamard's determinant inequality

$$|\det d\chi^{-1}(x)| \leq \|d\chi^{-1}(x)\|^n \leq \beta^{-n/2}$$

and the lower bound $\alpha \leq \frac{|\chi^{-1}(x)|^2}{|x|^2}$, $x \in \mathbb{R}^n_*$. If $\alpha \geq 1$ then $(1+\epsilon)\alpha - \epsilon = \alpha + \epsilon(\alpha - 1) \geq \alpha > 0$ for all ϵ . If $\alpha < 1$, then $(1+\epsilon)\alpha - \epsilon > 0$ if $\epsilon < \alpha/(1-\alpha)$ e.g., $\epsilon = \alpha/2(1-\alpha)$, which in turn gives $(1+\epsilon)\alpha - \epsilon = \alpha/2$. In conclusion, there exist an $\epsilon > 0$ such that

$$\begin{aligned} & \int |\det d\chi^{-1}(x)|^{1+\epsilon} \left(\frac{M(\chi^{-1}(x))}{M(x)} \right)^{1+\epsilon} M(x) dx \\ &\leq (2\pi)^{-n/2} |\det d\chi^{-1}(x)|^{1+\epsilon} \int \exp \left(-\frac{\alpha |x|^2}{4} \right) dx = \left(\frac{\alpha}{2} \right)^{n/2}. \end{aligned}$$

For the second inequality,

$$\begin{aligned}
& \int \left(\frac{M(y)}{M(\chi^{-1}(y)) |\det d\chi^{-1}(y)|} \right)^{1+\epsilon} M(\chi^{-1}(y)) |\det d\chi^{-1}(y)| \, dy \\
&= \int M(y)^{1+\epsilon} M(\chi^{-1}(y))^{-\epsilon} |\det d\chi^{-1}(y)|^{-\epsilon} \, dy \\
&= \int M(\chi(x))^{1+\epsilon} M(x)^{-\epsilon} |\det d\chi^{-1}(\chi(x))|^{-\epsilon} |\det d\chi(x)| \, dx \\
&= \int |\det d\chi(x)|^{1+\epsilon} M(\chi(x))^{1+\epsilon} M(x)^{-\epsilon} \, dx .
\end{aligned}$$

As the last term is equal to the expression in the previous case with χ^{-1} replaced by χ , the same proof applies with the bounds α and β exchanged. \square

Remark 3 While the moment condition for proving $p \in \mathcal{E}(M)$ has been repeatedly used, nonetheless the results above have some interest. The first one is an example where an explicit bound for the different norms on the fibers of the statistical bundle is derived. The second case is the starting point for the study of transformation models where a group of transformation χ_θ is given.

5 Weighted Orlicz-Sobolev Model Space

We proceed in this section to the extension of our discussion of translation statistical models to statistical models of the Gaussian space whose densities are differentiable. We restrict to generalities and refer to previous work in [10] for examples of applications, such as the discussion of Hyvärinen divergence. This is a special type of divergence between densities that involves an L^2 -distance between gradients of densities [9] which has multiple applications. In particular, it is related with the improperly called Fisher information in [22, p. 49].

We are led to consider a case classical weighted Orlicz-Sobolev spaces which is not treated in much detail in standard monographs such as [1]. The analysis of the finite dimensional Gaussian space i.e. the space of square-integrable random variables on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), M)$ is a well developed subject. Some of the result below could be read as special case of that theory. We refer to P. Malliavin's textbook [11, Ch. 5] and to D. Nualart's monograph [14].

5.1 Orlicz-Sobolev Spaces with Gaussian Weight

The first definitions are taken from our [10].

Definition 2 The exponential and the mixture Orlicz-Sobolev-Gauss (OSG) spaces are, respectively,

$$W^{1,(\cosh-1)}(M) = \{f \in L^{(\cosh-1)}(M) \mid \partial_j f \in L^{(\cosh-1)}(M)\} , \quad (19)$$

$$W^{1,(\cosh-1)_*}(M) = \{f \in L^{(\cosh-1)_*}(M) \mid \partial_j f \in L^{(\cosh-1)_*}(M)\} , \quad (20)$$

where ∂_j , $j = 1, \dots, n$, is the partial derivative in the sense of distributions.

As $\phi \in C_0^\infty(\mathbb{R}^n)$ implies $\phi M \in C_0^\infty(\mathbb{R}^n)$, for each $f \in W^{1,(\cosh-1)_*}(M)$ we have, in the sense of distributions, that

$$\langle \partial_j f, \phi \rangle_M = \langle \partial_j f, \phi M \rangle = -\langle f, \partial_j(\phi M) \rangle = \langle f, M(X_j - \partial_j)\phi \rangle = \langle f, \delta_j \phi \rangle_M ,$$

with $\delta_j \phi = (X_j - \partial_j)\phi$. The *Stein operator* δ_i acts on $C_0^\infty(\mathbb{R}^n)$.

The meaning of both operators ∂_j and $\delta_j = (X_j - \partial_j)$ when acting on square-integrable random variables of the Gaussian space is well known, but here we are interested in the action on OSG-spaces. Let us denote by $C_p^\infty(\mathbb{R}^n)$ the space of infinitely differentiable functions with polynomial growth. Polynomial growth implies the existence of all M -moments of all derivatives, hence $C_p^\infty(\mathbb{R}^n) \subset W^{1,(\cosh-1)_*}(M)$. If $f \in C_p^\infty(\mathbb{R}^n)$, then the distributional derivative and the ordinary derivative are equal and moreover $\delta_j f \in C_p^\infty(\mathbb{R}^n)$. For each $\phi \in C_0^\infty(\mathbb{R}^n)$ we have $\langle \phi, \delta_j f \rangle_M = \langle \partial_j \phi, f \rangle_M$.

The OSG spaces $W_{(\cosh-1)}^1(M)$ and $W_{(\cosh-1)_*}^1(M)$ are both Banach spaces, see [12, Sec. 10]. In fact, both the product functions $(u, x) \mapsto (\cosh-1)(u)M(x)$ and $(u, x) \mapsto (\cosh-1)_*(u)M(x)$ are ϕ -functions according the Musielak's definition. The norm on the OSG-spaces are the graph norms,

$$\|f\|_{W_{(\cosh-1)}^1(M)} = \|f\|_{L^{(\cosh-1)}(M)} + \sum_{j=1}^n \|\partial_j f\|_{L^{(\cosh-1)}(M)} \quad (21)$$

and

$$\|f\|_{W_{(\cosh-1)_*}^1(M)} = \|f\|_{L^{(\cosh-1)}(M)} + \sum_{j=1}^n \|\partial_j f\|_{L^{(\cosh-1)}(M)} . \quad (22)$$

Because of Proposition 9, see also [20, Th. 4.7], for each $p \in \mathcal{E}(M)$, we have both equalities and isomorphisms $L^{(\cosh-1)}(p \cdot M) = L^{(\cosh-1)}(M)$ and $L^{(\cosh-1)_*}(p \cdot M) = L^{(\cosh-1)_*}(M)$. It follows

$$\begin{aligned} W^{1,(\cosh-1)}(M) &= W^{1,(\cosh-1)}(p \cdot M) \\ &= \{f \in L^{(\cosh-1)}(p \cdot M) \mid \partial_j f \in L^{(\cosh-1)}(p \cdot M)\} , \end{aligned} \quad (23)$$

$$\begin{aligned} W^{1,(\cosh-1)_*}(M) &= W^{1,(\cosh-1)_*}(p \cdot M) \\ &= \{f \in L^{(\cosh-1)_*}(p \cdot M) \mid \partial_j f \in L^{(\cosh-1)_*}(p)\} , \end{aligned} \quad (24)$$

and equivalent graph norms for any density $p \in \mathcal{E}(M)$. The OSG spaces are compatible with the structure of the maximal exponential family $\mathcal{E}(M)$. In particular, as all Gaussian densities of a given dimension belong into the same exponential manifold, one could have defined the OSG spaces with respect to any of such densities.

We review some relations between OSG-spaces and ordinary Sobolev spaces. For all $R > 0$

$$(2\pi)^{-\frac{n}{2}} \geq M(x) \geq M(x)(|x| < R) \geq (2\pi)^{-\frac{n}{2}} e^{-\frac{R^2}{2}} (|x| < R), \quad x \in \mathbb{R}^n.$$

Proposition 18 *Let $R > 0$ and let Ω_R denote the open sphere of radius R .*

1. *We have the continuous mappings*

$$W^{1,(\cosh-1)}(\mathbb{R}^n) \subset W^{1,(\cosh-1)}(M) \rightarrow W^{1,p}(\Omega_R), \quad p \geq 1.$$

2. *We have the continuous mappings*

$$W^{1,p}(\mathbb{R}^n) \subset W^{1,(\cosh-1)_*}(\mathbb{R}^n) \subset W^{1,(\cosh-1)_*}(M) \rightarrow W^{1,1}(\Omega_R), \quad p > 1.$$

3. *Each $u \in W^{1,(\cosh-1)}(M)$ is a.s. Hölder of all orders on each $\overline{\Omega}_R$ and hence a.s. continuous. The restriction $W^{1,(\cosh-1)}(M) \rightarrow C(\overline{\Omega}_R)$ is compact.*

- Proof*
1. From the inequality on M and from $(\cosh-1)(y) \geq y^{2n}/(2n)!$.
 2. From the inequality on M and from $y^2/2 \geq (\cosh-1)_*(y)$ and $\cosh(1) - 1 + (\cosh-1)_*(y) \geq |y|$.
 3. It is the Sobolev's embedding theorem [2, Ch. 9].

Let us consider now the extension of the ∂_j operator to the OSG-spaces and its relation with the translation operator.

The operator given by the ordinary partial derivative $\partial_j : C_p^\infty(\mathbb{R}^n) \rightarrow C_p^\infty(\mathbb{R}^n) \subset L^{(\cosh-1)_*}(M)$ is closable. In fact, if both $f_n \rightarrow 0$ and $\partial_j f_n \rightarrow \eta$ in $L^{(\cosh-1)_*}(M)$, then for all $\phi \in C_0^\infty(\mathbb{R}^n)$,

$$\langle \phi, \eta \rangle_M = \lim_{n \rightarrow \infty} \langle \phi, \partial_j f_n \rangle_M = \lim_{n \rightarrow \infty} \langle \delta\phi, f_n \rangle_M = 0,$$

hence $\eta = 0$. The same argument shows that $\partial_j : C_0^\infty(\mathbb{R}^n) \rightarrow C_0^\infty(\mathbb{R}^n) \subset L^{(\cosh-1)}(M)$ is closable.

For $f \in L^{(\cosh-1)}(M)$ we define $\tau_h f(x) = f(x - h)$ and it holds $\tau_h f \in L^{(\cosh-1)}(M)$ because of Proposition 12(I). For each given $f \in W^{1,(\cosh-1)}(M)$ we denote by $\partial_j f \in W^{1,(\cosh-1)}(M)$, $j = 1, \dots, n$ its distributional partial derivatives and write $\nabla f = (\partial_j f : j = 1, \dots, n)$.

Proposition 19 (Continuity and directional derivative)

1. For each $f \in W^{1,(\cosh-1)}(M)$, each unit vector $h \in S^n$, and all $t \in \mathbb{R}$, it holds

$$f(x + th) - f(x) = t \int_0^1 \sum_{j=1}^n \partial_j f(x + sth) h_j \, ds .$$

Moreover, $|t| \leq \sqrt{2}$ implies

$$\|f(x + th) - f(x)\|_{L^{(\cosh-1)}(M)} \leq 2t \|\nabla f\|_{L^{(\cosh-1)}(M)} ,$$

especially, $\lim_{t \rightarrow 0} \|f(x + th) - f(x)\|_{L^{(\cosh-1)}(M)} = 0$ uniformly in h .

2. For each $f \in W^{1,(\cosh-1)}(M)$ the mapping $h \mapsto \tau_h f$ is differentiable from \mathbb{R}^n to $L^{\infty-0}(M)$ with gradient ∇f at $h = 0$.
3. For each $f \in W^{1,(\cosh-1)}(M)$ and each $g \in L^{(\cosh-1)_*}(M)$, the mapping $h \mapsto \langle \tau_h f, g \rangle_M$ is differentiable. Conversely, if $f \in L^{(\cosh-1)}(M)$ and $h \mapsto \tau_h f$ is weakly differentiable, then $f \in W^{1,(\cosh-1)}(M)$
4. If $\partial_j f \in C_0^{(\cosh-1)}(M)$, $j = 1, \dots, n$, then strong differentiability in $L^{(\cosh-1)}(M)$ holds.

Proof 1. Recall that for each $g \in C_0^\infty(\mathbb{R}^n)$ we have

$$\langle \partial_j f, g \rangle_M = -\langle f, \partial_j(gM) \rangle = \langle f, \delta_j g \rangle_M , \quad \delta_j g = X_j g - \partial_j g \in C_0^\infty(\mathbb{R}^n) .$$

We show the equality $\tau_{-th} f - f = t \int_0^1 \tau_{-sth}(\nabla f) \cdot h \, ds$ in the scalar product with a generic $g \in C_0^\infty(\mathbb{R}^n)$:

$$\begin{aligned} \langle \tau_{-th} f - f, g \rangle_M &= \int f(x + th)g(x)M(x) \, dx - \int f(x)g(x)M(x) \, dx \\ &= \int f(x)g(x - th)M(x - th) \, dx - \int f(x)g(x)M(x) \, dx \\ &= \int f(x)(g(x - th)M(x - th) - g(x)M(x)) \, dx \\ &= -t \int f(x) \int_0^1 \sum_{j=1}^n \partial_j(gM)(x - sth)h_j \, ds \, dx \\ &= -t \int_0^1 \int f(x) \sum_{j=1}^n \partial_j(gM)(x - sth)h_j \, dx \, ds \\ &= t \int_0^1 \int \sum_{j=1}^n \partial_j f(x)h_j \, g(x - sth)M(x - sth) \, dx \, ds \\ &= t \int_0^1 \int \sum_{j=1}^n \partial_j f(x + sth)h_j \, g(x)M(x) \, dx \, ds \end{aligned}$$

$$= \left\langle t \int_0^1 \tau_{-sth}(\nabla f) \cdot h \, ds, g \right\rangle_M .$$

If $|t| \leq \sqrt{\log 2}$ then the translation sth is small, $|sth| \leq \sqrt{\log 2}$ so that, according to Proposition 12(1), we have $\|\tau_{-sth}(\nabla f \cdot h)\|_{L^{(\cosh^{-1})}(M)} \leq 2 \|\nabla f \cdot h\|_{L^{(\cosh^{-1})}(M)}$ and the thesis follows.

2. We want to show that the following limit holds in all $L^\alpha(M)$ -norms, $\alpha > 1$:

$$\lim_{t \rightarrow 0} \frac{\tau_{-th}f - f}{t} = \sum_{j=1}^n h_j \partial_j f .$$

Because of the identity in the previous Item, we need to show the limit

$$\lim_{t \rightarrow 0} \int \left| \int_0^1 (\tau_{-sth}(\nabla f(x) \cdot h) - \nabla f(x) \cdot h) \, ds \right|^\alpha M(x) dx = 0 .$$

The Jensen's inequality gives

$$\begin{aligned} & \int \left| \int_0^1 (\tau_{-sth}(\nabla f(x) \cdot h) - \nabla f(x) \cdot h) \, ds \right|^\alpha M(x) dx \\ & \leq \int_0^1 \int |\tau_{-sth}(\nabla f(x) \cdot h) - \nabla f(x) \cdot h|^\alpha M(x) dx \, ds \end{aligned}$$

and the result follows because translations are bounded and the continuous in $L^\alpha(M)$.

3. We have

$$\begin{aligned} & \left\langle \int_0^1 (\tau_{(-sth)}f - f) \, ds, g \right\rangle_M = \int_0^1 \langle \tau_{(-sth)}f - f, g \rangle_M \, ds \\ & = \int_0^1 \langle f, \tau_{(-sth)}^*g - g \rangle_M \, ds . \end{aligned}$$

Conclusion follows because $y \mapsto \tau_y^*g$ is bounded continuous.

Assume now $f \in L^{(\cosh^{-1})}(M)$ and $h \mapsto \tau_h f$ is weakly differentiable. Then there exists $f_1, \dots, f_n \in L^{(\cosh^{-1})}(M)$ such that for each $\phi \in C_0^\infty(\mathbb{R}^n)$

$$\begin{aligned} \langle f_j, \phi M \rangle &= \langle f_j, \phi \rangle_M = \frac{d}{dt} \langle \tau_{-te_j}f, \phi \rangle_M = \frac{d}{dt} \langle \tau_{-te_j}f, \phi M \rangle \\ &= \frac{d}{dt} \langle f, \tau_{te_j}(\phi M) \rangle = -\langle f, \partial_j(\phi M) \rangle . \end{aligned}$$

The distributional derivative holds because ϕM is the generic element of $C_0^\infty(\mathbb{R}^n)$.

4. For each $\rho > 0$ Jensen's inequality implies

$$\begin{aligned} & \left\| \int_0^1 (\tau_{-sth}(\nabla f \cdot h) - \nabla f \cdot h) ds M(x) dx \right\|_{L^{(\cosh^{-1})}(M)} \\ & \leq \int_0^1 \|(\tau_{-sth}(\nabla f \cdot h) - \nabla f \cdot h) M(x) dx\|_{L^{(\cosh^{-1})}(M)} ds. \end{aligned}$$

As in Proposition 12(1) we choose $|t| \leq \sqrt{\log 2}$ to get $|ste_j| \leq \sqrt{\log 2}$, $0 \leq s \leq 1$, so that $\|\tau_{-sth} \nabla f \cdot h\|_{L^{(\cosh^{-1})}(M)} \leq 2 \|\nabla f \cdot h\|_{L^{(\cosh^{-1})}(M)}$, hence the integrand is bounded by $\|\nabla f \cdot h\|_{L^{(\cosh^{-1})}(M)}$. The convergence for each s follows from the continuity of the translation on $C_0^{(\cosh^{-1})}(M)$.

Notice that in Item 2. of the proposition we could have derived a stronger differentiability if the mapping $h \mapsto \tau_h \nabla f$ were continuous in $L^{(\cosh^{-1})}(M)$. That, and other similar observations, lead to the following definition.

Definition 3 The *Orlicz-Sobolev-Gauss exponential class* is

$$C_0^{1,(\cosh^{-1})}(M) = \left\{ f \in W^{1,(\cosh^{-1})}(M) \mid f, \partial_j f \in C_0^{(\cosh^{-1})}(M), j = 1, \dots, n \right\}$$

The following density results will be used frequently in approximation arguments. We denote by $(\omega_n)_{n \in \mathbb{N}}$ a sequence of mollifiers.

Proposition 20 (Calculus in $C_0^{1,(\cosh^{-1})}(M)$)

1. For each $f \in C_0^{1,(\cosh^{-1})}(M)$ the sequence $f * \omega_n, n \in \mathbb{N}$, belongs to $C^\infty(\mathbb{R}^n) \cap W^{1,(\cosh^{-1})}(M)$. Precisely, for each n and $j = 1, \dots, n$, we have the equality $\partial_j(f * \omega_n) = (\partial_j f) * \omega_n$; the sequences $f * \omega_n$, respectively $\partial_j f * \omega_n$, $j = 1, \dots, n$, converge to f , respectively $\partial_j f$, $j = 1, \dots, n$, strongly in $L^{(\cosh^{-1})}(M)$.
2. Same statement is true if $f \in W^{1,(\cosh^{-1})*}(M)$.
3. Let be given $f \in C_0^{1,(\cosh^{-1})}(M)$ and $g \in W^{1,(\cosh^{-1})*}(M)$. Then $fg \in W^{1,1}(M)$ and $\partial_j(fg) = \partial_j fg + f \partial_j g$.
4. Let be given $F \in C^1(\mathbb{R})$ with $\|F'\|_\infty < \infty$. For each $U \in C_0^{1,(\cosh^{-1})}(M)$, we have $F \circ U, F' \circ U \partial_j U \in C_0^{(\cosh^{-1})}(M)$ and $\partial_j F \circ U = F' \circ U \partial_j U$, in particular $F(U) \in C_0^{1,(\cosh^{-1})}(M)$.

Proof 1. We need only to note that the equality $\partial_j(f * \omega_n) = (\partial_j f) * \omega_n$ is true for $f \in W^{1,(\cosh^{-1})}(M)$. Indeed, the sequence $f * \omega_n$ belongs to $C^\infty(\mathbb{R}^n) \cap L^{(\cosh^{-1})}(M)$ and converges to f in $L^{(\cosh^{-1})}(M)$ -norm according from Proposition 14. The sequence $\partial_j f * \omega_n = (\partial_j f) * \omega_n$ converges to $\partial_j f$ in $L^{(\cosh^{-1})}(M)$ -norm because of the same theorem.

2. Same proof.
3. Note that $fg, \partial_j fg + f \partial_j g \in L^1(M)$. The following convergence in $L^1(M)$ holds

$$\partial_j f g + f \partial_j g = \lim_{n \rightarrow \infty} \partial_j f * \omega_n g * \omega_n + f * \omega_n \partial_j g * \omega_n = \lim_{n \rightarrow \infty} \partial_j f * \omega_n g * \omega_n ,$$

so that for all $\phi \in C_0^\infty(\mathbb{R}^n)$

$$\begin{aligned}\langle \partial_j f g + f \partial_j g, \phi \rangle &= \lim_{n \rightarrow \infty} \langle \partial_j f * \omega_n g * \omega_n, \phi \rangle \\ &= \lim_{n \rightarrow \infty} -\langle f * \omega_n g * \omega_n, \partial_j \phi \rangle = -\langle f g, \partial_j \phi \rangle.\end{aligned}$$

It follows that the distributional partial derivative of the product is $\partial_j f g = \partial_j f g + f \partial_j g$, in particular belongs to $L^1(M)$, hence $f g \in W^{1,1}(M)$.

4. From the assumption on F we have $|F(U)| \leq |F(0)| + \|F'\|_\infty |U|$. It follows $F \circ U \in L^{(\cosh^{-1})}(M)$ because

$$\begin{aligned}&\int (\cosh - 1) (\rho F(U(x))) M(x) dx \\ &\leq \frac{1}{2} (\cosh - 1) (2\rho F(0)) + \frac{1}{2} \int (\cosh - 1) (2\rho \|F'\|_\infty U(x)) M(x) dx,\end{aligned}$$

and $\rho \|F(U)\|_{L^{(\cosh^{-1})}(M)} \leq 1$ if both

$$(\cosh - 1)(2\rho F(0)) \leq 1, \quad 2\rho \|F'\|_\infty \|U\|_{L^{(\cosh^{-1})}(M)} \leq 1.$$

In the same way we show that $F' \circ U \partial_j U \in L^{(\cosh^{-1})}(M)$. Indeed,

$$\begin{aligned}&\int (\cosh - 1) (\rho F'(U(x)) \partial_j U(x)) M(x) dx \\ &\leq \int (\cosh - 1) (\rho \|F'\|_\infty \partial_j U(x)) M(x) dx,\end{aligned}$$

so that $\rho \|F' \circ U \partial_j U\|_{L^{(\cosh^{-1})}(M)} \leq 1$ if $\rho \|F'\|_\infty \|\partial_j U(x)\|_{L^{(\cosh^{-1})}(M)} = 1$. Because of the Item (1) the sequence $U * \omega_n$ belongs to C^∞ and converges strongly in $L^{(\cosh^{-1})}(M)$ to U , so that from

$$\|F \circ (U * \omega_n) - F \circ U\|_{L^{(\cosh^{-1})}(M)} \leq \|F'\|_\infty \|U * \omega_n - U\|_{L^{(\cosh^{-1})}(M)}$$

we see that $F \circ (U * \omega_n) \rightarrow F \circ U$ in $L^{(\cosh^{-1})}(M)$. In the same way,

$$\begin{aligned}&\|F' \circ (U * \omega_n) \partial_j (U * \omega_n) - F' \circ U \partial_j U\|_{L^{(\cosh^{-1})}(M)} \\ &\leq \|F' \circ (U * \omega_n) (\partial_j (U * \omega_n) - \partial_j U)\|_{L^{(\cosh^{-1})}(M)} \\ &\quad + \|(F' \circ (U * \omega_n) - F' \circ U) \partial_j U\|_{L^{(\cosh^{-1})}(M)} \\ &\leq \|F'\|_\infty \|\partial_j (U * \omega_n) - \partial_j U\|_{L^{(\cosh^{-1})}(M)} \\ &\quad + \|(F' \circ (U * \omega_n) - F' \circ U) \partial_j U\|_{L^{(\cosh^{-1})}(M)}.\end{aligned}$$

The first term goes clearly to 0, while the second term requires consideration. Note the bound

$$|(F' \circ (U \circ \omega_n) - F' \circ U) \partial_j U| \leq 2 \|F'\|_\infty |\partial_j U| ,$$

so that the sequence $(F' \circ (U \circ \omega_n) - F' \circ U) \partial_j U$ goes to zero in probability and is bounded by a function in $C_0^{(\cosh^{-1})}(M)$. This in turn implies the convergence in $L^{(\cosh^{-1})}(M)$.

Finally we check that the distributional derivative of $F \circ U$ is $F' \circ U \partial_j U$: for each $\phi \in C_0^\infty(\mathbb{R}^n)$

$$\begin{aligned} \langle \partial_j F \circ U, \phi M \rangle &= -\langle F \circ U, \partial_j(\phi M) \rangle \\ &= -\langle F \circ U, \delta_j \phi \rangle_M \\ &= \lim_{n \rightarrow \infty} \langle F \circ (U * \omega_n), \delta_j \phi \rangle_M \\ &= \lim_{n \rightarrow \infty} \langle \partial_j F \circ (U * \omega_n), \phi \rangle_M \\ &= \lim_{n \rightarrow \infty} \langle F' \circ (U * \omega_n) \partial_j (U * \omega_n), \phi \rangle_M \\ &= \langle F' \circ U \partial_j U, \phi \rangle_M \\ &= \langle F' \circ U \partial_j U, \phi M \rangle . \end{aligned}$$

We conclude our presentation by re-stating a technical result from [10, Prop. 15], where the assumptions were not sufficient for the stated result.

Proposition 21 1. If $U \in \mathcal{S}_M$ and $f_1, \dots, f_m \in L^{(\cosh^{-1})}(M)$, then $f_1 \cdots f_m e^{U-K_M(U)} \in L^\gamma(M)$ for some $\gamma > 1$, hence it is in $L^{(\cosh^{-1})_*}(M)$.
2. If $U \in \mathcal{S}_M \cap C_0^{1,(\cosh^{-1})}(M)$ and $f \in C_0^{1,(\cosh^{-1})}(M)$, then

$$f e^{u-K_M(u)} \in W^{1,(\cosh^{-1})_*}(M) \cap C(\mathbb{R}^n) ,$$

and its distributional partial derivatives are $(\partial_j f + f \partial_j u) e^{u-K_M(u)}$

Proof 1. From We know that $e^{U-K_M(U)} \cdot M \in \mathcal{E}(M)$ and $e^{U-K_M(U)} \in L^{1+\varepsilon}(M)$ for some $\varepsilon > 0$. From that, let us prove that $f_1 \cdots f_m e^{U-K_M(U)} \in L^\gamma(M)$ for some $\gamma > 1$. According to classical (m+1)-term Fenchel-Young inequality,

$$\begin{aligned} &|f_1(x) \cdots f_n(x)| e^{U(x)-K_M(U)} \\ &\leq \sum_{i=1}^m \frac{1}{\alpha_i} |f_i(x)|^{\alpha_i} + \frac{1}{\beta} |e^{U(x)-K_M(U)}|^\beta \\ &, \alpha_1, \dots, \alpha_m, \beta > 1, \sum_{i=1}^m \frac{1}{\alpha_i} + \frac{1}{\beta} = 1, x \in \mathbb{R}^n . \end{aligned}$$

Since $(\cosh^{-1})_*$ is convex, we have

$$\begin{aligned} & \mathbb{E}_M [(\cosh - 1)_*(|f_1 \cdots f_m| e^{U - K_M(U)})] \\ & \leq \sum_{i=1}^m \frac{1}{\alpha_i} \mathbb{E}_M [(\cosh - 1)_*(|f_i|^{\alpha_i})] + \frac{1}{\beta} \mathbb{E}_M [(\cosh - 1)_*(e^{\beta(U - K_M(U))})]. \end{aligned}$$

Since $f_1, \dots, f_m \in L^{(\cosh - 1)}(M) \subset \cap_{\alpha > 1} L^\alpha(M)$, one has $|f_i|^{\alpha_i} \in L^{(\cosh - 1)_*}(M)$, for $i = 1, \dots, m$ and all $\alpha_i > 1$, hence $\mathbb{E}_M [(\cosh - 1)_*(|f_i|^{\alpha_i})] < \infty$ for $i = 1, \dots, m$ and all $\alpha_i > 1$. By choosing $1 < \beta < 1 + \varepsilon$ one has $e^{\beta(U(x) - K_M(U))} \in L^\gamma(M) \subset L^{(\cosh - 1)_*}(M)$, $\gamma = \frac{1+\varepsilon}{\beta}$, so that $\mathbb{E}_M [(\cosh - 1)_*(e^{\beta(U - K_M(U))})] < \infty$. This proves that $(\cosh - 1)_*(f_1 \cdots f_m e^{U - K_M(U)}) \in L^1(M)$, which implies $f_1 \cdots f_m e^{U(x) - K_M(U)} \in L^{(\cosh - 1)_*}(M)$.

2. From the previous item we know $f e^{U - K_M(U)} \in L^{(\cosh - 1)_*}(M)$. For each $j = 1, \dots, n$ from Proposition 20(3) we have the distributional derivative $\partial_j(f e^U) = \partial_j f e^U + f \partial_j e^{U - K_M(U)}$ we need to show a composite function derivation, namely $\partial_j e^{U - K_M(U)} = \partial_j u e^{U - K_M(U)}$. Let $\chi \in C_0^\infty(\mathbb{R}^n)$ be a cut-off equal to 1 on the ball of radius 1, zero outside the ball of radius 2, derivative bounded by 2, and for $n \in \mathbb{N}$ consider the function $x \mapsto F_n(x) = \chi(x/n) e^x$ which is $C^\infty(\mathbb{R}^n)$ and whose derivative is bounded:

$$F'_n(x) = \left(\frac{1}{n} \chi'(x/n) + \chi(x/n) \right) e^x \leq \left(\frac{2}{n} + 1 \right) e^{2n}.$$

As Proposition 20(4) applies, we have $\partial_j F_n(U) = F'_n(U) \partial_j U \in C_0^{(\cosh - 1)}(M)$. Finally, for each $\phi \in C_0^\infty(\mathbb{R}^n)$,

$$\begin{aligned} \langle \partial_j e^U, \phi \rangle &= - \langle e^U, \partial_j \phi \rangle \\ &= - \lim_{n \rightarrow \infty} \langle F_n(U), \partial_j \phi \rangle \\ &= \lim_{n \rightarrow \infty} \langle \partial F_n(U), \phi \rangle \\ &= \lim_{n \rightarrow \infty} \left\langle \left(\frac{1}{n} \chi'(U/n) + \chi(U/n) \right) \partial_j U e^U, \phi \right\rangle \\ &= \langle \partial_j U e^U, \phi \rangle. \end{aligned}$$

Remark 4 As a particular case of the above proposition, we see that $U \in \mathcal{S}_M \cap C_0^{1,(\cosh - 1)}(M)$ implies

$$e^{U - K_M(U)} \in W^{1,(\cosh - 1)_*}(M) \quad \text{with} \quad \nabla e^{U - K_M(U)} = \nabla u e^{U - K_M(U)}.$$

6 Conclusions

In this paper we have given a self-contained expositions of the Exponential Affine Manifold on the Gaussian space. The Gaussian assumption allows to discuss topics that are not available in the general case, where the geometry of the sample space has no role.

In particular, we have focused on the action of translations on the probability densities of the manifold and on properties of their derivatives. Other related results, such as Poincaré-type inequalities, have been discussed.

Intended applications are those already discussed in [10], in particular Hyvärinen divergence and other statistical divergences involving derivatives, together with their gradient flows.

Acknowledgements The author acknowledges support from the de Castro Statistics Initiative, Collegio Carlo Alberto, Moncalieri, Italy. He is a member of INdAM/GNAMPA.

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Pure and Applied Mathematics (Amsterdam), vol. 140, 2nd edn. Elsevier/Academic Press, Amsterdam (2003)
2. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext. Springer, New York (2011)
3. Brigo, D., Hanzon, B., Le Gland, F.: Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli* **5**(3), 495–534 (1999)
4. Brigo, D., Pistone, G.: Optimal approximations of the Fokker-Planck-Kolmogorov equation: projection, maximum likelihood, eigenfunctions and Galerkin methods (2017). [arXiv:1603.04348v2](https://arxiv.org/abs/1603.04348v2)
5. Brigo, D., Pistone, G.: Projection based dimensionality reduction for measure valued evolution equations in statistical manifolds. In: Nielsen, F., Critchley, F., Dodson, C. (eds.) Computational Information Geometry. For Image and Signal Processing. Signals and Communication Technology, pp. 217–265. Springer, Berlin (2017)
6. Cena, A., Pistone, G.: Exponential statistical manifold. *Ann. Inst. Stat. Math.* **59**(1), 27–56 (2007). <https://doi.org/10.1007/s10463-006-0096-y>
7. Dellacherie, C., Meyer P.A.: Probabilités et potentiel: Chapitres I à IV, édition entièrement refondue. Hermann, Paris (1975)
8. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **1**(2), 325–347 (1998). <https://doi.org/10.1142/S021902579800017X>
9. Hyvärinen, A.: Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.* **6**, 695–709 (2005)
10. Lods, B., Pistone, G.: Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy* **17**(6), 4323–4363 (2015). <https://doi.org/10.3390/e17064323>
11. Malliavin, P.: Integration and Probability. Graduate Texts in Mathematics, vol. 157. Springer, New York (1995). <https://doi.org/10.1007/978-1-4612-4202-4>. (with the collaboration of Hélène Airault, Leslie Kay and Gérard Letac, edited and translated from the French by Kay, with a foreword by Mark Pinsky)
12. Musielak, J.: Orlicz Spaces and Modular Spaces. Lecture Notes in Mathematics, vol. 1034. Springer, Berlin (1983)

13. Nourdin, I., Peccati, G.: Normal Approximations with Malliavin Calculus. Cambridge Tracts in Mathematics, vol. 192. Cambridge University Press, Cambridge (2012). <https://doi.org/10.1017/CBO9781139084659>. (from Stein's method to universality)
14. Nualart, D.: The Malliavin Calculus and Related Topics. Probability and its Applications (New York), 2nd edn. Springer, Berlin (2006)
15. Pistone, G.: Examples of the application of nonparametric information geometry to statistical physics. *Entropy* **15**(10), 4042–4065 (2013). <https://doi.org/10.3390/e15104042>
16. Pistone, G.: Nonparametric information geometry. Geometric Science of Information. Lecture Notes in Computer Science, vol. 8085, pp. 5–36. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40020-9_3
17. Pistone, G.: Translations in the exponential Orlicz space with Gaussian weight. In: Nielsen, F., Barbaresco, F. (eds.) Geometric Science of Information. Third International Conference, GSI 2017, Paris, France, November 7–9, 2017, Proceedings. Lecture Notes in Computer Science, vol. 10589, pp. 569–576. Springer, Berlin (2017)
18. Pistone, G., Rogantin, M.P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli* **5**(4), 721–760 (1999). <https://doi.org/10.2307/3318699>
19. Pistone, G., Sempì, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **23**(5), 1543–1561 (1995). <https://doi.org/10.1214/aos/1176324311>
20. Santacroce, M., Siri, P., Trivellato, B.: New results on mixture and exponential models by Orlicz spaces. *Bernoulli* **22**(3), 1431–1447 (2016). <https://doi.org/10.3150/15-BEJ698>
21. Stroock, D.W.: Partial Differential Equations for Probabilists. Cambridge Studies in Advanced Mathematics, vol. 112. Cambridge University Press, Cambridge (2008). <https://doi.org/10.1017/CBO9780511755255>
22. Villani, C.: Entropy production and convergence to equilibrium. *Entropy Methods for the Boltzmann Equation. Lecture Notes in Mathematics*, vol. 1916, pp. 1–70. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-73705-6_1

Congruent Families and Invariant Tensors



Lorenz Schwachhöfer, Nihat Ay, Jürgen Jost and Hông Vân Lê

Abstract A classical result of Chentsov states that – up to constant multiples – the only 2-tensor field of a statistical model which is invariant under congruent Markov morphisms is the Fisher metric and the only invariant 3-tensor field is the Amari–Chentsov tensor. We generalize this result for arbitrary degree n , showing that any family of n -tensors which is invariant under congruent Markov morphisms is algebraically generated by the canonical tensor fields defined in Ay, Jost, Lê, Schwachhöfer (Bernoulli, 24:1692–1725, 2018, [4]).

Keywords Chentsov’s theorem · Sufficient statistic · Congruent Markov kernel · Statistical model

2010 Mathematics Subject Classification primary: 62B05, 62B10, 62B86 · secondary: 53C99

L. Schwachhöfer (✉)
TU Dortmund University, Dortmund, Germany
e-mail: lschwach@math.tu-dortmund.de

N. Ay · J. Jost
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany
e-mail: nay@mis.mpg.de

J. Jost
e-mail: jjost@mis.mpg.de

H. Vân Lê
Academy of Sciences of the Czech Republic, Prague, Czech Republic
e-mail: hvle@math.cas.cz

1 Introduction

The main task of *Information geometry* is to use differential geometric methods in probability theory in order to gain insight into the structure of families of probability measures or, slightly more general, finite measures on some (finite or infinite) sample space Ω . In fact, one of the key themes of differential geometry is to identify quantities that do not depend on how we parametrize our objects, but that depend only on their intrinsic structure. And since in information geometry, we not only have the structure of the parameter space, the classical object of differential geometry, but also the sample space on which the probability measures live, we should also look at invariance properties with respect to the latter. That is what we shall systematically do in this contribution.

When parametrizing such a family by a manifold M , there are two classically known symmetric tensor fields on the parameter space M . The first is a quadratic form (i.e., a Riemannian metric), called the *Fisher metric* \mathbf{g}^F , and the second is a 3-tensor, called the *Amari–Chentsov tensor* \mathbf{T}^{AC} . The Fisher metric was first suggested by Rao [18], followed by Jeffreys [14], Efron [13] and then systematically developed by Chentsov and Morozova [9, 10] and [17]; the Amari–Chentsov tensor and its significance was discovered by Amari [1, 2] and Chentsov [11]. If the family is given by a positive density function $\mathbf{p}(\xi) = p(\cdot; \xi)\mu$ w.r.t. some fixed background measure μ on Ω and $p : \Omega \times M \rightarrow (0, \infty)$ differentiable in the ξ -direction, then the score

$$\int_{\Omega} \partial_V \log p(\cdot; \xi) d\mathbf{p}(\xi) = 0 \quad (1)$$

vanishes, while the Fisher metric \mathbf{g}^F and the Amari–Chentsov tensor \mathbf{T}^{AC} associated to a parametrized measure model are given by

$$\begin{aligned} \mathbf{g}^F(V, W) &:= \int_{\Omega} \partial_V \log p(\cdot; \xi) \partial_W \log p(\cdot; \xi) d\mathbf{p}(\xi) \\ \mathbf{T}^{AC}(V, W, U) &:= \int_{\Omega} \partial_V \log p(\cdot; \xi) \partial_W \log p(\cdot; \xi) \partial_U \log p(\cdot; \xi) d\mathbf{p}(\xi). \end{aligned} \quad (2)$$

Of course, this naturally suggests to consider analogous tensors for arbitrary degree n . The tensor fields in (2) have some remarkable properties. On the one hand, they may be defined independently of the particular choice of a parametrization and thus are naturally defined from the differential geometric point of view. Their most important property from the point of view of statistics is that these tensors are invariant under sufficient statistics or, more general under congruent Markov morphisms. In fact, these tensor fields are characterized by this invariance property. This was shown in the case of finite sample spaces by Chentsov in [10] and for an arbitrary sample space by the authors of the present article in [3].

The question addressed in this article is to classify all tensor fields which are invariant under sufficient statistics and congruent Markov morphisms. In order to do this, we first have to make this invariance condition precise.

Observe that both [3, 10] require the family to be of the form $\mathbf{p}(\xi) = p(\cdot; \xi)\mu$ with $p > 0$, which in particular implies that all these measures are equivalent, i.e., have the same null sets. Later, in [4, 5], the authors of this article introduced a more general notion of a *parametrized measure model* as a map $\mathbf{p} : M \rightarrow \mathcal{M}(\Omega)$ from a (finite or infinite dimensional) manifold M into the space $\mathcal{M}(\Omega)$ of finite measures which is continuously Fréchet-differentiable when regarded as a map into the Banach lattice $\mathcal{S}(\Omega) \supset \mathcal{M}(\Omega)$ of *signed* finite measures. Such a model neither requires the existence of a measure dominating all measures $\mathbf{p}(\xi)$, nor does it require all these measures to be equivalent.

Furthermore, for each $r \in (0, 1]$ there is a well defined Banach lattice $\mathcal{S}^r(\Omega)$ of r -th powers of finite signed measures, whose nonnegative elements are denoted by $\mathcal{M}^r(\Omega) \subset \mathcal{S}^r(\Omega)$, and for each integer $n \in \mathbb{N}$, there is a *canonical n-tensor* on $\mathcal{S}^{1/n}(\Omega)$ given by

$$L_n^\Omega(\nu_1, \dots, \nu_n) := n^n (\nu_1 \cdots \nu_n)(\Omega), \quad (3)$$

where $\nu_1 \cdots \nu_n \in \mathcal{S}(\Omega)$ is a signed measure. The multiplication on the right hand side of (3) refers to the multiplication of roots of measures, cf. [4, (2.11)], see also (6). A parametrized measure model $\mathbf{p} : M \rightarrow \mathcal{M}(\Omega)$ is called *k-integrable* for $k \geq 1$ if the map

$$\mathbf{p}^{1/k} : M \longrightarrow \mathcal{M}^{1/k}(\Omega) \subset \mathcal{S}^{1/k}(\Omega), \quad \xi \longmapsto \mathbf{p}(\xi)^{1/k}$$

is continuously Fréchet differentiable, cf. [4, Definition 4.4]. In this case, we define the *canonical n-tensor of the model* as the pull-back $\tau_{(M, \Omega, \mathbf{p})}^n := (\mathbf{p}^{1/n})^* L_n^\Omega$ for all $n \leq k$. If the model is of the form $\mathbf{p}(\xi) = p(\cdot; \xi)\mu$ with a positive density function $p > 0$, then

$$\tau_{(M, \Omega, \mathbf{p})}^n(V_1, \dots, V_n) := \int_{\Omega} \partial_{V_1} \log p(\cdot; \xi) \cdots \partial_{V_n} \log p(\cdot; \xi) d\mathbf{p}(\xi), \quad (4)$$

so that $\mathbf{g}^F = \tau_{(M, \Omega, \mathbf{p})}^2$ and $\mathbf{T}^{AC} = \tau_{(M, \Omega, \mathbf{p})}^3$ by (2). The condition of *k*-integrability ensures that the integral in (4) exists for $n \leq k$.

A Markov kernel $K : \Omega \rightarrow \mathcal{P}(\Omega')$ induces a bounded linear map $K_* : \mathcal{S}(\Omega) \rightarrow \mathcal{S}(\Omega')$, called the *Markov morphism associated to K*. This Markov kernel is called *congruent*, if there is a statistic $\kappa : \Omega' \rightarrow \Omega$ such that $\kappa_* K_* \mu = \mu$ for all $\mu \in \mathcal{S}(\Omega)$.

We may associate to K the map $K_r : \mathcal{S}^r(\Omega) \rightarrow \mathcal{S}^r(\Omega')$ by $K_r(\mu_r) = (K_*(\mu_r^{1/r}))^r$, where $\mu_r^{1/r} \in \mathcal{S}(\Omega)$. While K_r is not Fréchet differentiable in general, we still can define in a natural way the *formal differential* dK_r and hence the pullback $K_r^* \Theta_{\Omega; r}^n$ for any covariant n -tensor on $\mathcal{S}^r(\Omega')$ which yields a covariant n -tensor on $\mathcal{S}^r(\Omega)$.

It is not hard to show that for the canonical tensor fields we have the identity $K_{1/n}^* L_n^\Omega = L_n^\Omega$ for any congruent Markov kernel $K : \Omega \rightarrow \mathcal{P}(\Omega')$, whence we may say that the canonical n -tensors L_n^Ω on $\mathcal{S}^{1/n}(\Omega)$ form a *congruent family*. Evidently, any tensor field which is given by linear combinations of tensor products of canonical

tensors and permutations of the argument is also a congruent family, and the families of this type are said to be *algebraically generated* by L_Ω^n .

Our main result is that these exhaust the possible invariant families of covariant tensor fields:

Theorem 1.1 *Let $(\Theta_{\Omega;r}^n)$ be a family of covariant n -tensors on $\mathcal{S}^r(\Omega)$ for each measurable space Ω . Then this family is invariant under congruent Markov morphisms if and only if it is algebraically generated by the canonical tensors L_m^Ω with $m \leq 1/r$.*

In particular, on each k -integrable parametrized measure model (M, Ω, \mathbf{p}) any tensor field which is invariant under congruent Markov morphisms is algebraically generated by the canonical tensor fields $\tau_{(M, \Omega, \mathbf{p})}^m$, $m \leq k$.

We shall show that this conclusion already holds if the family is invariant under congruent Markov morphisms $K : I \rightarrow \mathcal{P}(\Omega)$ with finite I . Also, observe that the theorems of Campbell [8] on invariant 2-tensors and of Chentsov [11, Theorem 11.1] on invariant 2- and 3-tensors on families of probability measures are special cases of Theorem 1.1.

Campbell's theorem on invariant 2-tensors as well as [3, Theorem 2.10] on invariant 3-tensors cover the case where the measures no longer need to be probability measures. In such a situation, the analogue of the score (1) no longer needs to vanish, and it provides a nontrivial 1-tensor.

Let us comment on the relation of our results to those of Bauer et al. [6, 7]. Assuming that the sample space Ω is a manifold (with boundary or even with corners), the space $\text{Dens}_+(\Omega)$ of (*smooth*) densities on Ω is defined as the set of all measures of the form $\mu = f \text{vol}_g$, where $f > 0$ is a smooth function with finite integral, vol_g being the volume form of some Riemannian metric g on M . Thus, $\text{Dens}_+(\Omega)$ is a Fréchet manifold, and regarding a diffeomorphism $K : \Omega \rightarrow \Omega$ as a congruent statistic, the induced maps $K_r : \text{Dens}_+(\Omega)^r \rightarrow \text{Dens}_+(\Omega)^r$ are diffeomorphisms of Fréchet manifolds. The main result in [6] states that for $\dim \Omega \geq 2$ any 2-tensor field which is invariant under diffeomorphisms is a multiple of the Fisher metric. Likewise, the space of diffeomorphism invariant n -tensors for arbitrary n [7] is generated by the canonical tensors. Thus, when restricting to parametrized measure models $\mathbf{p} : M \rightarrow \text{Dens}_+(\Omega) \subset \mathcal{M}(\Omega)$ whose image lies in the space of densities and which are differentiable w.r.t. the Fréchet manifold structure on $\text{Dens}_+(\Omega)$, then the invariance of a tensor field under diffeomorphisms rather than under arbitrary congruent Markov morphisms already implies that the tensor field is algebraically generated by the canonical tensors. Considering invariance under diffeomorphisms is natural in the sense that they can be regarded as the natural analogues of permutations of a finite sample space. In our more general setting, however, the concept of a diffeomorphism is no longer meaningful, and we need to consider invariance under a larger class of transformations, the congruent Markov morphisms.

We would also like to mention the work of J. Dowty [12] who showed a version of the Chentsov theorem for exponential families using the central limit theorem.

This paper is structured as follows. In Sect. 2 we recall from [4] the definition of a parametrized measure model, roots of measures and congruent Markov kernels, and

furthermore we give an explicit description of the space of covariant families which are algebraically generated by the canonical tensors. In Sect. 3 we recall the notion of congruent families of tensor fields and show that the canonical tensors and hence tensors which are algebraically generated by these are congruent. Then we show that these exhaust all invariant families of tensor fields on *finite* sample spaces Ω in Sect. 4, and finally, in Sect. 5, by reducing the general case to the finite case through step function approximations, we obtain the classification result Theorem 5.1 which implies Theorem 1.1 as a simplified version.

This paper is mainly an elaboration of parts of the monograph [5] where a more comprehensive treatment of this result is presented.

2 Preliminary Results

2.1 The Space of (Signed) Finite Measures and Their Powers

Let (Ω, Σ) be a measurable space, that is an arbitrary set Ω together with a sigma algebra Σ of subsets of Ω . Regarding the sigma algebra Σ on Ω as fixed, we let

$$\begin{aligned}\mathcal{P}(\Omega) &:= \{\mu : \mu \text{ a probability measure on } \Omega\} \\ \mathcal{M}(\Omega) &:= \{\mu : \mu \text{ a finite measure on } \Omega\} \\ \mathcal{S}(\Omega) &:= \{\mu : \mu \text{ a signed finite measure on } \Omega\} \\ \mathcal{S}_a(\Omega) &:= \{\mu \in \mathcal{S}(\Omega) : \int_{\Omega} d\mu = a\}.\end{aligned}\tag{5}$$

Clearly, $\mathcal{P}(\Omega) \subset \mathcal{M}(\Omega) \subset \mathcal{S}(\Omega)$, and $\mathcal{S}_0(\Omega), \mathcal{S}(\Omega)$ are real vector spaces, whereas $\mathcal{S}_a(\Omega)$ is an affine space with linear part $\mathcal{S}_0(\Omega)$. In fact, both $\mathcal{S}_0(\Omega)$ and $\mathcal{S}(\Omega)$ are Banach spaces whose norm is given by the total variation of a signed measure, defined as

$$\|\mu\| := \sup \sum_{i=1}^n |\mu(A_i)|$$

where the supremum is taken over all finite partitions $\Omega = A_1 \dot{\cup} \dots \dot{\cup} A_n$ with disjoint sets $A_i \in \Sigma$. Here, the symbol $\dot{\cup}$ stands for the disjoint union of sets. In particular,

$$\|\mu\| = \mu(\Omega) \quad \text{for } \mu \in \mathcal{M}(\Omega).$$

In [4], for each $r \in (0, 1]$ the space $\mathcal{S}^r(\Omega)$ of *r-th powers of measures on Ω* is defined. We shall not repeat the formal definition here, but we recall the most important features of these spaces.

Each $\mathcal{S}^r(\Omega)$ is a Banach lattice whose norm we denote by $\|\cdot\|_{\mathcal{S}^r(\Omega)}$, and $\mathcal{M}^r(\Omega) \subset \mathcal{S}^r(\Omega)$ denotes the spaces of nonnegative elements. Moreover, $\mathcal{S}^1(\Omega) = \mathcal{S}(\Omega)$ in a canonical way. For $r, s, r+s \in (0, 1]$ there is a bilinear product

$$\cdot : \mathcal{S}^r(\Omega) \times \mathcal{S}^s(\Omega) \longrightarrow \mathcal{S}^{r+s}(\Omega) \quad \text{such that} \quad \|\mu_r \cdot \mu_s\|_{\mathcal{S}^{r+s}(\Omega)} \leq \|\mu_r\|_{\mathcal{S}^r(\Omega)} \|\mu_s\|_{\mathcal{S}^s(\Omega)}, \quad (6)$$

and for $0 < k < 1/r$ there is a exponentiating map $\pi^k : \mathcal{S}^r(\Omega) \rightarrow \mathcal{S}^{kr}(\Omega)$ which is continuous for $k < 1$ and a Fréchet- C^1 -map for $k \geq 1$.

In order to understand these objects more concretely, let $\mu \in \mathcal{M}(\Omega)$ be a measure, so that $\mu^r := \pi^r(\mu) \in \mathcal{S}^r(\Omega)$. Then for all $\phi \in L^{1/r}(\Omega, \mu)$ we have $\phi\mu^r \in \mathcal{S}^r(\Omega)$, and $\phi\mu^r \in \mathcal{M}^r(\Omega)$ if and only if $\phi \geq 0$. The inclusion

$$\mathcal{S}^r(\Omega; \mu) := \{\phi\mu^r \mid \phi \in L^{1/r}(\Omega, \mu)\} \hookrightarrow \mathcal{S}^r(\Omega)$$

is an isometric inclusion of Banach spaces, and the elements of $\mathcal{S}^r(\Omega, \mu)$ are said to be *dominated by μ* . We also define

$$\mathcal{S}_0^r(\Omega; \mu) := \{\phi\mu^r \mid \phi \in L^{1/r}(\Omega, \mu), \mathbb{E}_\mu(\phi) = 0\} \subset \mathcal{S}^r(\Omega; \mu).$$

Moreover,

$$(\phi\mu^r) \cdot (\psi\mu^s) = (\phi\psi)\mu^{r+s}, \quad \pi^k(\phi\mu^r) := \text{sign}(\phi)|\phi|^k\mu^{rk}, \quad (7)$$

where $\phi \in L^{1/r}(\Omega, \mu)$ and $\psi \in L^{1/s}(\Omega, \mu)$. The Fréchet derivative of π^k at $\mu_r \in \mathcal{S}^r(\Omega)$ is given by

$$d_{\mu_r}\pi^k(\nu_r) = k |\mu|^{k-1} \cdot \nu_r. \quad (8)$$

Furthermore, for an integer $n \in \mathbb{N}$, we have the *canonical n-tensor on $\mathcal{S}^{1/n}(\Omega)$* , given by

$$L_\Omega^n(\mu_1, \dots, \mu_n) := n^n \int_\Omega d(\mu_1 \cdots \mu_n) \quad \text{for } \mu_i \in \mathcal{S}^{1/n}(\Omega), \quad (9)$$

which is a symmetric n -multilinear form, where we regard the product $\mu_1 \cdots \mu_n$ as an element of $\mathcal{S}^1(\Omega) = \mathcal{S}(\Omega)$. For instance, for $n = 2$ the bilinear form

$$\langle \cdot; \cdot \rangle := \frac{1}{4} L_\Omega^2(\cdot, \cdot)$$

equips $\mathcal{S}^{1/2}(\Omega)$ with a Hilbert space structure with induced norm $\|\cdot\|_{\mathcal{S}^{1/2}(\Omega)}$.

2.2 Parametrized Measure Models

Recall from [4] that a parametrized measure model is a triple (M, Ω, \mathbf{p}) consisting of a (finite or infinite dimensional) manifold M and a map $\mathbf{p} : M \rightarrow \mathcal{M}(\Omega)$ which is Fréchet-differentiable when regarded as a map into $\mathcal{S}(\Omega)$ (cf. [4, Definition 4.1]). If $\mathbf{p}(\xi) \in \mathcal{P}(\Omega)$ for all $\xi \in M$, then (M, Ω, \mathbf{p}) is called a *statistical model*. Moreover, (M, Ω, \mathbf{p}) is called *k-integrable*, if $\mathbf{p}^{1/k} : M \rightarrow \mathcal{S}^{1/k}(\Omega)$ is also Fréchet inte-

grable (cf. [15, Definition 2.6]). For a parametrized measure model, the differential $d_\xi \mathbf{p}(v) \in \mathcal{S}(\Omega)$ with $v \in T_\xi M$ is always dominated by $\mathbf{p}(\xi) \in \mathcal{M}(\Omega)$, and we define the *logarithmic derivative* (cf. [4, Definition 4.3]) as the Radon–Nikodym derivative

$$\partial_v \log \mathbf{p}(\xi) := \frac{d\{d_\xi \mathbf{p}(v)\}}{d\mathbf{p}(\xi)} \in L^1(\Omega, \mathbf{p}(\xi)). \quad (10)$$

Then \mathbf{p} is k -integrable if and only if $\partial_v \log \mathbf{p} \in L^k(\Omega, \mathbf{p}(\xi))$ for all $v \in T_\xi M$, and the function $v \mapsto \|\partial_v \log \mathbf{p}\|_{\mathbf{p}(\xi)}$ on TM is continuous (cf. [15, Theorem 2.7]). In this case, the Fréchet derivative of $\mathbf{p}^{1/k}$ is given as

$$d_\xi \mathbf{p}^{1/k}(v) = \frac{1}{k} \partial_v \log \mathbf{p}(\xi) \mathbf{p}^{1/k}. \quad (11)$$

2.3 Congruent Markov Morphisms

Definition 2.1 A *Markov kernel* between two measurable spaces (Ω, \mathfrak{B}) and (Ω', \mathfrak{B}') is a map $K : \Omega \rightarrow \mathcal{P}(\Omega')$ associating to each $\omega \in \Omega$ a probability measure on Ω' such that for each fixed measurable $A' \subset \Omega'$ the map

$$\Omega \longrightarrow [0, 1], \quad \omega \longmapsto K(\omega)(A') =: K(\omega; A')$$

is measurable for all $A' \in \mathfrak{B}'$. The linear map

$$K_* : \mathcal{S}(\Omega) \longrightarrow \mathcal{S}(\Omega'), \quad K_* \mu(A') := \int_{\Omega} K(\omega; A') d\mu(\omega) \quad (12)$$

is called the *Markov morphism induced by K* .

Evidently, a Markov morphism maps $\mathcal{M}(\Omega)$ to $\mathcal{M}(\Omega')$, and

$$\|K_* \mu\| = \|\mu\| \quad \text{for all } \mu \in \mathcal{M}(\Omega), \quad (13)$$

so that K_* also maps $\mathcal{P}(\Omega)$ to $\mathcal{P}(\Omega')$. For any $\mu \in \mathcal{S}(\Omega)$, $\|K_* \mu\| \leq \|\mu\|$, whence K_* is bounded.

Example 2.1 A measurable map $\kappa : \Omega \rightarrow \Omega'$, called a *statistic*, induces a Markov kernel by setting $K^\kappa(\omega) := \delta_{\kappa\omega} \in \mathcal{P}(\Omega')$. In this case,

$$K_*^\kappa \mu(A') = \int_{\Omega} K^\kappa(\omega; A') d\mu(\omega) = \int_{\kappa^{-1}(A')} d\mu = \mu(\kappa^{-1} A') = \kappa_* \mu(A'),$$

whence $K_*^\kappa \mu = \kappa_* \mu$ is the push-forward of (signed) measures on Ω to (signed) measures on Ω' .

Definition 2.2 A Markov kernel $K : \Omega \rightarrow \mathcal{P}(\Omega')$ is called *congruent w.r.t. to the statistic* $\kappa : \Omega' \rightarrow \Omega$ if

$$\kappa_* K(\omega) = \delta_\omega \quad \text{for all } \omega \in \Omega,$$

or, equivalently, if K_* is a right inverse of κ_* , i.e., $\kappa_* K_* = \text{Id}_{\mathcal{S}(\Omega)}$. It is called *congruent* if it is congruent w.r.t. some statistic $\kappa : \Omega' \rightarrow \Omega$.

This notion was introduced by Chentsov in the case of finite sample spaces [11], but the natural generalization in Definition 2.2 to arbitrary sample spaces has been treated in [3, 4, 16].

Example 2.2 A statistic $\kappa : \Omega \rightarrow I$ between finite sets induces a partition

$$\Omega = \bigcup_{i \in I} \Omega_i, \quad \text{where} \quad \Omega_i = \kappa^{-1}(i).$$

In this case, a Markov kernel $K : I \rightarrow \mathcal{P}(\Omega)$ is κ -congruent if and only if

$$K(i)(\Omega_j) = K(i; \Omega_j) = 0 \quad \text{for all } i \neq j \in I.$$

If (M, Ω, \mathbf{p}) is a parametrized measure model and $K : \Omega \rightarrow \mathcal{P}(\Omega')$ a Markov kernel, then $(M, \mathbf{p}', \Omega')$ with $\mathbf{p}' := K_* \mathbf{p} : M \rightarrow \mathcal{M}(\Omega') \subset \mathcal{S}(\Omega')$ is again a parametrized measure model. In this case, we have the following result.

Proposition 2.1 ([4, Theorem 3.3]) *Let $K_* : \mathcal{S}(\Omega) \rightarrow \mathcal{S}(\Omega')$ be a Markov morphism induced by the Markov kernel $K : \Omega \rightarrow \mathcal{P}(\Omega')$, let $\mathbf{p} : M \rightarrow \mathcal{M}(\Omega)$ be a k -integrable parametrized measure model and $\mathbf{p}' := K_* \mathbf{p} : M \rightarrow \mathcal{M}(\Omega')$. Then \mathbf{p}' is also k -integrable, and*

$$\|\partial_v \log \mathbf{p}'(\xi)\|_{L^k(\Omega', \mathbf{p}'(\xi))} \leq \|\partial_v \log \mathbf{p}(\xi)\|_{L^k(\Omega, \mathbf{p}(\xi))}. \quad (14)$$

2.4 Tensor Algebras

In this section we shall provide the algebraic background on tensor algebras. Let V be a vector space over a commutative field \mathbb{F} , and let V^* be its dual. The *tensor algebra of V^** is defined as

$$\mathbf{T}(V^*) := \bigoplus_{n=0}^{\infty} \otimes^n V^*,$$

where

$$\otimes^n V^* = \{\tau^n : \underbrace{V \times \cdots \times V}_{n \text{ times}} \longrightarrow \mathbb{F} \mid \tau^n \text{ is } n\text{-multilinear}\}.$$

In particular, $\otimes^0 V^* := \mathbb{F}$ and $\otimes^1 V^* := V^*$. Then $\mathbf{T}(V^*)$ is a graded associative unital algebra, where the product $\otimes : \otimes^n V^* \times \otimes^m V^* \rightarrow \otimes^{n+m} V^*$ is defined as

$$(\tau_1^n \otimes \tau_2^m)(v_1, \dots, v_{n+m}) := \tau_1^n(v_1, \dots, v_n) \cdot \tau_2^m(v_{n+1}, \dots, v_{n+m}). \quad (15)$$

By convention, the multiplication with elements of $\otimes^0 V^* = \mathbb{F}$ is the scalar multiplication, so that $1 \in \mathbb{F}$ is the unit of $\mathbf{T}(V^*)$. Observe that $\mathbf{T}(V^*)$ is non-commutative.

There is a linear action of S_n , the permutation group of n elements, on $\otimes^n V^*$ given by

$$(P_\sigma \tau^n)(v_1, \dots, v_n) := \tau^n(v_{\sigma^{-1}(1)}, \dots, v_{\sigma^{-1}(n)}) \quad (16)$$

for $\sigma \in S_n$ and $\tau^n \in \otimes^n V^*$. Indeed, the identity $P_{\sigma_1}(P_{\sigma_2} \tau^n) = P_{\sigma_1 \sigma_2} \tau^n$ is easily verified. We call a tensor $\tau^n \in \otimes^n V^*$ *symmetric*, if $P_\sigma \tau^n = \tau^n$ for all $\sigma \in S_n$, and we let

$$\odot^n V^* := \{\tau^n \in \otimes^n V^* \mid \tau^n \text{ is symmetric}\}$$

the *n-fold symmetric power of V^** . Evidently, $\odot^n V^* \subset \otimes^n V^*$ is a linear subspace.

A *unital subalgebra* of $\mathbf{T}(V^*)$ is a linear subspace $\mathcal{A} \subset \mathbf{T}(V^*)$ containing $\mathbb{F} = \otimes^0 V^*$ which is closed under tensor products, i.e. such that $\tau_1, \tau_2 \in \mathcal{A}$ implies that $\tau_1 \otimes \tau_2 \in \mathcal{A}$. We call such a subalgebra *graded* if

$$\mathcal{A} = \bigoplus_{n=0}^{\infty} \mathcal{A}_n \quad \text{with } \mathcal{A}_n := \mathcal{A} \cap \otimes^n V^*,$$

and a graded subalgebra $\mathcal{A} \subset \mathbf{T}(V)$ is called *permutation invariant* if \mathcal{A}_n is preserved by the action of S_n on $\mathcal{A}_n \subset \otimes^n V^*$.

Definition 2.3 Let $\mathcal{S} \subset \mathbf{T}(V^*)$ be an arbitrary subset. The intersection of all permutation invariant unital subalgebras of $\mathbf{T}(V^*)$ containing \mathcal{S} is called the *permutation invariant subalgebra generated by \mathcal{S}* and is denoted by $\mathcal{A}_{\text{perm}}(\mathcal{S})$.

Observe that $\mathcal{A}_{\text{perm}}(\mathcal{S})$ is the smallest permutation invariant unital subalgebra of $\mathbf{T}(V^*)$ which contains \mathcal{S} .

Example 2.3 Evidently, $\mathcal{A}_{\text{perm}}(\emptyset) = \mathbb{F}$. To see another example, let $\tau^1 \in V^*$. If we let $\mathcal{A}_0 := \mathbb{F}$ and $\mathcal{A}_n := \mathbb{F}(\underbrace{\tau^1 \otimes \dots \otimes \tau^1}_{n \text{ times}})$ for $n \geq 1$, then $\mathcal{A}_{\text{perm}}(\tau^1) = \bigoplus_{n=0}^{\infty} \mathcal{A}_n$. In

fact, $\mathcal{A}_{\text{perm}}(\tau^1)$ is even commutative and isomorphic to the algebra of polynomials over \mathbb{F} in one variable.

For $n \in \mathbb{N}$, we denote by **Part**(n) the collection of partitions $\mathbf{P} = \{P_1, \dots, P_r\}$ of $\{1, \dots, n\}$, that is, $\bigcup_k P_k = \{1, \dots, n\}$, and these sets are pairwise disjoint. We denote the number r of sets in the partition by $|\mathbf{P}|$.

Given a partition $\mathbf{P} = \{P_1, \dots, P_r\} \in \text{Part}(n)$, we associate to it a bijective map

$$\pi_{\mathbf{P}} : \biguplus_{i \in \{1, \dots, r\}} (\{i\} \times \{1, \dots, n_i\}) \longrightarrow \{1, \dots, n\}, \quad (17)$$

where $n_i := |P_i|$, such that $\pi_{\mathbf{P}}(\{i\} \times \{1, \dots, n_i\}) = P_i$. This map is well defined, up to permutation of the elements in P_i .

Part(n) is partially ordered by the relation $\mathbf{P} \leq \mathbf{P}'$ if \mathbf{P} is a subdivision of \mathbf{P}' . This ordering has the partition $\{\{1\}, \dots, \{n\}\}$ into singleton sets as its minimum and $\{\{1, \dots, n\}\}$ as its maximum.

Consider now a subset of $\mathbf{T}(V^*)$ of the form

$$\mathcal{S} := \{\tau^1, \tau^2, \tau^3, \dots\} \text{ containing one symmetric tensor } \tau^n \in \odot^n V^* \text{ for each } n \in \mathbb{N}. \quad (18)$$

For a partition $\mathbf{P} \in \mathbf{Part}(n)$ with the associated map $\pi_{\mathbf{P}}$ from (17) we define $\tau^{\mathbf{P}} \in \otimes^n V^*$ as

$$\tau^{\mathbf{P}}(v_1, \dots, v_n) := \prod_{i=1}^r \tau^{n_i}(v_{\pi_{\mathbf{P}}(i,1)}, \dots, v_{\pi_{\mathbf{P}}(i,n_i)}). \quad (19)$$

Observe that this definition is independent of the choice of the bijection $\pi_{\mathbf{P}}$, since τ^{n_i} is symmetric.

Example 2.4 (1) If $\mathbf{P} = \{\{1, \dots, n\}\}$ is the trivial partition, then

$$\tau^{\mathbf{P}} = \tau^n.$$

(2) If $\mathbf{P} = \{\{1\}, \dots, \{n\}\}$ is the partition into singletons, then

$$\tau^{\mathbf{P}}(v_1, \dots, v_n) = \tau^1(v_1) \cdots \tau^1(v_n).$$

(3) To give a concrete example, let $n = 5$ and $\mathbf{P} = \{\{1, 3\}, \{2, 5\}, \{4\}\}$. Then

$$\tau^{\mathbf{P}}(v_1, \dots, v_5) = \tau^2(v_1, v_3) \cdot \tau^2(v_2, v_5) \cdot \tau^1(v_4).$$

We can now present the main result of this section.

Proposition 2.2 Let $\mathcal{S} \subset \mathbf{T}(V^*)$ be given as in (18). Then the permutation invariant subalgebra generated by \mathcal{S} equals

$$\mathcal{A}_{\text{perm}}(\mathcal{S}) = \mathbb{F} \oplus \bigoplus_{n=1}^{\infty} \text{span} \left\{ \tau^{\mathbf{P}} \mid \mathbf{P} \in \mathbf{Part}(n) \right\}. \quad (20)$$

Proof Let us denote the right hand side of (20) by $\mathcal{A}'_{\text{perm}}(\mathcal{S})$, so that we wish to show that $\mathcal{A}_{\text{perm}}(\mathcal{S}) = \mathcal{A}'_{\text{perm}}(\mathcal{S})$.

By Example 2.4.(1), $\tau^n \in \mathcal{A}'_{\text{perm}}(\mathcal{S})$ for all $n \in \mathbb{N}$, whence $\mathcal{S} \subset \mathcal{A}'_{\text{perm}}(\mathcal{S})$. Furthermore, by (19) we have

$$\tau^{\mathbf{P}} \otimes \tau^{\mathbf{P}'} = \tau^{\mathbf{P} \cup \mathbf{P}'},$$

where $\mathbf{P} \cup \mathbf{P}' \in \mathbf{Part}(n+m)$ is the partition of $\{1, \dots, n+m\}$ obtained by regarding $\mathbf{P} \in \mathbf{Part}(n)$ and $\mathbf{P}' \in \mathbf{Part}(m)$ as partitions of $\{1, \dots, n\}$ and $\{n+1, \dots, n+m\}$, respectively. Moreover, if $\sigma \in S_n$ is a permutation and $\mathbf{P} = \{P_1, \dots, P_r\}$ a partition, then the definition in (19) implies that

$$P_\sigma(\tau^{\mathbf{P}}) = \tau^{\sigma^{-1}\mathbf{P}}, \quad \text{where } \sigma^{-1}(\{P_1, \dots, P_r\}) := \{\sigma^{-1}P_1, \dots, \sigma^{-1}P_r\}.$$

That is, $\mathcal{A}'_{\text{perm}}(\mathcal{S}) \subset \mathbf{T}(V^*)$ is a permutation invariant unital subalgebra of $\mathbf{T}(V^*)$ contain \mathcal{S} , whence $\mathcal{A}_{\text{perm}}(\mathcal{S}) \subset \mathcal{A}'_{\text{perm}}(\mathcal{S})$.

For the converse, observe that for a partition $\mathbf{P} = \{P_1, \dots, P_r\} \in \mathbf{Part}(n)$, we may – after applying a permutation of $\{1, \dots, n\}$ – assume that

$$P_1 = \{1, \dots, k_1\}, P_2 = \{k_1 + 1, \dots, k_1 + k_2\}, \dots, P_r = \{n - k_r + 1, \dots, n\},$$

with $k_i = |P_i|$, and in this case, (15) and (19) implies that

$$\tau^{\mathbf{P}} = (\tau^{k_1}) \otimes (\tau^{k_2}) \otimes \cdots \otimes (\tau^{k_r}) \in \mathcal{A}_{\text{perm}}(\mathcal{S}),$$

so that any permutation invariant subalgebra containing \mathcal{S} also must contain $\tau^{\mathbf{P}}$ for all partitions, and this shows that $\mathcal{A}'_{\text{perm}}(\mathcal{S}) \subset \mathcal{A}_{\text{perm}}(\mathcal{S})$. \square

2.5 Tensor Fields

Recall that a (*covariant*) n -tensor field¹ Ψ on a manifold M is a collection of n -multilinear forms Ψ_p on $T_p M$ for all $p \in M$ such that for continuous vector fields X^1, \dots, X^n on M the function

$$p \longmapsto \Psi_p(X_p^1, \dots, X_p^n)$$

is continuous. This notion can also be adapted to the case where M has a weaker structure than that of a manifold. The examples we have in mind are the subsets $\mathcal{P}^r(\Omega) \subset \mathcal{M}^r(\Omega)$ of $\mathcal{S}^r(\Omega)$ for an arbitrary measurable space Ω and $r \in (0, 1]$, which fail to be manifolds. Nevertheless, there is a natural notion of *tangent cone at* μ_r of these sets which is the collection of the derivatives of all curves in $\mathcal{M}^r(\Omega)$ (in $\mathcal{P}^r(\Omega)$, respectively) through μ_r . These cones were determined in [4, Proposition 2.1] as

$$T_{\mu^r} \mathcal{M}^r(\Omega) = \mathcal{S}^r(\Omega; \mu) \quad \text{and} \quad T_{\mu^r} \mathcal{P}^r(\Omega) = \mathcal{S}_0^r(\Omega; \mu).$$

¹Since we do not consider non-covariant n -tensor fields in this paper, we shall suppress the attribute *covariant*.

with $\mu \in \mathcal{M}(\Omega)$ ($\mu \in \mathcal{P}(\Omega)$, respectively). Then in analogy to the notion for general manifolds, we can now define the notion of n -tensor fields on $\mathcal{M}^r(\Omega)$ and $\mathcal{P}^r(\Omega)$ as follows.

Definition 2.4 Let Ω be a measurable space and $r \in (0, 1]$. A *vector field on $\mathcal{M}^r(\Omega)$* is a continuous map $X : \mathcal{M}^r(\Omega) \rightarrow \mathcal{S}^r(\Omega)$ such that $X_{\mu^r} \in T_{\mu^r} \mathcal{M}^r(\Omega)$ for all $\mu^r \in \mathcal{M}^r(\Omega)$. The notion of a vector field on $\mathcal{P}^r(\Omega)$ is defined analogously.

A (*covariant*) n -*tensor field on $\mathcal{M}^r(\Omega)$* is a collection of n -multilinear forms Ψ_{μ^r} on $T_{\mu^r} \mathcal{M}^r(\Omega)$ for all $\mu^r \in \mathcal{M}^r(\Omega)$ such that for continuous vector fields X^1, \dots, X^n on $\mathcal{M}^r(\Omega)$ the function

$$\mu^r \longmapsto \Psi_{\mu^r}(X_{\mu^r}^1, \dots, X_{\mu^r}^n)$$

is continuous. The notion of vector fields and n -tensor fields on $\mathcal{P}^r(\Omega)$ is defined analogously.

If Ψ, Ψ' are tensor fields of degree n and m , respectively, and $\sigma \in S_n$ is a permutation, then the pointwise tensor product $\Psi \otimes \Psi'$ and the permutation $P_\sigma \Psi$ defined in (15) and (16) are tensor fields of degree $n+m$ and n , respectively. Moreover, for a differentiable map $f : N \rightarrow M$ the *pull-back of Ψ under f* is the tensor field on N defined by

$$f^* \Psi(v_1, \dots, v_n) := \Psi(df(v_1), \dots, df(v_n)). \quad (21)$$

Evidently, we have

$$f^*(\Psi \otimes \Psi') = (f^*\Psi) \otimes (f^*\Psi') \quad \text{and} \quad P_\sigma(f^*\Psi) = f^*(P_\sigma \Psi). \quad (22)$$

For instance, if (M, Ω, \mathbf{p}) is a k -integrable parametrized measure model, then by (11), $d_\xi \mathbf{p}^{1/k}(v) \in \mathcal{S}^{1/k}(\Omega; \mu) = T_{\mathbf{p}^{1/k}(\xi)} \mathcal{M}^{1/k}(\Omega)$, so that for any n -tensor field Ψ on $\mathcal{M}^{1/k}(\Omega)$ the pull-back

$$(\mathbf{p}^{1/k})^* \Psi(v_1, \dots, v_n) := \Psi(d\mathbf{p}^{1/k}(v_1), \dots, d\mathbf{p}^{1/k}(v_n))$$

is well defined. The same holds if $\mathbf{p} : M \rightarrow \mathcal{P}(\Omega)$ is a statistical model and Ψ is an n -tensor field on $\mathcal{P}^{1/k}(\Omega)$. Moreover, (22) holds in this context as well when replacing f by $\mathbf{p}^{1/k}$.

Definition 2.5 Let Ω be a measurable space, $n \in \mathbb{N}$ an integer and $0 < r \leq 1/n$. Then the *canonical n -tensor field on $\mathcal{S}^r(\Omega)$* is defined as the pull-back

$$\tau_{\Omega';r}^n := (\pi^{1/nr})^* L_\Omega^n \quad (23)$$

with the symmetric n -tensor L_Ω^n on $\mathcal{S}^{1/n}(\Omega)$ defined in (9). The definition of the pullback in (21) and the formula for the Fréchet-derivative of $\pi^{1/nr}$ in (8) now imply by a straightforward calculation that

$$(\tau_{\Omega;r}^n)_{\mu_r}(\nu_1, \dots, \nu_n) := \begin{cases} \frac{1}{r^n} \int_{\Omega} d(\nu_1 \cdot \dots \cdot \nu_n \cdot |\mu_r|^{1/r-n}) & \text{if } r < 1/n, \\ L_{\Omega}^n(\nu_1, \dots, \nu_n) & \text{if } r = 1/n, \end{cases} \quad (24)$$

where $\mu_r \in \mathcal{S}^r(\Omega)$ and $\nu_i \in T_{\mu_r} \mathcal{S}^r(\Omega) \subset \mathcal{S}^r(\Omega)$.

Furthermore, if (M, Ω, \mathbf{p}) is a k -integrable parametrized measure model, $k := 1/r \geq n$, then we define the *canonical n -tensor field* of (M, Ω, \mathbf{p}) as the pull-back

$$\tau_{(M, \Omega, \mathbf{p})}^n := (\mathbf{p}^{1/k})^* \tau_{\Omega;r}^n = (\mathbf{p}^{1/n}) L_{\Omega}^n. \quad (25)$$

In this case, (11) implies that for $v_1, \dots, v_n \in T_{\xi} M$

$$\tau_{(M, \Omega, \mathbf{p})}^n(v_1, \dots, v_n) = \int_{\Omega} \partial_{v_1} \log \mathbf{p}(\xi) \cdots \partial_{v_n} \log \mathbf{p}(\xi) \, d\mathbf{p}(\xi). \quad (26)$$

Example 2.5 (1) The canonical 1-tensor of (M, Ω, \mathbf{p}) is given as

$$(\tau_{(M, \Omega, \mathbf{p})}^1)_{\mu}(v) = \int_{\Omega} \partial_{v_1} \log \mathbf{p}(\xi) \, d\mathbf{p}(\xi) = \partial_v \|\mathbf{p}(\xi)\|.$$

Thus, on a statistical model (i.e., if $\mathbf{p}(\xi) \in \mathcal{P}(\Omega)$ for all ξ) $\tau_{(M, \Omega, \mathbf{p})}^1 \equiv 0$.

- (2) The canonical 2-tensor $\tau_{(M, \Omega, \mathbf{p})}^2$ is called the *Fisher metric* of the model and is often simply denoted by \mathbf{g} . It is defined only if the model is 2-integrable.
- (3) The canonical 3-tensor $\tau_{(M, \Omega, \mathbf{p})}^3$ is called the *Amari–Chentsov tensor* of the model. It is often simply denoted by \mathbf{T} and is defined only if the model is 3-integrable.

3 Congruent Families of Tensor Fields

The question we wish to address in this section is to characterize families of n -tensor fields on $\mathcal{M}^r(\Omega)$ (on $\mathcal{P}^r(\Omega)$, respectively) for measurable spaces Ω which are unchanged under congruent Markov morphisms.

First of all, we need to clarify what is meant by this. The problem we have is that a given Markov kernel $K : \Omega \rightarrow \mathcal{P}(\Omega)$ induces the bounded linear Markov morphism $K_* : \mathcal{S}(\Omega) \rightarrow \mathcal{S}(\Omega')$ which maps $\mathcal{P}(\Omega)$ and $\mathcal{M}(\Omega)$ to $\mathcal{P}(\Omega')$ and $\mathcal{M}(\Omega')$, respectively, but there is no induced differentiable map from $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$ to $\mathcal{P}^r(\Omega')$ and $\mathcal{M}^r(\Omega')$, respectively, if $r < 1$. The best we can do is to make the following definition.

Definition 3.1 Let $K : \Omega \rightarrow \mathcal{P}(\Omega')$ be a Markov kernel with the associated Markov morphism $K_* : \mathcal{S}(\Omega) \rightarrow \mathcal{S}(\Omega')$ from (12). For $r \in (0, 1]$ we define

$$K_r : \mathcal{S}^r(\Omega) \rightarrow \mathcal{S}^r(\Omega'), \quad K_r := \pi^r K_* \pi^{1/r}, \quad (27)$$

which maps $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$ to $\mathcal{P}^r(\Omega')$ and $\mathcal{M}^r(\Omega')$, respectively.

Since $r \leq 1$, it follows that $\pi^{1/r}$ is a Fréchet- C^1 -map and K_* is linear. However, π^r is continuous but not differentiable for $r < 1$, whence the same holds for K_r .

Nevertheless, let us for the moment pretend that K_r was differentiable. Then, when rewriting (27) as $\pi^{1/r} K_r = K_* \pi^{1/r}$, the chain rule and (11) would imply that

$$|K_r \mu_r|^{1/r-1} \cdot (d_{\mu_r} K_r \nu_r) = K_*(|\mu_r|^{1/r-1} \cdot \nu_r) \quad (28)$$

for all $\mu_r, \nu_r \in \mathcal{S}^r(\Omega)$.

On the other hand, as K_r maps $\mathcal{M}^r(\Omega)$ to $\mathcal{M}^r(\Omega')$, its differential at $\mu^r \in \mathcal{M}^r(\Omega)$ for $\mu \in \mathcal{M}(\Omega)$ would restrict to a linear map

$$d_{\mu^r} K_r : T_{\mu^r} \mathcal{M}^r(\Omega) = \mathcal{S}^r(\Omega, \mu) \longrightarrow T_{\mu^r} \mathcal{M}^r(\Omega') = \mathcal{S}^r(\Omega, \mu'),$$

where $\mu' := K_* \mu \in \mathcal{M}(\Omega')$. This together with (28) implies that the restriction of $d_{\mu_r} K_r$ to $\mathcal{S}^r(\Omega, \mu)$ must be given as

$$d_{\mu^r} K_r : \mathcal{S}^r(\Omega, \mu) \longrightarrow \mathcal{S}^r(\Omega', \mu'), \quad d_{\mu^r} K_r(\phi \mu^r) = \frac{d\{K_*(\phi \mu)\}}{d\mu'} \mu'^r. \quad (29)$$

Indeed, by [4, Theorem 3.3], (29) defines a bounded linear map $d_{\mu^r} K_r$. In fact, it is shown in that reference that

$$\|d_{\mu^r} K_r(\phi \mu^r)\|_{\mathcal{S}^r(\Omega', \mu')} = \left\| \frac{d\{K_*(\phi \mu)\}}{d\mu'} \right\|_{L^{1/r}(\Omega', \mu')} \leq \|\phi\|_{L^{1/r}(\Omega, \mu)} = \|\phi \mu^r\|_{\mathcal{S}^r(\Omega, \mu)}.$$

Definition 3.2 For $\mu \in \mathcal{M}(\Omega)$, the bounded linear map (29) is called the *formal derivative* of K_r at μ .

If (M, Ω, \mathbf{p}) is a k -integrable parametrized measure model, then so is $(M, \Omega', \mathbf{p}')$ with $\mathbf{p}' := K_* \mathbf{p}$ by Proposition 2.1. In this case, we may also write

$$\mathbf{p}'^{1/k} = K_{1/k} \mathbf{p}^{1/k}. \quad (30)$$

Proposition 3.1 *The formal derivative of K_r defined in (29) satisfies the identity*

$$d_\xi \mathbf{p}'^{1/k} = (d_{\mathbf{p}(\xi)^{1/k}} K_{1/k})(d_\xi \mathbf{p}^{1/k})$$

for all $\xi \in M$ which may be regarded as the chain rule applied to the derivative of (30).

Proof For $v \in T_\xi M$, $\xi \in M$ we calculate

$$\begin{aligned}
(d_{\mathbf{p}(\xi)^{1/k}} K_{1/k})(d_\xi \mathbf{p}^{1/k}(v)) &\stackrel{(11)}{=} \frac{1}{k} (d_{\mathbf{p}^{1/k}(\xi)} K_{1/k})(\partial_v \log \mathbf{p}(\xi) \mathbf{p}(\xi)^{1/k}) \\
&\stackrel{(29)}{=} \frac{1}{k} \frac{d\{K_*(\partial_v \log \mathbf{p}(\xi) \mathbf{p}(\xi))\}}{d\{\mathbf{p}'(\xi)\}} \mathbf{p}'(\xi)^{1/k} \\
&= \frac{1}{k} \frac{d\{K_*(d_\xi \mathbf{p}(v))\}}{d\{\mathbf{p}'(\xi)\}} \mathbf{p}'(\xi)^{1/k} \\
&= \frac{1}{k} \frac{d\{d_\xi \mathbf{p}'(v)\}}{d\{\mathbf{p}'(\xi)\}} \mathbf{p}'(\xi)^{1/k} \\
&= \frac{1}{k} \partial_v \log \mathbf{p}'(\xi) \mathbf{p}'(\xi)^{1/k} \stackrel{(11)}{=} d_\xi \mathbf{p}'^{1/k}(v),
\end{aligned}$$

which shows the assertion. \square

Our definition of formal derivatives is just strong enough to define the pullback of tensor fields on the space of probability measures in analogy to (21).

Definition 3.3 (*Pullback of tensors by a Markov morphism*) Let $K : \Omega \rightarrow \mathcal{P}(\Omega')$ be a Markov kernel, and let Ψ^n be an n -tensor field on $\mathcal{M}^r(\Omega')$ (on $\mathcal{P}^r(\Omega')$, respectively), cf. Definition 2.4. Then the *pull-back tensor under K* is defined as the covariant n -tensor $K_r^* \Psi^n$ on $\mathcal{M}^r(\Omega)$ (on $\mathcal{P}^r(\Omega)$, respectively) given as

$$K_r^* \Psi^n(V_1, \dots, V_n) := \Psi^n(dK_r(V_1), \dots, dK_r(V_n))$$

with the formal derivative dK_r from (29).

Evidently, $K_r^* \Psi^n$ is again a covariant n -tensor on $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$, respectively, since dK_r is continuous. Moreover, Proposition 3.1 implies that for a parametrized measure model (M, Ω, \mathbf{p}) and the induced model $(M, \Omega', \mathbf{p}')$ with $\mathbf{p}' = K_* \mathbf{p}$ we have the identity

$$\mathbf{p}'^* \Psi^n = \mathbf{p}^* K_r^* \Psi^n \quad (31)$$

for any covariant n -tensor field Ψ^n on $\mathcal{P}^r(\Omega)$ or $\mathcal{M}^r(\Omega)$, respectively.

With this, we can now give a definition of congruent families of tensor fields.

Definition 3.4 (*Congruent families of tensors*) Let $r \in (0, 1]$, and let $(\Theta_{\Omega';r}^n)$ be a collection of covariant n -tensors on $\mathcal{P}^r(\Omega)$ (on $\mathcal{M}^r(\Omega)$, respectively) for each measurable space Ω .

This collection is said to be a *congruent family of n -tensors of regularity r* if for any congruent Markov kernel $K : \Omega \rightarrow \Omega'$ we have

$$K_r^* \Theta_{\Omega';r}^n = \Theta_{\Omega;r}^n.$$

The following gives an important example of such families.

Proposition 3.2 *The restriction of the canonical n -tensors L_{Ω}^n (9) to $\mathcal{P}^{1/n}(\Omega)$ and $\mathcal{M}^{1/n}(\Omega)$, respectively, yield a congruent family of n -tensors. Likewise, then canonical n -tensors $(\tau_{\Omega;r}^n)$ on $\mathcal{P}^r(\Omega)$ and $\mathcal{M}^r(\Omega)$, respectively, with $r \leq 1/n$ yield congruent families of n -tensors.*

Proof Let $K : \Omega \rightarrow \mathcal{P}(\Omega')$ be a Markov kernel which is congruent w.r.t. the statistic $\kappa : \Omega' \rightarrow \Omega$ (cf. Definition 2.2). For $\mu \in \mathcal{M}(\Omega)$ let $\mu' := K_*\mu \in \mathcal{M}(\Omega')$, so that $\kappa_*\mu' = \kappa_*K_*\mu = \mu$. Let $\nu_{1/n}^i = \phi_i\mu^{1/n} \in T_{\mu'}\mathcal{M}^{1/n}(\Omega) = \mathcal{S}^{1/n}(\Omega, \mu')$, with $\phi_i \in L^{1/n}(\Omega, \mu)$, $i = 1, \dots, n$, and define $\phi'_i \in L^{1/n}(\Omega', \mu')$ by

$$K_*(\phi_i\mu) = \phi'_i\mu'.$$

By the κ -congruency of K , this implies that

$$\phi_i\mu = \kappa_*K_*(\phi_i\mu) = \kappa_*(\phi'_i\mu') = (\kappa^*\phi'_i)\kappa_*\mu' = (\kappa^*\phi'_i)\kappa_*K_*\mu = (\kappa^*\phi'_i)\mu,$$

where $\kappa^*\phi(\cdot) := \phi(\kappa(\cdot))$, so that

$$\kappa^*\phi'_i = \phi_i.$$

Then

$$\begin{aligned} (K_{1/n}^*L_{\Omega'}^n)_{\mu^{1/n}}(\nu_{1/n}^1, \dots, \nu_{1/n}^n) &= L_{\Omega'}^n \left((d_{\mu^{1/n}}K_{1/n})(\phi_1\mu^{1/n}), \dots, (d_{\mu^{1/n}}K_{1/n})(\phi_n\mu^{1/n}) \right) \\ &\stackrel{(29)}{=} L_{\Omega'}^n(\phi'_1\mu'^{1/n}, \dots, \phi'_n\mu'^{1/n}) \\ &\stackrel{(9)}{=} n^n \int_{\Omega'} \phi'_1 \cdots \phi'_n d\mu' \\ &= n^n \int_{\Omega} \kappa^*(\phi'_1 \cdots \phi'_n) d(\kappa_*\mu') \\ &= n^n \int_{\Omega} \phi_1 \cdots \phi_n d\mu = L_{\Omega}^n(\nu_{1/n}^1, \dots, \nu_{1/n}^n). \end{aligned}$$

This shows that (L_{Ω}^n) is a congruent family of n -tensors. For $r \leq 1/n$, observe that by (27) we have

$$K_r = \pi^{rn} K_{1/n} \pi^{1/rn} \implies K_r^* = (\pi^{1/rn})^* K_{1/n}^* (\pi^{rn})^*$$

and hence,

$$K_r^* \tau_{\Omega';r}^n \stackrel{(23)}{=} (\pi^{1/rn})^* K_{1/n}^* (\pi^{rn})^* (\pi^{1/rn})^* L_{\Omega'}^n = (\pi^{1/rn})^* K_{1/n}^* L_{\Omega'}^n = (\pi^{1/rn})^* L_{\Omega}^n \stackrel{(23)}{=} \tau_{\Omega;r}^n,$$

showing the congruency of the family $\tau_{\Omega;r}^n$ as well. \square

By (22) and Definition 3.4, it follows that tensor products and permutations of congruent families of tensors yield again such families. Moreover, since

$$\|K_r(\mu_r)\|_{\mathcal{S}^r(\Omega')} = \|K_*\mu_r^{1/r}\|_{\mathcal{S}(\Omega')} \stackrel{(13)}{=} \|\mu_r^{1/r}\|_{\mathcal{S}(\Omega)},$$

multiplying a congruent family with a continuous function depending only on $\|\mu_r^{1/r}\|_{\mathcal{S}(\Omega)} = \|\mu_r^{1/r}\|$ yields again a congruent family of tensors. Therefore, defining for a partition $\mathbf{P} \in \text{Part}(n)$ with the associated map $\pi_{\mathbf{P}}$ from (17) the tensor $\tau_{\mathbf{P}}^{\mathbf{P}} \in \otimes^n V^*$ as

$$(\tau_{\Omega;r}^{\mathbf{P}})_{\mu_r}(v_1, \dots, v_n) := \prod_{i=1}^r (\tau_{\Omega;r}^{n_i})_{\mu_r}(v_{\pi_{\mathbf{P}}(i,1)}, \dots, v_{\pi_{\mathbf{P}}(i,n_i)}), \quad (32)$$

this together with Proposition 2.2 yields the following.

Proposition 3.3 *For $r \in (0, 1]$,*

$$(\tilde{\Theta}_{\Omega;r}^n)_{\mu_r} = \sum_{\mathbf{P}} a_{\mathbf{P}}(\|\mu_r^{1/r}\|)(\tau_{\Omega;r}^{\mathbf{P}})_{\mu_r}, \quad (33)$$

is a congruent family of n -tensor fields on $\mathcal{M}^r(\Omega)$, where the sum is taken over all partitions $\mathbf{P} = \{P_1, \dots, P_l\} \in \text{Part}(n)$ with $|P_i| \leq 1/r$ for all i , and where $a_{\mathbf{P}} : (0, \infty) \rightarrow \mathbb{R}$ are continuous functions. Furthermore,

$$\Theta_{\Omega;r}^n = \sum_{\mathbf{P}} c_{\mathbf{P}} \tau_{\Omega;r}^{\mathbf{P}}, \quad (34)$$

is a congruent family of n -tensor fields on $\mathcal{P}^r(\Omega)$, where the sum is taken over all partitions $\mathbf{P} = \{P_1, \dots, P_l\} \in \text{Part}(n)$ with $1 < |P_i| \leq 1/r$ for all i , and where the $c_{\mathbf{P}} \in \mathbb{R}$ are constants.

In the light of Proposition 2.2, it is reasonable to use the following terminology.

Definition 3.5 The congruent families of n -tensors on $\mathcal{M}^r(\Omega)$ and $\mathcal{P}^r(\Omega)$ given in (33) and (34), respectively, are called the families which are *algebraically generated by the canonical tensors*.

4 Congruent Families on Finite Sample Spaces

In this section, we wish to apply our discussion of the previous sections to the case where the sample space Ω is assumed to be a finite set, in which case it is denoted by I rather than Ω .

The simplification of this case is due to the fact that in this case the spaces $\mathcal{S}^r(I)$ are finite dimensional. Indeed, we have

$$\begin{aligned} \mathcal{S}(I) &= \left\{ \mu = \sum_{i \in I} \mu_i \delta_i \mid \mu_i \in \mathbb{R} \right\}, \\ \mathcal{M}(I) &= \left\{ \mu \in \mathcal{S}(I) \mid \mu_i \geq 0 \right\}, \quad \mathcal{P}(I) = \left\{ \mu \in \mathcal{S}(I) \mid \mu_i \geq 0, \sum_i \mu_i = 1 \right\}, \\ \mathcal{M}_+(I) &:= \left\{ \mu \in \mathcal{S}(I) \mid \mu_i > 0 \right\}, \quad \mathcal{P}_+(I) := \left\{ \mu \in \mathcal{S}(I) \mid \mu_i > 0, \sum_i \mu_i = 1 \right\}, \end{aligned} \quad (35)$$

where δ_i denotes the Dirac measure supported at $i \in I$. The norm on $\mathcal{S}(I)$ is then

$$\left\| \sum_{i \in I} \mu_i \delta_i^r \right\| = \sum_{i \in I} |\mu_i|.$$

The space $\mathcal{S}^r(I)$ is then given as

$$\begin{aligned} \mathcal{S}^r(I) &= \{\mu_r = \sum_{i \in I} \mu_i \delta_i^r \mid \mu_i \in \mathbb{R}\}, \\ \mathcal{M}^r(I) &= \{\mu_r \in \mathcal{S}^r(I) \mid \mu_i \geq 0\}, \quad \mathcal{P}(I) = \left\{ \mu_r \in \mathcal{S}^r(I) \mid \mu_i \geq 0, \sum_i \mu_i^{1/r} = 1 \right\}, \\ \mathcal{M}_+^r(I) &= \{\mu_r \in \mathcal{S}^r(I) \mid \mu_i > 0\}, \quad \mathcal{P}_+(I) = \left\{ \mu_r \in \mathcal{S}^r(I) \mid \mu_i > 0, \sum_i \mu_i^{1/r} = 1 \right\}. \end{aligned} \quad (36)$$

The sets $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I) \subset \mathcal{S}(I)$ are manifolds of dimension $|I|$ and $|I| - 1$, respectively. Indeed, $\mathcal{M}_+(I) \subset \mathcal{S}(I)$ is an open subset, whereas $\mathcal{P}_+(I)$ is an open subset of the affine hyperplane $\mathcal{S}_1(I)$, cf (5). In particular, we have

$$T_\mu \mathcal{P}_+(I) = \mathcal{S}_0(I) \quad \text{and} \quad T_\mu \mathcal{M}_+(I) = \mathcal{S}(I).$$

The norm on $\mathcal{S}^r(I)$ is given as

$$\left\| \sum_{i \in I} \mu_i \delta_i^r \right\|_{\mathcal{S}^r(I)} = \sum_{i \in I} |\mu_i|^{1/r},$$

and the product \cdot and the exponentiating map $\pi^k : \mathcal{S}^r(I) \rightarrow \mathcal{S}^{kr}(I)$ from above are given as

$$\left(\sum_i \mu_i \delta_i^r \right) \cdot \left(\sum_i \nu_i \delta_i^s \right) = \sum_i \mu_i \nu_i \delta_i^{r+s}, \quad \pi^k \left(\sum_{i \in I} \mu_i \delta_i^r \right) = \sum_i \text{sign}(\mu_i) |\mu_i|^k \delta_i^{kr}. \quad (37)$$

Evidently, π^k maps $\mathcal{M}_+^r(I)$ and $\mathcal{P}_+^r(I)$ to $\mathcal{M}_+^{kr}(I)$ and $\mathcal{P}_+^{kr}(I)$, respectively, and the restriction of π^k to these sets is differentiable even if $k < 1$.

A Markov kernel between the finite sets $I = \{1, \dots, m\}$ and $I' = \{1, \dots, n\}$ is determined by the $(n \times m)$ -Matrix $(K_{i'}^i)_{i,i'}$ by

$$K(\delta_i) = K^i = \sum_{i'} K_{i'}^i \delta_{i'},$$

where $K_{i'}^i \geq 0$ and $\sum_{i'} K_{i'}^i = 1$ for all $i \in I$. Therefore, by linearity,

$$K_* \left(\sum_i x_i \delta_i \right) = \sum_{i,i'} K_{i'}^i x_i \delta_{i'}.$$

In particular, $K_*(\mathcal{P}_+(I)) \subset K_*(\mathcal{P}_+(I'))$ and $K_*(\mathcal{M}_+(I)) \subset K_*(\mathcal{M}_+(I'))$.

If $\kappa : I' \rightarrow I$ is a statistic between finite sets (cf. Example 2.2) and if we denote the induced partition by $A_i := \kappa^{-1}(i) \subset I'$, then a Markov kernel $K : I \rightarrow \mathcal{P}(I')$

given by the matrix $(K_{i'}^i)_{i,i'}$ as above is κ -congruent if and only if

$$K_{i'}^i = 0 \quad \text{whenever } i' \notin A_i.$$

Since $(\delta_i^r)_{i \in I}$ is a basis of $\mathcal{S}^r(\Omega)$, we can describe any n -tensor Ψ^n on $\mathcal{S}^r(I)$ by defining for all multiindices $\vec{i} := (i_1, \dots, i_n) \in I^n$ the component functions

$$\psi^{\vec{i}}(\mu_r) := (\Psi^n)_{\mu_r}(\delta_{i_1}^r, \dots, \delta_{i_n}^r) =: (\Psi^n)_{\mu_r}(\delta_{\vec{i}}), \quad (38)$$

which are real valued functions depending continuously on $\mu_r \in \mathcal{S}^r(I)$. Thus, by (24), the component functions of the canonical tensor $\tau_{\Omega;r}^n$ from (38) are given as

$$\mu_r = \sum_{i \in I} m_i \delta_i^r \in \mathcal{S}^r(I) \quad \implies \quad \theta_{I;r}^{\vec{i}}(\mu_r) = \begin{cases} |m_i|^{1/r-n} & \text{if } \vec{i} = (i, \dots, i), \\ 0 & \text{otherwise.} \end{cases} \quad (39)$$

Remark 4.1 Observe that $\theta_{I;r}^{\vec{i}}$ is continuous on $\mathcal{M}_+^r(I)$ and hence $\tau_{I;r}^n = (\pi^{1/nr})^* L_I^n$ is well-defined on $\mathcal{M}_+^r(I)$ even if $r > 1/n$, as on this set $m_i > 0$. This reflects the fact that the restriction $\pi^{1/nr} : \mathcal{M}_+^r(I) \rightarrow \mathcal{S}^{1/n}(I)$ is differentiable for any $r > 0$ by (37).

In particular, for $r = 1$, when restricting to $\mathcal{M}_+(I)$ or $\mathcal{P}_+(I)$, the canonical tensor fields

$$(\tau_{I;1}^n) =: (\tau_I^n) \quad \text{and} \quad (\tau_{I;1}^{\mathbf{P}}) =: (\tau_I^{\mathbf{P}})$$

yield a congruent family of n -tensors on $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I)$, respectively, as is verified as in the proof of Proposition 3.2. Therefore, the families of n -tensor fields

$$(\tilde{\Theta}_I^n)_{\mu_r} = \sum_{\mathbf{P} \in \mathbf{Part}(n)} a_{\mathbf{P}}(\|\mu_r^{1/r}\|)(\tau_I^{\mathbf{P}})_{\mu_r}, \quad (40)$$

on $\mathcal{M}_+(I)$ and

$$\Theta_I^n = \sum_{\mathbf{P} \in \mathbf{Part}(n), |P_i| > 1} c_{\mathbf{P}} \tau_I^{\mathbf{P}} \quad (41)$$

on $\mathcal{P}_+(I)$ are congruent, where in contrast to (33) and (34) we need not restrict the sum to partitions with $|P_i| \leq 1/r$ for all i . In analogy to Definition 3.5 we call these the families of congruent tensors *algebraically generated by the canonical n -tensors* $\{\tau_I^n\}$.

The main result of this section (Theorem 4.1) will be that (40) and (41) are the only families of congruent n -tensor fields which are defined on $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I)$, respectively, for all *finite* sets I . In order to do this, we first deal with congruent families on $\mathcal{M}_+(I)$ only.

A multiindex $\vec{i} = (i_1, \dots, i_n) \in I^n$ induces a partition $\mathbf{P}(\vec{i})$ of the set $\{1, \dots, n\}$ into the equivalence classes of the relation $k \sim l \Leftrightarrow i_k = i_l$. For instance, for $n = 6$ and pairwise distinct elements $i, j, k \in I$, the partition induced by $\vec{i} := (j, i, i, k, j, i)$ is

$$\mathbf{P}(\vec{i}) = \{\{1, 5\}, \{2, 3, 6\}, \{4\}\}.$$

Since the canonical n -tensors τ_I^n are symmetric by definition, it follows that for any partition $\mathbf{P} \in \mathbf{Part}(n)$ we have by (32)

$$(\tau_I^\mathbf{P})_\mu(\delta_{\vec{i}}) \neq 0 \iff \mathbf{P} \leq \mathbf{P}(\vec{i}). \quad (42)$$

Lemma 4.1 *In (40) and (41) above, $a_{\mathbf{P}} : (0, \infty) \rightarrow \mathbb{R}$ and $c_{\mathbf{P}}$ are uniquely determined.*

Proof To show the first statement, let us assume that there are functions $a_{\mathbf{P}} : (0, \infty) \rightarrow \mathbb{R}$ such that

$$\sum_{\mathbf{P} \in \mathbf{Part}(n)} a_{\mathbf{P}}(\|\mu\|)(\tau_I^\mathbf{P})_\mu = 0 \quad (43)$$

for all finite sets I and $\mu \in \mathcal{M}_+(I)$, but there is a partition \mathbf{P}_0 with $a_{\mathbf{P}_0} \not\equiv 0$. In fact, we pick \mathbf{P}_0 to be minimal with this property, and choose a multiindex $\vec{i} \in I^n$ with $\mathbf{P}(\vec{i}) = \mathbf{P}_0$. Then

$$\begin{aligned} 0 &= \sum_{\mathbf{P} \in \mathbf{Part}(n)} a_{\mathbf{P}}(\|\mu\|)(\tau_I^\mathbf{P})_\mu(\delta_{\vec{i}}) \stackrel{(42)}{=} \sum_{\mathbf{P} \leq \mathbf{P}_0} a_{\mathbf{P}}(\|\mu\|)(\tau_I^\mathbf{P})_\mu(\delta_{\vec{i}}) \\ &= a_{\mathbf{P}_0}(\|\mu\|)(\tau_I^{\mathbf{P}_0})_\mu(\delta_{\vec{i}}), \end{aligned}$$

where the last equation follows since $a_{\mathbf{P}} \equiv 0$ for $\mathbf{P} < \mathbf{P}_0$ by the minimality assumption on \mathbf{P}_0 .

But $(\tau_I^{\mathbf{P}_0})_\mu(\delta_{\vec{i}}) \neq 0$ again by (42), since $\mathbf{P}(\vec{i}) = \mathbf{P}_0$, so that $a_{\mathbf{P}_0}(\|\mu\|) = 0$ for all μ , contradicting $a_{\mathbf{P}_0} \not\equiv 0$.

Thus, (43) occurs only if $a_{\mathbf{P}} \equiv 0$ for all \mathbf{P} , showing the uniqueness of the functions $a_{\mathbf{P}}$ in (40).

The uniqueness of the constants $c_{\mathbf{P}}$ in (41) follows similarly, but we have to account for the fact that $\delta_i \notin \mathcal{S}_0(I) = T_\mu \mathcal{P}_+(I)$. In order to get around this, let I be a finite set and $J := \{0, 1, 2\} \times I$. For $i \in I$, we define

$$V_i := 2\delta_{(0,i)} - \delta_{(1,i)} - \delta_{(2,i)} \in \mathcal{S}_0(J),$$

and for a multiindex $\vec{i} = (i_1, \dots, i_n) \in I^n$ we let

$$(\tau_J^\mathbf{P})_\mu(V^{\vec{i}}) := (\tau_J^\mathbf{P})_\mu(V_{i_1}, \dots, V_{i_n}).$$

Multiplying this term out, we see that $(\tau_J^{\mathbf{P}})_\mu(V^{\vec{i}})$ is a linear combination of terms of the form $(\tau_J^{\mathbf{P}})_\mu(\delta_{(a_1, i_1)}, \dots, \delta_{(a_n, i_n)})$, where $a_i \in \{0, 1, 2\}$. Thus, from (42) we conclude that

$$(\tau_J^{\mathbf{P}})_\mu(V^{\vec{i}}) \neq 0 \quad \text{only if } \mathbf{P} \leq \mathbf{P}(\vec{i}). \quad (44)$$

Moreover, if $\mathbf{P}(\vec{i}) = \{P_1, \dots, P_r\}$ with $|P_i| = k_i$, and $\mu_0 := 1/|J| \sum \delta_{(a,i)} \in \mathcal{P}_+(J)$, then

$$\begin{aligned} (\tau_J^{k_i})_{\mu_0}(V_i, \dots, V_i) &\stackrel{(39)}{=} 2^{k_i} (\tau_J^{k_i})_{\mu_0}(\delta_{(0,i)}, \dots, \delta_{(0,i)}) \\ &\quad + (-1)^{k_i} (\tau_J^{k_i})_{\mu_0}(\delta_{(1,i)}, \dots, \delta_{(1,i)}) + (-1)^{k_i} (\tau_J^{k_i})_{\mu_0}(\delta_{(2,i)}, \dots, \delta_{(2,i)}) \\ &\stackrel{(39)}{=} (2^{k_i} + 2(-1)^{k_i}) |J|^{k_i - 1}. \end{aligned}$$

Thus, by (19) we have

$$(\tau_J^{\mathbf{P}(\vec{i})})_{\mu_0}(V^{\vec{i}}) = \prod_{i=1}^r (\tau_J^{k_i})_{\mu_0}(V_i, \dots, V_i) = \prod_{i=1}^r (2^{k_i} + 2(-1)^{k_i}) |J|^{k_i - 1} = |J|^{n-r} \prod_{i=1}^r (2^{k_i} + 2(-1)^{k_i}).$$

In particular, since $2^{k_i} + 2(-1)^{k_i} > 0$ for all $k_i \geq 2$ we conclude that

$$(\tau_J^{\mathbf{P}(\vec{i})})_{\mu_0}(V^{\vec{i}}) \neq 0, \quad (45)$$

as long as $\mathbf{P}(\vec{i})$ does not contain singleton sets.

With this, we can now proceed as in the previous case: assume that

$$\sum_{\mathbf{P} \in \mathbf{Part}(n), |P_i| \geq 2} c_{\mathbf{P}} \tau_I^{\mathbf{P}} = 0 \quad \text{when restricted to } \mathcal{P}_+(I) \quad (46)$$

for constants $c_{\mathbf{P}}$ which do not all vanish, and we let \mathbf{P}_0 be minimal with $c_{\mathbf{P}_0} \neq 0$. Let $\vec{i} = (i_1, \dots, i_n) \in I^n$ be a multiindex with $\mathbf{P}(\vec{i}) = \mathbf{P}_0$, and let $J := \{0, 1, 2\} \times I$ be as above. Then

$$\begin{aligned} 0 &= \sum_{\mathbf{P} \in \mathbf{Part}(n), |P_i| \geq 2} c_{\mathbf{P}} (\tau_J^{\mathbf{P}})_{\mu_0}(V^{\vec{i}}) \stackrel{(44)}{=} \sum_{\mathbf{P} \leq \mathbf{P}_0, |P_i| \geq 2} c_{\mathbf{P}} (\tau_J^{\mathbf{P}})_{\mu_0}(V^{\vec{i}}) \\ &= c_{\mathbf{P}_0} (\tau_J^{\mathbf{P}_0})_{\mu_0}(V^{\vec{i}}), \end{aligned}$$

where the last equality follows by the assumption that \mathbf{P}_0 is minimal. But $(\tau_J^{\mathbf{P}_0})_{\mu_0}(V^{\vec{i}}) \neq 0$ by (45), whence $c_{\mathbf{P}_0} = 0$, contradicting the choice of \mathbf{P}_0 .

This shows that (46) can happen only if all $c_{\mathbf{P}} = 0$, and this completes the proof. \square

The main result of this section is the following.

Theorem 4.1 (Classification of congruent families of n -tensors) *The class of congruent families of n -tensors on $\mathcal{M}_+(I)$ and $\mathcal{P}_+(I)$, respectively, for finite sets I is the*

class algebraically generated by the canonical n -tensors $\{\tau_I^n\}$. That is, these families are the ones given in (40) and (41), respectively.

The rest of this section will be devoted to its proof which is split up into several lemmas.

Lemma 4.2 *Let τ_I^P be the canonical n -tensor from Eq. 32, and define the center*

$$c_I := \frac{1}{|I|} \sum_i \delta_i \in \mathcal{P}_+(I). \quad (47)$$

Then for any $\lambda > 0$ we have

$$(\tau_I^P)_{\lambda c_I}(\delta_{\vec{i}}) = \begin{cases} \left(\frac{|I|}{\lambda}\right)^{n-|P|} & \text{if } P \leq P(\vec{i}), \\ 0 & \text{otherwise.} \end{cases} \quad (48)$$

Proof For $\mu = \lambda c_I$, $\lambda > 0$, the components μ_i of μ all equal $\mu_i = \lambda/|I|$, whence in this case we have for all multiindices \vec{i} with $P \leq P(\vec{i})$

$$(\tau_I^P)_{\lambda c_I}(\delta_{\vec{i}}) = \prod_{i=1}^r \theta_{I; \lambda c_I}^{i, \dots, i} \stackrel{(39)}{=} \prod_{i=1}^r \left(\frac{|I|}{\lambda}\right)^{k_i-1} = \left(\frac{|I|}{\lambda}\right)^{k_1+\dots+k_r-r} = \left(\frac{|I|}{\lambda}\right)^{n-|P|}$$

showing (48). If $P \not\leq P(\vec{i})$, the claim follows from (42). \square

Now let us suppose that $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ is a congruent family of n -tensors on $\mathcal{M}_+(I)$, and define $\theta_{I, \mu}^{\vec{i}}$ as in (38) and $c_I \in \mathcal{P}_+(I)$ as in (47).

Lemma 4.3 *Let $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ and $\theta_{I, \mu}^{\vec{i}}$ be as before, and let $\lambda > 0$. If $\vec{i}, \vec{j} \in I^n$ are multiindices with $P(\vec{i}) = P(\vec{j})$, then*

$$\theta_{I, \lambda c_I}^{\vec{i}} = \theta_{I, \lambda c_I}^{\vec{j}}.$$

Proof If $P(\vec{i}) = P(\vec{j})$, then there is a permutation $\sigma : I \rightarrow I$ such that $\sigma(i_k) = j_k$ for $k = 1, \dots, n$. We define the congruent Markov kernel $K : I \rightarrow \mathcal{P}(I)$ by $K^i := \delta_{\sigma(i)}$. Then evidently, $K_* c_I = c_I$, and Definition 3.4 implies

$$\begin{aligned} \theta_{I, \lambda c_I}^{\vec{i}} &= (\tilde{\Theta}_I^n)_{\lambda c_I}(\delta_{i_1}, \dots, \delta_{i_n}) \\ &= (\tilde{\Theta}_I^n)_{K_*(\lambda c_I)}(K_* \delta_{i_1}, \dots, K_* \delta_{i_n}) \\ &= (\tilde{\Theta}_I^n)_{\lambda c_I}(\delta_{j_1}, \dots, \delta_{j_n}) = \theta_{I, \lambda c_I}^{\vec{j}}, \end{aligned}$$

which shows the claim. \square

By virtue of this lemma, we may define

$$\theta_{I,\lambda c_I}^{\mathbf{P}} := \theta_{I,\lambda c_I}^{\vec{i}}, \quad \text{where } \vec{i} \in I^n \text{ is a multiindex with } \mathbf{P}(\vec{i}) = \mathbf{P}.$$

Lemma 4.4 Let $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ and $\theta_{I,\lambda c_I}^{\mathbf{P}}$ be as before, and suppose that $\mathbf{P}_0 \in \mathbf{Part}(n)$ is a partition such that

$$\theta_{I,\lambda c_I}^{\mathbf{P}} = 0 \quad \text{for all } \mathbf{P} < \mathbf{P}_0, \lambda > 0 \text{ and } I. \quad (49)$$

Then there is a continuous function $f_{\mathbf{P}_0} : (0, \infty) \rightarrow \mathbb{R}$ such that

$$\theta_{I,\lambda c_I}^{\mathbf{P}_0} = f_{\mathbf{P}_0}(\lambda) |I|^{n-|\mathbf{P}_0|}. \quad (50)$$

Proof Let I, J be finite sets, and let $I' := I \times J$. We define the Markov kernel

$$K : I \longrightarrow \mathcal{P}(I'), \quad i \longmapsto \frac{1}{|J|} \sum_{j \in J} \delta_{(i,j)}$$

which is congruent w.r.t. the canonical projecton $\kappa : I' \rightarrow I$. Then $K_* c_I = c_{I'}$ is easily verified. Moreover, if $\vec{i} = (i_1, \dots, i_n) \in I^n$ is a multiindex with $\mathbf{P}(\vec{i}) = \mathbf{P}_0$, then

$$\begin{aligned} \theta_{I,\lambda c_I}^{\mathbf{P}_0} &= (\tilde{\Theta}_I^n)_{\lambda c_I}(\delta_{i_1}, \dots, \delta_{i_n}) \\ &\stackrel{\text{Definition 3.4}}{=} (\tilde{\Theta}_{I'}^n)_{K_*(\lambda c_I)}(K_* \delta_{i_1}, \dots, K_* \delta_{i_n}) \\ &= (\tilde{\Theta}_{I'}^n)_{\lambda c_{I'}} \left(\frac{1}{|J|} \sum_{j_1 \in J} \delta_{(i_1, j_1)}, \dots, \frac{1}{|J|} \sum_{j_n \in J} \delta_{(i_n, j_n)} \right) \\ &= \frac{1}{|J|^n} \sum_{(j_1, \dots, j_n) \in J^n} \theta_{I', \lambda c_{I'}}^{\mathbf{P}((i_1, j_1), \dots, (i_n, j_n))}. \end{aligned}$$

Observe that $\mathbf{P}((i_1, j_1), \dots, (i_n, j_n)) \leq \mathbf{P}(\vec{i}) = \mathbf{P}_0$. If $\mathbf{P}((i_1, j_1), \dots, (i_n, j_n)) < \mathbf{P}_0$, then $\theta_{I', \lambda c_{I'}}^{\mathbf{P}((i_1, j_1), \dots, (i_n, j_n))} = 0$ by (49).

Moreover, there are $|J|^{\mathbf{P}_0|}$ multiindices $(j_1, \dots, j_n) \in J^n$ for which $\mathbf{P}((i_1, j_1), \dots, (i_n, j_n)) = \mathbf{P}_0$, and since for all of these $\theta_{I', \lambda c_{I'}}^{\mathbf{P}((i_1, j_1), \dots, (i_n, j_n))} = \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0}$, we obtain

$$\theta_{I,\lambda c_I}^{\mathbf{P}_0} = \frac{1}{|J|^n} \sum_{(j_1, \dots, j_n) \in J^n} \theta_{I', \lambda c_{I'}}^{\mathbf{P}((i_1, j_1), \dots, (i_n, j_n))} = \frac{|J|^{\mathbf{P}_0|}}{|J|^n} \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0} = \frac{1}{|J|^{n-|\mathbf{P}_0|}} \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0},$$

and since $|I'| = |I| |J|$, it follows that

$$\frac{1}{|I|^{n-|\mathbf{P}_0|}} \theta_{I, \lambda c_I}^{\mathbf{P}_0} = \frac{1}{|I|^{n-|\mathbf{P}_0|}} \left(\frac{1}{|J|^{n-|\mathbf{P}_0|}} \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0} \right) = \frac{1}{|I'|^{n-|\mathbf{P}_0|}} \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0}.$$

Interchanging the roles of I and J in the previous arguments, we also get

$$\frac{1}{|J|^{n-|\mathbf{P}_0|}} \theta_{J, \lambda c_J}^{\mathbf{P}_0} = \frac{1}{|I'|^{n-|\mathbf{P}_0|}} \theta_{I', \lambda c_{I'}}^{\mathbf{P}_0} = \frac{1}{|I|^{n-|\mathbf{P}_0|}} \theta_{I, \lambda c_I}^{\mathbf{P}_0},$$

whence $f_{\mathbf{P}_0}(\lambda) := \frac{1}{|I|^{n-|\mathbf{P}_0|}} \theta_{I, \lambda c_I}^{\mathbf{P}_0}$ is indeed independent of the choice of the finite set I . \square

Lemma 4.5 *Let $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ and $\lambda > 0$ be as before. Then there is a congruent family $\{\tilde{\Psi}_I^n : I \text{ finite}\}$ of the form (40) such that*

$$(\tilde{\Theta}_I^n - \tilde{\Psi}_I^n)_{\lambda c_I} = 0 \quad \text{for all finite sets } I \text{ and all } \lambda > 0.$$

Proof For a congruent family of n -tensors $\{\tilde{\Theta}_I^n : I \text{ finite}\}$, we define

$$N(\{\tilde{\Theta}_I^n\}) := \{\mathbf{P} \in \mathbf{Part}(n) : (\tilde{\Theta}_I^n)_{\lambda c_I}(\delta_{\vec{i}}) = 0 \text{ whenever } \mathbf{P}(\vec{i}) \leq \mathbf{P}\}.$$

If $N(\{\tilde{\Theta}_I^n\}) \subsetneq \mathbf{Part}(n)$, then let

$$\mathbf{P}_0 = \{P_1, \dots, P_r\} \in \mathbf{Part}(n) \setminus N(\{\tilde{\Theta}_I^n\})$$

be a minimal element, i.e., such that $\mathbf{P} \in N(\{\tilde{\Theta}_I^n\})$ for all $\mathbf{P} < \mathbf{P}_0$. In particular, for this partition (49) and hence (50) holds. Let

$$(\tilde{\Theta}'_I^n)_\mu := (\tilde{\Theta}_I^n)_\mu - \|\mu\|^{n-|\mathbf{P}_0|} f_{\mathbf{P}_0}(\|\mu\|) (\tau_I^{\mathbf{P}_0})_\mu \quad (51)$$

with the function $f_{\mathbf{P}_0}$ from (50). Then $\{\tilde{\Theta}'_I^n : I \text{ finite}\}$ is again a family of n -tensors.

Let $\mathbf{P} \in N(\{\tilde{\Theta}_I^n\})$ and \vec{i} be a multiindex with $\mathbf{P}(\vec{i}) \leq \mathbf{P}$. If $(\tau_I^{\mathbf{P}_0})_{\lambda c_I}(\delta_{\vec{i}}) \neq 0$, then by Lemma 4.2 we would have $\mathbf{P}_0 \leq \mathbf{P}(\vec{i}) \leq \mathbf{P} \in N(\{\tilde{\Theta}_I^n\})$ which would imply that $\mathbf{P}_0 \in N(\{\tilde{\Theta}_I^n\})$, contradicting the choice of \mathbf{P}_0 .

Thus, $(\tau_I^{\mathbf{P}_0})_{\lambda c_I}(\delta_{\vec{i}}) = 0$ and hence $(\tilde{\Theta}'_I^n)_{\lambda c_I}(\delta_{\vec{i}}) = 0$ whenever $\mathbf{P}(\vec{i}) \leq \mathbf{P}$, showing that $\mathbf{P} \in N(\{\tilde{\Theta}'_I^n\})$.

Thus, what we have shown is that $N(\{\tilde{\Theta}_I^n\}) \subset N(\{\tilde{\Theta}'_I^n\})$. On the other hand, if $\mathbf{P}(\vec{i}) = \mathbf{P}_0$, then again by Lemma 4.2

$$(\tau_I^{\mathbf{P}_0})_{\lambda c_I}(\delta_{\vec{i}}) = \left(\frac{|I|}{\lambda} \right)^{n-|\mathbf{P}_0|},$$

and since $\|\lambda c_I\| = \lambda$, it follows that

$$\begin{aligned}
(\tilde{\Theta}'_I^n)_{\lambda c_I}(\delta_{\vec{i}}) &\stackrel{(51)}{=} (\tilde{\Theta}_I^n)_{\lambda c_I}(\delta_{\vec{i}}) - \lambda^{n-|\mathbf{P}_0|} f_{\mathbf{P}_0}(\lambda) (\tau_I^{\mathbf{P}_0})_{\lambda c_I}(\delta_{\vec{i}}) \\
&= \theta_{I, \lambda c_I}^{\mathbf{P}_0} - \lambda^{n-|\mathbf{P}_0|} f_{\mathbf{P}_0}(\lambda) \left(\frac{|I|}{\lambda} \right)^{n-|\mathbf{P}_0|} \\
&= \theta_{I, \lambda c_I}^{\mathbf{P}_0} - f_{\mathbf{P}_0}(\lambda) |I|^{n-|\mathbf{P}_0|} \stackrel{(50)}{=} 0.
\end{aligned}$$

That is, $(\tilde{\Theta}'_I^n)_{\lambda c_I}(\delta_{\vec{i}}) = 0$ whenever $\mathbf{P}(\vec{i}) = \mathbf{P}_0$. If \vec{i} is a multiindex with $\mathbf{P}(\vec{i}) < \mathbf{P}_0$, then $\mathbf{P}(\vec{i}) \in N(\{\tilde{\Theta}'_I^n\})$ by the minimality of \mathbf{P}_0 , so that $\tilde{\Theta}_I^n(\delta_{\vec{i}}) = 0$. Moreover, $(\tau_I^{\mathbf{P}_0})_{\lambda c_I}(\delta_{\vec{i}}) = 0$ by Lemma 4.2, whence

$$(\tilde{\Theta}'_I^n)_{\lambda c_I}(\delta_{\vec{i}}) = 0 \quad \text{whenever } \mathbf{P}(\vec{i}) \leq \mathbf{P}_0,$$

showing that $\mathbf{P}_0 \in N(\{\tilde{\Theta}'_I^n\})$. Therefore,

$$N(\{\tilde{\Theta}_I^n\}) \subsetneq N(\{\tilde{\Theta}'_I^n\}).$$

What we have shown is that given a congruent family of n -tensors $\{\tilde{\Theta}_I^n\}$ with $N(\{\tilde{\Theta}_I^n\}) \subsetneq \mathbf{Part}(n)$, we can enlarge $N(\{\tilde{\Theta}_I^n\})$ by subtracting a multiple of the canonical tensor of some partition. Repeating this finitely many times, we conclude that for some congruent family $\{\tilde{\Psi}_I^n\}$ of the form (40)

$$N(\{\tilde{\Theta}_I^n - \tilde{\Psi}_I^n\}) = \mathbf{Part}(n),$$

and this implies by definition that $(\tilde{\Theta}_I^n - \tilde{\Psi}_I^n)_{\lambda c_I} = 0$ for all I and all $\lambda > 0$. \square

Lemma 4.6 *Let $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ be a congruent family of n -tensors such that $(\tilde{\Theta}_I^n)_{\lambda c_I} = 0$ for all I and $\lambda > 0$. Then $\tilde{\Theta}_I^n = 0$ for all I .*

Proof Consider $\mu \in \mathcal{M}_+(I)$ such that $\pi_I(\mu) = \mu/\|\mu\| \in \mathcal{P}_+(I)$ has rational coefficients, i.e.

$$\mu = \|\mu\| \sum_i \frac{k_i}{n} \delta_i$$

for some $k_i, n \in \mathbb{N}$ and $\sum_{i \in I} k_i = n$. Let

$$I' := \bigcup_{i \in I} (\{i\} \times \{1, \dots, k_i\}),$$

so that $|I'| = n$, and consider the congruent Markov kernel

$$K : i \longmapsto \frac{1}{k_i} \sum_{j=1}^{k_i} \delta_{(i,j)}.$$

Then

$$K_*\mu = \|\mu\| \sum_i \frac{k_i}{n} \left(\frac{1}{k_i} \sum_{j=1}^{k_i} \delta_{(i,j)} \right) = \|\mu\| \frac{1}{n} \sum_i \sum_{j=1}^{k_i} \delta_{(i,j)} = \|\mu\| c_{I'}.$$

Thus, Definition 3.4 implies

$$(\tilde{\Theta}_I^n)_\mu(V_1, \dots, V_n) = \underbrace{(\tilde{\Theta}_{I'}^n)_{\|\mu\|c_{I'}}(K_*V_1, \dots, K_*V_n)}_{=0} = 0,$$

so that $(\tilde{\Theta}_I^n)_\mu = 0$ whenever $\pi_I(\mu)$ has rational coefficients. But these μ form a dense subset of $\mathcal{M}_+(I)$, whence $(\tilde{\Theta}_I^n)_\mu = 0$ for all $\mu \in \mathcal{M}_+(I)$, which completes the proof. \square

We are now ready to prove the main result in this section.

Proof of Theorem 4.1 Let $\{\tilde{\Theta}_I^n : I \text{ finite}\}$ be a congruent family of n -tensors. By Lemma 4.5 there is a congruent family $\{\tilde{\Psi}_I^n : I \text{ finite}\}$ of the form (40) such that $(\tilde{\Theta}_I^n - \tilde{\Psi}_I^n)_{\lambda c_I} = 0$ for all finite I and all $\lambda > 0$.

Since $\{\tilde{\Theta}_I^n - \tilde{\Psi}_I^n : I \text{ finite}\}$ is again a congruent family, Lemma 4.6 implies that $\tilde{\Theta}_I^n - \tilde{\Psi}_I^n = 0$ and hence $\tilde{\Theta}_I^n = \tilde{\Psi}_I^n$ is of the form (40), showing the statement of Theorem 4.1 for n -tensors on $\mathcal{M}_+(I)$.

To show the second part, let us consider for a finite set I the inclusion and projection

$$\iota_I : \mathcal{P}_+(I) \hookrightarrow \mathcal{M}_+(I), \quad \text{and} \quad \pi_I : \begin{aligned} \mathcal{M}_+(I) &\longrightarrow \mathcal{P}_+(I) \\ \mu &\longmapsto \frac{\mu}{\|\mu\|} \end{aligned}$$

Evidently, π_I is a left inverse of ι_I , i.e., $\pi_I \iota_I = Id_{\mathcal{P}_+(I)}$, and by (13) it follows that K_* commutes both with π_I and ι_I .

Thus, if $\{\Theta_I^n : I \text{ finite}\}$ is a congruent family of n -tensors on $\mathcal{P}_+(I)$, then

$$\tilde{\Theta}_I^n := \pi_I^* \Theta_I^n$$

yields a congruent families of n -tensors on $\mathcal{M}_+(I)$ and by the first part of the theorem must be of the form (40). But then,

$$\Theta_I^n = \iota_I^* \tilde{\Theta}_I^n = \sum_{\mathbf{P}} c_{\mathbf{P}}(\tau_I^n)|_{\mathcal{P}_+(I)},$$

where $c_{\mathbf{P}} = a_{\mathbf{P}}(1)$. Since $(\tau_I^n)|_{\mathcal{P}_+(I)} = 0$ if \mathbf{P} contains a singleton set, it follows that Θ_I^n is of the form (41). \square

5 Congruent Families on Arbitrary Sample Spaces

In this section, we wish to generalize the classification result for congruent families on finite sample spaces (Theorem 4.1) to the case of arbitrary sample spaces. As it turns out, we show that even in this case, congruent families of tensor fields are algebraically generated by the canonical tensor fields. More precisely, we have the following result.

Theorem 5.1 (Classification of congruent families) *For $0 < r \leq 1$, let $(\Theta_{\Omega;r}^n)$ be a family of covariant n -tensors on $\mathcal{M}^r(\Omega)$ (on $\mathcal{P}^r(\Omega)$, respectively) for each measurable space Ω . Then the following are equivalent:*

- (1) $(\Theta_{\Omega;r}^n)$ is a congruent family of covariant n -tensors of regularity r .
- (2) For each congruent Markov morphism $K : I \rightarrow \mathcal{P}(\Omega)$ for a finite set I , we have $K_r^* \Theta_{\Omega;r}^n = \Theta_{I;r}^n$.
- (3) $\Theta_{\Omega;r}^n$ is of the form (33) (of the form (34), respectively) for uniquely determined continuous functions $a_{\mathbf{P}}$ (constants $c_{\mathbf{P}}$, respectively).

In the light of Definition 3.5, we may reformulate the equivalence of the first and the third statement as follows:

Corollary 5.1 *The space of congruent families of covariant n -tensors on $\mathcal{M}^r(\Omega)$ and $\mathcal{P}^r(\Omega)$, respectively, is algebraically generated by the canonical n -tensors $\tau_{\Omega;r}^n$ for $n \leq 1/r$.*

Proof of Theorem 5.1 We already showed in Proposition 3.3 that the tensors (33) and (34), respectively, are congruent families, hence the third statement implies the first. The first immediately implies the second by the definition of the congruency of tensors. Thus, it remains to show that the second statement implies the third.

We shall give the proof only for the families $(\Theta_{\Omega;r}^n)$ of covariant n -tensors on $\mathcal{M}^r(\Omega)$, as the proof for families on $\mathcal{P}^r(\Omega)$ is analogous.

Observe that for finite sets I , the space $\mathcal{M}_+^r(I) \subset \mathcal{S}^r(I)$ is an open subset and hence a manifold, and the restrictions $\pi^\alpha : \mathcal{M}_+^r(I) \rightarrow \mathcal{M}_+^{r\alpha}(I)$ are diffeomorphisms not only for $\alpha \geq 1$ but for all $\alpha > 0$. Thus, given the congruent family $(\Theta_{\Omega;r}^n)$, we define for each finite set I the tensor

$$\Theta_I^n := (\pi^r)^* \Theta_{I;r}^n \quad \text{on } \mathcal{M}_+(I).$$

Then for each congruent Markov kernel $K : I \rightarrow \mathcal{P}(J)$ with I, J finite we have

$$\begin{aligned} K^* \Theta_J^n &= K^*(\pi^r)^* \Theta_{J;r}^n = (\pi^r K_*)^* \Theta_{J;r}^n \\ &\stackrel{(3.1)}{=} (K_r \pi^r)^* \Theta_{J;r}^n = (\pi^r)^* K_r^* \Theta_{J;r}^n = (\pi^r)^* \Theta_{I;r}^n \\ &= \Theta_I^n. \end{aligned}$$

Thus, the family (Θ_I^n) on $\mathcal{M}_+(I)$ is a congruent family of covariant n -tensors on finite sets, whence by Theorem 4.1

$$(\Theta_I^n)_\mu = \sum_{\mathbf{P} \in \text{Part}(n)} a_{\mathbf{P}}(\|\mu\|) (\tau_I^{\mathbf{P}})_\mu$$

for uniquely determined functions $a_{\mathbf{P}}$, whence on $\mathcal{M}_+^r(I)$,

$$\begin{aligned} \Theta_{I;r}^n &= (\pi^{1/r})^* \Theta_I^n \\ &= \sum_{\mathbf{P} \in \text{Part}(n)} a_{\mathbf{P}}(\|\mu_r^{1/r}\|) (\pi^{1/r})^* \tau_I^{\mathbf{P}} \\ &= \sum_{\mathbf{P} \in \text{Part}(n)} a_{\mathbf{P}}(\|\mu_r^{1/r}\|) \tau_{I;r}^{\mathbf{P}}. \end{aligned}$$

By our assumption, $\Theta_{I;r}^n$ must be a covariant n -tensor on $\mathcal{M}(I)$, whence it must extend continuously to the boundary of $\mathcal{M}_+(I)$.

But by (39) it follows that $\tau_{I;r}^{n_i}$ has a singularity at the boundary of $\mathcal{M}(I)$, unless $n_i \leq 1/r$. From this it follows that $\Theta_{I;r}^n$ extends to all of $\mathcal{M}(I)$ if and only if $a_{\mathbf{P}} \equiv 0$ for all partitions $\mathbf{P} = \{P_1, \dots, P_i\}$ where $|P_i| > 1/r$ for some i .

Thus, $\Theta_{I;r}^n$ must be of the form (33) for all finite sets I . Let

$$\Psi_{\Omega;r}^n := \Theta_{\Omega;r}^n - \sum_{\mathbf{P}} a_{\mathbf{P}}(\|\mu_r^{1/r}\|) \tau_{\Omega;r}^n$$

for the previously determined functions $a_{\mathbf{P}}$, so that $(\Psi_{\Omega;r}^n)$ is a congruent family of covariant n -tensors, and $\Psi_{I;r}^n = 0$ for every finite I .

We assert that this implies that $\Psi_{\Omega;r}^n = 0$ for all Ω , which shows that $\Theta_{\Omega;r}^n$ is of the form (33) for all Ω , which will complete the proof.

To see this, let $\mu_r \in \mathcal{M}^r(\Omega)$ and $\mu := \mu_r^{1/r} \in \mathcal{M}(\Omega)$. Moreover, let $V_j = \phi_j \mu_r \in \mathcal{S}^r(\Omega, \mu_r)$, $j = 1, \dots, n$, such that the ϕ_j are step functions. That is, there is a finite partition $\Omega = \bigcup_{i \in I} \Omega_i$ such that

$$\phi_j = \sum_{i \in I} \phi_j^i \chi_{\Omega_i}$$

for $\phi_j^i \in \mathbb{R}$ and $m_i := \mu(\Omega_i) > 0$.

Let $\kappa : \Omega \rightarrow I$ be the statistic $\kappa(\Omega_i) = \{i\}$, and $K : I \rightarrow \mathcal{P}(\Omega)$, $K(i) := 1/m_i \chi_{\Omega_i} \mu$. Then clearly, K is κ -congruent, and $\mu = K_* \mu'$ with $\mu' := \sum_{i \in I} m_i \delta_i \in \mathcal{M}_+(I)$. Thus, by (29)

$$d_{\mu'} K_r \left(\sum_{i \in I} \phi_j^i m_i^r \delta_i^r \right) = \sum_{i \in I} \phi_j^i \chi_{\Omega_i} \mu_r = \phi_j \mu_r = V_j,$$

whence if we let $V'_j := \sum_{i \in I} \phi_j^i m_i^r \delta_i^r \in \mathcal{S}^r(I)$, then

$$\begin{aligned}\Psi_{\Omega;r}^n(V_1, \dots, V_n) &= \Psi_{\Omega;r}^n(dK_r(V'_1), \dots, dK_r(V'_n)) \\ &= K_r^* \Psi_{\Omega;r}^n(V'_1, \dots, V'_n) \\ &= \Psi_{I;r}^n(V'_1, \dots, V'_n) = 0,\end{aligned}$$

since by the congruence of the family $(\Psi_{\Omega;r}^n)$ we must have $K_r^* \Psi_{\Omega;r}^n = \Psi_{I;r}^n$, and $\Psi_{I;r}^n = 0$ by assumption as I is finite.

That is, $\Psi_{\Omega;r}^n(V_1, \dots, V_n) = 0$ whenever $V_j = \phi_j \mu_r \in \mathcal{S}^r(\Omega, \mu_r)$ with step functions ϕ_j . But the elements V_j of this form are dense in $\mathcal{S}^r(\Omega, \mu_r)$, hence the continuity of $\Psi_{\Omega;r}^n$ implies that $\Psi_{\Omega;r}^n = 0$ for all Ω as claimed. \square

As two special cases of this result, we obtain the following.

Corollary 5.2 (Generalization of Chentsov's theorem) (1) Let (Θ_Ω^2) be a congruent family of 2-tensors on $\mathcal{P}^{1/2}(\Omega)$. Then up to a constant, this family is the Fisher metric. That is, there is a constant $c \in \mathbb{R}$ such that for all Ω ,

$$\Theta_\Omega^2 = c \mathbf{g}_F.$$

In particular, if (M, Ω, \mathbf{p}) is a 2-integrable statistical model, then

$$\mathbf{p}^* \Theta_\Omega^2 = c \mathbf{g}_M$$

is – up to a constant – the Fisher metric of the model.

(2) Let (Θ_Ω^3) be a congruent family of 3-tensors on $\mathcal{P}^{1/3}(\Omega)$. Then up to a constant, this family is the Amari–Chentsov tensor. That is, there is a constant $c \in \mathbb{R}$ such that for all Ω ,

$$\Theta_\Omega^3 = c \mathbf{T}.$$

In particular, if (M, Ω, \mathbf{p}) is a 3-integrable statistical model, then

$$\mathbf{p}^* \Theta_\Omega^3 = c \mathbf{T}_M$$

is – up to a constant – the Amari–Chentsov tensor of the model.

Corollary 5.3 (Generalization of Campbell's theorem) Let (Θ_Ω^2) be a congruent family of 2-tensors on $\mathcal{M}^{1/2}(\Omega)$. Then there are continuous functions $a, b : (0, \infty) \rightarrow \mathbb{R}$ such that

$$(\Theta_\Omega^2)_{\mu^{1/2}}(V_1, V_2) = a(\|\mu\|) \mathbf{g}_F(V_1, V_2) + b(\|\mu\|) \tau_{\Omega;1/2}^1(V_1) \tau_{\Omega;1/2}^1(V_2).$$

In particular, if (M, Ω, \mathbf{p}) is a 2-integrable parametrized measure model, then

$$\begin{aligned}\mathbf{p}^*(\Theta_\Omega^2)_\xi(V_1, V_2) &= a(\|\mathbf{p}(\xi)\|) \int_\Omega \partial_{V_1} \log \mathbf{p}(\xi) \partial_{V_2} \log \mathbf{p}(\xi) d\mathbf{p}(\xi) \\ &\quad + b(\|\mathbf{p}(\xi)\|) (\partial_{V_1} \|\mathbf{p}(\xi)\|) (\partial_{V_2} \|\mathbf{p}(\xi)\|).\end{aligned}$$

While the above results show that for small n there is a unique family of congruent n -tensors, this is no longer true for larger n . For instance, for $n = 4$ Theorem 5.1 implies that any restricted congruent family of invariant 4-tensors on $\mathcal{P}^r(\Omega)$, $0 < r \leq 1/4$, is of the form

$$\begin{aligned}\Theta_{\Omega}^4(V_1, \dots, V_4) = & c_0 \tau_{\Omega;r}^4(V_1, \dots, V_4) \\ & + c_1 \tau_{\Omega;r}^2(V_1, V_2) \tau_{\Omega;r}^2(V_3, V_4) \\ & + c_2 \tau_{\Omega;r}^2(V_1, V_3) \tau_{\Omega;r}^2(V_2, V_3) \\ & + c_3 \tau_{\Omega;r}^2(V_1, V_4) \tau_{\Omega;r}^2(V_2, V_4),\end{aligned}$$

so that the space of congruent families on $\mathcal{P}^r(\Omega)$ is already 4-dimensional in this case. Evidently, this dimension rapidly increases with n .

Acknowledgements This work was mainly carried out at the Max Planck Institute for Mathematics in the Sciences in Leipzig, and we are grateful for the excellent working conditions provided at that institution. The research of H.V. Lê was supported by the GAČR project 18-01953J and RVO: 67985840. L. Schwachhöfer acknowledges partial support by grant SCHW893/5-1 of the Deutsche Forschungsgemeinschaft.

References

1. Amari, S.: Theory of information spaces. A geometrical foundation of statistics. POST RAAG Report 106 (1980)
2. Amari, S.: Differential geometry of curved exponential families curvature and information loss. *Ann. Stat.* **10**, 357–385 (1982)
3. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **162**, 327–364 (2015)
4. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Parametrized measure models. *Bernoulli* **24**(3), 1692–1725 (2018)
5. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: *Information Geometry*, Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin (2017)
6. Bauer, M., Bruveris, M., Michor, P.: Uniqueness of the Fisher-Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.* **48**(3), 499–506 (2016)
7. Bauer, M., Bruveris, M., Michor, P. In: Presentation at the Fourth Conference on Information Geometry and its Applications. (IGAIA IV, 2016), Liblice, Czech Republic
8. Campbell, L.L.: An extended Chentsov characterization of a Riemannian metric. *Proc. Am. Math. Soc.* **98**, 135–141 (1986)
9. Chentsov, N.: Category of mathematical statistics. *Dokl. Acad. Nauk. USSR* **164**, 511–514 (1965)
10. Chentsov, N.: Algebraic foundation of mathematical statistics. *Math. Operationsforsch. Stat. Ser. Stat.* **9**, 267–276 (1978)
11. Chentsov, N.: Statistical Decision Rules and Optimal Inference. Moscow, Nauka (1972) (in Russian); English translation in: Translation of Mathematical Monograph, vol. 53. American Mathematical Society, Providence, (1982)
12. Dowty, J.: Chentsov's theorem for exponential families. [arXiv:1701.08895](https://arxiv.org/abs/1701.08895)
13. Efron, B.: Defining the curvature of a statistical problem (with applications to second order efficiency), with a discussion by Rao, C.R., Pierce, D.A., Cox, D.R., Lindley, D.V., LeCam, L.,

- Ghosh, J.K., Pfanzagl, J., Keiding, N., Dawid, A.P., Reeds, J., and with a reply by the author. *Ann. Statist.* **3**, 1189–1242 (1975)
- 14. Jeffreys, H.: An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. Ser. A.* **186**, 453–461 (1946)
 - 15. Jost, J., Lê, H.V., Schwachhöfer, L.: The Cramér-Rao inequality on singular statistical models I. [arXiv:1703.09403](https://arxiv.org/abs/1703.09403) (2017)
 - 16. Lê, H.V.: The uniqueness of the Fisher metric as information metric. *Ann. Inst. Stat. Math.* **69**, 879–896 (2017)
 - 17. Morozova, E., Chentsov, N.: Natural geometry on families of probability laws, *Itogi Nauki i Techniki, Current problems of mathematics, Fundamental directions*, vol. 83, pp. 133–265. Moscow (1991)
 - 18. Rao, C.R.: Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–89 (1945)

Nonlinear Filtering and Information Geometry: A Hilbert Manifold Approach



Nigel J. Newton

Abstract 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 represented in terms of stochastic differential equations for the posterior distribution of the signal. The natural “state space” for a filter is a sufficiently rich family of probability measures having a suitable topology, and the manifolds of infinite-dimensional Information Geometry are obvious candidates. After some discussion of these, the paper goes on to summarise recent results that “lift” the equations of nonlinear filtering to a Hilbert manifold, M . Apart from providing a starting point for the development of approximations, this gives insight into the information-theoretic properties of filters, which are related to their quadratic variation in the Fisher–Rao metric. A new result is proved on the regularity of a multi-objective measure of approximation errors on M .

Keywords Bayes’ formula · Fisher–Rao metric · Infinite dimensions · Information geometry · Information theory · Nonlinear filtering · Quadratic variation

1 Introduction

Let $(X_t, Y_t, t \geq 0)$ be an $m + d$ -vector Markov diffusion process satisfying the Itô stochastic differential equations

$$X_t = X_0 + \int_0^t BX_s ds + \Gamma V_t \quad \text{and} \quad Y_t = \int_0^t CX_s ds + W_t, \quad (1)$$

N. J. Newton (✉)
University of Essex, Colchester CO4 3SQ, UK
e-mail: njn@essex.ac.uk
URL: <https://www1.essex.ac.uk/csee/staff/>

where X_0 has the m -variate Gaussian distribution with mean vector \bar{X}_0 and covariance matrix Σ_0 , and $(V_t, W_t, t \geq 0)$ is an $r + d$ -vector standard Brownian motion, independent of X_0 . (See, for example, [1].) Suppose that X_t represents the state of a physical system at time t , which can be observed only indirectly, through the process Y . For example, X_t might be the 6-vector of position and velocity components of an object moving in 3-dimensional space, which is to be tracked on the basis of partial (if $\text{rank}(C) < m$) and noise-corrupted observations. In such situations it is useful to compute, at each time t , the conditional probability distribution for X_t given the current observation history $(Y_s, 0 \leq s \leq t)$. Since the equations in (1) are linear, this is Gaussian, and so characterised by its mean vector \bar{X}_t and covariance matrix Σ_t . These satisfy the celebrated Kalman–Bucy filtering equations [2],

$$\begin{aligned}\bar{X}_t &= \bar{X}_0 + \int_0^t B\bar{X}_s ds + \int_0^t \Sigma_s C^*(dY_s - C\bar{X}_s ds), \\ \dot{\Sigma}_t &= B\Sigma_t + \Sigma_t B^* + \Gamma\Gamma^* - \Sigma_t C^*C\Sigma_t,\end{aligned}\tag{2}$$

where C^* indicates the matrix transpose.

The term “filtering” has its origins in the engineering literature. Since Σ_t does not depend on Y , the only statistic to be computed by the Kalman–Bucy filter is the mean vector \bar{X}_t , and this has linear dynamics. So the filter admits a spectral interpretation, according to which (2) “filters out” some of the spectral components of Y , passing only those that are known (by virtue of its dynamics) to be present in X .

If the dynamics of the *signal process* X or *observation process* Y in (1) are replaced by nonlinear dynamics then the estimation problem becomes one of *nonlinear filtering*, in which the observation-conditional distribution of the signal can rarely be expressed in terms of a finite number of statistics. More importantly, even if only a few statistics of X_t (such as its conditional mean) are needed, they can rarely be expressed in terms of a finite-dimensional differential equation such as (2). We are then obliged to compute, or at least approximate, the entire observation-conditional distribution of the signal X_t . It is then natural to ask: what is a suitable “state-space” for such a nonlinear filter? This should be a sufficiently large set of probability measures on the range space of X_t , endowed with a topology suited to approximations. It is clear that *Information Geometry* should have a role to play in this respect; in fact, one of the most successful implementations of the Kalman–Bucy filter is the so-called *information filter*, in which the natural coordinate system of the exponential manifold of non-singular Gaussian measures on \mathbb{R}^m is used as a parametrisation. (See (16) below.) The aim of this paper is to investigate information geometric representations for nonlinear filters in a more general setting.

We consider a general problem in which $(X_t \in \mathbb{X}, t \geq 0)$ is a Markov signal process taking values in a complete separable metric space \mathbb{X} . This setup includes Markov chains (for which \mathbb{X} is finite), Markov diffusion processes (for which $\mathbb{X} = \mathbb{R}^m$) and function-valued processes (for which \mathbb{X} is a Banach space, e.g. $C([0, 1]; \mathbb{R}^m)$). The observation process $(Y_t \in \mathbb{R}^d, t \geq 0)$ is of the “signal-plus-white-noise” type,

$$Y_t = \int_0^t h_s(X_s) ds + W_t, \quad (3)$$

where $h : [0, \infty) \times \mathbb{X} \rightarrow \mathbb{R}^d$ is a Borel measurable function, and $(W_t \in \mathbb{R}^d, t \geq 0)$ is a standard d -vector Brownian motion, independent of X . Both processes are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ comprising an abstract sample space Ω , a σ -algebra of events (subsets of Ω) \mathcal{F} , and a probability measure $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$. Let $(\mathcal{Y}_t \subset \mathcal{F}, t \geq 0)$ be the *filtration* generated by Y , augmented by the \mathbb{P} -null sets of \mathcal{F} . (For any t , \mathcal{Y}_t is the σ -algebra of events determined by the history $(Y_s, 0 \leq s \leq t)$.) The essential feature of this scenario is that non-overlapping increments of Y are X -conditionally independent of each other. Together with the Markov property of X , this enables the \mathcal{Y}_t -conditional distribution of X_t (its $(Y_s, 0 \leq s \leq t)$ -conditional distribution) to be interpreted as a *prior* distribution in the subsequent Bayesian problem of estimating the future of X . This is what leads to recursive formulae such as (2).

Let \mathcal{X} be the Borel σ -algebra of subsets of \mathbb{X} (the smallest σ -algebra containing the open sets), and let $\mathcal{P}(\mathcal{X})$ be the space of probability measures on \mathcal{X} . (This is of infinite dimension unless \mathbb{X} is finite.) The nonlinear filter furnishes a time-evolving *regular conditional probability distribution* for X_t ; this is a (\mathcal{Y}_t) -adapted, $\mathcal{P}(\mathcal{X})$ -valued stochastic process $(\Pi_t \in \mathcal{P}(\mathcal{X}), t \geq 0)$ having appropriate regularity properties. (See, for example, [3, 4].) Practical implementations of the filter can be based on (\mathcal{Y}_t) -adapted approximations to Π that take values in finite-dimensional subsets of $\mathcal{P}(\mathcal{X})$, $(\hat{\Pi}_t \in \mathcal{Q} \subset \mathcal{P}(\mathcal{X}), t \geq 0)$, and this introduces the issue of approximation errors.

Single estimation objectives, such as mean-square error minimisation in the estimate of a real-valued variate $f(X_t)$, induce their own specific measures of approximation error on $\mathcal{P}(\mathcal{X})$. If several such variates are to be estimated then a more generic measure of approximation error, such as the *L^2 metric on densities* may be useful: if μ is a (reference) probability measure on \mathcal{X} , π_t and $\hat{\pi}_t$ are densities of Π_t and $\hat{\Pi}_t$ with respect to μ , and $\pi_t, \hat{\pi}_t, f \in L^2(\mu)$ then, according to the Cauchy–Schwarz–Bunyakovsky inequality,

$$\mathbb{E}(E_{\Pi_t}f - E_{\hat{\Pi}_t}f)^2 \leq \mathbb{E}E_\mu f^2 E_\mu (\pi_t - \hat{\pi}_t)^2, \quad (4)$$

where, for any $P \in \mathcal{P}(\mathcal{X})$, E_P represents expectation (integration) with respect to P , and \mathbb{E} represents expectation with respect to \mathbb{P} . This *approximation error* is one of two components in the mean-square *estimation error* incurred when $E_{\hat{\Pi}_t}f$ is used as an estimate of $f(X_t)$:

$$\mathbb{E}(f(X_t) - E_{\hat{\Pi}_t}f)^2 = \mathbb{E}E_{\Pi_t}(f - E_{\Pi_t}f)^2 + \mathbb{E}(E_{\Pi_t}f - E_{\hat{\Pi}_t}f)^2. \quad (5)$$

The first term on the right-hand side here is the *statistical error* arising from the limitations of the observation Y ; the second term is the approximation error arising from the use of $\hat{\Pi}_t$ instead of Π_t . In the construction of approximations it is appropriate to consider the size of the second term *relative to that of the first*—there is no point in

approximating $E_{\Pi_t} f$ with great accuracy if it is itself a poor estimate of $f(X_t)$. This reasoning leads to the (somewhat extreme) multi-objective measure of mean-square approximation errors $\mathcal{D}_{MO}(\hat{\Pi}_t | \Pi_t)$, where

$$\begin{aligned}\mathcal{D}_{MO}(Q|P) &:= \frac{1}{2} \sup_{f \in L^2(P)} \frac{(E_Q f - E_P f)^2}{E_P(f - E_P f)^2} \\ &= \frac{1}{2} \sup_{f \in F} (E_P(dQ/dP - 1)f)^2 \\ &= \frac{1}{2} \|dQ/dP - 1\|_{L^2(P)}^2.\end{aligned}\quad (6)$$

Here, F is the subset of $L^2(P)$ whose members have unit variances, and dQ/dP is the density (Radon–Nikodym derivative) of Q with respect to P . ($\mathcal{D}_{MO}(Q|P) = \infty$ if no such density exists.)

\mathcal{D}_{MO} is Pearson's χ^2 -divergence; it is also the α -divergence of [5] with $\alpha = -3$. Although extreme, it illustrates an important feature of multi-objective approximation criteria—they require probabilities of events that are small to be approximated with greater absolute accuracy than those that are large. This is true of \mathcal{D}_{MO} since F contains the scaled indicator functions $f_B = (P(B)(1 - P(B))^{-1/2}\mathbf{1}_B$, where $B \in \mathcal{X}$, so that if $P(B)$ is small

$$(E_Q f_B - E_P f_B)^2 \approx P(B)^{-1} (Q(B) - P(B))^2. \quad (7)$$

The posterior distribution of nonlinear filtering Π_t plays two roles: (i) it is used to compute estimates of the signal at time t ; (ii) it summarises the observations prior to time t in the Bayesian problem of estimating X at all later times. Some nonlinear filtering problems of current interest, such as the tracking of many objects, are inherently multi-objective; however, even if a particular problem has only one objective (such as the estimation of a single variate $f(X_t)$), the need to achieve this at times beyond t places additional constraints on the accuracy of $\hat{\Pi}$ at time t . An event $B \in \mathcal{X}$ for which $\Pi_t(B)$ is small may be the precursor of later events that are supported by new observations, and so small probabilities should not be ignored. Apart from its lack of multi-objectivity in this sense, the L^2 norm on densities introduces boundaries, which can create problems for numerical methods.

The remainder of the paper is structured as follows. Section 2 reviews the use of information geometry in nonlinear filtering, and outlines two infinite-dimensional statistical manifolds that can be used in this context: the exponential Orlicz manifold of [6–9], and the Hilbert manifold of [10, 11]. Section 3 reviews results from [12], which “lifts” the equations of nonlinear filtering to the Hilbert manifold and studies their information theoretic properties. The multi-objective criterion in (6) can be made less extreme if the set of functions over which the mean-square error is minimised is further restricted. This is the subject of Sect. 4, which contains a new result on the regularity of a milder multi-objective measure of approximation errors on the Hilbert manifold.

2 Nonlinear Filtering and Information Geometry

We begin by reviewing the classical finite-dimensional exponential model as it is the inspiration behind later works on the infinite-dimensional theory, and provides an example of the connection between filtering and information geometry. It is developed pedagogically in [5, 13]. Let $(\mathbb{X}, \mathcal{X}, \mu)$ be as defined in Sect. 1, and let λ be a measure (not necessarily a probability measure) defined on \mathcal{X} that is mutually absolutely continuous with respect to μ . Let $(\xi_i, i = 1, \dots, n)$ be real-valued random variables defined on \mathbb{X} , with the following properties: (i) the random variables $(1, \xi_1, \dots, \xi_n)$ represent linearly independent members of $L^0(\mu)$, i.e. $\mu(\alpha + y^* \xi = 0) = 1$ if and only if $\alpha = 0$ and $\mathbb{R}^n \ni y = 0$, where ξ is the n -vector random variable with components ξ_i ; (ii) $\int_{\mathbb{X}} \exp(y^* \xi) d\lambda < \infty$ for all y in a non-empty open subset $G \subset \mathbb{R}^n$. For each $y \in G$, let P_y be the probability measure on \mathcal{X} with density

$$\frac{dP_y}{d\lambda} = \exp(y^* \xi - c(y)), \quad (8)$$

where $c(y) := \log \int \exp(y^* \xi) d\lambda$, and let $N := \{P_y \in \mathcal{P}(\mathcal{X}) : y \in G\}$. It follows from (i) that $G \ni y \mapsto P_y \in N$ is a bijection. Let $\theta : N \rightarrow G$ be its inverse; then (N, θ) is an exponential statistical manifold with an atlas comprising the single chart θ . Let $\eta : N \rightarrow \mathbb{R}^n$ be defined by $\eta(P) = E_P \xi$; then $G_\eta := \eta(N)$ is open and $\eta \circ \theta^{-1} : G \rightarrow G_\eta$ is diffeomorphic, and so η is another global chart, dubbed *expectation parameters* in [5].

The tangent space at $P \in N$, $T_P N$, is the linear space of *derivations* at P , and is spanned by the vectors $(\partial_i, 1 \leq i \leq n)$ where, for any $f \in C^1(N; \mathbb{R})$, $\partial_i f := \partial(f \circ \theta^{-1})/\partial y_i$. The tangent bundle, TN , is a smooth $2n$ -dimensional manifold, trivialised by the charts $\Theta : TN \rightarrow G \times \mathbb{R}^n$ and $H : TN \rightarrow G_\eta \times \mathbb{R}^n$, where

$$\Theta(P, U) = (\theta(P), U\theta) \quad \text{and} \quad H(P, U) = (\eta(P), U\eta). \quad (9)$$

When endowed with the *Fisher–Rao metric*, N becomes a Riemannian manifold; the Fisher–Rao metric at $P \in N$, expressed in θ -coordinates, is

$$g_{ij}(P) = E_P(\xi_i - E_P \xi_i)(\xi_j - E_P \xi_j). \quad (10)$$

The charts in (9) are intimately connected with the *Kullback–Leibler (KL)-divergence* between pairs of probability measures $P, Q \in \mathcal{P}(\mathcal{X})$:

$$\mathcal{D}_1(P|Q) = \mathcal{D}_{-1}(Q|P) := E_P dQ/dP \log(dQ/dP) \in [0, \infty]. \quad (11)$$

When restricted to N , $\mathcal{D}_{\pm 1}$ is finite and of class C^∞ . The charts θ and η are Fenchel–Legendre adjoint variables with respect to $\mathcal{D}_{\pm 1}$ in the sense that, for any $P \in N$, $\mathcal{D}_1(\theta^{-1}|P)$ and $\mathcal{D}_{-1}(\eta^{-1}|P)$ are strictly convex functions,

$$\begin{aligned}\mathcal{D}_{-1}(Q|P) &= \max_{y \in G} \{(\eta(Q) - \eta(P))^*(y - \theta(P)) - \mathcal{D}_1(\theta^{-1}(y)|P)\}, \\ \mathcal{D}_1(Q|P) &= \max_{z \in G_\eta} \{(\theta(Q) - \theta(P))^*(z - \eta(P)) - \mathcal{D}_{-1}(\eta^{-1}(z)|P)\},\end{aligned}\quad (12)$$

and the unique maximisers are $\theta(Q)$ and $\eta(Q)$, respectively. N is dually flat with respect to a pair of connections and their associated covariant derivatives $\nabla^{(1)}$ and $\nabla^{(-1)}$, and θ and η are the associated affine charts [5]: for any differentiable vector fields $\mathbf{U}, \mathbf{V} \in \Gamma TN$,

$$(\nabla_{\mathbf{U}}^{(1)} \mathbf{V})\theta = \mathbf{U}(\mathbf{V}\theta) \quad \text{and} \quad (\nabla_{\mathbf{U}}^{(-1)} \mathbf{V})\eta = \mathbf{U}(\mathbf{V}\eta). \quad (13)$$

Consider the instance of N in which $\mathbb{X} = \mathbb{R}^m$, λ is Lebesgue measure and $n = m(m+3)/2$. Let \mathbb{S} be the set of symmetric $m \times m$ real matrices, let $\mathbb{S}^+ \subset \mathbb{S}$ be the subset of positive definite matrices, and let $(\alpha, \beta) : \mathbb{R}^n \rightarrow \mathbb{R}^m \times \mathbb{S}$ be the linear bijection defined by

$$\begin{aligned}\alpha_i(y) &:= y_i \quad \text{for } 1 \leq i \leq m \\ \beta_{kl}(y) &:= y_i \quad \text{for } 1 \leq l \leq k \leq m, \text{ with } i = m + (k-1)k/2 + l.\end{aligned}\quad (14)$$

If $\xi : \mathbb{X} \rightarrow \mathbb{R}^n$ is defined by

$$y^* \xi(x) = \alpha(y)^* x - \frac{1}{2} x^* \beta(y) x, \quad (15)$$

and $G := (\alpha, \beta)^{-1}(\mathbb{R}^m, \mathbb{S}^+)$ then N comprises all non-singular Gaussian measures on \mathbb{R}^m . (The measure with coordinates $y \in G$ has mean vector $\beta(y)^{-1} \alpha(y)$ and covariance matrix $\beta(y)^{-1}$.) The process of posterior distributions of the Kalman–Bucy filter of (2) satisfies the following Stratonovich stochastic differential equation on N [12]:

$$\circ d\Pi_t = \left(\mathbf{U}(\Pi_t) - \frac{1}{2} \sum_{k=1}^d \nabla_{\mathbf{V}_k}^{(-1)} \mathbf{V}_k(\Pi_t) \right) dt + \sum_{k=1}^d \mathbf{V}_k(\Pi_t) \circ d\nu_{k,t}, \quad (16)$$

where ν is the *innovations process*

$$\nu_t = Y_t - \int_0^t \int_{\mathbb{R}^m} Cx \Pi_s(dx) ds. \quad (17)$$

The vector-fields $\mathbf{U}, \mathbf{V}_k \in \Gamma TN$ are as follows,

$$\begin{aligned}\mathbf{U}\eta &= (\alpha, \beta)^{-1}(B\alpha(\eta), B\beta(\eta) + \beta(\eta)B^* + \Gamma\Gamma^*) \\ \mathbf{V}_k\theta &= (\alpha, \beta)^{-1}(C^*\mathbf{e}_k, 0) \quad \text{for } 1 \leq k \leq d,\end{aligned}\quad (18)$$

where $(\mathbf{e}_k, 1 \leq k \leq d)$ is the coordinate orthonormal basis in \mathbb{R}^d . As (18) shows, \mathbf{U} is η -affine and \mathbf{V}_k is θ -affine (in fact θ -constant). This connection between filtering and equivalents of the maps η and θ is retained in the infinite-dimensional setting.

Apart from the Kalman–Bucy filter, there are a number of *nonlinear* filters whose posterior distributions evolve according to (16), on finite-dimensional exponential manifolds with appropriately defined vector fields. (See Section 4 in [12].) Among these are the Wonham–Shiryayev filter for finite-state Markov chains [14, 15], and Beneš’ classical examples [16]. However, in order to apply these ideas to more typical nonlinear filtering problems, we must use appropriate infinite-dimensional statistical manifolds.

If, in the construction of N , the probability measure μ is used as the reference λ then requirement (ii) (that $\int_{\mathbb{X}} \exp(y^* \xi) d\lambda < \infty$ on an open set $G \subset \mathbb{R}^n$) becomes a statement regarding the existence of moment generating functions with non-empty open domains for the random variables $y^* \xi$, $y \in G$. This property underpins the smoothness of the KL-divergence on N , thereby enabling the construction of the Fisher–Rao metric and α -covariant derivatives. Beginning with this observation, G. Pistone and his co-workers developed an infinite-dimensional statistical manifold of all probability measures that are mutually absolutely continuous with respect to μ [6–9]. Each P in this manifold is associated with its own local chart whose domain is a *maximal exponential model* containing P , $\mathcal{E}(P)$. The chart is

$$s_P(Q) := \log dQ/dP - E_P \log dQ/dP. \quad (19)$$

As (11) shows, the KL-divergence is a bilinear function in the density dQ/dP and its log (loosely regarded as belonging to dual spaces of measurable functions). In order to reproduce the smoothness properties enjoyed by the KL-divergence on N , the charts of any infinite-dimensional manifold must “control” both dQ/dP and its log. That in (19) controls $\log dQ/dP$ directly, but the density is controlled only indirectly (through the exponential function). This requires a model space with a strong topology—the exponential Orlicz space.

The exponential Orlicz manifold is a maximal generalisation of the manifold N . It comprises disjoint maximal exponential models, each having its own atlas of compatible global charts [6]. The KL-divergence is of class C^∞ on each of these. In the context of continuous-time nonlinear filtering, it is appropriate to work with a single maximal exponential model $\mathcal{E}(\mu)$. This contains all probability measures on \mathcal{X} having strictly positive densities, p , with respect to μ , for which

$$E_\mu \exp(\epsilon |\log p|) < \infty \quad \text{for some } \epsilon > 0. \quad (20)$$

If, for example, $\mathbb{X} = \mathbb{R}$ and μ is the standard Gaussian measure, then $\mathcal{E}(\mu)$ contains probability measures with densities of the form $c \exp(-\alpha x^2)$ for positive constants α , but not those with densities of the form $c \exp(-\alpha x^4)$.

The role of the chart θ on N is played on $\mathcal{E}(\mu)$ by the charts s_P of (19), and that of the chart η is played by the “mean parameters” of [8]. Like s_P , these are defined locally at each $P \in \mathcal{E}(\mu)$, by $\eta_P(Q) = dQ/dP - 1$. They take values in the pre-dual

to the exponential Orlicz space. However, despite being injective, the η_P are not homeomorphic and so cannot be used as charts. This prevents the construction of the $\nabla^{(-1)}$ covariant derivative on the tangent bundle $T\mathcal{E}(\mu)$. (Since the chart s_P is affine for $\nabla^{(1)}$, no such problem exists for the latter.) In [7] the authors define a type of α -parallel transport for each $\alpha \in [-1, 1]$. These are defined on “statistical bundles” that differ from the tangent bundle, each modelled on an appropriate Lebesgue or Orlicz space.

The exponential Orlicz manifold was considered as an ambient manifold in the construction of finite-dimensional approximations to measure-valued differential equations, including those of nonlinear filtering, in [17] and the references therein. A method was proposed of projecting the equations onto finite-dimensional submanifolds using the Fisher–Rao metric, thus optimising the resulting finite-dimensional equations with respect to the KL-divergence. The idea was developed in the context of a Hilbert space of square-root densities in [18], and comparisons were made with moment matching techniques. However, the KL-divergence is discontinuous on this Hilbert space [10], and so it is not amenable to the study of multi-objective approximation errors.

A different approach to infinite-dimensional information geometry was developed in [10, 11]. The principal idea is to use a “balanced chart” that directly controls the density $dP/d\mu$ and its log, and so enable the use of the simplest infinite-dimensional model space, the Hilbert space. The Hilbert manifold M comprises all probability measures $P \in \mathcal{P}(\mathcal{X})$ having densities p with respect to μ , for which

$$E_\mu p^2 < \infty \quad \text{and} \quad E_\mu \log^2 p < \infty. \quad (21)$$

Let H be the Hilbert space of centred (zero mean), square-integrable random variables on $(\mathbb{X}, \mathcal{X}, \mu)$; then the map $\phi : M \rightarrow H$, defined by

$$\phi(P) = p - 1 + \log p - E_\mu \log p, \quad (22)$$

is a bijection onto H (Proposition 2.1 in [10]), and so can be considered to be a chart that induces the structure of a Hilbert space on M . The inverse of ϕ takes the form $\phi^{-1}(a) = \psi(\rho(a))$, where $\psi : \mathbb{R} \rightarrow (0, \infty)$ is the inverse of the function $(0, \infty) \ni y \mapsto y - 1 + \log y \in \mathbb{R}$, and $\rho : H \rightarrow L^2(\mu)$ is defined as follows:

$$\rho(a) = a + Z(a). \quad (23)$$

(Here $Z : H \rightarrow (-\infty, 0]$ is the unique function for which $E_\mu \psi(a + Z(a)) = 1$.)

M is a *generalised exponential model* (in the sense of [19]), in which ψ replaces the exponential function of $\mathcal{E}(\mu)$. ψ is convex and has bounded derivatives of all orders. Members of M have densities that are restricted both in their behaviour when small, where the balanced chart becomes the log-density, and in their behaviour when large, where the balanced chart becomes the density.

A tangent vector U at $P \in M$ is an equivalence class of differentiable curves at P : two curves $(P_t \in M, -\epsilon < t < \epsilon)$ and $(Q_t \in M, -\epsilon < t < \epsilon)$ being equivalent at

P if $P_0 = Q_0 = P$ and $(d/dt)\phi(P_t)|_{t=0} = (d/dt)\phi(Q_t)|_{t=0}$. We denote the tangent space at P by $T_P M$ and the tangent bundle by TM . The latter is globally trivialised by the chart $\Phi : TM \rightarrow H \times H$, where

$$\Phi(P, U) = (\phi(P), (d/dt)\phi(P_t)) \quad \text{for } (P_t, -\epsilon < t < \epsilon) \in U. \quad (24)$$

Let $m : M \rightarrow H$ and $e : M \rightarrow H$ be as follows:

$$m(P) = p - 1 \quad \text{and} \quad e(P) = \log p - E_\mu \log p. \quad (25)$$

Like the map η_P on $\mathcal{E}(\mu)$, m and e are injective but not homeomorphic, and so cannot be used as charts. The same is true of the maps induced on the tangent bundle: $TN \ni (P, U) \mapsto (m(P), Um) \in H^2$ and $TN \ni (P, U) \mapsto (e(P), Ue) \in H^2$ are injective but not homeomorphic. (See Section 3 in [10].) Nevertheless, they can be used to define ± 1 -parallel transport on Hilbert statistical bundles in the same way as η_P on $\mathcal{E}(\mu)$. (This, along with α -parallel transports for $\alpha \in (-1, 1)$, is established via the embedding of M in a Hilbert manifold of finite measures in [11].) Another interpretation of a tangent vector $U \in T_P M$ is that it represents a *signed measure* \tilde{U} on \mathcal{X} having a density $d\tilde{U}/d\mu = u_m := Um$; this has the following properties: $u_m \in H$ and $u_m/p \in L^2(\mu)$.

Like η and θ of N , the maps m and e are adjoint variables in a Fenchel–Legendre transform involving the KL-divergence (Proposition 4.2 in [10]). In particular,

$$\begin{aligned} \mathcal{D}_1(Q|P) + \mathcal{D}_{-1}(Q|P) &= \langle m(Q) - m(P), e(Q) - e(P) \rangle_H, \\ \|m(Q) - m(P)\|_H^2 + \|e(Q) - e(P)\|_H^2 &\leq \|\phi(Q) - \phi(P)\|_H^2, \end{aligned} \quad (26)$$

and these lead to the global bound

$$\mathcal{D}_1(Q|P) + \mathcal{D}_{-1}(Q|P) \leq \frac{1}{2} \|\phi(Q) - \phi(P)\|_H^2. \quad (27)$$

Since \mathbb{X} is separable, H is of countable dimension, and any $(P, U) \in TM$ can be represented in terms of its components:

$$\phi_i(P) := \langle \phi(P), \xi_i \rangle_H \quad \text{and} \quad u_i := \langle U\phi, \xi_i \rangle_H, \quad (28)$$

where $(\xi_i \in H, i \in \mathbb{N})$ is an orthonormal basis for H . In particular, the tangent space $T_P M$ is spanned by the vectors $(D_i, i \in \mathbb{N})$, where $(P, D_i) = \Phi^{-1}(\phi(P), \xi_i)$.

The KL-divergence is of class C^1 on $M \times M$. It also admits a mixed second derivative, enabling the construction of the Fisher–Rao metric on TM ; this takes the form

$$\langle U, V \rangle_P = \langle Um, Ve \rangle_H = \sum_{i,j=1}^{\infty} G(P)_{ij} u_i v_j, \quad (29)$$

where $G(P)_{ij} := \langle D_i, D_j \rangle_P = \langle D_i m, D_j e \rangle_H$. (The series here converge since the Fisher-Rao metric is bounded by the model space metric on all fibres of the tangent bundle.) The tensor of (29) is a *weak* Riemannian metric: it is positive definite, but the maps $\langle U, \cdot \rangle_P : T_P M \rightarrow \mathbb{R}$, for $U \in T_P M$, do not account for all the continuous linear maps from $T_P M$ to \mathbb{R} . The use of the balanced chart of (22) with the Lebesgue model space $L^\lambda(\mu)$ (for $\lambda > 2$) leads to a Banach manifold on which the KL-divergence is of class $C^{\lceil \lambda \rceil - 1}$ [11]; this allows α -covariant derivatives to be constructed on statistical bundles, as in [7].

In the context of nonlinear filtering, the Hilbert manifold has a number of advantages:

- its image through the chart is the entire Hilbert space H ;
- it admits the global bound (27), and its natural inner product is an upper bound on the Fisher-Rao metric on all fibres of the tangent bundle;
- it admits the L^2 - l^2 isometry described above;
- it fits within the most elegant theory of stochastic calculus, the L^2 theory;
- its chart is balanced between the two fundamental affine structures of nonlinear filtering, and affords them equal importance (26).

As sets, neither $\mathcal{E}(\mu)$ nor M is contained within the other: $\mathcal{E}(\mu)$ places a more stringent integrability constraint on $\log p$, whereas M places a more stringent constraint on the density p . The submanifold $\mathcal{E}_2(\mu) \subset \mathcal{E}(\mu)$, comprising those $P \in \mathcal{E}(\mu)$ for which $2s_\mu(P) \in s_\mu(\mathcal{E}(\mu))$, is a subset of M but typically not a topological embedding, as the example in which $\mathbb{X} = (0, 1)$ and μ is Lebesgue measure illustrates. For each $n \in \mathbb{N}$, let P_n be the probability measure on $(0, 1)$ with density $p_n = \exp(a_n)/E_\mu \exp(a_n)$, where $a_n(x) = -\mathbf{1}_{(n^{-1}, 1)}(x)x^{-1/3}$. Then $P_n \in \mathcal{E}_2(\mu)$ for all n ; however, while the sequence (P_n) converges in M (to the probability measure whose density is the pointwise limit of p_n), it does not converge in $\mathcal{E}_2(\mu)$. Both $\mathcal{E}(\mu)$ and M admit large classes of embedded finite-dimensional submanifolds on which numerical calculations can be performed.

3 Nonlinear Filtering on the Hilbert Manifold

This section develops the equations of nonlinear filtering on M , and investigates their properties. It is based on [12], where detailed hypotheses and proofs can be found. In particular, the signal X is a Markov process evolving in the complete separable metric space $(\mathbb{X}, \mathcal{X}, \mu)$ of Sect. 1, and the observation process Y satisfies (3). The distribution of X_t is assumed to have a density with respect to μ , $p_t = dP_t/d\mu$, satisfying a differential equation of the Kolmogorov forward (Fokker-Planck) type:

$$\frac{\partial p_t}{\partial t} = \mathcal{A}_t p_t, \quad (30)$$

where $(\mathcal{A}_t, t \geq 0)$ is a family of linear operators on an appropriate class of functions $f : \mathbb{X} \rightarrow \mathbb{R}$. For example, if $\mathbb{X} = \mathbb{R}^m$, X is a time-homogeneous m -dimensional diffusion process with drift vector $b(X_t)$ and diffusion matrix $a(X_t)$, and μ is mutually absolutely continuous with respect to Lebesgue measure (with density r) then, for any $f \in C^2(\mathbb{X}; \mathbb{R})$,

$$\mathcal{A}_t f = \mathcal{A} f = \frac{1}{2r} \sum_{i,j=1}^m \frac{\partial^2(a_{ij}rf)}{\partial x_i \partial x_j} - \frac{1}{r} \sum_{i=1}^m \frac{\partial(b_i rf)}{\partial x_i}. \quad (31)$$

The nonlinear filtering literature contains many results establishing, in particular instances of X , that the posterior distribution Π_t also has a density with respect to μ , π_t , and that this satisfies an Itô stochastic differential equation of the form

$$\pi_t = p_0 + \int_0^t \mathcal{A}_s \pi_s ds + \int_0^t \pi_s (h_s - \bar{h}_s)^* d\nu_s, \quad (32)$$

where $\bar{h}_t := E_{\Pi_t} h_t$, and $(\nu_t, t \geq 0)$ is the *innovations process*

$$\nu_t := Y_t - \int_0^t \bar{h}_s ds. \quad (33)$$

Rigorous results of this nature are full of technical detail, which will not be given here. In the instance that X is a diffusion process, (32) is called the *Kushner–Stratonovich equation* [3, 4]; if X is a finite-state jump process then it becomes the Shirayev–Wonham equation [14, 15]. The innovations process is a standard d -dimensional Brownian Motion with respect to the observation history (\mathcal{Y}_t) [4]. Its future increments are independent of that history, and so carry only information about X that is not already known by the filter.

The main result in [12] (Proposition 3.1) “lifts” (32) to the Hilbert manifold M . It is based on numerous technical hypotheses, which will not be stated here. Under these, Π remains on M at all times, and $\phi(\Pi)$ satisfies the following infinite-dimensional Itô stochastic differential equation:

$$\phi(\Pi_t) = \phi(P_0) + \int_0^t (u_s - \zeta_s) ds + \sum_{k=1}^d \int_0^t v_{k,s} d\nu_{k,s}, \quad (34)$$

where

$$u_t := \Lambda(1 + \pi_t^{-1})\mathcal{A}_t \pi_t, \quad \zeta_t := \frac{1}{2} \Lambda |h_t - \bar{h}_t|^2, \quad v_{k,t} := \Lambda(1 + \pi_t)(h_t - \bar{h}_t)_k. \quad (35)$$

(See [20] for a pedagogic treatment of the Itô calculus in infinite dimensions.) Here, $\Lambda : \mathcal{L}^0(\mathbb{X}; \mathbb{R}) \rightarrow H$ is an operator that lifts \mathcal{X} -measurable functions to H :

$$\begin{aligned} \Lambda f &\ni f - E_\mu f & \text{if } E_\mu f^2 < \infty \\ \Lambda f &= 0 & \text{otherwise.} \end{aligned} \quad (36)$$

(We need to distinguish between the *equivalence class* of zero-mean square-integrable functions, H , and the set of instances of such functions, $\mathcal{L}_0^2(\mathbb{X}; \mathbb{R})$, since stochastic processes comprising the latter may not have suitable measurability properties.)

The processes u , ζ and v_k of (35) are not expressed in terms of vector fields of M , and so (34) does not represent an explicit way of computing Π . However, if the observation function h is bounded then ζ and v_k can be represented in terms of time-dependent, locally-Lipschitz-continuous vector fields $\mathbf{Z}, \mathbf{V}_k \in \Gamma TM$ with ϕ -representations $\mathbf{z}, \mathbf{v}_k : [0, \infty) \times H \rightarrow H$, where

$$\begin{aligned} \mathbf{z}_t(a) &= \mathbf{Z}_t(\phi^{-1}(a))\phi = \Lambda \left(\frac{1}{2}|h_t - E_Ph_t|^2 + (1+p)(h_t - E_Ph_t)^*E_Ph_t \right) \\ \mathbf{v}_{k,t}(a) &= \mathbf{V}_{k,t}(\phi^{-1}(a))\phi = \Lambda(1+p)(h_t - E_Ph_t)_k, \end{aligned} \quad (37)$$

and $P = \phi^{-1}(a)$. Except in special cases, u presents more of a problem. This is because the *infinitesimal* characterisation of P in (30) is typically dependent on the topology of \mathbb{X} . In order to establish an evolution equation for Π , it would be necessary to incorporate this topology into a statistical manifold. Ongoing work in this area can be found in [21–23] and other papers in this volume. However, for the purposes of constructing *approximations* to Π based on finite-dimensional submanifolds $N \subset M$, it suffices to consider *projections* of u , ζ and v_k onto TN . Provided these can be represented in terms of suitably regular vector fields of N then they form the basis for explicit approximations. In this context, the value of defining exact and approximate posterior distributions on a common infinite-dimensional statistical manifold is that it enables multi-objective approximation errors to be quantified, thereby allowing optimal choices to be made of submanifolds of a given class and dimension, for particular filtering problems. If direct computations of the infinite-dimensional equations were required, then it would almost certainly be better to base them on equations of *Zakai* type [3], which are satisfied by *un-normalised* versions of the conditional density. A suitable Hilbert manifold of finite measures is developed in [11].

Apart from providing a starting point for the development of approximations, (34) can be used to study the information-theoretic properties of nonlinear filters, which are of interest in the *stochastic dynamics* approach to nonequilibrium statistical mechanics [24–26]. The central quantities of information theory are the *mutual information* between two random variables, and its conditional variant. If $U : \Omega \rightarrow \mathbb{U}$ and $V : \Omega \rightarrow \mathbb{V}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in complete, separable metric spaces \mathbb{U} and \mathbb{V} , respectively, and $\mathcal{G} \subset \mathcal{F}$ is a sub- σ -algebra of \mathcal{F} , then the \mathcal{G} -conditional mutual information between U and V is defined as follows,

$$I(U; V | \mathcal{G}) := \mathcal{D}_{-1}(P_{UV|\mathcal{G}} | P_{U|\mathcal{G}} \otimes P_{V|\mathcal{G}}), \quad (38)$$

where $P_{UV|\mathcal{G}}$ is the regular \mathcal{G} -conditional joint distribution of U and V , and $P_{U|\mathcal{G}}$ and $P_{V|\mathcal{G}}$ are the regular \mathcal{G} -conditional marginals. (The conditional mutual information is often defined to be the mean value of this quantity [27].) The (unconditional) mutual information is recovered from (38) if \mathcal{G} is the trivial σ -algebra $\{\emptyset, \Omega\}$.

The mutual information between the signal X and observation history $(Y_s, 0 \leq s \leq t)$ was first computed in [28]. We state, here, a conditional variant of this result (see Eqs. (3.21) and (3.22) in [12]):

$$I(X; Y_s^t | \mathcal{Y}_s) = \frac{1}{2} \int_s^t E_\mu \pi_r |h_r - \bar{h}_r|^2 dr, \quad (39)$$

where $Y_s^t := (Y_r - Y_s, s \leq r \leq t)$. The integrand here is the (random) rate at which the nonlinear filter acquires new information about X through Y . This is related to the *quadratic variation* of the filter process Π in the Fisher–Rao metric. If a stochastic process Z evolves in a Banach space then its quadratic variation can be defined by the following limit (where it exists)

$$[Z]_t := \lim_{\delta \downarrow 0} \sum_{n=1}^{\lceil \delta^{-1} t \rceil} \|Z_{n\delta} - Z_{(n-1)\delta}\|^2. \quad (40)$$

Greater care has to be exercised with manifold-valued processes. The usual definition of quadratic variation involves the *horizontal lift* of a stochastic process to the frame bundle according to a suitable connection, followed by its *anti-development* into a process taking values in the model space [29]. Since the Hilbert manifold, M , is covered by a single chart, the representation $\phi(\Pi)$ is itself the anti-development of Π with respect to the horizontal lift associated with the natural (chart-induced) connection on M . Π_t can be represented in terms of its components $(\phi_i(\Pi_t), i \in \mathbb{N})$, as defined in (28). Its quadratic variation in the Fisher–Rao metric can then be expressed in terms of the quadratic covariations of these components:

$$[\Pi]_t := \int_0^t \sum_{i,j=1}^{\infty} G(\Pi_s)_{ij} d[\phi_i(\Pi), \phi_j(\Pi)]_s. \quad (41)$$

where G is as defined in (29). A straightforward calculation establishes the connection between the information supply of (39) and $[\Pi]$ (Proposition 3.2 in [12]):

$$I(X; Y_s^t | \mathcal{Y}_s) = \frac{1}{2} \mathbb{E}([\Pi]_t - [\Pi]_s | \mathcal{Y}_s). \quad (42)$$

Over a short time interval $[s, s + \delta]$ the observation process “supplies” to the filter the small quantity of new information $([\Pi]_{s+\delta} - [\Pi]_s)/2$. If the filter process Π were differentiable then the filter would “know” this information at time s , thereby contradicting its novelty. Since it satisfies a stochastic differential equation driven

by the continuous martingale ν , Π is not differentiable; its sample paths are fractals, and the information supply to the filter is dependent on the quadratic variation of these sample paths.

4 The Ψ Multi-objective Criterion

This section develops a less extreme version of the multi-objective measure of approximation errors of (6), and investigates its properties on the Hilbert manifold M . A milder measure is obtained if additional constraints are placed on membership of the set F in (6). For example, we might consider the subset

$$F_{n,K} := \{f \in F : \|f\|_{L^{2n}(P)} \leq K\}, \quad (43)$$

for some $n \in \mathbb{N}$ and a suitable constant $1 < K < \infty$. In the context of the indicator functions of (7), this reduces the scaling factor from $P(B)^{-1}$ to $P(B)^{-1/n}$.

Consider the example in which $\mathbb{X} = \mathbb{R}^m$ and P is a Gaussian measure with non-singular covariance matrix Σ . The set $F_{n,K}$ then includes functions of the type $f(x) = \exp(\alpha|x|^2)$ for sufficiently small $\alpha > 0$, and the effect of increasing n is only to reduce the permissible α in inverse proportion. The fact that these multi-objective criteria include functions with this high rate of growth suggests that they are still too sharp. Even if an idealised nonlinear filtering problem has posterior distributions with Gaussian tails, the underlying physical problem is unlikely to be modelled accurately in the outer reaches of \mathbb{R}^m . In view of this, it would seem natural to further restrict membership of the set F of (6), for example by requiring functions to be bounded in an exponential Orlicz space:

$$F_{\Phi,K} := \{f \in F : \|f\|_{L^\Phi(P)} \leq K\}. \quad (44)$$

The norm here is the *Luxembourg* norm on the exponential Orlicz space $L^\Phi(P)$:

$$\|f\|_{L^\Phi(P)} := \inf \{t > 0 : E_P \Phi(t^{-1}f) \leq 1\}, \quad (45)$$

where $\Phi : \mathbb{R} \rightarrow [0, \infty)$ is the following Young function:

$$\Phi(y) := \cosh(y) - 1. \quad (46)$$

(See, for example, [30].) In the Gaussian example, this would constrain the rate of growth of f to be at most quadratic, and $F_{\Phi,K}$ would contain quadratic functions if K were sufficiently large. (If $m = 1$ and P is the standard Gaussian measure, then the function $f(x) = x^2/\sqrt{3}$ has unit variance and $\|f\|_{L^\Phi(P)} < 3/2$.) The resulting multi-objective measure of approximation errors is:

$$\bar{\mathcal{M}}(Q, P) := \frac{1}{2} \sup_{f \in F_{\phi, K}} (E_P(dQ/dP - 1)f)^2, \quad (47)$$

which cannot be expressed as succinctly as was D_{MO} in (6). However, $\bar{\mathcal{M}}$ can be bounded by another multi-objective criterion defined in terms of the complementary Orlicz space $L^\Psi(P)$.

Definition 1 (*The Ψ multi-objective criterion*) For any $P, Q \in \mathcal{P}(\mathcal{X})$, let

$$\mathcal{M}(Q, P) := \begin{cases} \frac{1}{2} \|dQ/dP - 1\|_{L^\Psi(P)}^2 & \text{if } Q \ll P \\ +\infty & \text{otherwise,} \end{cases} \quad (48)$$

where $\|\cdot\|_{L^\Psi(P)}$ is the Luxembourg norm on the Orlicz space $L^\Psi(P)$, defined by

$$\|g\|_{L^\Psi(P)} := \inf \{t > 0 : E_P \Psi(t^{-1}g) \leq 1\}, \quad (49)$$

and $\Psi : \mathbb{R} \rightarrow [0, \infty)$ is the complementary Young function:

$$\Psi(z) := z \sinh^{-1}(z) - \sqrt{z^2 + 1} + 1. \quad (50)$$

It follows from the generalised Holder inequality [30] that

$$\bar{\mathcal{M}}(Q, P) \leq 2 \sup_{f \in F_{\phi, K}} \|f\|_{L^\Psi(P)}^2 \|q/p - 1\|_{L^\Psi(P)}^2 \leq 4K^2 \mathcal{M}(Q, P). \quad (51)$$

The following proposition, which is presented here for the first time, establishes the regularity of \mathcal{M} on M .

Proposition 1 (i) For any $P \in M$, any sequence $(Q_n, P_n) \rightarrow (P, P)$ with $Q_n \neq P_n$, and any $0 \leq \gamma < 2$

$$\|\phi(Q_n) - \phi(P_n)\|_H^{-\gamma} \mathcal{M}(Q_n, P_n) \rightarrow 0. \quad (52)$$

(ii) $\mathcal{M} \in C^1(M \times M; [0, \infty))$, and has the following derivative:

$$(V, U)\mathcal{M}_{Q, P} = \begin{cases} 0 & \text{if } Q = P \\ \|q/p - 1\|_{L^\Psi(P)} \frac{E_\mu \sinh^{-1}(\theta)(v_m - u_m)}{E_P \sqrt{\theta^2 + 1}} \\ -\|q/p - 1\|_{L^\Psi(P)}^2 \frac{E_\mu \sqrt{\theta^2 + 1} u_m}{E_P \sqrt{\theta^2 + 1}} & \text{otherwise,} \end{cases} \quad (53)$$

where $\theta := \|q/p - 1\|_{L^\Psi(P)}^{-1}(q/p - 1)$, u_m and v_m are the m -representations of the tangent vectors $U \in T_P M$ and $V \in T_Q M$,

$$u_m = Um_P = \frac{p}{1+p} \rho_{\phi(P)}^{(1)} U \phi_P, \quad v_m = Vm_Q = \frac{q}{1+q} \rho_{\phi(Q)}^{(1)} V \phi_Q, \quad (54)$$

and $\rho : H \rightarrow L^2(\mu)$ is as defined in (23). (See Section 3 in [10].)

Proof We shall make use of the inequality: for any $\sigma \in \mathbb{R}$ and $\tau \in (0, \infty)$,

$$|\sinh^{-1}(\sigma/\tau)| = -\log \tau + \log \left(|\sigma| + \sqrt{\sigma^2 + \tau^2} \right) \leq |\log \tau| + 2|\sigma| + \tau. \quad (55)$$

Using this with $\sigma = q(x) - p(x)$ and $\tau = tp(x)$ we see that, for any $P, Q \in M$ and any $t > 0$

$$E_P \Psi(t^{-1}(q/p - 1)) < \infty,$$

so that \mathcal{M} is finite on $M \times M$.

Let P_n, Q_n and γ be as in (i), and let $h_n := 2\|q_n/p_n - 1\|_{L^\psi(P)}$. Suppose that there exists an $\epsilon > 0$ for which $\|q_n - p_n\|_H^{-\gamma} h_n^2 > \epsilon$ infinitely often, and let $\alpha_n := (\epsilon \|q_n - p_n\|_H^\gamma)^{-1/2}$. Restricting attention to the subsequence for which this is so, we have

$$\begin{aligned} E_{P_n} \Psi(h_n^{-1}(q_n/p_n - 1)) &\leq E_{P_n} \Psi(\alpha_n(q_n/p_n - 1)) \\ &\leq \alpha_n E_\mu |q_n - p_n| \sinh^{-1}(\alpha_n |q_n/p_n - 1|) \\ &\leq \alpha_n \|q_n - p_n\|_H \| |\log p_n| + 2\alpha_n |q_n - p_n| + p_n \|_{L^2(\mu)} \\ &\rightarrow 0, \end{aligned}$$

where we have used the Cauchy–Schwarz–Bunyakovsky inequality and (55) with $\sigma = \alpha_n(q_n - p_n)$ and $\tau = p_n$ in the third step. However, this contradicts the definition of $\|\cdot\|_{L^\psi(P)}$, and so no such ϵ exists and

$$(\|b_n - a\|_H + \|a_n - a\|_H)^{-\gamma} h_n^2 \leq \|b_n - a_n\|_H^{-\gamma} h_n^2 \leq \|q_n - p_n\|_H^{-\gamma} h_n^2 \rightarrow 0,$$

where $a = \phi(P)$, $a_n = \phi(P_n)$, $b_n = \phi(Q_n)$ and we have used the triangle inequality in the first step, and the inequality in (26) in the second step. This proves part (i) and (53) in the special case that $Q = P$.

Let $F : (0, \infty)^3 \rightarrow \mathbb{R}$ and $G : L^2(\mu) \times L^2(\mu) \times (0, \infty) \rightarrow \mathbb{R}$ be as follows:

$$F(y, z, t) := tz \left(\Psi(t^{-1}(y/z - 1)) - 1 \right) \quad \text{and} \quad G(\tilde{b}, \tilde{a}, t) := E_\mu F(\psi(\tilde{b}), \psi(\tilde{a}), t).$$

The partial derivatives of F are:

$$\begin{aligned} \frac{\partial F}{\partial y} &= \sinh^{-1}(t^{-1}(y/z - 1)), \quad \frac{\partial F}{\partial z} = -\frac{\partial F}{\partial y} - t\sqrt{t^{-2}(y/z - 1)^2 + 1}, \\ \frac{\partial F}{\partial t} &= -z\sqrt{t^{-2}(y/z - 1)^2 + 1}. \end{aligned}$$

Let $\tilde{a}_n, \tilde{b}_n \in L^2(\mu)$ and $t_n \in (0, \infty)$ be sequences converging to \tilde{a}, \tilde{b} and t , respectively, and let $f_i : L^2(\mu) \times L^2(\mu) \times (0, \infty) \rightarrow L^0(\mu)$ ($i = 1, 2, 3, 4$) be defined as follows:

$$\begin{aligned} f_1(\tilde{b}, \tilde{a}, t) &:= \frac{\partial F}{\partial y}(\psi(\tilde{b}), \psi(\tilde{a}), t)\psi'(\tilde{b}), \quad f_2(\tilde{b}, \tilde{a}, t) := -\frac{\partial F}{\partial y}(\psi(\tilde{b}), \psi(\tilde{a}), t)\psi'(\tilde{a}), \\ f_3(\tilde{b}, \tilde{a}, t) &:= \left(\frac{\partial F}{\partial y} + \frac{\partial F}{\partial z} \right)(\psi(\tilde{b}), \psi(\tilde{a}), t)\psi'(\tilde{a}), \quad f_4(\tilde{b}, \tilde{a}, t) := \frac{\partial F}{\partial t}(\psi(\tilde{b}), \psi(\tilde{a}), t). \end{aligned}$$

In order to show that $G \in C^1(L^2(\mu) \times L^2(\mu) \times (0, \infty); \mathbb{R})$ with derivative

$$G_{\tilde{b}, \tilde{a}, t}^{(1)}(\tilde{v}, \tilde{u}, s) = E_\mu f_1(\tilde{b}, \tilde{a}, t)\tilde{v} + E_\mu(f_2 + f_3)(\tilde{b}, \tilde{a}, t)\tilde{u} + E_\mu f_4(\tilde{b}, \tilde{a}, t)s, \quad (56)$$

it suffices to show that $f_i(\tilde{b}_n, \tilde{a}_n, t_n) \rightarrow f_i(\tilde{b}, \tilde{a}, t)$ in $L^2(\mu)$, for $i = 1, 2, 3, 4$. (This allows the mean-value theorem to be applied to $F(\psi(\tilde{b}(x)), \psi(\tilde{a}(x)), t)$, and the resulting remainder term to be bounded by means of the Cauchy–Schwarz–Bunyakovsky inequality.) Now

$$f_1(\tilde{b}_n, \tilde{a}_n, t_n) - f_1(\tilde{b}, \tilde{a}, t) = f_{11} + f_{12} + f_{13}(\tilde{q}_n, \tilde{p}_n, t_n) - f_{13}(\tilde{q}, \tilde{p}, t),$$

where

$$\begin{aligned} f_{11} &:= \frac{\tilde{q}_n}{1 + \tilde{q}_n} (\log(t\tilde{p}) - \log(t_n\tilde{p}_n)), \quad f_{12} := \log(t\tilde{p}) \left(\frac{\tilde{q}}{1 + \tilde{q}} - \frac{\tilde{q}_n}{1 + \tilde{q}_n} \right), \\ f_{13}(y, z, s) &:= \frac{y}{1 + y} \log \left(y - z + \sqrt{(y - z)^2 + (sz)^2} \right), \end{aligned}$$

and $\tilde{p}, \tilde{q}, \tilde{p}_n$ and \tilde{q}_n are the densities of finite measures on \mathcal{X} . (\tilde{P} is a probability measure if and only if $\tilde{a} \in \rho(H)$. See Section 4 in [11].) Since ψ is Lipschitz continuous (with constant 1) and $\log \psi(\tilde{a}_n) = \tilde{a}_n + 1 - \psi(\tilde{a}_n)$, both \tilde{p}_n and $\log \tilde{p}_n$ converge in $L^2(\mu)$ (to \tilde{p} and \tilde{q} , respectively). So $f_{11} \rightarrow 0$ in $L^2(\mu)$, and since $f_{12} \rightarrow 0$ in probability and is dominated by the square-integrable function $2|\log(t\tilde{p})|$, $f_{12} \rightarrow 0$ in $L^2(\mu)$ as well. Similarly, $f_{13}(\tilde{q}_n, \tilde{p}_n, t_n) \rightarrow f_{13}(\tilde{q}, \tilde{p}, t)$ in probability and f_{13} can be bounded in the following way:

$$\frac{y}{1 + y} \log \left(\frac{2s^2y}{s^2 + 1} \right) \leq f_{13}(y, z, s) \leq 2|y - z|^{1/2} + (sz)^{1/2} \quad \text{for all } z \in (0, \infty),$$

and so

$$\sup_n E_\mu f_{13}(\tilde{q}_n, \tilde{p}_n, t_n)^4 < \infty,$$

and the de-la-Vallée-Poussin theorem completes the proof that $f_1(\tilde{b}_n, \tilde{a}_n, t_n) \rightarrow f_1(\tilde{b}, \tilde{a}, t)$ in $L^2(\mu)$. The proof that $f_2(\tilde{b}_n, \tilde{a}_n, t_n) \rightarrow f_2(\tilde{b}, \tilde{a}, t)$ in $L^2(\mu)$ is similar. Now

$$f_3(\tilde{b}_n, \tilde{a}_n, t_n) - f_3(\tilde{b}, \tilde{a}, t) = \left(\frac{1}{1 + \tilde{p}} - \frac{1}{1 + \tilde{p}_n} \right) \sqrt{(\tilde{q} - \tilde{p})^2 + (t\tilde{p})^2} + \xi_n, \quad (57)$$

where

$$\xi_n := \frac{1}{1 + \tilde{p}_n} \left(\sqrt{(\tilde{q} - \tilde{p})^2 + (t\tilde{p})^2} - \sqrt{(\tilde{q}_n - \tilde{p}_n)^2 + (t_n\tilde{p}_n)^2} \right).$$

The first term on the right-hand side of (57) converges to zero in probability, and since it is dominated by the square-integrable function $2\sqrt{(\tilde{q} - \tilde{p})^2 + (t\tilde{p})^2}$, it also converges in $L^2(\mu)$; moreover

$$\begin{aligned} |\xi_n| &\leq |(\tilde{q} - \tilde{p}) - (\tilde{q}_n - \tilde{p}_n)| + |t\tilde{p} - t_n\tilde{p}_n| \\ &\leq |\tilde{q} - \tilde{q}_n| + |\tilde{p} - \tilde{p}_n| + t|\tilde{p} - \tilde{p}_n| + |t - t_n|\tilde{p}_n, \end{aligned}$$

which completes the proof that $f_3(\tilde{b}_n, \tilde{a}_n, t_n) \rightarrow f_3(\tilde{b}, \tilde{a}, t)$ in $L^2(\mu)$. The proof that $f_4(\tilde{b}_n, \tilde{a}_n, t_n) \rightarrow f_4(\tilde{b}, \tilde{a}, t)$ in $L^2(\mu)$ is similar.

We have thus shown that $G \in C^1(L^2(\mu) \times L^2(\mu) \times (0, \infty); \mathbb{R})$ with derivative as in (56). It is shown in Proposition 4.1 in [11] that the map ρ is of class C^1 on H , and so $G(\rho(b), \rho(a), \cdot) \in C^1(H \times H \times (0, \infty); \mathbb{R})$. For any $b \neq a \in H$, the monotone convergence theorem shows that

$$\lim_{t \uparrow \infty} t^{-1} G(\rho(b), \rho(a), t) = -1;$$

furthermore, Jensen's inequality shows that

$$G(\rho(b), \rho(a), t) \geq t(\Psi(t^{-1}E_\mu|q - p|) - 1) \rightarrow \infty \text{ as } t \downarrow 0.$$

Together with the fact that $\partial G / \partial t = E_\mu f_4(\rho(b), \rho(a), t) < -1$, these limits show that there is a unique value of t (namely $\|q/p - 1\|_{L^\Psi(P)}$) for which $G(\rho(b), \rho(a), t) = 0$. The implicit mapping theorem completes the proof of (53) for the case $Q \neq P$.

It remains to show that, for any $a \in H$ and any sequence $(b_n, a_n) \rightarrow (a, a)$, $\mathcal{M}(\phi^{-1}, \phi^{-1})_{b_n, a_n}^{(1)} \rightarrow 0$. Let $h_n := \|q_n/p_n - 1\|_{L^\Psi(P)}$, and let $\theta_n := h_n^{-1}(q_n/p_n - 1)$. Using (55) with $\sigma = q_n - p_n$ and $\tau = h_n p_n$, we obtain

$$\begin{aligned} \|\sinh^{-1}(\theta_n)\|_{L^2(\mu)} &\leq \|\log(h_n p_n)\|_{L^2(\mu)} + 2\|q_n - p_n\|_{L^2(\mu)} + h_n \|p_n\|_{L^2(\mu)} \\ &\leq K_1(1 + |\log h_n|); \end{aligned}$$

furthermore

$$\left\| \sqrt{\theta_n^2 + 1} \frac{p_n}{1 + p_n} \right\|_{L^2(\mu)} \leq \|h_n^{-1}(q_n - p_n)^2 + p_n^2\|_{L^1(\mu)}^{1/2} \leq K_2(1 + h_n^{-1}).$$

Proposition 4.1(iii) in [11] and the inequality in (26) show that

$$\sup_n \|\rho_{a_n}^{(1)}\|_{H^*} \leq K_3 \quad \text{and} \quad \|(m \circ \phi^{-1})_a^{(1)}\|_{H^*} \leq 1 \quad \text{for all } a \in H,$$

and so

$$\left\| \mathcal{M}(\phi^{-1}, \phi^{-1})_{b_n, a_n}^{(1)} \right\|_{(H \times H)^*} \leq h_n K_1(1 + |\log h_n|) + h_n^2 K_2 K_3(1 + h_n^{-1}) \rightarrow 0,$$

which completes the proof.

It follows from Proposition 1 that $\mathcal{M}(\phi^{-1}, \phi^{-1})$ (and hence $\bar{\mathcal{M}}(\phi^{-1}, \phi^{-1})$) is Lipschitz continuous on compact subsets of $H \times H$ and so, for any compact set $B \subset M$, there exists a $K_B < \infty$ such that

$$\bar{\mathcal{M}}(Q, P) \leq 4K^2 \mathcal{M}(Q, P) \leq K_B \|\phi(Q) - \phi(P)\|_H \quad \text{for all } P, Q \in B. \quad (58)$$

This complements the convergence result of Proposition 1(i). Since \mathcal{M} is the square of the norm on the pre-dual space of the maximal exponential model $\mathcal{E}(\mu)$, it seems likely that a similar result (at least with respect to Q for fixed P) also holds there; however, this is not investigated here.

Acknowledgements The author thanks the anonymous referee for carefully reading the paper and suggesting a number of improvements in its presentation.

References

1. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Springer, New York (1991)
2. Bucy, R.S., Joseph, P.D.: Filtering for Stochastic Processes with Applications to Guidance, vol. 326, 2nd edn. AMS Chelsea Publishing, Providence (1987)
3. Crisan, D., Rozovskii, B.: The Oxford Handbook of Nonlinear Filtering. Oxford University Press, Oxford (2011)
4. Liptser, R.S., Shirayev, A.N.: Statistics of Random Processes I—General Theory. Springer, New York (2001)
5. Amari, S.-I., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, Providence (2000)
6. Cena, A., Pistone, G.: Exponential statistical manifold. Ann. Inst. Stat. Math. **59**, 27–56 (2007)
7. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **1**, 325–347 (1998)
8. Pistone, G., Rogantin, M.P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli **5**, 721–760 (1999)

9. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **23**, 1543–1561 (1995)
10. Newton, N.J.: An infinite dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **263**, 1661–1681 (2012)
11. Newton, N.J.: Infinite dimensional statistical manifolds based on a balanced chart. *Bernoulli* **22**, 711–731 (2016)
12. Newton, N.J.: Information geometric nonlinear filtering. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **18**, 1550014 (2015). <https://doi.org/10.1142/S0219025715500149>
13. Chentsov, N.N.: Statistical Decision Rules and Optimal Inference, Translations of Mathematical Monographs, vol. 53. American Mathematical Society, Providence (1982)
14. Shirayev, A.N.: Stochastic equations of nonlinear filtering of jump Markov processes. *Problemy Peredachi Informatsii II* **3**, 3–22 (1966)
15. Wonham, W.M.: Some applications of stochastic differential equations to optimal nonlinear filtering. *SIAM J. Control* **2**, 347–369 (1965)
16. Beneš, V.E.: Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5**, 65–92 (1981)
17. Brigo, D., Pistone, G.: Dimensionality reduction for measure valued evolution equations in statistical manifolds. In: Nielsen, F., Critchley, F., Dodson, C.T.J. (eds.) Computational Information Geometry for Image and Signal Processing, pp. 217–265. Springer, Berlin (2017)
18. Brigo, D., Hanzon, B., Le Gland, F.: Approximate nonlinear filtering on exponential manifolds of densities. *Bernoulli* **5**, 495–534 (1999)
19. Naudts, J.: Generalised Thermostatistics. Springer, New York (2011)
20. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Cambridge University Press, Cambridge (1992)
21. Bruveris, M., Michor, P.W.: Geometry of the Fisher-Rao metric on the space of smooth densities on a compact manifold (2016). [arXiv:1607.04450](https://arxiv.org/abs/1607.04450)
22. Lods, B., Pistone, G.: Information geometry formalism for the spatially homogeneous Boltzmann equation. *Entropy* **17**, 4323–4363 (2015)
23. Newton, N.J.: Manifolds of differentiable densities (2016). [arXiv:1608.03979](https://arxiv.org/abs/1608.03979)
24. Lebowitz, J.J., Spohn, H.: A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.* **95**, 333–366 (1999)
25. Mitter, S.K., Newton, N.J.: Information and entropy flow in the Kalman-Bucy filter. *J. Stat. Phys.* **118**, 145–167 (2005)
26. Newton, N.J.: Interactive statistical mechanics and nonlinear filtering. *J. Stat. Phys.* **133**, 711–737 (2008)
27. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (2006)
28. Duncan, T.E.: On the calculation of mutual information. *SIAM J. Appl. Math.* **19**, 215–220 (1970)
29. Elworthy, K.D.: Stochastic Differential Equations on Manifolds. London Mathematical Society Lecture Notes in Mathematics, vol. 20. Cambridge University Press, Cambridge (1987)
30. Rao, M.M., Ren, Z.D.: Theory of Orlicz Spaces. Monographs and Textbooks in Pure and Applied Mathematics, vol. 146. M. Dekker, New York (1991)

Infinite-Dimensional Log-Determinant Divergences III: Log-Euclidean and Log-Hilbert–Schmidt Divergences



Hà Quang Minh

Abstract We introduce a family of Log-Determinant (Log-Det) divergences on the set of symmetric, positive definite (SPD) matrices that includes the Log-Euclidean distance as a special case. This is then generalized to a family of Log-Det divergences on the set of positive definite Hilbert–Schmidt operators on a Hilbert space, with the Log-Hilbert–Schmidt distance being a special case. The divergences introduced here are novel both in the finite and infinite-dimensional settings. We also generalize the Power Euclidean distances to the infinite-dimensional setting, which we call the Extended Power Hilbert–Schmidt distances. While these families include the Log-Euclidean and Log-Hilbert–Schmidt distances, respectively, as special cases, they do not satisfy the same invariances as the latter distances, in contrast to the Log-Det divergences. In the case of RKHS covariance operators, we provide closed form formulas for all of the above divergences and distances in terms of the corresponding Gram matrices.

Keywords Log-determinant divergences · SPD matrices · Positive definite operators · Log-Euclidean distance · Log-Hilbert–Schmidt distance · Power Euclidean distance · Power Hilbert–Schmidt distance

1 Introduction and Motivation

The current work is a continuation of the author’s previous and ongoing work [7–11] on the generalization of the different families of distances and divergences on the set of symmetric, positive definite (SPD) matrices to the infinite-dimensional setting. The family of concern in this work is the Alpha-Beta Log-Determinant divergences.

Throughout the paper, we denote by $\text{Sym}(n)$, $n \in \mathbb{N}$, the set of real-valued symmetric matrices of size $n \times n$, $\text{Sym}^+(n)$ the set of $n \times n$ symmetric, positive semi-definite matrices, and $\text{Sym}^{++}(n)$ the set of $n \times n$ symmetric, positive definite

H. Q. Minh (✉)

RIKEN Center for Advanced Intelligence Project, Tokyo, Japan
e-mail: minh.haquang@riken.jp

matrices. We recall that on the set $\text{Sym}^{++}(n)$, for a pair $\alpha > 0, \beta > 0$ fixed, the Alpha-Beta Log-Determinant divergences $D^{(\alpha,\beta)}(A, B)$ is defined to be [4]

$$D^{(\alpha,\beta)}(A, B) = \frac{1}{\alpha\beta} \log \det \left[\frac{\alpha(AB^{-1})^\beta + \beta(AB^{-1})^{-\alpha}}{\alpha + \beta} \right]. \quad (1)$$

This family includes the Alpha Log-Det divergences in [3] as a special case. For $\alpha = \beta$, $D^{(\alpha,\alpha)}(A, B)$ is a family of symmetric divergences on $\text{Sym}^{++}(n)$, with $\alpha = 1/2$ corresponding to the symmetric Stein divergence [13] and the limiting case $\alpha = \beta = 0$,

$$\lim_{\alpha \rightarrow 0} D^{(\alpha,\alpha)}(A, B) = \frac{1}{2} \|\log(B^{-1/2}AB^{-1/2})\|_F^2, \quad (2)$$

where \log denotes the principal matrix logarithm and F denotes the Frobenius norm. The quantity $\|\log(B^{-1/2}AB^{-1/2})\|_F$ is precisely the Riemannian distance between A and B under the *affine-invariant Riemannian metric* on $\text{Sym}^{++}(n)$ (see e.g. [2, 12]). In [8], we generalize the Alpha Log-Determinant divergences from [3] and in [7], the Alpha-Beta Log-Determinant divergences $D^{(\alpha,\beta)}(A, B)$, to the set of positive definite trace class operators on an infinite-dimensional Hilbert space \mathcal{H} . Subsequently, in [11], this formulation is extended to the entire Hilbert manifold of positive definite Hilbert–Schmidt operators on \mathcal{H} . The resulting family of divergences then includes the infinite-dimensional affine-invariant Riemannian distance in [6] as a special case.

Another commonly used Riemannian metric on $\text{Sym}^{++}(n)$ is the Log-Euclidean metric [1], under which the Riemannian distance is the *Log-Euclidean distance*

$$d_{\log E}(A, B) = \|\log(A) - \log(B)\|_F. \quad (3)$$

This distance is faster to compute than the affine-invariant Riemannian distance and can be used to define many positive definite kernels on $\text{Sym}^{++}(n)$, e.g. Gaussian kernel. Its infinite-dimensional generalization is the *Log-Hilbert–Schmidt distance* between positive definite Hilbert–Schmidt operators, as formulated in [9].

Contributions of the current work. The current paper answers the following questions: (i) What is the family of divergences on $\text{Sym}^{++}(n)$, comparable to that defined in Eq. (1), that include the squared Log-Euclidean distance as a limiting case, comparable to that given in Eq. (2)? (ii) What is its infinite-dimensional generalization?

In [5], the authors introduced a parametrized family of distances, the so-called power Euclidean distances, that include the Log-Euclidean distance as a limiting case. However, the power Euclidean distances do *not* satisfy many invariance properties enjoyed by the Log-Euclidean distance, including scale invariance and inversion invariance (see Sect. 1.1 below for detail).

In this work, we first introduce a family of divergences on $\text{Sym}^{++}(n)$, inspired by the Log-Determinant divergences in Eq. (1), that share the same invariance properties as the Log-Euclidean distance and that includes the squared Log-Euclidean distance

as a limiting case. We call this family *Log-Euclidean divergences*. We then generalize this family to the Hilbert manifold of positive definite Hilbert–Schmidt operators, with the resulting family of divergences called *Log-Hilbert–Schmidt divergences*. This latter family includes the squared Log-Hilbert–Schmidt distance in [9] as a limiting case. For comparison, we also generalize the power Euclidean distances to the infinite-dimensional setting. We call this family *Extended Power Hilbert–Schmidt distances*, which include the Log-Hilbert–Schmidt distance as a limiting case. While not enjoying the same invariance properties, these distances are simpler to formulate than the Log-Det-based divergences above and thus they should be of interest in their own right. In the setting of *RKHS covariance operators*, we provide closed form formulas for all the the distances and divergences discussed above.

1.1 Background: Power Euclidean Distances

In [5], the authors introduced the following parametrized family of distances on $\text{Sym}^{++}(n)$,

$$d_{E,\alpha}(A, B) = \left\| \frac{A^\alpha - B^\alpha}{\alpha} \right\|_F, \quad A, B \in \text{Sym}^{++}(n), \quad \alpha \in \mathbb{R}, \alpha \neq 0. \quad (4)$$

If $A, B \in \text{Sym}^+(n)$, that is A, B are positive semi-definite, then we need to restrict to $\alpha > 0$. A key property of the family $d_{E,\alpha}(A, B)$ is that it includes the Log-Euclidean distance as the limiting case $\alpha \rightarrow 0$ (for completeness, we give the proofs in Sect. 5).

Lemma 1 *Let $n \in \mathbb{N}$ be fixed. Assume that $A, B \in \text{Sym}^{++}(n)$. Then for $\alpha \in \mathbb{R}$,*

$$\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - B^\alpha}{\alpha} \right\|_F = \|\log(A) - \log(B)\|_F = d_{\log E}(A, B). \quad (5)$$

The limit in Lemma 1 is a direct consequence of the following limit.

Lemma 2 *Let $n \in \mathbb{N}$ be fixed. Let $A \in \text{Sym}^{++}(n)$ be fixed. Then for $\alpha \in \mathbb{R}$,*

$$\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - I}{\alpha} - \log(A) \right\|_F = 0. \quad (6)$$

For $A, B \in \text{Sym}^+(n)$, the limit in Lemma 1 is generally not valid, since we have

Lemma 3 *Assume that $A \in \text{Sym}^+(n)$ but $A \notin \text{Sym}^{++}(n)$, that is A has at least one zero eigenvalue. Then*

$$\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - I}{\alpha} \right\|_F = \infty. \quad (7)$$

While including the Log-Euclidean distance as the limiting case, the power Euclidean distances are essentially Euclidean distances and do *not* satisfy the same invariance properties as the Log-Euclidean distance. Specifically,

1. $d_{\log E}(A, B)$ is *scale invariant*, whereas $d_{E,\alpha}(A, B)$ is *not* scale invariant, with

$$d_{E,\alpha}(sA, sB) = s^\alpha d_{E,\alpha}(A, B), \quad s \in \mathbb{R}, s > 0. \quad (8)$$

2. $d_{\log E}(A, B)$ is *inversion-invariant*, whereas $d_{E,\alpha}(A, B)$ is *dually invariant* with respect to inversion, with

$$d_{E,\alpha}(A^{-1}, B^{-1}) = \left\| \frac{A^{-\alpha} - B^{-\alpha}}{\alpha} \right\|_F = d_{E,-\alpha}(A, B). \quad (9)$$

It is obvious that both scale invariance and inversion invariance only hold in the limiting case $\alpha = 0$. In the following, we introduce a family of divergences on $\text{Sym}^{++}(n)$ that includes the Log-Euclidean distance as a special case and at the same time satisfy all the above invariance properties of the Log-Euclidean distance.

2 The Finite-Dimensional Case: Log-Euclidean Divergences

We recall that the Log-Euclidean distance as formulated in [1] is the Riemannian distance associated with the bi-invariant Riemannian metric on $\text{Sym}^{++}(n)$, considered as a Lie group, under the following commutative Lie group operation

$$\begin{aligned} \odot : \text{Sym}^{++}(n) \times \text{Sym}^{++}(n) &\rightarrow \text{Sym}^{++}(n) \\ A \odot B &= \exp(\log(A) + \log(B)). \end{aligned} \quad (10)$$

Along with the commutative operation \odot , the following power operation can be defined

$$\begin{aligned} \circledast : \mathbb{R} \times \text{Sym}^{++}(n) &\rightarrow \text{Sym}^{++}(n) \\ \lambda \circledast A &= \exp(\lambda \log(A)) = A^\lambda. \end{aligned} \quad (11)$$

Equipped with the operations (\odot, \circledast) , the set $(\text{Sym}^{++}(n), \odot, \circledast)$ becomes a vector space, with \odot acting as vector addition and \circledast acting as scalar multiplication. By the commutativity of \odot , for any $\alpha \in \mathbb{R}$, we have

$$(A \odot B)^\alpha = \exp[\alpha \log(A \odot B)] = \exp[\alpha \log(A) + \alpha \log(B)] = A^\alpha \odot B^\alpha. \quad (12)$$

We first have the following result.

Lemma 4 For any pair $A, B \in \text{Sym}^{++}(n)$,

$$\det(A \odot B) = \det(A) \det(B). \quad (13)$$

The definition of the Alpha-Beta Log-Det $D^{(\alpha, \beta)}(A, B)$ in Eq. (1) is based on the standard matrix product AB^{-1} , which is noncommutative. If we replace this product with the commutative product $A \odot B^{-1}$, we arrive at the following definition.

Definition 1 (*Alpha-Beta Log-Euclidean divergences*) Let $\alpha > 0, \beta > 0$ be fixed. The (α, β) -Log-Euclidean divergence between $A, B \in \text{Sym}^{++}(n)$ is defined to be

$$D_{\odot}^{(\alpha, \beta)}(A, B) = \frac{1}{\alpha\beta} \log \det \left[\frac{\alpha(A \odot B^{-1})^{\beta} + \beta(A \odot B^{-1})^{-\alpha}}{\alpha + \beta} \right] \quad (14)$$

$$= \frac{1}{\alpha\beta} \sum_{j=1}^n \log \left(\frac{\alpha\lambda_j^{\beta} + \beta\lambda_j^{-\alpha}}{\alpha + \beta} \right), \quad (15)$$

where $\{\lambda_j\}_{j=1}^n$ denote the eigenvalues of $A \odot B^{-1}$.

Theorem 1 (*Limiting cases*) For $\alpha > 0$ fixed,

$$\lim_{\beta \rightarrow 0} D_{\odot}^{(\alpha, \beta)}(A, B) = \frac{1}{\alpha^2} \{ \text{tr}[(A^{-1} \odot B)^{\alpha} - I] - \alpha \log \det(A^{-1} \odot B) \}. \quad (16)$$

For $\beta > 0$ fixed,

$$\lim_{\alpha \rightarrow 0} D_{\odot}^{(\alpha, \beta)}(A, B) = \frac{1}{\beta^2} \{ \text{tr}[(B^{-1} \odot A)^{\beta} - I] - \beta \log \det(B^{-1} \odot A) \}. \quad (17)$$

Definition 2 (*Limiting cases*) Motivated by Theorem 1, we define

$$D_{\odot}^{(\alpha, 0)}(A, B) = \lim_{\beta \rightarrow 0} D_{\odot}^{(\alpha, \beta)}(A, B), \quad D_{\odot}^{(0, \beta)}(A, B) = \lim_{\alpha \rightarrow 0} D_{\odot}^{(\alpha, \beta)}(A, B). \quad (18)$$

Remark 1 By the commutativity of the \odot operation, $A \odot B^{-1}$ is symmetric and $A \odot B^{-1} = B^{-1/2} \odot A \odot B^{-1/2}$, so that these two expressions can be used interchangeably.

Special case: Log-Euclidean distance. For $\alpha = \beta$, we obtain

$$\begin{aligned} D_{\odot}^{(\alpha, \alpha)}(A, B) &= \frac{1}{\alpha^2} \log \det \left(\frac{(A \odot B^{-1})^{\alpha} + (A \odot B^{-1})^{-\alpha}}{2} \right) \\ &= \frac{1}{\alpha^2} \sum_{j=1}^n \log \left(\frac{\lambda_j^{\alpha} + \lambda_j^{-\alpha}}{2} \right). \end{aligned} \quad (19)$$

Theorem 2 (Log-Euclidean distance) *For any pair $A, B \in \text{Sym}^{++}(n)$,*

$$\lim_{\alpha \rightarrow 0} D_{\odot}^{(\alpha, \alpha)}(A, B) = \frac{1}{2} \|\log(A) - \log(B)\|_F^2. \quad (20)$$

2.1 Properties of the Log-Euclidean Divergences

The following results show that the Log-Euclidean divergences, as defined in Definitions 1 and 2, satisfy all the invariance properties of the Log-Euclidean distance.

Theorem 3 (Positivity)

$$D_{\odot}^{(\alpha, \beta)}(A, B) \geq 0, \quad (21)$$

$$D_{\odot}^{(\alpha, \beta)}(A, B) = 0 \iff A = B. \quad (22)$$

Theorem 4 (Invariance under Lie group operation) *For all $A, B, C \in \text{Sym}^{++}(n)$,*

$$D_{\odot}^{(\alpha, \beta)}[(A \odot C), (B \odot C)] = D_{\odot}^{(\alpha, \beta)}(A, B). \quad (23)$$

Theorem 5 (Dual symmetry)

$$D_{\odot}^{(\alpha, \beta)}(B, A) = D_{\odot}^{(\beta, \alpha)}(A, B). \quad (24)$$

In particular, for $\alpha = \beta$, $D_{\odot}^{(\alpha, \alpha)}(A, B)$ is symmetric, that is

$$D_{\odot}^{(\alpha, \alpha)}(A, B) = D_{\odot}^{(\alpha, \alpha)}(B, A). \quad (25)$$

Theorem 6 (Scale invariance)

$$D_{\odot}^{(\alpha, \beta)}(sA, sB) = D_{\odot}^{(\alpha, \beta)}(A, B), \quad \forall s > 0. \quad (26)$$

Theorem 7 (Dual-invariance under inversion)

$$D_{\odot}^{(\alpha, \beta)}(A^{-1}, B^{-1}) = D_{\odot}^{(\beta, \alpha)}(A, B). \quad (27)$$

In particular, for $\alpha = \beta$, $D_{\odot}^{(\alpha, \alpha)}$ is invariant under inversion, that is

$$D_{\odot}^{(\alpha, \alpha)}(A^{-1}, B^{-1}) = D_{\odot}^{(\alpha, \alpha)}(A, B). \quad (28)$$

Theorem 8 (Unitary invariance) *For any invertible $C \in \mathbb{R}^{n \times n}$ with $C^T C = I$,*

$$D_{\odot}^{(\alpha, \beta)}(CAC^{-1}, CBC^{-1}) = D_{\odot}^{(\alpha, \beta)}(A, B). \quad (29)$$

3 The Infinite-Dimensional Case: Log-Hilbert–Schmidt Divergences

We now generalize the Alpha-Beta Log-Euclidean divergences on $\text{Sym}^{++}(n)$ to the infinite-dimensional setting. The abstract formulation that is presented here parallels [7], which defines the Alpha-Beta Log-Det divergences between positive definite trace class operators on a Hilbert space \mathcal{H} , and [11], which generalizes the formulation in [7] to the entire Hilbert manifold of positive definite Hilbert–Schmidt operators on \mathcal{H} .

Throughout the following, let \mathcal{H} be a real, separable Hilbert space, with $\dim(\mathcal{H}) = \infty$, unless stated explicitly otherwise. Let $\text{Tr}(\mathcal{H})$ denote the set of trace class operators on \mathcal{H} . In [8], we define the set of *extended (unitized) trace class operators* on \mathcal{H} to be

$$\text{Tr}_X(\mathcal{H}) = \{A + \gamma I : A \in \text{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}. \quad (30)$$

This set becomes a Banach algebra under the *extended trace class norm* $\|A + \gamma I\|_{\text{tr}_X} = \|A\|_{\text{tr}} + |\gamma| = \text{tr}|A| + |\gamma|$. For $(A + \gamma I) \in \text{Tr}_X(\mathcal{H})$, its *extended trace* is defined to be

$$\text{tr}_X(A + \gamma I) = \text{tr}(A) + \gamma, \quad \text{with } \text{tr}_X(I) = 1. \quad (31)$$

Along with the extended trace, in [8] we defined the *extended Fredholm determinant*

$$\det_X(A + \gamma I) = \gamma \det[(A/\gamma) + I], \quad (A + \gamma I) \in \text{Tr}_X(\mathcal{H}), \gamma \neq 0, \quad (32)$$

where \det on the right hand side denotes the Fredholm determinant. The set of *positive definite (unitized) trace class operators* $\mathcal{PC}_1(\mathcal{H}) \subset \text{Tr}_X(\mathcal{H})$ is then defined to be

$$\mathcal{PC}_1(\mathcal{H}) = \{A + \gamma I > 0 : A^* = A, A \in \text{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}. \quad (33)$$

Let $\text{HS}(\mathcal{H})$ denote the set of Hilbert–Schmidt operators on \mathcal{H} . In [6], the author considered the following set of extended (unitized) Hilbert–Schmidt operators

$$\text{HS}_X(\mathcal{H}) = \{A + \gamma I : A \in \text{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}, \quad (34)$$

which can be equipped with the *extended Hilbert–Schmidt inner product* $\langle \cdot, \cdot \rangle_{\text{HS}_X}$,

$$\langle A + \gamma I, B + \mu I \rangle_{\text{HS}_X} = \langle A, B \rangle_{\text{HS}} + \gamma \mu = \text{tr}(A^* B) + \gamma \mu, \quad (35)$$

along with the associated *extended Hilbert–Schmidt norm*

$$\|A + \gamma I\|_{\text{HS}_X}^2 = \|A\|_{\text{HS}}^2 + \gamma^2 = \text{tr}(A^* A) + \gamma^2, \quad \text{with } \|I\|_{\text{HS}_X} = 1. \quad (36)$$

The set of *positive definite (unitized) Hilbert–Schmidt operators* is then defined to be

$$\mathcal{PC}_2(\mathcal{H}) = \{A + \gamma I > 0 : A = A^*, A \in \text{HS}(\mathcal{H}), \gamma \in \mathbb{R}\} \subset \text{HS}_X(\mathcal{H}). \quad (37)$$

The set $\mathcal{PC}_2(\mathcal{H})$ is an open subset of the Hilbert space $\text{HS}_X(\mathcal{H})$ and thus forms a Hilbert manifold. Clearly $\mathcal{PC}_1(\mathcal{H})$ is a strict subset of $\mathcal{PC}_2(\mathcal{H})$ when $\dim(\mathcal{H}) = \infty$.

3.1 Log-Hilbert–Schmidt Divergences Between Positive Definite Trace Class Operators

In [9], we define the following generalizations of the operations \odot and \circledast to the set $\mathcal{PC}_2(\mathcal{H})$ of positive definite Hilbert–Schmidt operators on \mathcal{H} ,

$$\begin{aligned} \odot : \mathcal{PC}_2(\mathcal{H}) \times \mathcal{PC}_2(\mathcal{H}) &\rightarrow \mathcal{PC}_2(\mathcal{H}), \\ (A + \gamma I) \odot (B + \mu I) &= \exp(\log(A + \gamma I) + \log(B + \mu I)), \end{aligned} \quad (38)$$

$$\begin{aligned} \circledast : \mathbb{R} \times \mathcal{PC}_2(\mathcal{H}) &\rightarrow \mathcal{PC}_2(\mathcal{H}), \\ \lambda \circledast (A + \gamma I) &= \exp(\lambda \log(A + \gamma I)) = (A + \gamma I)^\lambda. \end{aligned} \quad (39)$$

The subset $\mathcal{PC}_1(\mathcal{H}) \subset \mathcal{PC}_2(\mathcal{H})$ is in fact closed under these operations. In [7], we show that for $(A + \gamma I) \in \mathcal{PC}_1(\mathcal{H})$, $(A + \gamma I)^\lambda \in \mathcal{PC}_1(\mathcal{H}) \forall \lambda \in \mathbb{R}$, in particular $(A + \gamma I)^{-1} \in \mathcal{PC}_1(\mathcal{H})$. The following result shows that $\mathcal{PC}_1(\mathcal{H})$ is also closed under \odot .

Lemma 5 *Assume that $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_1(\mathcal{H})$. Then $(A + \gamma I) \odot (B + \mu I)^{-1} = Z + \frac{\gamma}{\mu} I \in \mathcal{PC}_1(\mathcal{H})$ for some operator $Z \in \text{Tr}(\mathcal{H})$.*

With $Z + \frac{\gamma}{\mu} I = (A + \gamma I) \odot (B + \mu I)^{-1} \in \mathcal{PC}_1(\mathcal{H})$, we are ready to define the following generalization of the Log-Euclidean divergences to $\mathcal{PC}_1(\mathcal{H})$.

Definition 3 (*Alpha-Beta Log-Hilbert–Schmidt divergences between positive definite trace class operators*) Assume that $\dim(\mathcal{H}) = \infty$. Let $\alpha > 0, \beta > 0$ be fixed. Let $r \in \mathbb{R}$, $r \neq 0$ be fixed. For $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_1(\mathcal{H})$, the (α, β) -Log-Hilbert–Schmidt divergence $D_{r, \odot}^{(\alpha, \beta)}[(A + \gamma I), (B + \mu I)]$ is defined to be

$$\begin{aligned} & D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] \\ &= \frac{1}{\alpha\beta} \log \left[\left(\frac{\gamma}{\mu} \right)^{r(\delta - \frac{\alpha}{\alpha+\beta})} \det_X \left(\frac{\alpha(Z + \frac{\gamma}{\mu} I)^{r(1-\delta)} + \beta(Z + \frac{\gamma}{\mu} I)^{-r\delta}}{\alpha + \beta} \right) \right], \end{aligned} \quad (40)$$

where $Z + \frac{\gamma}{\mu} I = (A + \gamma I) \odot (B + \mu I)^{-1}$, $Z \in \text{Tr}(\mathcal{H})$, and $\delta = \frac{\alpha\gamma^r}{\alpha\gamma^r + \beta\mu^r}$.

Theorem 9 (Limiting cases) Let $\alpha > 0$ be fixed. Assume that $r = r(\beta)$ is smooth, with $r(0) = r(\beta = 0)$. Then

$$\begin{aligned} \lim_{\beta \rightarrow 0} D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] &= \frac{r(0)}{\alpha^2} \left[\left(\frac{\mu}{\gamma} \right)^{r(0)} - 1 \right] \log \frac{\mu}{\gamma} \\ &+ \frac{1}{\alpha^2} \text{tr}_X([(A + \gamma I)^{-1} \odot (B + \mu I)]^{r(0)} - I) \\ &- \frac{1}{\alpha^2} \left(\frac{\mu}{\gamma} \right)^{r(0)} \log \det_X [(A + \gamma I)^{-1} \odot (B + \mu I)]^{r(0)}. \end{aligned} \quad (41)$$

Let $\beta > 0$ be fixed. Assume that $r = r(\alpha)$ is smooth, with $r(0) = r(\alpha = 0)$. Then

$$\begin{aligned} \lim_{\alpha \rightarrow 0} D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] &= \frac{r(0)}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^{r(0)} - 1 \right] \log \frac{\gamma}{\mu} \\ &+ \frac{1}{\beta^2} \text{tr}_X([(B + \mu I)^{-1} \odot (A + \gamma I)]^{r(0)} - I) \\ &- \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^{r(0)} \log \det_X [(B + \mu I)^{-1} \odot (A + \gamma I)]^{r(0)}. \end{aligned} \quad (42)$$

Motivated by Theorem 9, the following are the definitions of $D_{r,\odot}^{(\alpha,0)}$ and $D_{r,\odot}^{(0,\beta)}$.

Definition 4 (Limiting cases) Assume that $\dim(\mathcal{H}) = \infty$. Let $\alpha > 0, \beta > 0, r \neq 0$ be fixed. For $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_1(\mathcal{H})$, the Alpha-Beta Log-Hilbert-Schmidt divergence $D_{r,\odot}^{(\alpha,0)}[(A + \gamma I), (B + \mu I)]$ is defined to be

$$\begin{aligned} D_{r,\odot}^{(\alpha,0)}[(A + \gamma I), (B + \mu I)] &= \frac{r}{\alpha^2} \left[\left(\frac{\mu}{\gamma} \right)^r - 1 \right] \log \frac{\mu}{\gamma} \\ &+ \frac{1}{\alpha^2} \text{tr}_X([(A + \gamma I)^{-1} \odot (B + \mu I)]^r - I) \\ &- \frac{1}{\alpha^2} \left(\frac{\mu}{\gamma} \right)^r \log \det_X [(A + \gamma I)^{-1} \odot (B + \mu I)]^r. \end{aligned} \quad (43)$$

Similarly, $D_{r,\odot}^{(0,\beta)}[(A + \gamma I), (B + \mu I)]$ is defined to be

$$\begin{aligned}
D_{r,\odot}^{(0,\beta)}[(A + \gamma I), (B + \mu I)] &= \frac{r}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] \log \frac{\gamma}{\mu} \\
&\quad + \frac{1}{\beta^2} \text{tr}_X [(B + \mu I)^{-1} \odot (A + \gamma I)]^r - I \\
&\quad - \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \log \det_X [(B + \mu I)^{-1} \odot (A + \gamma I)]^r.
\end{aligned} \tag{44}$$

Special case: Log-Euclidean divergences. For $\gamma = \mu$ and $r = \alpha + \beta$, we have for $\alpha > 0, \beta > 0$,

$$\begin{aligned}
D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \gamma I)] \\
= \frac{1}{\alpha\beta} \log \det_X \left(\frac{\alpha[(A + \gamma I) \odot (B + \gamma I)^{-1}]^\beta + \beta[(A + \gamma I) \odot (B + \gamma I)^{-1}]^{-\alpha}}{\alpha + \beta} \right),
\end{aligned}$$

For $A, B \in \text{Sym}^{++}(n)$, we recover the Log-Euclidean divergences by setting $\gamma = 0$. The same holds true for the limiting cases ($\alpha = 0, \beta > 0$) and ($\alpha > 0, \beta = 0$).

3.2 Log-Hilbert–Schmidt Divergences Between Positive Definite Hilbert–Schmidt Operators

We now present the generalization of the Log-Hilbert–Schmidt divergences $D_{r,\odot}^{(\alpha,\beta)}$ from the set $\mathcal{PC}_1(\mathcal{H})$ to the entire Hilbert manifold $\mathcal{PC}_2(\mathcal{H})$ of positive definite Hilbert–Schmidt operators on \mathcal{H} . In the case ($\alpha > 0, \beta > 0$), this generalization is made possible by the following result.

Lemma 6 Assume that $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$. Let $\alpha > 0, \beta > 0$ be fixed. Let $p, q \in \mathbb{R}$ be such that $p\alpha(\gamma/\mu)^p = q\beta(\gamma/\mu)^{-q}$. Then for $Z + \frac{\gamma}{\mu}I = (A + \gamma I) \odot (B + \mu I)^{-1} = (B + \mu I)^{-1/2} \odot (A + \gamma I) \odot (B + \mu I)^{-1/2}$, we have

$$\frac{\alpha \left(Z + \frac{\gamma}{\mu} I \right)^p + \beta \left(Z + \frac{\gamma}{\mu} I \right)^{-q}}{\alpha + \beta} \in \mathcal{PC}_1(\mathcal{H}). \tag{45}$$

Lemma 6 implies that the formula for $D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]$ as given in Eq.(40) remains valid in the general case $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$. In the cases ($\alpha > 0, \beta = 0$), ($\alpha = 0, \beta > 0$), we need to use the *extended Hilbert–Carleman determinant* from [11],

$$\det_{2X}(A + \gamma I) = \det_X[(A + \gamma I) \exp(-A/\gamma)], \quad (A + \gamma I) \in \mathcal{PC}_2(\mathcal{H}). \tag{46}$$

The following is the general definition of $D_{r,\odot}^{(\alpha,\beta)}$ on $\mathcal{PC}_2(\mathcal{H})$.

Definition 5 (*Alpha-Beta Log-Hilbert–Schmidt divergences between positive definite Hilbert–Schmidt operators*) Let $\alpha > 0, \beta > 0, r \neq 0$ be fixed. For $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$,

- (i) $D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]$ is defined as in Eq. (40).
- (ii) $D_{r,\odot}^{(\alpha,0)}[(A + \gamma I), (B + \mu I)]$ is defined to be

$$\begin{aligned} D_{r,\odot}^{(\alpha,0)}[(A + \gamma I), (B + \mu I)] &= \frac{1}{\alpha^2} \left[\left(\frac{\mu}{\gamma} \right)^r - 1 \right] \left(1 + r \log \frac{\mu}{\gamma} \right) \\ &\quad - \frac{1}{\alpha^2} \left(\frac{\mu}{\gamma} \right)^r \log \det_{2X}([(A + \gamma I)^{-1} \odot (B + \mu I)]^r). \end{aligned} \quad (47)$$

- (iii) $D_{r,\odot}^{(0,\beta)}[(A + \gamma I), (B + \mu I)]$ is defined to be

$$\begin{aligned} D_{r,\odot}^{(0,\beta)}[(A + \gamma I), (B + \mu I)] &= \frac{1}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] \left(1 + r \log \frac{\gamma}{\mu} \right) \\ &\quad - \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \log \det_{2X}([(B + \mu I)^{-1} \odot (A + \gamma I)]^r). \end{aligned} \quad (48)$$

The squared Log-Hilbert–Schmidt distance $\|\log(A + \gamma I) - \log(B + \mu I)\|_{HS_X}^2$ is then precisely twice the limit of $D_{2\alpha,\odot}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)]$ as $\alpha \rightarrow 0$.

Theorem 10 (Limiting case: Log-Hilbert–Schmidt distance) *Assume that $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$. Assume that $r = r(\alpha)$ is smooth, with $r(0) = 0, r'(0) \neq 0$, and $r(\alpha) \neq 0$ for $\alpha \neq 0$. Then*

$$\lim_{\alpha \rightarrow 0} D_{r,\odot}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)] = \frac{[r'(0)]^2}{8} \|\log(A + \gamma I) - \log(B + \mu I)\|_{HS_X}^2. \quad (49)$$

In particular, for $r = 2\alpha$,

$$\lim_{\alpha \rightarrow 0} D_{2\alpha,\odot}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)] = \frac{1}{2} \|\log(A + \gamma I) - \log(B + \mu I)\|_{HS_X}^2. \quad (50)$$

3.3 Properties of the Log-Hilbert–Schmidt Divergences

Assume in the following that $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$. The following properties generalize those of the Log-Euclidean divergences in Sect. 2.1.

Theorem 11 (Positivity)

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] \geq 0, \quad (51)$$

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] = 0 \iff A = B, \gamma = \mu. \quad (52)$$

Theorem 12 (Invariance under Lie group operation) For any $(C + \nu I) \in \mathcal{PC}_2(\mathcal{H})$,

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I) \odot (C + \nu I), (B + \mu I) \odot (C + \nu I)] = D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]. \quad (53)$$

Theorem 13 (Dual symmetry)

$$D_{r,\odot}^{(\beta,\alpha)}[(B + \mu I), (A + \gamma I)] = D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]. \quad (54)$$

In particular, for $\beta = \alpha$, we have

$$D_{r,\odot}^{(\alpha,\alpha)}[(B + \mu I), (A + \gamma I)] = D_{r,\odot}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)]. \quad (55)$$

Theorem 14 (Scale invariance)

$$D_{r,\odot}^{(\alpha,\beta)}[s(A + \gamma I), s(B + \mu I)] = D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)], \quad s > 0. \quad (56)$$

Theorem 15 (Dual invariance under inversion)

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] = D_{r,\odot}^{(\beta,\alpha)}[(A + \gamma I), (B + \mu I)]. \quad (57)$$

In particular, for $\alpha = \beta$,

$$D_{r,\odot}^{(\alpha,\alpha)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] = D_{r,\odot}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)], \quad (58)$$

so that $D_{r,\odot}^{(\alpha,\alpha)}$ is inversion invariant. Furthermore,

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] = D_{-r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]. \quad (59)$$

Theorem 16 (Invariance under unitary transformations) For any $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$ and any $C \in \mathcal{L}(\mathcal{H})$, with $CC^* = C^*C = I$,

$$D_{r,\odot}^{(\alpha,\beta)}[C(A + \gamma I)C^*, C(B + \mu I)C^*] = D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)]. \quad (60)$$

3.4 The Log-Hilbert–Schmidt Divergences Between RKHS Covariance Operators

We now derive explicit expressions for the Log–Hilbert–Schmidt divergences between RKHS covariance operators. Let \mathcal{X} be a separable topological space and K be a continuous positive definite kernel on $\mathcal{X} \times \mathcal{X}$. Then the reproducing kernel Hilbert space (RKHS) \mathcal{H}_K induced by K is separable ([14], Lemma 4.33). Let $\Phi : \mathcal{X} \rightarrow \mathcal{H}_K$ be the corresponding feature map, so that

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_K} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}. \quad (61)$$

Let $\mathbf{X} = [x_1, \dots, x_m]$ be a data matrix randomly sampled from \mathcal{X} according to a Borel probability distribution ρ , where $m \in \mathbb{N}$ is the number of observations. The feature map Φ on \mathbf{X} defines the bounded linear operator

$$\Phi(\mathbf{X}) : \mathbb{R}^m \rightarrow \mathcal{H}_K, \quad \Phi(\mathbf{X})\mathbf{b} = \sum_{j=1}^m b_j \Phi(x_j), \quad \mathbf{b} \in \mathbb{R}^m. \quad (62)$$

Informally, the operator $\Phi(\mathbf{X})$ can also be viewed as the (potentially infinite) mapped feature matrix $\Phi(\mathbf{X}) = [\Phi(x_1), \dots, \Phi(x_m)]$ of size $\dim(\mathcal{H}_K) \times m$ in the feature space \mathcal{H}_K , with the j th column being $\Phi(x_j)$. The corresponding empirical covariance operator for $\Phi(\mathbf{X})$ (RKHS covariance operators are described in more detail in [8]) is defined to be

$$C_{\Phi(\mathbf{X})} = \frac{1}{m} \Phi(\mathbf{X}) J_m \Phi(\mathbf{X})^* : \mathcal{H}_K \rightarrow \mathcal{H}_K, \quad (63)$$

where $\Phi(\mathbf{X})^* : \mathcal{H}_K \rightarrow \mathbb{R}^m$ is the adjoint operator of $\Phi(\mathbf{X})$ and J_m is the centering matrix, defined by $J_m = I_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T$ with $\mathbf{1}_m = (1, \dots, 1)^T \in \mathbb{R}^m$.

Let $\mathbf{X} = [x_i]_{i=1}^m$, $\mathbf{Y} = [y_i]_{i=1}^m$, $m \in \mathbb{N}$, be two random data matrices sampled from \mathcal{X} according to two Borel probability distributions and $C_{\Phi(\mathbf{X})}$, $C_{\Phi(\mathbf{Y})}$ be the corresponding covariance operators induced by the kernel K . Let us derive the explicit expression for

$$D_{r,\odot}^{(\alpha, \beta)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})]. \quad (64)$$

We first consider the scenario where $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ are two compact operators between two separable Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$, such that $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are trace class operators. In this case, we can express $D_{r,\odot}^{(\alpha, \beta)}[(AA^* + \gamma I_{\mathcal{H}_2}), (BB^* + \mu I_{\mathcal{H}_2})]$ in terms of quantities involving the operators $A^*A, B^*B, A^*B : \mathcal{H}_1 \rightarrow \mathcal{H}_1$. We need the following technical result from [9].

Lemma 7 ([9]) *Assume that $\mathcal{H}_1, \mathcal{H}_2$ are two separable Hilbert spaces. Let $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be a compact operator. Let $\{\lambda_k(A^*A)\}_{k=1}^{N_A}$, $1 \leq N_A \leq \infty$, be the nonzero eigenvalues of $A^*A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$, with corresponding orthonormal eigenvectors $\{\phi_k(A^*A)\}_{k=1}^{N_A}$. Then the nonzero eigenvalues of the operator $AA^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are*

precisely the same $\{\lambda_k(A^*A)\}_{k=1}^{N_A}$, with corresponding orthonormal eigenvectors $\left\{\frac{A\phi_k(A^*A)}{\sqrt{\lambda_k(A^*A)}}\right\}_{k=1}^{N_A}$. Then $\log(I_{\mathcal{H}_1} + A^*A) : \mathcal{H}_1 \rightarrow \mathcal{H}_1$ admits the orthogonal spectral decomposition

$$\log(I_{\mathcal{H}_1} + A^*A) = \sum_{k=1}^{N_A} \log(1 + \lambda_k(A^*A)) \phi_k(A^*A) \otimes \phi_k(A^*A), \quad (65)$$

and $\log(I_{\mathcal{H}_2} + AA^*) : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ admits the orthogonal spectral decomposition

$$\log(I_{\mathcal{H}_2} + AA^*) = \sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A^*A))}{\lambda_k(A^*A)} (A\phi_k(A^*A)) \otimes (A\phi_k(A^*A)). \quad (66)$$

Remark 2 If $\lambda_k(A^*A) = 0$, then the corresponding eigenvector $\phi_k(A^*A)$ satisfies

$$\begin{aligned} \|A\phi_k(A^*A)\|^2 &= \langle A\phi_k(A^*A), A\phi_k(A^*A) \rangle = \langle \phi_k(A^*A), A^*A\phi_k(A^*A) \rangle \\ &= \lambda_k(A^*A)\|\phi_k(A^*A)\|^2 = 0 \Rightarrow A\phi_k(A^*A) = 0. \end{aligned} \quad (67)$$

Thus the expansion in Eq. (66) is the same as

$$\log(I_{\mathcal{H}_2} + AA^*) = \sum_{k=1}^{\infty} \frac{\log(1 + \lambda_k(A^*A))}{\lambda_k(A^*A)} (A\phi_k(A^*A)) \otimes (A\phi_k(A^*A)).$$

However, for our current purposes, we aim to take a factor A and A^* out from the expansion (see below) and thus we employ Eq. (66). \square

The expansion in Eq. (66) can be alternatively re-written as follows.

Lemma 8 Let $A : \mathcal{H} \rightarrow \mathcal{H}$ be a self-adjoint, positive, compact operator on \mathcal{H} . Let $\{\lambda_k(A)\}_{k=1}^{N_A}$, $1 \leq N_A \leq \infty$, be the strictly positive eigenvalues of A , with corresponding orthonormal eigenvectors $\{\phi_k(A)\}_{k=1}^{N_A}$. Then the orthogonal spectral decomposition

$$g(A) = \sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A))}{\lambda_k(A)} \phi_k(A) \otimes \phi_k(A) \quad (68)$$

defines a bounded, self-adjoint, positive operator on \mathcal{H} . The operator $g(A)$ satisfies

$$Ag(A) = g(A)A = \log(I + A). \quad (69)$$

In particular, if $\dim(\mathcal{H}) = d < \infty$, then A is a $d \times d$ SPD matrix and $g(A)$ can be computed as follows. Let $A = U_A \Sigma_A U_A^T$ be the (reduced) singular value decomposition of A , where $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_{N_A})$ is a diagonal matrix whose main diagonal

entries are the nonzero eigenvalues $\{\lambda_k\}_{k=1}^{N_A}$ of A , and U_A is an orthogonal matrix of size $d \times N_A$. Then $g(A)$ is the $d \times d$ matrix given by

$$g(A) = U_A \log(\Sigma_A + I_{N_A}) \Sigma_A^{-1} U_A^T. \quad (70)$$

With the operator-valued function g defined in Eq. (68), we obtain

Corollary 1 *Assume the hypothesis of Lemma 7. Then*

$$\log(I_{\mathcal{H}_2} + AA^*) = Ag(A^*A)A^*. \quad (71)$$

Having expressed $\log(I_{\mathcal{H}_2} + AA^*)$ in terms of $g(A^*A) : \mathcal{H}_1 \rightarrow \mathcal{H}_1$, we now have

Proposition 1 *Let $\mathcal{H}_1, \mathcal{H}_2$ be separable Hilbert spaces. Let $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be compact operators such that $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are trace class operators. Assume that $\dim(\mathcal{H}_2) = \infty$. Then for any $\alpha > 0, \beta > 0, r \in \mathbb{R}, r \neq 0$, for any $\gamma > 0, \mu > 0$,*

$$\begin{aligned} & D_{r,\odot}^{(\alpha,\beta)}[(AA^* + \gamma I_{\mathcal{H}_2}), (BB^* + \mu I_{\mathcal{H}_2})] \\ &= \frac{r \left(\delta - \frac{\alpha}{\alpha+\beta} \right)}{\alpha\beta} \log \left(\frac{\gamma}{\mu} \right) + \frac{1}{\alpha\beta} \log \left(\frac{\alpha \left(\frac{\gamma}{\mu} \right)^p + \beta \left(\frac{\gamma}{\mu} \right)^{-q}}{\alpha + \beta} \right) \\ &+ \frac{1}{\alpha\beta} \log \det \left(\frac{\alpha \left(\frac{\gamma}{\mu} \right)^p \exp(pC) + \beta \left(\frac{\gamma}{\mu} \right)^{-q} \exp(-qC)}{\alpha \left(\frac{\gamma}{\mu} \right)^p + \beta \left(\frac{\gamma}{\mu} \right)^{-q}} \right) \end{aligned} \quad (72)$$

where $\delta = \frac{\alpha\gamma^r}{\alpha\gamma^r + \beta\mu^r}$, $p = r(1 - \delta)$, $q = r\delta$, and the operator C is given by

$$C = \begin{pmatrix} \log \left(I_{\mathcal{H}_1} + \frac{A^*A}{\gamma} \right) & -\frac{A^*B}{\sqrt{\gamma}\mu} g \left(\frac{B^*B}{\mu} \right) \\ \frac{B^*A}{\sqrt{\gamma}\mu} g \left(\frac{A^*A}{\gamma} \right) & -\log \left(I_{\mathcal{H}_1} + \frac{B^*B}{\mu} \right) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2. \quad (73)$$

Let us now apply Proposition 1 to the RKHS setting. Let $K[\mathbf{X}]$, $K[\mathbf{Y}]$, and $K[\mathbf{X}, \mathbf{Y}]$ be the $m \times m$ Gram matrices defined by

$$(K[\mathbf{X}])_{ij} = K(x_i, x_j), (K[\mathbf{Y}])_{ij} = K(y_i, y_j), (K[\mathbf{X}, \mathbf{Y}])_{ij} = K(x_i, y_j). \quad (74)$$

for $1 \leq i, j \leq m$. By definition of the feature map Φ , as given in Eq. (61), we have

$$\begin{aligned} K[\mathbf{X}] &= \Phi(\mathbf{X})^* \Phi(\mathbf{X}), \quad K[\mathbf{Y}] = \Phi(\mathbf{Y})^* \Phi(\mathbf{Y}), \\ K[\mathbf{X}, \mathbf{Y}] &= \Phi(\mathbf{X})^* \Phi(\mathbf{Y}), \quad K[\mathbf{Y}, \mathbf{X}] = \Phi(\mathbf{Y})^* \Phi(\mathbf{X}). \end{aligned} \quad (75)$$

Let $A = \frac{1}{\sqrt{m}} \Phi(\mathbf{x}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$, $B = \frac{1}{\sqrt{m}} \Phi(\mathbf{y}) J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$, so that

$$\begin{aligned} AA^* &= C_{\Phi(\mathbf{X})}, \quad BB^* = C_{\Phi(\mathbf{Y})}, \quad A^*A = \frac{1}{m} J_m K[\mathbf{X}] J_m, \quad B^*B = \frac{1}{m} J_m K[\mathbf{Y}] J_m, \\ A^*B &= \frac{1}{m} J_m K[\mathbf{X}, \mathbf{Y}] J_m, \quad B^*A = \frac{1}{m} J_m K[\mathbf{Y}, \mathbf{X}] J_m. \end{aligned} \quad (76)$$

Applying Proposition 1 with A and B as defined above, we obtain

Theorem 17 *Let $\alpha > 0, \beta > 0, r \neq 0$ be fixed. Assume that $\dim(\mathcal{H}_K) = \infty$. Then*

$$\begin{aligned} D_{r, \odot}^{(\alpha, \beta)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})], \quad \gamma > 0, \mu > 0 \\ = \frac{r \left(\delta - \frac{\alpha}{\alpha + \beta} \right)}{\alpha \beta} \log \left(\frac{\gamma}{\mu} \right) + \frac{1}{\alpha \beta} \log \left(\frac{\alpha \left(\frac{\gamma}{\mu} \right)^p + \beta \left(\frac{\gamma}{\mu} \right)^{-q}}{\alpha + \beta} \right) \\ + \frac{1}{\alpha \beta} \log \det \left(\frac{\alpha \left(\frac{\gamma}{\mu} \right)^p \exp(pC) + \beta \left(\frac{\gamma}{\mu} \right)^{-q} \exp(-qC)}{\alpha \left(\frac{\gamma}{\mu} \right)^p + \beta \left(\frac{\gamma}{\mu} \right)^{-q}} \right) \end{aligned} \quad (77)$$

where $\delta = \frac{\alpha \gamma^r}{\alpha \gamma^r + \beta \mu^r}$, $p = r(1 - \delta)$, $q = r\delta$, and C is the $2m \times 2m$ matrix given by

$$C = \begin{pmatrix} \log \left(I_m + \frac{J_m K[\mathbf{X}] J_m}{\gamma m} \right) & - \frac{J_m K[\mathbf{X}, \mathbf{Y}] J_m}{\sqrt{\gamma \mu m}} g \left(\frac{J_m K[\mathbf{Y}] J_m}{\mu m} \right) \\ \frac{J_m K[\mathbf{Y}, \mathbf{X}] J_m}{\sqrt{\gamma \mu m}} g \left(\frac{J_m K[\mathbf{X}] J_m}{\gamma m} \right) & - \log \left(I_m + \frac{J_m K[\mathbf{Y}] J_m}{\mu m} \right) \end{pmatrix}. \quad (78)$$

Furthermore, for $\alpha = \beta$, $r = r(\alpha)$ smooth with $r(0) = 0$, $r'(0) \neq 0$, $r(\alpha) \neq 0$ for $\alpha \neq 0$, we verify directly that

$$\begin{aligned} \lim_{\alpha \rightarrow 0} D_{r, \odot}^{(\alpha, \alpha)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})] &= \frac{[r'(0)]^2}{8} [\text{tr}(C^2) + (\log \gamma - \log \mu)^2] \\ &= \frac{[r'(0)]^2}{8} \|\log(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})\|_{\text{HS}_X}^2. \end{aligned} \quad (79)$$

Remark 3 We use a slightly different notation compared to Theorem 6 in [9], whose matrices A^*A , B^*B , and A^*B differ from those in the current work by a multiplicative factor of $\frac{1}{\gamma}$, $\frac{1}{\mu}$, and $\frac{1}{\sqrt{\gamma \mu}}$, respectively. An equivalent expression for $\|\log(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})\|_{\text{HS}_X}^2$, is given in [9] (Theorem 11) and Theorem 20 below.

We now obtain the closed form formula for $D_{r, \odot}^{(0, \beta)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})]$, which gives the formula for $D_{r, \odot}^{(\alpha, 0)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})]$ via dual symmetry.

Proposition 2 Assume the hypothesis of Proposition 1. Then

$$\begin{aligned} D_{r,\odot}^{(0,\beta)}[(AA^* + \gamma I_{\mathcal{H}_2}), (BB^* + \mu I_{\mathcal{H}_2})] \\ = -\frac{r}{\beta^2} \log \frac{\gamma}{\mu} + \frac{1}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] + \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \text{tr}[\exp(rC) - I_{\mathcal{H}_1}] \\ - \frac{r}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \left[\log \det \left(\frac{A^*A}{\gamma} + I_{\mathcal{H}_1} \right) - \log \det \left(\frac{B^*B}{\mu} + I_{\mathcal{H}_1} \right) \right], \end{aligned} \quad (80)$$

where the operator $C : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2$ is as defined in Proposition 1.

Applying Proposition 2, we immediately obtain

Theorem 18 Assume that $\dim(\mathcal{H}_K) = \infty$. Let $\beta > 0$ be fixed. Then

$$\begin{aligned} D_{r,\odot}^{(0,\beta)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})] &\quad \gamma > 0, \mu > 0 \\ = -\frac{r}{\beta^2} \log \frac{\gamma}{\mu} + \frac{1}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] + \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \text{tr}[\exp(rC) - I_{2m}] \\ - \frac{r}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \left[\log \det \left(\frac{J_m K[\mathbf{X}] J_m}{\gamma} + I_m \right) - \log \det \left(\frac{J_m K[\mathbf{Y}] J_m}{\mu} + I_m \right) \right], \end{aligned} \quad (81)$$

where C is the $2m \times 2m$ matrix as defined in Theorem 17.

4 Extended Power-Hilbert–Schmidt Distances

In this section, we present the generalization of the power Euclidean distances in Sect. 1.1 to the set of positive definite Hilbert–Schmidt operators $\mathcal{PC}_2(\mathcal{H})$.

Definition 6 (Extended power-Hilbert–Schmidt distance) For $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$, $\alpha \in \mathbb{R}$, $\alpha \neq 0$, define

$$d_{\text{HS}_X, \alpha}[(A + \gamma I), (B + \mu I)] = \left\| \frac{(A + \gamma I)^\alpha - (B + \mu I)^\alpha}{\alpha} \right\|_{\text{HS}_X}. \quad (82)$$

The following is the generalization of Lemma 1 to the infinite-dimensional setting.

Theorem 19 For any pair $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$,

$$\lim_{\alpha \rightarrow 0} \left\| \frac{(A + \gamma I)^\alpha - (B + \mu I)^\alpha}{\alpha} \right\|_{\text{HS}_X} = \|\log(A + \gamma I) - \log(B + \mu I)\|_{\text{HS}_X}. \quad (83)$$

Remark 4 In Definition 6, for $\alpha > 0$, we can also define, for two positive Hilbert–Schmidt operators A, B , not necessarily positive definite, the following distance

$$d_{\text{HS},\alpha}(A, B) = \left\| \frac{A^\alpha - B^\alpha}{\alpha} \right\|_{\text{HS}}. \quad (84)$$

While this is a valid distance on $\text{HS}(\mathcal{H}) \cap \text{Sym}^+(\mathcal{H})$, we will *not* have a limit of the form given in Theorem 19, since $\log(A)$ is *unbounded* even when A is strictly positive.

Theorem 19 is based on the following property, which generalizes Lemma 2.

Lemma 9 *Assume that $A + I > 0$, where $A \in \text{HS}(\mathcal{H})$. Then*

$$\lim_{\alpha \rightarrow 0} \left\| \frac{(A + I)^\alpha - I}{\alpha} - \log(A + I) \right\|_{\text{HS}} = 0. \quad (85)$$

Similarly, for $A + \gamma I > 0$,

$$\lim_{\alpha \rightarrow 0} \left\| \frac{(A + \gamma I)^\alpha - \gamma^\alpha I}{\alpha} - \log\left(\frac{A}{\gamma} + I\right) \right\|_{\text{HS}} = 0, \quad \forall \gamma > 0. \quad (86)$$

In terms of the Hilbert–Schmidt norm and the trace operation, the extended power-Hilbert–Schmidt distance decomposes as follows.

Proposition 3 *For any two operators $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_2(\mathcal{H})$, for any $\alpha \in \mathbb{R}$,*

$$\begin{aligned} & \| (A + \gamma I)^\alpha - (B + \mu I)^\alpha \|_{\text{HS}}^2 = \\ &= \gamma^{2\alpha} \left\| \left(\frac{A}{\gamma} + I \right)^\alpha - I \right\|_{\text{HS}}^2 - 2\gamma^\alpha \mu^\alpha 2\text{tr} \left[\left(\left(\frac{A}{\gamma} + I \right)^\alpha - I \right) \left(\left(\frac{B}{\mu} + I \right)^\alpha - I \right) \right] \\ &+ \mu^{2\alpha} \left\| \left(\frac{B}{\mu} + I \right)^\alpha - I \right\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2. \end{aligned} \quad (87)$$

4.1 The Case of RKHS Covariance Operators

In this section, we present the closed form expression for the extended power-Hilbert–Schmidt distance between RKHS covariance operators, that is

$$d_{\text{HS}_{X,\alpha}}[(C_{\Phi(X)} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(Y)} + \mu I_{\mathcal{H}_K})], \quad (88)$$

where $C_{\Phi(X)}$ and $C_{\Phi(Y)}$ are two RKHS covariance operators as defined in Sect. 3.4.

We first define the following function $h_\alpha(A)$, $\alpha > 0$, for a self-adjoint, positive compact operator A on \mathcal{H} , which plays a role similar to $g(A)$ in Sect. 3.4.

Lemma 10 *Let $A : \mathcal{H} \rightarrow \mathcal{H}$ be a self-adjoint, positive, compact operator on \mathcal{H} . Let $\{\lambda_k(A)\}_{k=1}^{N_A}$, $1 \leq N_A \leq \infty$, be the nonzero eigenvalues of A , with corresponding*

orthonormal eigenvectors $\{\phi_k(A)\}_{k=1}^{N_A}$. For $\alpha > 0$, the orthogonal spectral decomposition

$$h_\alpha(A) = \sum_{k=1}^{N_A} \frac{(1 + \lambda_k(A))^\alpha - 1}{\lambda_k(A)} \phi_k(A) \otimes \phi_k(A) \quad (89)$$

defines a bounded, self-adjoint, positive operator on \mathcal{H} . Similarly, for $\alpha < 0$, the following orthogonal spectral decomposition

$$-h_\alpha(A) = \sum_{k=1}^{N_A} \frac{1 - (1 + \lambda_k(A))^\alpha}{\lambda_k(A)} \phi_k(A) \otimes \phi_k(A) \quad (90)$$

defines a bounded, self-adjoint, positive operator on \mathcal{H} . The operator $h_\alpha(A)$ satisfies

$$Ah_\alpha(A) = h_\alpha(A)A = (I + A)^\alpha - I. \quad (91)$$

In particular, for $\alpha = 1$, we have for any positive, self-adjoint compact operator A ,

$$Ah_1(A) = h_1(A)A = \sum_{k=1}^{N_A} \lambda_k(A) \phi_k(A) \otimes \phi_k(A) = A. \quad (92)$$

Corollary 2 Assume that \mathcal{H}_1 and \mathcal{H}_2 are two separable Hilbert spaces. Let $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be a compact operator. Then for $\alpha > 0$,

$$(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2} = Ah_\alpha(A^*A)A^*, \quad (93)$$

where h_α is as defined in Eq. (89). The special case $\alpha = 1$ is recovered by the property

$$Ah_1(A^*A) = A, \quad h_1(A^*A)A^* = A^*. \quad (94)$$

The special case $\alpha = -1$ is recovered by the property

$$Ah_{-1}(A^*A) = -A(A^*A + I_{\mathcal{H}_1})^{-1}, \quad h_{-1}(A^*A)A^* = -(A^*A + I_{\mathcal{H}_1})^{-1}A^*. \quad (95)$$

Corollary 3 Assume that \mathcal{H}_1 and \mathcal{H}_2 are two separable Hilbert spaces. Let $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be two compact operators such that $A^*A, B^*B : \mathcal{H}_1 \rightarrow \mathcal{H}_1$ are Hilbert–Schmidt. Then $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are also Hilbert–Schmidt and for any $\alpha \in \mathbb{R}$,

$$\begin{aligned} \|(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}\|_{\text{HS}(\mathcal{H}_2)} &= \|(A^*A + I_{\mathcal{H}_1})^\alpha - I_{\mathcal{H}_1}\|_{\text{HS}(\mathcal{H}_1)}, \\ \langle [(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}], [(BB^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}] \rangle_{\text{HS}(\mathcal{H}_2)} & \end{aligned} \quad (96)$$

$$= \langle h_\alpha(A^*A)A^*B, A^*Bh_\alpha(B^*B) \rangle_{\text{HS}(\mathcal{H}_1)}. \quad (97)$$

Proposition 4 Assume that \mathcal{H}_1 and \mathcal{H}_2 are two separable Hilbert spaces. Let $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be two compact operators such that $A^*A, B^*B : \mathcal{H}_1 \rightarrow \mathcal{H}_1$ are Hilbert–Schmidt. Then $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are also Hilbert–Schmidt and for any $\gamma > 0$, $\mu > 0$, $\alpha \in \mathbb{R}$,

$$\begin{aligned} & \| (AA^* + \gamma I_{\mathcal{H}_2})^\alpha - (BB^* + \mu I_{\mathcal{H}_2})^\alpha \|_{\text{HS}}^2 = \| (A^*A + \gamma I_{\mathcal{H}_1})^\alpha - \gamma^\alpha I_{\mathcal{H}_1} \|_{\text{HS}}^2 \\ & \quad - 2(\gamma\mu)^{\alpha-1} \left\langle h_\alpha \left(\frac{A^*A}{\gamma} \right) A^*B, A^*B h_\alpha \left(\frac{B^*B}{\mu} \right) \right\rangle_{\text{HS}} \\ & \quad + \| (B^*B + \mu I_{\mathcal{H}_1})^\alpha - \mu I_{\mathcal{H}_1} \|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2. \end{aligned} \quad (98)$$

In particular, for $\alpha = 1$, Proposition 4 gives

$$\begin{aligned} & \| (AA^* + \gamma I_{\mathcal{H}_2}) - (BB^* + \mu I_{\mathcal{H}_2}) \|_{\text{HS}}^2 \\ &= \| AA^* \|_{\text{HS}}^2 + \| BB^* \|_{\text{HS}}^2 - 2\text{tr}(AA^*BB^*) + (\gamma - \mu)^2 \\ &= \| A^*A \|_{\text{HS}}^2 + \| B^*B \|_{\text{HS}}^2 - 2\langle A^*B, A^*B \rangle_{\text{HS}} + (\gamma - \mu)^2 \\ &= \| A^*A \|_{\text{HS}}^2 + \| B^*B \|_{\text{HS}}^2 - 2\| A^*B \|_{\text{HS}}^2 + (\gamma - \mu)^2. \end{aligned} \quad (99)$$

As in Sect. 3.4, we now apply Proposition 4 to the setting $\mathcal{H}_1 = \mathbb{R}^m$, $\mathcal{H}_2 = \mathcal{H}_K$, $A = \frac{1}{\sqrt{m}}\Phi(\mathbf{X})J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$, $B = \frac{1}{\sqrt{m}}\Phi(\mathbf{Y})J_m : \mathbb{R}^m \rightarrow \mathcal{H}_K$, so that

$$A^*A = \frac{1}{m} J_m K[\mathbf{X}]J_m, \quad B^*B = \frac{1}{m} J_m K[\mathbf{Y}]J_m, \quad A^*B = \frac{1}{m} J_m K[\mathbf{X}, \mathbf{Y}]J_m.$$

Let N_A, N_B be the strictly positive eigenvalues of A^*A and B^*B , respectively. Consider the following (reduced) singular value decompositions of A^*A and B^*B

$$A^*A = U_A \Sigma_A U_A^T, \quad B^*B = U_B \Sigma_B U_B^T, \quad (100)$$

where U_A is a matrix of size $m \times N_A$ with $U_A^T U_A = I_{N_A}$, U_B is a matrix of size $m \times N_B$ with $U_B^T U_B = I_{N_B}$, Σ_A is a diagonal matrix of size $N_A \times N_A$ whose main diagonal entries are the nonzero eigenvalues of A^*A , Σ_B is a diagonal matrix of size $N_B \times N_B$ whose main diagonal entries are the nonzero eigenvalues of B^*B . Define

$$\begin{aligned} D_{AB,\alpha} &= \mathbf{1}_{N_A}^T [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] \Sigma_A^{-1} (U_A^T A^* B U_B \circ U_A^T A^* B U_B) \\ &\quad \times [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} \mathbf{1}_{N_B}, \end{aligned} \quad (101)$$

where \circ denotes the Hadamard (element-wise) matrix product. Likewise, define

$$\begin{aligned} C_{AB} &= \lim_{\alpha \rightarrow 0} \frac{D_{AB,\alpha}}{\alpha^2} = \mathbf{1}_{N_A}^T [\log(\Sigma_A/\gamma + I_{N_A})] \Sigma_A^{-1} (U_A^T A^* B U_B \circ U_A^T A^* B U_B) \\ &\quad \times [\log(\Sigma_B/\mu + I_{N_B})] \Sigma_B^{-1} \mathbf{1}_{N_B}. \end{aligned} \quad (102)$$

The following is our main result in this section.

Theorem 20 Assume that $\dim(\mathcal{H}_K) = \infty$. Then for any $\alpha \in \mathbb{R}$,

$$\begin{aligned} & \| (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})^\alpha - (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})^\alpha \|_{\text{HS}_{\mathbf{X}}}^2 = \text{tr}[((\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A})^2] \\ & + \text{tr}[((\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B})^2] - 2D_{AB,\alpha} + (\gamma^\alpha - \mu^\alpha)^2. \end{aligned} \quad (103)$$

As $\alpha \rightarrow 0$, as expected, we have

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \| (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})^\alpha - (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})^\alpha \|_{\text{HS}_{\mathbf{X}}}^2 \\ & = \text{tr}[\log^2(\Sigma_A/\gamma + I_{N_A})] + \text{tr}[\log^2(\Sigma_B/\mu + I_{N_B})] - 2C_{AB} + (\log \gamma - \log \mu)^2 \\ & = \| \log(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K}) \|_{\text{HS}_{\mathbf{X}}}^2. \end{aligned} \quad (104)$$

5 Proofs for the Main Results

5.1 Proofs for the Power Euclidean Distances

Proof (of Lemma 2) Let $A = U \Lambda U^T$ denote the spectral decomposition for A , where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $U = [\mathbf{u}_1, \dots, \mathbf{u}_n]$, with $\{\lambda_k\}_{k=1}^n$ being the eigenvalues of A and $\{\mathbf{u}_k\}_{k=1}^n$ the corresponding orthonormal eigenvectors. Then $A^\alpha = U \Lambda^\alpha U^T$, $\log(A) = U \log(\Lambda) U^T$, and

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - I}{\alpha} - \log(A) \right\|_F^2 = \lim_{\alpha \rightarrow 0} \text{tr} \left[\frac{\Lambda^\alpha - I}{\alpha} - \log(\Lambda) \right]^2 \\ & = \lim_{\alpha \rightarrow 0} \sum_{k=1}^n \left(\frac{\lambda_k^\alpha - 1}{\alpha} - \log \lambda_k \right)^2 = 0, \quad \text{by Lemma 19.} \end{aligned} \quad \square$$

Proof (of Lemma 3) Let $\{\lambda_j\}_{j=1}^n$ be the eigenvalues of A . Without loss of generality, assume that $\lambda_j > 0$, $1 \leq j \leq n-1$, and $\lambda_n = 0$. Then we have

$$\left\| \frac{A^\alpha - I}{\alpha} \right\|_F^2 = \sum_{j=1}^n \frac{(\lambda_j^\alpha - 1)^2}{\alpha^2} = \frac{1}{\alpha^2} + \sum_{j=1}^{n-1} \frac{(\lambda_j^\alpha - 1)^2}{\alpha^2}.$$

Thus $\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - I}{\alpha} \right\|_F^2 = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} + \sum_{j=1}^{n-1} (\log \lambda_j)^2 = \infty$. \square

Proof (of Lemma 1) By Lemma 2, we have

$$\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - I}{\alpha} - \log(A) \right\|_F = 0, \quad \lim_{\alpha \rightarrow 0} \left\| \frac{B^\alpha - I}{\alpha} - \log(B) \right\|_F = 0.$$

It then follows from Lemma 18 that $\lim_{\alpha \rightarrow 0} \left\| \frac{A^\alpha - B^\alpha}{\alpha} \right\|_F = \| \log(A) - \log(B) \|_F$. \square

5.2 Proofs for the Log-Hilbert–Schmidt Divergences

Since the Log-Euclidean divergences are special cases of the Log-Hilbert–Schmidt divergences, we present proofs for the latter. To prove Lemma 5, we first prove the following technical results.

Lemma 11 ([8]) Let B be a constant with $0 < B < 1$. Then for all $|x| \leq B$,

$$|\log(1+x)| \leq \frac{1}{1-B}|x|. \quad (105)$$

Lemma 12 (i) Assume that $(A + I) \in \mathcal{PC}_1(\mathcal{H})$. Then $\log(A + I) \in \text{Tr}(\mathcal{H})$. Conversely, if $B \in \text{Tr}(\mathcal{H})$, then $\exp(B) = A + I$, for some $A \in \text{Tr}(\mathcal{H})$.

(ii) Assume that $(A + I) \in \mathcal{PC}_2(\mathcal{H})$. Then $\log(A + I) \in \text{HS}(\mathcal{H})$. Conversely, if $B \in \text{HS}(\mathcal{H})$, then $\exp(B) = A + I$, for some $A \in \text{HS}(\mathcal{H})$.

Proof (i) For the first part, assume that the eigenvalues of A are $\{\lambda_k\}_{k=1}^{\infty}$, $\lambda_k + 1 > 0 \forall k \in \mathbb{N}$, with corresponding orthonormal eigenvectors $\{\phi_k\}_{k=1}^{\infty}$. Then $\log(A + I) = \sum_{k=1}^{\infty} \log(\lambda_k + 1)\phi_k \otimes \phi_k$. If $A \in \text{Tr}(\mathcal{H})$, then $\sum_{k=1}^{\infty} |\lambda_k| < \infty$. For any constant $0 < \epsilon < 1$, there exists $N(\epsilon) \in \mathbb{N}$ such that $|\lambda_k| \leq \epsilon \forall k > N(\epsilon)$. Thus by Lemma 11

$$\|\log(A + I)\|_{\text{tr}} = \sum_{k=1}^{\infty} |\log(\lambda_k + 1)| \leq \sum_{k=1}^{N(\epsilon)} |\log(\lambda_k + 1)| + \frac{1}{1-\epsilon} \sum_{k=N(\epsilon)+1}^{\infty} |\lambda_k| < \infty.$$

Thus $\log(A + I) \in \text{Tr}(\mathcal{H})$. Conversely, if $B \in \text{Tr}(\mathcal{H})$, then $\exp(B) = I + \sum_{k=1}^{\infty} \frac{B^k}{k!}$, with $\sum_{k=1}^{\infty} \frac{B^k}{k!} \in \text{Tr}(\mathcal{H})$, since $\left\| \sum_{k=1}^{\infty} \frac{B^k}{k!} \right\|_{\text{tr}} \leq \sum_{k=1}^{\infty} \frac{\|B\|_{\text{tr}}^k}{k!} = \exp(\|B\|_{\text{tr}}) - 1 < \infty$.

(ii) The proof of the second part is entirely analogous. □

Lemma 13 (i) Assume that $(A + I), (B + I) \in \mathcal{PC}_1(\mathcal{H})$. Then $(A + I) \odot (B + I) = C + I > 0$, for some $C \in \text{Tr}(\mathcal{H})$.

(ii) Assume that $(A + I), (B + I) \in \mathcal{PC}_2(\mathcal{H})$. Then $(A + I) \odot (B + I) = C + I > 0$, for some $C \in \text{HS}(\mathcal{H})$.

Proof (i) For the first part, we have by Lemma 12 that for $(A + I), (B + I) \in \mathcal{PC}_1(\mathcal{H})$, $\log(A + I) \in \text{Tr}(\mathcal{H})$, $\log(B + I) \in \text{Tr}(\mathcal{H})$, so that again by Lemma 12,

$$(A + I) \odot (B + I) = \exp[\log(A + I) + \log(B + I)] = C + I > 0,$$

where $C = \sum_{k=1}^{\infty} \frac{[\log(A+I)+\log(B+I)]^k}{k!} \in \text{Tr}(\mathcal{H})$.

(ii) The proof of the second part is entirely analogous. □

Proof (of Lemma 5) For $(A + \gamma I), (B + \mu I) \in \mathcal{PC}_1(\mathcal{H})$,

$$\begin{aligned} (A + \gamma I) \odot (B + \mu I)^{-1} &= \exp[\log(A + \gamma I) - \log(B + \mu I)] \\ &= \exp[\log(I + (A/\gamma)) - \log(I + (B/\mu)) + (\log \gamma - \log \mu)I] \\ &= \frac{\gamma}{\mu} \exp[\log(I + (A/\gamma)) - \log(I + (B/\mu))] = \frac{\gamma}{\mu} [(I + (A/\gamma)) \odot (I + (B/\mu))^{-1}]. \end{aligned}$$

By Lemma 13, $(I + (A/\gamma)) \odot (I + (B/\mu))^{-1} = C + I$ for some $C \in \text{Tr}(\mathcal{H})$. Thus

$$(A + \gamma I) \odot (B + \mu I)^{-1} = \frac{\gamma}{\mu} (C + I) = Z + \frac{\gamma}{\mu} I > 0, \quad \text{where } Z = \frac{\gamma}{\mu} C. \quad \square$$

The proofs for the subsequent results in Sect. 3.1 parallel those in [7] and the proofs in Sects. 3.2 and 3.3 parallel those in [11]. Thus for the results in these sections, we only prove, for illustration, the Dual Invariance property. We then present the proofs of the divergences between RKHS covariance operators in Sect. 3.4.

Proof (of Theorem 15 - Dual invariance under inversion) For simplicity, we prove

$$D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] = D_{r,\odot}^{(\beta,\alpha)}[(A + \gamma I), (B + \mu I)]$$

when $\alpha > 0, \beta > 0$. The proofs for the other cases are similar. We have

$$(A + \gamma I)^{-1} = \frac{1}{\gamma} I - \frac{A}{\gamma} (A + \gamma I)^{-1}, \quad (B + \mu I)^{-1} = \frac{1}{\mu} I - \frac{B}{\mu} (B + \mu I)^{-1}.$$

Write $\delta = \delta(\alpha, \beta) = \frac{\alpha\gamma^r}{\alpha\gamma^r + \beta\mu^r}$. Then $\delta(\beta, \alpha) = \frac{\beta\gamma^r}{\beta\gamma^r + \alpha\mu^r}$. Let $Z_2 + \frac{\mu}{\gamma} I = (A + \gamma I)^{-1} \odot (B + \mu I) = (Z + \frac{\gamma}{\mu} I)^{-1}$, where $Z_2 \in \text{HS}(\mathcal{H})$. By definition,

$$\begin{aligned} &D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] \\ &= \frac{1}{\alpha\beta} \log \left(\frac{1/\gamma}{1/\mu} \right)^{\delta_2 - \frac{\alpha}{\alpha+\beta}} + \frac{1}{\alpha\beta} \log \det \left(\frac{\alpha (Z_2 + \frac{\mu}{\gamma} I)^{r(1-\delta_2)} + \beta (Z_2 + \frac{\mu}{\gamma} I)^{-r\delta_2}}{\alpha + \beta} \right), \end{aligned}$$

where $\delta_2 = \frac{\alpha(1/\gamma)^r}{\alpha(1/\gamma)^r + \beta(1/\mu)^r} = \frac{\alpha\mu^r}{\alpha\mu^r + \beta\gamma^r} = 1 - \delta(\beta, \alpha)$ and $1 - \delta_2 = \delta(\beta, \alpha)$. Also $\delta_2 - \frac{\alpha}{\alpha+\beta} = 1 - \delta(\beta, \alpha) - \frac{\alpha}{\alpha+\beta} = -(\delta(\beta, \alpha) - \frac{\beta}{\alpha+\beta})$. It thus follows that

$$\begin{aligned} &D_{r,\odot}^{(\alpha,\beta)}[(A + \gamma I)^{-1}, (B + \mu I)^{-1}] \\ &= \frac{1}{\alpha\beta} \log \left(\frac{\gamma}{\mu} \right)^{\delta(\beta,\alpha) - \frac{\beta}{\alpha+\beta}} + \frac{1}{\alpha\beta} \log \det \left(\frac{\beta (Z + \frac{\gamma}{\mu} I)^{r(1-\delta(\beta,\alpha))} + \alpha (Z + \frac{\gamma}{\mu} I)^{-r\delta(\beta,\alpha)}}{\alpha + \beta} \right) \\ &= D_{r,\odot}^{(\beta,\alpha)}[(A + \gamma I), (B + \mu I)]. \quad \square \end{aligned}$$

Proof (of Lemma 8) The operator $g(A)$ in Eq.(68) is well-defined and bounded by the limits $\lim_{k \rightarrow \infty} \lambda_k(A) = 0$ and $\lim_{x \rightarrow 0} \frac{\log(1+x)}{x} = 1$. Furthermore, it is obvious that $g(A)$ is self-adjoint and positive, since $\log(1+x) > 0 \forall x > 0$. \square

Proof (of Corollary 1) For any $w \in \mathcal{H}_2$, we have

$$\begin{aligned} (A\phi_k(A^*A)) \otimes (A\phi_k(A^*A))w &= \langle (A\phi_k(A^*A)), w \rangle_{\mathcal{H}_2} (A\phi_k(A^*A)) \\ &= \langle \phi_k(A^*A), A^*w \rangle_{\mathcal{H}_1} (A\phi_k(A^*A)) = A(\phi_k(A^*A) \otimes \phi_k(A^*A))A^*w. \end{aligned}$$

By Lemma 8, the following operator

$$g(A^*A) = \sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A^*A))}{\lambda_k(A^*A)} \phi_k(A^*A) \otimes \phi_k(A^*A)$$

is bounded, self-adjoint, and positive. From Eq.(66) and the previous expression,

$$\begin{aligned} \log(I_{\mathcal{H}_2} + AA^*) &= \sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A^*A))}{\lambda_k(A^*A)} A\phi_k(A^*A) \otimes A\phi_k(A^*A) \\ &= A \left[\sum_{k=1}^{N_A} \frac{\log(1 + \lambda_k(A^*A))}{\lambda_k(A^*A)} \phi_k(A^*A) \otimes \phi_k(A^*A) \right] A^* = Ag(A^*A)A^*. \quad \square \end{aligned}$$

In the following result, we use the fact that for two separable Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ and two compact operators $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ and $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$, the nonzero eigenvalues of $AB : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ and $BA : \mathcal{H}_1 \rightarrow \mathcal{H}_1$ are the same.

Lemma 14 *Assume that \mathcal{H}_1 and \mathcal{H}_2 are two separable Hilbert spaces. Let $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be two compact operators. Then the nonzero eigenvalues of the operator $\log(I_{\mathcal{H}_2} + AA^*) - \log(I_{\mathcal{H}_2} + BB^*) : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are the same as those of the operator*

$$\begin{pmatrix} \log(I_{\mathcal{H}_1} + A^*A) & -A^*Bg(B^*B) \\ B^*Ag(A^*A) & -\log(I_{\mathcal{H}_1} + B^*B) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2, \quad (106)$$

where \mathcal{H}_1^2 denotes the direct sum of \mathcal{H}_1 with itself.

Proof We note that $\mathcal{H}_1^2 = \mathcal{H}_1 \oplus \mathcal{H}_1 = \{(h_1, h_2) : h_1, h_2 \in \mathcal{H}_1\}$, with inner product

$$\langle (h_1, h_2), (k_1, k_2) \rangle_{\mathcal{H}_1^2} = \langle h_1, k_1 \rangle_{\mathcal{H}_1} + \langle h_2, k_2 \rangle_{\mathcal{H}_1}.$$

For two operators $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$, we can define the operator $[A \ B] : \mathcal{H}_1^2 \rightarrow \mathcal{H}_2$ by

$$[A \ B] \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = Ah_1 + Bh_2, \quad h_1, h_2 \in \mathcal{H}_1,$$

along with the adjoint operator $[A \ B]^* = \begin{pmatrix} A^* \\ B^* \end{pmatrix} : \mathcal{H}_2 \rightarrow \mathcal{H}_1^2$, defined by

$$[A \ B]^* w = \begin{pmatrix} A^* \\ B^* \end{pmatrix} w = \begin{pmatrix} A^* w \\ B^* w \end{pmatrix}, \quad w \in \mathcal{H}_2.$$

Thus we can write

$$\begin{aligned} \log(I_{\mathcal{H}_2} + AA^*) - \log(I_{\mathcal{H}_2} + BB^*) &= Ag(A^*A)A^* - Bg(B^*B)B^* \\ &= [Ag(A^*A) - Bg(B^*B)] \begin{pmatrix} A^* \\ B^* \end{pmatrix}, \text{ where } [Ag(A^*A) - Bg(B^*B)] : \mathcal{H}_1^2 \rightarrow \mathcal{H}_2. \end{aligned}$$

From the discussion preceding the lemma, the nonzero eigenvalues of $\log(I_{\mathcal{H}_2} + AA^*) - \log(I_{\mathcal{H}_2} + BB^*) : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are the same as the nonzero eigenvalues of the operator

$$\begin{aligned} \begin{pmatrix} A^* \\ B^* \end{pmatrix} [Ag(A^*A) - Bg(B^*B)] &= \begin{pmatrix} A^*Ag(A^*A) & -A^*Bg(B^*B) \\ B^*Ag(A^*A) & -B^*Bg(B^*B) \end{pmatrix} \\ &= \begin{pmatrix} \log(I_{\mathcal{H}_1} + A^*A) & -A^*Bg(B^*B) \\ B^*Ag(A^*A) & -\log(I_{\mathcal{H}_1} + B^*B) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2. \end{aligned} \quad \square$$

Lemma 15 Let $\mathcal{H}_1, \mathcal{H}_2$ be separable Hilbert spaces. Let $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be compact operators such that $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are trace class operators. Then for any $\alpha > 0, \beta > 0$ and any $p, q \in \mathbb{R}$,

$$\begin{aligned} \det \left(\frac{\alpha[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^p + \beta[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^q}{\alpha + \beta} \right) \\ = \det \left[\frac{\alpha \exp(pC) + \beta \exp(qC)}{\alpha + \beta} \right], \end{aligned} \quad (107)$$

where the operator $C : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2$ is the trace class operator defined by

$$C = \begin{pmatrix} \log(I_{\mathcal{H}_1} + A^*A) & -A^*Bg(B^*B) \\ B^*Ag(A^*A) & -\log(I_{\mathcal{H}_1} + B^*B) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2, \quad (108)$$

Proof (of Lemma 15) We have by definition

$$(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1} = \exp[\log(AA^* + I_{\mathcal{H}_2}) - \log(BB^* + I_{\mathcal{H}_2})].$$

Thus if $\{\lambda_k\}_{k=1}^\infty$ are the eigenvalues of $\log(AA^* + I_{\mathcal{H}_2}) - \log(BB^* + I_{\mathcal{H}_2})$, then the eigenvalues of $(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}$ are $\{e^{\lambda_k}\}_{k=1}^\infty$ and

$$\det \left(\frac{\alpha[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^p + \beta[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^q}{\alpha + \beta} \right) \\ = \prod_{k=1}^{\infty} \frac{\alpha e^{p\lambda_k} + \beta e^{q\lambda_k}}{\alpha + \beta} = \prod_{k=1, \lambda_k \neq 0}^{\infty} \frac{\alpha e^{p\lambda_k} + \beta e^{q\lambda_k}}{\alpha + \beta}.$$

By Lemma 14, the nonzero eigenvalues of the operator $\log(I_{\mathcal{H}_2} + AA^*) - \log(I_{\mathcal{H}_2} + BB^*) : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are the same as the nonzero eigenvalues of the operator

$$C = \begin{pmatrix} \log(I_{\mathcal{H}_1} + A^*A) & -A^*Bg(B^*B) \\ B^*Ag(A^*A) & -\log(I_{\mathcal{H}_1} + B^*B) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2,$$

from which the desired result follows. \square

Proof (of Proposition 1) Let $Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2} = (AA^* + \gamma I_{\mathcal{H}_2}) \odot (BB^* + \mu I_{\mathcal{H}_2})^{-1}$, with $Z \in \text{Tr}(\mathcal{H}_2)$ and $\frac{\mu}{\gamma}Z + I_{\mathcal{H}_2} = (\frac{AA^*}{\gamma} + I_{\mathcal{H}_2}) \odot (\frac{BB^*}{\mu} + I_{\mathcal{H}_2})^{-1}$. Then

$$D_{r,\odot}^{(\alpha,\beta)}[(AA^* + \gamma I_{\mathcal{H}_2}), (BB^* + \mu I_{\mathcal{H}_2})] \\ = \frac{r\left(\delta - \frac{\alpha}{\alpha+\beta}\right)}{\alpha\beta} \log\left(\frac{\gamma}{\mu}\right) + \frac{1}{\alpha\beta} \log \det_X \left(\frac{\alpha\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^{r(1-\delta)} + \beta\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^{-r\delta}}{\alpha + \beta} \right) \\ = \frac{r\left(\delta - \frac{\alpha}{\alpha+\beta}\right)}{\alpha\beta} \log\left(\frac{\gamma}{\mu}\right) + \frac{1}{\alpha\beta} \log \left(\frac{\alpha\left(\frac{\gamma}{\mu}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q}}{\alpha + \beta} \right) \\ + \frac{1}{\alpha\beta} \log \det \left(\frac{\alpha\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^p + \beta\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^{-q}}{\alpha\left(\frac{\gamma}{\mu}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q}} \right)$$

with $p = r(1 - \delta)$ and $q = r\delta$. The determinant in the last term is

$$\det \left(\frac{\alpha\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^p + \beta\left(Z + \frac{\gamma}{\mu}I_{\mathcal{H}_2}\right)^{-q}}{\alpha\left(\frac{\gamma}{\mu}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q}} \right) \\ = \det \left(\frac{\alpha\left(\frac{\gamma}{\mu}\right)^p \left(\frac{\mu}{\gamma}Z + I_{\mathcal{H}_2}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q} \left(\frac{\mu}{\gamma}Z + I_{\mathcal{H}_2}\right)^{-q}}{\alpha\left(\frac{\gamma}{\mu}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q}} \right) \\ = \det \left(\frac{\alpha\left(\frac{\gamma}{\mu}\right)^p \exp(pC) + \beta\left(\frac{\gamma}{\mu}\right)^{-q} \exp(-qC)}{\alpha\left(\frac{\gamma}{\mu}\right)^p + \beta\left(\frac{\gamma}{\mu}\right)^{-q}} \right)$$

by Lemma 15, where the operator C is given by

$$C = \begin{pmatrix} \log\left(I_{\mathcal{H}_1} + \frac{A^*A}{\gamma}\right) & -\frac{A^*B}{\sqrt{\gamma\mu}}g\left(\frac{B^*B}{\mu}\right) \\ \frac{B^*A}{\sqrt{\gamma\mu}}g\left(\frac{A^*A}{\gamma}\right) & -\log\left(I_{\mathcal{H}_1} + \frac{B^*B}{\mu}\right) \end{pmatrix} : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2. \quad \square$$

Proof (of Theorem 17) The first part of the Theorem follows immediately from Proposition 1. For the limit, we note that for $\alpha = \beta$,

$$\begin{aligned} D_{r,\odot}^{(\alpha,\alpha)}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})] \\ = \frac{r(\delta - \frac{1}{2})}{\alpha^2} \log\left(\frac{\gamma}{\mu}\right) + \frac{1}{\alpha^2} \log \left(\frac{\left(\frac{\gamma}{\mu}\right)^p + \left(\frac{\gamma}{\mu}\right)^{-q}}{2} \right) \\ + \frac{1}{\alpha^2} \log \det \left(\frac{\left(\frac{\gamma}{\mu}\right)^p \exp(pC) + \left(\frac{\gamma}{\mu}\right)^{-q} \exp(-qC)}{\left(\frac{\gamma}{\mu}\right)^p + \left(\frac{\gamma}{\mu}\right)^{-q}} \right) \end{aligned}$$

By Lemma 20, we have

$$\begin{aligned} \lim_{\alpha \rightarrow 0} & \left[\frac{r(\delta - \frac{1}{2})}{\alpha^2} \log\left(\frac{\gamma}{\mu}\right) + \frac{1}{\alpha^2} \log \left(\frac{\left(\frac{\gamma}{\mu}\right)^p + \left(\frac{\gamma}{\mu}\right)^{-q}}{2} \right) \right] \\ &= \frac{[r'(0)]^2}{4} \log^2\left(\frac{\gamma}{\mu}\right) - \frac{[r'(0)]^2}{8} \log^2\left(\frac{\gamma}{\mu}\right) = \frac{[r'(0)]^2}{8} (\log \gamma - \log \mu)^2. \end{aligned}$$

Let $\{\lambda_k\}_{k=1}^{2m}$ be the eigenvalues of C , then we have by Lemma 21,

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \log \det \left(\frac{\left(\frac{\gamma}{\mu}\right)^p \exp(pC) + \left(\frac{\gamma}{\mu}\right)^{-q} \exp(-qC)}{\left(\frac{\gamma}{\mu}\right)^p + \left(\frac{\gamma}{\mu}\right)^{-q}} \right) \\ &= \lim_{\alpha \rightarrow 0} \sum_{j=1}^{2m} \frac{1}{\alpha^2} \log \left(\frac{\left(\frac{\gamma}{\mu}\right)^p \exp(p\lambda_j) + \left(\frac{\gamma}{\mu}\right)^{-q} \exp(-q\lambda_j)}{\left(\frac{\gamma}{\mu}\right)^p + \left(\frac{\gamma}{\mu}\right)^{-q}} \right) \\ &= \frac{[r'(0)]^2}{8} \sum_{j=1}^{2m} \lambda_j^2 = \frac{[r'(0)]^2}{8} \text{tr}(C^2). \end{aligned}$$

Combining all of these expressions, we obtain the desired limit. From the definition of C in Proposition 1, of $g(A)$ in Eq. (70) using the SVDs in Eq. (100), and C_{AB} in Eq. (102),

$$\begin{aligned}
\text{tr}(C^2) &= \text{tr} \left[\log^2 \left(I_{\mathcal{H}_1} + \frac{A^* A}{\gamma} \right) \right] + \text{tr} \left[\log^2 \left(I_{\mathcal{H}_1} + \frac{B^* B}{\mu} \right) \right] - \frac{2}{\gamma \mu} \text{tr} \left[B^* A g \left(\frac{A^* A}{\gamma} \right) A^* B g \left(\frac{B^* B}{\mu} \right) \right] \\
&= \text{tr} \left[\log^2 (I_{N_A} + \Sigma_A / \gamma) \right] + \text{tr} [\log^2 (I_{N_B} + \Sigma_B / \mu)] \\
&\quad - \frac{2}{\gamma \mu} \text{tr} [B^* A U_A \log(\Sigma_A / \gamma + I_{N_A}) (\Sigma_A / \gamma)^{-1} U_A^T A^* B U_B \log(\Sigma_B / \mu + I_{N_B}) (\Sigma_B / \mu)^{-1} U_B^T] \\
&= \text{tr} [\log^2 (I_{N_A} + \Sigma_A / \gamma)] + \text{tr} [\log^2 (I_{N_B} + \Sigma_B / \mu)] \\
&\quad - 2 \langle \log(\Sigma_A / \gamma + I_{N_A}) \Sigma_A^{-1} U_A^T A^* B U_B, U_A^T A^* B U_B \log(\Sigma_B / \mu + I_{N_B}) \Sigma_B^{-1} \rangle_F \\
&= \text{tr} [\log^2 (I_{N_A} + \Sigma_A / \gamma)] + \text{tr} [\log^2 (I_{N_B} + \Sigma_B / \mu)] - 2C_{AB}
\end{aligned}$$

Thus $\text{tr}(C^2) + (\log \gamma - \log \mu)^2 = \|\log(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})\|_{\text{HS}_{\mathbf{X}}}^2$ by Theorem 6 in [9], proving the last part of the theorem. \square

Lemma 16 *Let \mathcal{H} be a separable Hilbert space and $A \in \text{Tr}(\mathcal{H})$. Then*

$$\text{tr}_{\mathbf{X}}[\exp(A)] = 1 + \text{tr}[\exp(A) - I]. \quad (109)$$

Proof By Lemma 12, for $A \in \text{Tr}(\mathcal{H})$, we have $\exp(A) = B + I$ where $B = \exp(A) - I \in \text{Tr}(\mathcal{H})$. Thus it follows from the definition of the extended trace $\text{tr}_{\mathbf{X}}$ that

$$\text{tr}_{\mathbf{X}}[\exp(A)] = \text{tr}_{\mathbf{X}}[(\exp(A) - I) + I] = 1 + \text{tr}[\exp(A) - I].$$

\square

Lemma 17 *Let $\mathcal{H}_1, \mathcal{H}_2$ be separable Hilbert spaces. Let $A, B : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ be compact operators such that $AA^*, BB^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are trace class operators. Then*

$$\text{tr}_{\mathbf{X}}[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^r = 1 + \text{tr}[\exp(rC) - I_{\mathcal{H}_1^2}], \quad (110)$$

where $C : \mathcal{H}_1^2 \rightarrow \mathcal{H}_1^2$ is as defined in Lemma 15.

Proof By Lemma 14, the nonzero eigenvalues of the trace class operator $\log(AA^* + I_{\mathcal{H}_2}) - \log(BB^* + I_{\mathcal{H}_2}) : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are the same as those of the operator C . Thus

$$\begin{aligned}
&\text{tr}_{\mathbf{X}}[(AA^* + I_{\mathcal{H}_2}) \odot (BB^* + I_{\mathcal{H}_2})^{-1}]^r \\
&= \text{tr}_{\mathbf{X}}[\exp(r(\log(AA^* + I_{\mathcal{H}_2}) - \log(BB^* + I_{\mathcal{H}_2})))]] \\
&= 1 + \text{tr}[\exp(r(\log(AA^* + I_{\mathcal{H}_2}) - \log(BB^* + I_{\mathcal{H}_2}))) - I_{\mathcal{H}_2}] \\
&= 1 + \text{tr}[\exp(rC) - I_{\mathcal{H}_1^2}]. \quad \square
\end{aligned}$$

Proof (of Proposition 2) By Lemma 17, we have

$$\begin{aligned}
\text{tr}_{\mathbf{X}}[(BB^* + \mu I_{\mathcal{H}_2})^{-1} \odot (AA^* + \gamma I_{\mathcal{H}_2})]^r &= \left(\frac{\gamma}{\mu} \right)^r \text{tr}_{\mathbf{X}} \left[\left(\frac{BB^*}{\mu} + I_{\mathcal{H}_2} \right)^{-1} \odot \left(\frac{AA^*}{\gamma} + I_{\mathcal{H}_2} \right)^r \right] \\
&= \left(\frac{\gamma}{\mu} \right)^r (1 + \text{tr}[\exp(rC) - I_{\mathcal{H}_1^2}]). \text{ Furthermore,}
\end{aligned}$$

$$\begin{aligned}
& \log \det_X [(BB^* + \mu I_{\mathcal{H}_2})^{-1} \odot (AA^* + \gamma I_{\mathcal{H}_2})] \\
&= \log \det_X (AA^* + \gamma I_{\mathcal{H}_2}) - \log \det_X (BB^* + \mu I_{\mathcal{H}_2}) \\
&= \log \det \left(\frac{AA^*}{\gamma} + I_{\mathcal{H}_2} \right) - \log \det \left(\frac{BB^*}{\mu} + I_{\mathcal{H}_2} \right) + \log \gamma - \log \mu.
\end{aligned}$$

Thus we have by definition

$$\begin{aligned}
& D_{r,\odot}^{(0,\beta)} [(AA^* + \gamma I_{\mathcal{H}_2}), (BB^* + \mu I_{\mathcal{H}_2})] \\
&= \frac{r}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] \log \frac{\gamma}{\mu} + \frac{1}{\beta^2} \text{tr}_X [((BB^* + \mu I_{\mathcal{H}_2})^{-1} \odot (AA^* + \gamma I_{\mathcal{H}_2}))^r - I] \\
&\quad - \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \log \det_X [(BB^* + \mu I_{\mathcal{H}_2})^{-1} \odot (AA^* + \gamma I_{\mathcal{H}_2})]^r \\
&= \frac{r}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] \log \frac{\gamma}{\mu} + \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r (1 + \text{tr}[\exp(rC) - I_{\mathcal{H}_1^2}]) - \frac{1}{\beta^2} \\
&\quad - \frac{r}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \left[\log \det \left(\frac{A^* A}{\gamma} + I_{\mathcal{H}_1} \right) - \log \det \left(\frac{B^* B}{\mu} + I_{\mathcal{H}_1} \right) + \log \frac{\gamma}{\mu} \right] \\
&= -\frac{r}{\beta^2} \log \frac{\gamma}{\mu} + \frac{1}{\beta^2} \left[\left(\frac{\gamma}{\mu} \right)^r - 1 \right] + \frac{1}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \text{tr}[\exp(rC) - I_{\mathcal{H}_1^2}] \\
&\quad - \frac{r}{\beta^2} \left(\frac{\gamma}{\mu} \right)^r \left[\log \det \left(\frac{A^* A}{\gamma} + I_{\mathcal{H}_1} \right) - \log \det \left(\frac{B^* B}{\mu} + I_{\mathcal{H}_1} \right) \right]. \quad \square
\end{aligned}$$

5.3 Proofs for the Extended Power-Hilbert–Schmidt Distances

Proof (of Lemma 9) Let $\{\lambda_j\}_{j \in \mathbb{N}}$ be the eigenvalues of A with corresponding orthonormal eigenvectors $\{\phi_j\}_{j \in \mathbb{N}}$, then $A = \sum_{j=1}^{\infty} \lambda_j \phi_j \otimes \phi_j$, where $\phi_j \otimes \phi_j : \mathcal{H} \rightarrow \mathcal{H}$ is the rank-one operator defined by $(\phi_j \otimes \phi_j)w = \langle \phi_j, w \rangle \phi_j \forall w \in \mathcal{H}$. Thus $\log(A + I) = \sum_{j=1}^{\infty} \log(\lambda_j + 1) \phi_j \otimes \phi_j$. Let $\{e_k\}_{k=1}^{\infty}$ be any orthonormal basis for \mathcal{H} . Then

$$\begin{aligned}
\langle \phi_i \otimes \phi_i, \phi_j \otimes \phi_j \rangle_{\text{HS}} &= \sum_{k=1}^{\infty} \langle (\phi_i \otimes \phi_i)e_k, (\phi_j \otimes \phi_j)e_k \rangle = \sum_{k=1}^{\infty} \langle \phi_i, e_k \rangle \langle \phi_j, e_k \rangle \langle \phi_i, \phi_j \rangle \\
&= \delta_{ij} \sum_{k=1}^{\infty} |\langle \phi_i, e_k \rangle|^2 = \delta_{ij} |\phi_i|^2 = \delta_{ij}.
\end{aligned}$$

Thus $\{(\phi_j \otimes \phi_j)\}_{j=1}^{\infty}$ is an orthonormal system in the Hilbert space $\text{HS}(\mathcal{H})$. By Lebesgue's Monotone Convergence Theorem, we then have

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \left\| \frac{(A + I)^\alpha - I}{\alpha} - \log(A + I) \right\|_{\text{HS}}^2 &= \lim_{\alpha \rightarrow 0} \sum_{j=1}^{\infty} \left[\frac{(\lambda_j + 1)^\alpha - 1}{\alpha} - \log(\lambda_j + 1) \right]^2 \\ &= \sum_{j=1}^{\infty} \lim_{\alpha \rightarrow 0} \left[\frac{(\lambda_j + 1)^\alpha - 1}{\alpha} - \log(\lambda_j + 1) \right]^2 = 0, \end{aligned}$$

since $\lim_{\alpha \rightarrow 0} \frac{(\lambda_j + 1)^\alpha - 1}{\alpha} = \log(\lambda_j + 1), \forall j \in \mathbb{N}$ by Lemma 19. This proves Eq. (85).

For the last expression, Eq. (86), since $\lim_{\alpha \rightarrow 0} \gamma^\alpha = 1$,

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \left\| \frac{[(A + \gamma I)^\alpha - \gamma^\alpha I]}{\alpha} - \log\left(\frac{A}{\gamma} + I\right) \right\|_{\text{HS}} &= \lim_{\alpha \rightarrow 0} \left\| \frac{\gamma^\alpha [(A/\gamma + I)^\alpha - I]}{\alpha} - \log\left(\frac{A}{\gamma} + I\right) \right\|_{\text{HS}} \\ &= \lim_{\alpha \rightarrow 0} \left\| \frac{[(A/\gamma + I)^\alpha - I]}{\alpha} - \log\left(\frac{A}{\gamma} + I\right) \right\|_{\text{HS}} = 0 \quad \text{by Eq. (85).} \end{aligned} \quad \square$$

Proof (of Theorem 19) By definition of the extended Hilbert–Schmidt norm, we have

$$\begin{aligned} &\left\| \frac{(A + \gamma I)^\alpha - (B + \mu I)^\alpha}{\alpha} \right\|_{\text{HS}_X}^2 \\ &= \left\| \frac{[(A + \gamma I)^\alpha - \gamma^\alpha I] - [(B + \mu I)^\alpha - \mu^\alpha I]}{\alpha} + \frac{(\gamma^\alpha - \mu^\alpha)}{\alpha} I \right\|_{\text{HS}_X}^2 \\ &= \left\| \frac{[(A + \gamma I)^\alpha - \gamma^\alpha I] - [(B + \mu I)^\alpha - \mu^\alpha I]}{\alpha} \right\|_{\text{HS}}^2 + \frac{(\gamma^\alpha - \mu^\alpha)^2}{\alpha^2}. \end{aligned}$$

By L'Hopital's rule, we have $\lim_{\alpha \rightarrow 0} \frac{(\gamma^\alpha - \mu^\alpha)^2}{\alpha^2} = (\log \gamma - \log \mu)^2$. By Lemma 9,

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \left\| \frac{(A + \gamma I)^\alpha - \gamma^\alpha I}{\alpha} - \log\left(\frac{A}{\gamma} + I\right) \right\|_{\text{HS}} &= 0, \\ \lim_{\alpha \rightarrow 0} \left\| \frac{(B + \mu I)^\alpha - \mu^\alpha I}{\alpha} - \log\left(\frac{B}{\mu} + I\right) \right\|_{\text{HS}} &= 0. \end{aligned}$$

Thus it follows from Lemma 18 that

$$\lim_{\alpha \rightarrow 0} \left\| \frac{[(A + \gamma I)^\alpha - \gamma^\alpha I] - [(B + \mu I)^\alpha - \mu^\alpha I]}{\alpha} \right\|_{\text{HS}} = \left\| \log\left(\frac{A}{\gamma} + I\right) - \log\left(\frac{B}{\mu} + I\right) \right\|_{\text{HS}}.$$

Combining the previous limits, we obtain

$$\begin{aligned} \lim_{\alpha \rightarrow 0} \left\| \frac{(A + \gamma I)^\alpha - (B + \mu I)^\alpha}{\alpha} \right\|_{\text{HS}_X}^2 &= \left\| \log\left(\frac{A}{\gamma} + I\right) - \log\left(\frac{B}{\mu} + I\right) \right\|_{\text{HS}}^2 + (\log \gamma - \log \mu)^2 \\ &= \| \log(A + \gamma I) - \log(B + \mu I) \|_{\text{HS}_X}^2. \end{aligned} \quad \square$$

Proof (of Proposition 3) By definition of the extended Hilbert–Schmidt norm,

$$\begin{aligned}
& \|(A + \gamma I)^\alpha - (B + \mu I)^\alpha\|_{\text{HS}_X}^2 \\
&= \|[[(A + \gamma I)^\alpha - \gamma^\alpha I] - [(B + \mu I)^\alpha - \mu^\alpha I] + (\gamma^\alpha - \mu^\alpha)I\]\|_{\text{HS}_X}^2 \\
&= \|[[(A + \gamma I)^\alpha - \gamma^\alpha I] - [(B + \mu I)^\alpha - \mu^\alpha I]\]\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2 \\
&= \|(A + \gamma I)^\alpha - \gamma^\alpha I\|_{\text{HS}}^2 - 2\text{tr}[(A + \gamma I)^\alpha - \gamma^\alpha I](B + \mu I)^\alpha - \mu^\alpha I] \\
&\quad + \|(B + \mu I)^\alpha - \mu^\alpha I\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2 \\
&= \gamma^{2\alpha} \left\| \left(\frac{A}{\gamma} + I \right)^\alpha - I \right\|_{\text{HS}}^2 - 2\gamma^\alpha \mu^\alpha 2\text{tr} \left[\left(\left(\frac{A}{\gamma} + I \right)^\alpha - I \right) \left(\left(\frac{B}{\mu} + I \right)^\alpha - I \right) \right] \\
&\quad + \mu^{2\alpha} \left\| \left(\frac{B}{\mu} + I \right)^\alpha - I \right\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2.
\end{aligned}$$

□

Proof (of Lemma 10) Consider the function $f(x) = \frac{(1+x)^\alpha - 1}{x}$ for $x > 0$. By L'Hopital's rule, we have $\lim_{x \rightarrow 0} \frac{(1+x)^\alpha - 1}{x} = \lim_{x \rightarrow 0} \alpha(1+x)^{\alpha-1} = \alpha$. Thus the operator $h_\alpha(A)$ is always bounded. Furthermore, for $\alpha > 0$, we have $(1+x)^\alpha - 1 > 0 \forall x > 0$, so that $h_\alpha(A)$ is a self-adjoint and positive operator on \mathcal{H} . Similarly, when $\alpha < 0$, we have $(1+x)^\alpha - 1 < 0 \forall x > 0$, so that $-h_\alpha(A)$ is a self-adjoint, positive operator on \mathcal{H} . □

Proof (of Corollary 2) Let $\{\lambda_k(A^*A)\}_{k=1}^{N_A}$ be the nonzero eigenvalues of $A^*A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$, with corresponding orthonormal eigenvectors $\{\phi_k(A^*A)\}_{k=1}^{N_A}$. These are precisely the nonzero eigenvalues of $AA^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$. Then the nonzero eigenvalues of $(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2} : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are precisely $\{(1 + \lambda_k(A^*A))^\alpha - 1\}_{k=1}^{N_A}$, with corresponding orthonormal eigenvectors $\left\{ \frac{A\phi_k(A^*A)}{\sqrt{\lambda_k(A^*A)}} \right\}_{k=1}^{N_A}$. Thus we have the spectral decomposition

$$\begin{aligned}
& (AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2} = \sum_{k=1}^{N_A} \frac{(1 + \lambda_k(A^*A))^\alpha - 1}{\lambda_k(A^*A)} (A\phi_k(A^*A)) \otimes (A\phi_k(A^*A)) \\
&= A \left(\sum_{k=1}^{N_A} \frac{(1 + \lambda_k(A^*A))^\alpha - 1}{\lambda_k(A^*A)} \phi_k(A^*A) \otimes \phi_k(A^*A) \right) A^* \\
&= Ah_\alpha(A^*A)A^*, \quad \text{by Lemma 10.}
\end{aligned}$$

From the property $A\phi_k(A^*A) = 0$ for $\lambda_k(A^*A) = 0$, we have for $\alpha = 1$,

$$Ah_1(A^*A) = A \sum_{k=1}^{N_A} \phi_k(A^*A) \otimes \phi_k(A^*A) = A \sum_{k=1}^{\infty} \phi_k(A^*A) \otimes \phi_k(A^*A) = A.$$

By taking adjoint, we have $h_1(A^*A)A^* = A^*$. Thus it follows that $Ah_1(A^*A)A^* = AA^*$ as expected. Similarly, for $\alpha = -1$, from the identity $\frac{1-(1+x)^{-1}}{x} = \frac{1}{1+x}$, $x \neq 0, x \neq -1$,

$$\begin{aligned} Ah_{-1}(A^*A) &= -A \left[\sum_{k=1}^{N_A} \frac{1}{1 + \lambda_k(A^*A)} \phi_k(A^*A) \otimes \phi_k(A^*A) \right] \\ &= -A \left[\sum_{k=1}^{\infty} \frac{1}{1 + \lambda_k(A^*A)} \phi_k(A^*A) \otimes \phi_k(A^*A) \right] = -A(I_{\mathcal{H}_1} + A^*A)^{-1}. \end{aligned}$$

Consequently, by taking adjoint, we have $h_{-1}(A^*A)A^* = -(I_{\mathcal{H}_1} + A^*A)^{-1}A^*$. Thus

$$(AA^* + I_{\mathcal{H}_2})^{-1} - I_{\mathcal{H}_2} = Ah_{-1}(A^*A)A^* = -A(A^*A + I_{\mathcal{H}_1})^{-1}A^*$$

as can be directly verified. This completes the proof. \square

Proof (of Corollary 3) The first expression follows from the fact that the nonzero eigenvalues of $AA^* : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ are the same as the nonzero eigenvalues of $A^*A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$ and hence the nonzero eigenvalues of $(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}$ are the same as the nonzero eigenvalues of $(A^*A + I_{\mathcal{H}_1})^\alpha - I_{\mathcal{H}_1}$. It follows that

$$\| (AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2} \|_{\text{HS}(\mathcal{H}_2)} = \| (A^*A + I_{\mathcal{H}_1})^\alpha - I_{\mathcal{H}_1} \|_{\text{HS}(\mathcal{H}_1)}.$$

For the second expression on the inner product, we have by Corollary 2 that

$$\begin{aligned} &\langle [(AA^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}], [(BB^* + I_{\mathcal{H}_2})^\alpha - I_{\mathcal{H}_2}] \rangle_{\text{HS}(\mathcal{H}_2)} \\ &= \langle Ah_\alpha(A^*A)A^*, Bh_\alpha(B^*B)B^* \rangle_{\text{HS}(\mathcal{H}_2)} \\ &= \text{tr}[Ah_\alpha(A^*A)A^*Bh_\alpha(B^*B)B^*] = \text{tr}[B^*Ah_\alpha(A^*A)A^*Bh_\alpha(B^*B)] \\ &= \langle h_\alpha(A^*A)A^*B, A^*Bh_\alpha(B^*B) \rangle_{\text{HS}(\mathcal{H}_1)}. \end{aligned} \quad \square$$

Proof (of Proposition 4) By Proposition 3,

$$\begin{aligned} &\| (AA^* + \gamma I_{\mathcal{H}_2})^\alpha - (BB^* + \mu I_{\mathcal{H}_2})^\alpha \|_{\text{HS}_X}^2 = \gamma^{2\alpha} \left\| \left(\frac{AA^*}{\gamma} + I_{\mathcal{H}_2} \right)^\alpha - I_{\mathcal{H}_2} \right\|_{\text{HS}}^2 \\ &- 2\gamma^\alpha \mu^\alpha 2\text{tr} \left[\left(\left(\frac{AA^*}{\gamma} + I_{\mathcal{H}_2} \right)^\alpha - I_{\mathcal{H}_2} \right) \left(\left(\frac{BB^*}{\mu} + I_{\mathcal{H}_2} \right)^\alpha - I_{\mathcal{H}_2} \right) \right] \\ &+ \mu^{2\alpha} \left\| \left(\frac{BB^*}{\mu} + I_{\mathcal{H}_2} \right)^\alpha - I_{\mathcal{H}_2} \right\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2. \end{aligned}$$

By Corollary 3, the last expression becomes

$$\begin{aligned}
& \gamma^{2\alpha} \left\| \left(\frac{A^* A}{\gamma} + I_{\mathcal{H}_1} \right)^\alpha - I_{\mathcal{H}_1} \right\|_{\text{HS}}^2 - 2\gamma^\alpha \mu^\alpha 2 \left\langle h_\alpha \left(\frac{A^* A}{\gamma} \right) \frac{A^* B}{\sqrt{\gamma\mu}}, \frac{A^* B}{\sqrt{\gamma\mu}} h_\alpha \left(\frac{B^* B}{\mu} \right) \right\rangle_{\text{HS}} \\
& + \mu^{2\alpha} \left\| \left(\frac{B^* B}{\mu} + I_{\mathcal{H}_1} \right)^\alpha - I_{\mathcal{H}_1} \right\|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2 \\
& = \| (A^* A + \gamma I_{\mathcal{H}_1})^\alpha - \gamma^\alpha I_{\mathcal{H}_1} \|_{\text{HS}}^2 - 2(\gamma\mu)^{\alpha-1} \left\langle h_\alpha \left(\frac{A^* A}{\gamma} \right) A^* B, A^* B h_\alpha \left(\frac{B^* B}{\mu} \right) \right\rangle_{\text{HS}} \\
& + \| (B^* B + \mu I_{\mathcal{H}_1})^\alpha - \mu I_{\mathcal{H}_1} \|_{\text{HS}}^2 + (\gamma^\alpha - \mu^\alpha)^2.
\end{aligned}$$

□

Proof (of Theorem 20) With the given SVDs of $A^* A$ and $B^* B$, we have

$$\begin{aligned}
h_\alpha \left(\frac{A^* A}{\gamma} \right) &= \sum_{k=1}^{N_A} \frac{(1 + (1/\gamma)\lambda_k(A^* A))^\alpha - 1}{(1/\gamma)\lambda_k(A^* A)} \phi_k(A^* A) \otimes \phi_k(A^* A) \\
&= \frac{1}{\gamma^{\alpha-1}} \sum_{k=1}^{N_A} \frac{(\lambda_k(A^* A) + \gamma)^\alpha - \gamma^\alpha}{\lambda_k(A^* A)} \phi_k(A^* A) \otimes \phi_k(A^* A) \\
&= \frac{1}{\gamma^{\alpha-1}} U_A [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] \Sigma_A^{-1} U_A^T.
\end{aligned}$$

Similarly, $h_\alpha \left(\frac{B^* B}{\mu} \right) = \frac{1}{\mu^{\alpha-1}} U_B [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} U_B^T$. It follows that

$$\begin{aligned}
& (\gamma\mu)^{\alpha-1} \left\langle h_\alpha \left(\frac{A^* A}{\gamma} \right) A^* B, A^* B h_\alpha \left(\frac{B^* B}{\mu} \right) \right\rangle_F \\
& = \langle U_A [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] \Sigma_A^{-1} U_A^T A^* B, A^* B U_B [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} U_B^T \rangle_F \\
& = \text{tr}(B^* A U_A \Sigma_A^{-1} [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] U_A^T A^* B U_B [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} U_B^T) \\
& = \text{tr}(U_B^T B^* A U_A \Sigma_A^{-1} [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] U_A^T A^* B U_B [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1}) \\
& = \langle [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] \Sigma_A^{-1} U_A^T A^* B U_B, U_A^T A^* B U_B [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} \rangle_F \\
& = \mathbf{1}_{N_A}^T [(\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A}] \Sigma_A^{-1} (U_A^T A^* B U_B \circ U_A^T A^* B U_B) \\
& \quad \times [(\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B}] \Sigma_B^{-1} \mathbf{1}_{N_B} = D_{AB,\alpha}.
\end{aligned}$$

By Proposition 4, we then have

$$\begin{aligned}
& \|(C_{\Phi}(\mathbf{X}) + \gamma I_{\mathcal{H}_K})^\alpha - (C_{\Phi}(\mathbf{Y}) + \mu I_{\mathcal{H}_K})^\alpha\|_{\text{HS}_{\mathbf{X}}}^2 \\
& = \|(A^* A + \gamma I_m)^\alpha - \gamma^\alpha I_m\|_F^2 + \|(B^* B + \mu I_m)^\alpha - \mu^\alpha I_m\|_F^2 \\
& \quad - 2(\gamma\mu)^{\alpha-1} \left\langle h_\alpha \left(\frac{A^* A}{\gamma} \right) A^* B, A^* B h_\alpha \left(\frac{B^* B}{\mu} \right) \right\rangle_F + (\gamma^\alpha - \mu^\alpha)^2 \\
& = \text{tr}[((\Sigma_A + \gamma I_{N_A})^\alpha - \gamma^\alpha I_{N_A})^2] + \text{tr}[((\Sigma_B + \mu I_{N_B})^\alpha - \mu^\alpha I_{N_B})^2] \\
& \quad - 2D_{AB,\alpha} + (\gamma^\alpha - \mu^\alpha)^2.
\end{aligned}$$

By the three limits in Lemma 19, we have

$$\begin{aligned} & \lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \| (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})^\alpha - (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})^\alpha \|_{\text{HS}_{\mathbf{X}}}^2 \\ &= \text{tr}[\log^2(\Sigma_A/\gamma + I_{N_A})] + \text{tr}[\log^2(\Sigma_B/\mu + I_{N_B})] - 2C_{AB} + (\log \gamma - \log \mu)^2 \\ &= \| \log(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}) - \log(C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K}) \|_{\text{HS}_{\mathbf{X}}}^2, \end{aligned}$$

where the last equality follows from Theorem 6 in [9]. \square

5.4 Miscellaneous Technical Results

We collect here some technical results that are used in proving the main results.

Lemma 18 *Let E be a Banach space. Let $A, B, \{A_n\}_{n \in \mathbb{N}}, \{B_n\}_{n \in \mathbb{N}} \in E$ be such that $\lim_{n \rightarrow \infty} \|A_n - A\| = 0, \lim_{n \rightarrow \infty} \|B_n - B\| = 0$. Then*

$$\lim_{n \rightarrow \infty} \|(A_n - B_n) - (A - B)\| = 0, \quad \lim_{n \rightarrow \infty} \|A_n - B_n\| = \|A - B\|. \quad (111)$$

Proof This follows from the triangle inequality. \square

The limits in the following lemmas can be obtained by applying L'Hopital's rule.

Lemma 19 *Assume that $\gamma > 0, \mu > 0$ are fixed. Then for $\alpha \in \mathbb{R}$,*

$$\lim_{\alpha \rightarrow 0} \frac{\gamma^\alpha - \mu^\alpha}{\alpha} = \log \gamma - \log \mu. \quad (112)$$

In particular, for $\lambda \in \mathbb{R}, \lambda > 0, \lim_{\alpha \rightarrow 0} \frac{\lambda^\alpha - 1}{\alpha} = \log(\lambda)$, and for $\lambda \in \mathbb{R}$ such that $\lambda + \gamma > 0, \lim_{\alpha \rightarrow 0} \frac{(\lambda + \gamma)^\alpha - \gamma^\alpha}{\alpha} = \log\left(\frac{\lambda}{\gamma} + 1\right)$.

Lemma 20 ([7]) *Let $\gamma > 0$ be fixed. Let $\lambda > 0$ be fixed. Assume that $r = r(\alpha)$ is smooth, with $r(0) = 0$. Define $\delta = \frac{\gamma^\alpha}{\gamma^r + 1}, p = r(1 - \delta), q = r\delta$. Then*

$$\lim_{\alpha \rightarrow 0} \frac{r(\delta - \frac{1}{2})}{\alpha^2} = \frac{[r'(0)]^2}{4} \log \gamma. \quad (113)$$

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \log \left(\frac{\lambda^p + \lambda^{-q}}{2} \right) = \frac{[r'(0)]^2}{4} \left[-(\log \gamma)(\log \lambda) + \frac{1}{2}(\log \lambda)^2 \right]. \quad (114)$$

In particular, if $\gamma = \lambda$, then $\lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \log \left(\frac{\gamma^p + \gamma^{-q}}{2} \right) = -\frac{[r'(0)]^2}{8} (\log \gamma)^2$.

Lemma 21 Let $\gamma > 0$ be fixed. Let $\lambda \in \mathbb{R}$ be fixed. Assume that $r = r(\alpha)$ is smooth, with $r(0) = 0$. Define $\delta = \frac{\gamma^r}{\gamma^r + 1}$, $p = r(1 - \delta)$, $q = r\delta$. Then

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \log \left(\frac{\gamma^p e^{p\lambda} + \gamma^{-q} e^{-q\lambda}}{\gamma^p + \gamma^{-q}} \right) = \frac{[r'(0)]^2}{8} \lambda^2, \quad (115)$$

independent of γ . For $r = r(\alpha) = 2\alpha$, $\lim_{\alpha \rightarrow 0} \frac{1}{\alpha^2} \log \left(\frac{\gamma^p e^{p\lambda} + \gamma^{-q} e^{-q\lambda}}{\gamma^p + \gamma^{-q}} \right) = \frac{1}{2} \lambda^2$.

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29**(1), 328–347 (2007)
2. Bhatia, R.: Positive Definite Matrices. Princeton University Press, Princeton (2007)
3. Chebbi, Z., Moakher, M.: Means of Hermitian positive-definite matrices based on the log-determinant α -divergence function. *Linear Algebr. Appl.* **436**(7), 1872–1889 (2012)
4. Cichocki, A., Cruces, S., Amari, S.: Log-determinant divergences revisited: alpha-beta and gamma log-det divergences. *Entropy* **17**(5), 2988–3034 (2015)
5. Dryden, I., Koloydenko, A., Zhou, D.: Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3**, 1102–1123 (2009)
6. Larotonda, G.: Nonpositive curvature: a geometrical approach to Hilbert–Schmidt operators. *Differ. Geom. Appl.* **25**, 679–700 (2007)
7. Minh, H.: Infinite-dimensional log-determinant divergences II: alpha-beta divergences. [arXiv:1610.08087v2](https://arxiv.org/abs/1610.08087v2) (2016)
8. Minh, H.: Infinite-dimensional log-determinant divergences between positive definite trace class operators. *Linear Algebr. Appl.* **528**, 331–383 (2017)
9. Minh, H., San Biagio, M., Murino, V.: Log-Hilbert–Schmidt metric between positive definite operators on Hilbert spaces. In: Advances in Neural Information Processing Systems (NIPS), pp. 388–396 (2014)
10. Minh, H.Q.: Affine-invariant Riemannian distance between infinite-dimensional covariance operators. *Geometric Science of Information*, pp. 30–38 (2015)
11. Minh, H.Q.: Log-determinant divergences between positive definite Hilbert–Schmidt operators. *Geometric Science of Information* (2017)
12. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66**(1), 41–66 (2006)
13. Sra, S.: A new metric on the manifold of kernel matrices with application to matrix geometric means. In: Advances in Neural Information Processing Systems (NIPS), pp. 144–152 (2012)
14. Steinwart, I., Christmann, A.: Support Vector Machines. Springer Science & Business Media, Berlin (2008)

Part III

**Theoretical Aspects of Information
Geometry**

Entropy on Spin Factors



Peter Harremoës

Abstract Recently it has been demonstrated that the Shannon entropy or the von Neuman entropy are the only entropy functions that generate a local Bregman divergences as long as the state space has rank 3 or higher. In this paper we will study the properties of Bregman divergences for convex bodies of rank 2. The two most important convex bodies of rank 2 can be identified with the bit and the qubit. We demonstrate that if a convex body of rank 2 has a Bregman divergence that satisfies sufficiency then the convex body is spectral and if the Bregman divergence is monotone then the convex body has the shape of a ball. A ball can be represented as the state space of a spin factor, which is the most simple type of Jordan algebra. We also study the existence of recovery maps for Bregman divergences on spin factors. In general the convex bodies of rank 2 appear as faces of state spaces of higher rank. Therefore our results give strong restrictions on which convex bodies could be the state space of a physical system with a well-behaved entropy function.

Keywords Bregman divergence · Entropy · Monotonicity · Spin factor Sufficiency

1 Introduction

Although quantum physics has been around for more than a century the foundation of the theory is still somewhat obscure. Quantum theory operates at distances and energy levels that are very far from everyday experience and much of our intuition does not carry over to the quantum world. Nevertheless, the mathematical models of quantum physics have an impressive predictive power. These years many scientists try to contribute to the development of quantum computers and it becomes more important to pinpoint the nature of the quantum resources that are supposed to speed

P. Harremoës (✉)
Niels Brock Copenhagen Business College, Copenhagen, Denmark
e-mail: harremoes@ieee.org
URL: <http://peter.harremoes.dk>

up the processing of a quantum computer compared with a classic computer. There is also an interest in extending quantum physics to be able to describe gravity on the quantum level and maybe the foundation of quantum theory has to be modified in order to be able to describe gravity. Therefore the foundation of quantum theory is not only of philosophical interest, but it is also important for application of the existing theory and for extending the theory.

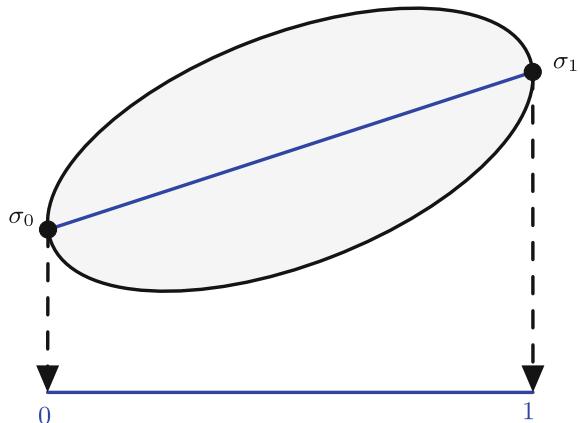
A computer has to consist of some components and the smallest component must be a memory cell. In a classical computer each memory cell can store one bit. In a quantum computer a memory cell can store one qubit. In this paper we will focus on such minimal memory cells and demonstrate that under certain assumptions any such memory cell can be represented as a so-called spin factor. We formalize the memory cell by requiring that the state space has rank 2. In some recent papers it was proved that a local Bregman divergence on a state space of rank at least 3 is proportional to information divergence and the state space must be spectral [8, 10]. Further, on a state space of rank at least 3 locality of a Bregman divergence is equivalent to the conditions called sufficiency and monotonicity. If the rank of the state space is 2 the situation is quite different. First of all the condition called locality reduce almost to a triviality. Therefore it is of interest to study sufficiency and monotonicity on state spaces of rank 2.

The paper is organized as follows. In the first part we study convex bodies and use mathematical terminology without reference to physics. The convex bodies may or may not correspond to state spaces of physical systems. In Sect. 2 some basic terminology regarding convex sets is established and the rank of a set is defined. In Sect. 3 regret and Bregman divergences are defined, but for a detailed motivation we refer to [8]. In Sect. 4 spectral sets are defined and it is proved that a spectral set of rank 2 has central symmetry. In Sect. 5 sufficiency of a regret function is defined and it is proved that a convex body of rank 2 with a regret function that satisfies sufficiency is spectral.

Spin factors are introduced in Sect. 6. Spin factors appear as sections of state spaces of physical systems described by density matrices on complex Hilbert spaces. Therefore we will borrow some terminology from physics. In Sect. 7 monotonicity of a Bregman divergence is introduced. It is proved that a convex body with a sufficient Bregman divergence that is monotone under dilations can be represented as a spin factor. For general spin factors we have not obtained a simple characterization of the monotone Bregman divergence, but some partial results are presented in Sect. 8. In Sect. 9 it is proved that equality in the inequality for a monotone Bregman divergence implies the existence of a recovery map.

In this paper we focus on finite dimensional convex bodies. Many of the results can easily be generalized to bounded convex set in separable Hilbert spaces, but that would require that topological considerations were taken into account.

Fig. 1 A retraction with orthogonal points σ_0 and σ_1 . The corresponding section is obtained by reversing the arrows



2 Convex Bodies of Rank 2

In this paper we will work within a category where the objects are *convex bodies*, i.e. finite dimensional convex compact sets. The morphisms will be *affinities*, i.e. affine maps between convex bodies. The convex bodies are candidates for state spaces of physical systems, so a point in a convex bodies might be interpreted as a state that represents our knowledge of the physical system. A convex combination $\sum p_i \cdot \sigma_i$ is interpreted as a state where the system is prepared in state σ_i with probability p_i . In classical physics the state space is a simplex and in the standard formalism of quantum physics the state space is isomorphic to the density matrices on a complex Hilbert space.

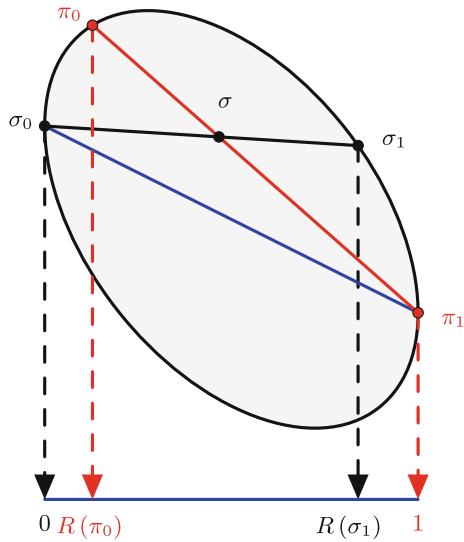
A bijective affinity will be called an *isomorphism*. Let \mathcal{K} and \mathcal{L} denote convex bodies. An affinity $S : \mathcal{K} \rightarrow \mathcal{L}$ is called a *section* if there exists an affinity $R : \mathcal{L} \rightarrow \mathcal{K}$ such that $R \circ S = id_{\mathcal{K}}$, and such an affinity R is called a *retraction*. Often we will identify a section $S : \mathcal{K} \rightarrow \mathcal{L}$ with the set $S(\mathcal{K})$ as a subset of \mathcal{L} . Note that the affinity $S \circ R : \mathcal{L} \rightarrow \mathcal{L}$ is *idempotent* and that any idempotent affinity determines a section/retraction pair. We say that σ_0 and σ_1 are *mutually singular* if there exists a section $S : [0, 1] \rightarrow \mathcal{K}$ such that $S(0) = \sigma_0$ and $S(1) = \sigma_1$. Such a section is illustrated on Fig. 1. A retraction $R : \mathcal{K} \rightarrow [0, 1]$ is a special case of a *test* [13, p. 15] (or an *effect* as it is often called in generalized probabilistic theories [2]). We say that $\sigma_0, \sigma_1 \in \mathcal{K}$ are *orthogonal* if σ_0 and σ_1 belong to a face \mathcal{F} of \mathcal{K} such that σ_0 and σ_1 are mutually singular in \mathcal{F} .

The following result was stated in [9] without a detailed proof.

Theorem 1 *If σ is a point in a convex body \mathcal{K} then σ can be written as a convex combination $\sigma = (1 - t) \cdot \sigma_0 + t \cdot \sigma_1$ where σ_0 and σ_1 orthogonal.*

Proof Without loss of generality we may assume that σ is an algebraically interior point of \mathcal{K} . For any σ_0 on the boundary of \mathcal{K} there exists a σ_1 on the boundary of \mathcal{K} and $t_{\sigma_0} \in]0, 1[$ such that $(1 - t_{\sigma_0}) \cdot \sigma_0 + t_{\sigma_0} \cdot \sigma_1 = \sigma$. Let R denote a retraction

Fig. 2 Illustration to the proof of Theorem 1



$R : \mathcal{K} \rightarrow [0, 1]$ such that $R(\sigma_0) = 0$. Let S denote a section corresponding to R such that $S(0) = \sigma_0$. Let π_1 denote the point $S(1)$. There exists a point π_0 on the boundary such that $\sigma = (1 - t_{\pi_0}) \cdot \pi_0 + t_{\pi_0} \cdot \pi_1$ (Fig. 2). Then

$$R(\sigma) = R((1 - t_{\pi_0}) \cdot \pi_0 + t_{\pi_0} \cdot \pi_1) \quad (1)$$

$$= (1 - t_{\pi_0}) \cdot R(\pi_0) + t_{\pi_0} \cdot R(\pi_1) \quad (2)$$

$$\geq t_{\pi_0} \quad (3)$$

and

$$R(\sigma) = R((1 - t_{\sigma_0}) \cdot \sigma_0 + t_{\sigma_0} \cdot \sigma_1) \quad (4)$$

$$= (1 - t_{\sigma_0}) \cdot 0 + t_{\sigma_0} \cdot R(\sigma_1) \quad (5)$$

$$= t_{\sigma_0} \cdot R(\sigma_1). \quad (6)$$

Therefore

$$t_{\sigma_0} \cdot R(\sigma_1) \geq t_{\pi_0}. \quad (7)$$

Since t_{σ_0} is a continuous function of σ_0 we may choose σ_0 such that t_{σ_0} is minimal, but if t_{σ_0} is minimal Inequality (7) implies that $R(\sigma_1) = 1$ so that σ_0 and σ_1 are orthogonal. \square

Iterated use of Theorem 1 leads to an extended version of Caratheodory's theorem [9, Thm. 2].

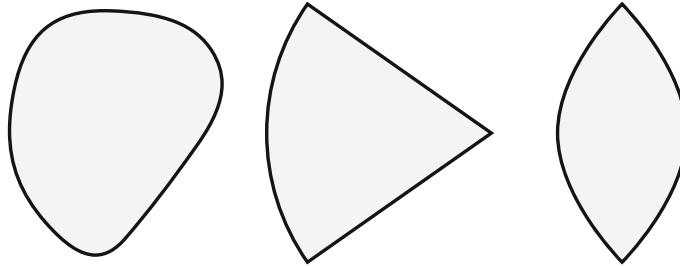


Fig. 3 Convex bodies of rank 2. The convex body to the left has a smooth strictly convex boundary so that any point on the boundary has exactly one orthogonal point. The body in the middle has a set of three extreme points that are orthogonal, but any point can be written as a convex combination of just two points. The convex body to the right is centrally symmetric without 1-dimensional proper faces, i.e. it is a spectral set

Theorem 2 (Orthogonal Caratheodory Theorem) *Let \mathcal{K} denote a convex body of dimension d . Then any point $\sigma \in \mathcal{K}$ has a decomposition $\sigma = \sum_{i=1}^n t_i \cdot \sigma_i$ where t_1^n is a probability vector and σ_i are orthogonal extreme points in \mathcal{K} and $n \leq d + 1$.*

The Caratheodory number of a convex body is the maximal number of extreme points needed to decompose a point into extreme points. We need a similar definition related to orthogonal decompositions.

Definition 3 The *rank* of a convex body \mathcal{K} is the maximal number of orthogonal extreme points needed in an orthogonal decomposition of a point in \mathcal{K} .

If \mathcal{K} has rank 1 then it is a singleton. Some examples of convex bodies of rank 2 are illustrated in Fig. 3. Clearly the Caratheodory number lower bounds the rank of a convex body. Figure 5 provides an example where the Carathodory number is different from the rank. The rest of this paper will focus on convex bodies of rank 2. Convex bodies of rank 2 satisfy *weak spectrality* as defined in [4].

If \mathcal{K} is a convex body it is sometimes convenient to consider the cone \mathcal{K}_+ generated by \mathcal{K} . The cone \mathcal{K}_+ consists of elements of the form $x \cdot \sigma$ where $x \geq 0$ and $\sigma \in \mathcal{K}$. Elements of the cone are called *positive elements* and such elements can be multiplied by positive constants via $x \cdot (y \cdot \sigma) = (x \cdot y) \cdot \sigma$ and can be added as follows.

$$x \cdot \rho + y \cdot \sigma = (x + y) \cdot \left(\frac{x}{x + y} \cdot \rho + \frac{y}{x + y} \cdot \sigma \right). \quad (8)$$

For a point $\sigma \in \mathcal{K}$ the *trace* of $x \cdot \sigma \in \mathcal{K}_+$ is defined by $\text{tr}[x \cdot \sigma] = x$. The cone \mathcal{K}_+ can be embedded in a real vector space by taking the affine hull of the cone and use the apex of the cone as origin of the vector space and the trace extends to a linear function on this vector space. In this way a convex body \mathcal{K} can be identified with the set of positive elements in a vector space with trace 1.

Lemma 4 Let \mathcal{K} be a convex body and let $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ be an affinity. Let $\Phi_\mu = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} e^{-\mu} \Phi^{on}$. Then $\Phi_\infty = \lim_{\mu \rightarrow \infty} \Phi_\mu$ is a retraction of \mathcal{K} onto the set of fix-points of Φ .

Proof Since \mathcal{K} is compact the affinity Φ has a fix-point that we will call s_0 . The affinity can be extended to a positive trace preserving affinity of the real vector space generated by \mathcal{K} into itself. Since Φ maps a convex body into itself all the eigenvalues of Φ are numerically upper bounded by 1. The affinity can be extended to a complexification of the vector space. On this complexification of the vector space there exist a basis in which the affinity Φ has the Jordan normal form with blocks of the form

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \ddots & \ddots & \ddots \\ 0 & 0 & \lambda & \ddots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \quad (9)$$

and Φ^n has blocks of the form

$$\begin{pmatrix} \lambda^n & \binom{n}{n-1} \lambda^{n-1} & \binom{n}{n-2} \lambda^{n-2} & \cdots & \binom{n}{n-\ell+2} \lambda^{n-\ell+2} & \binom{n}{n-\ell+1} \lambda^{n-\ell+1} \\ 0 & \lambda^n & \binom{n}{1} \lambda^{n-1} & \ddots & \ddots & \ddots \\ 0 & 0 & \lambda^n & \ddots & \ddots & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \cdots & \lambda^n & \binom{n}{n-1} \lambda^{n-1} \\ 0 & 0 & 0 & \cdots & 0 & \lambda^n \end{pmatrix}. \quad (10)$$

Now

$$\sum_{n=0}^{\infty} \frac{\mu^n}{n!} e^{-\mu} \binom{n}{n-j} \lambda^{n-j} = \sum_{n=j}^{\infty} \frac{\mu^n}{(n-j)! j!} e^{-\mu} \lambda^{n-j} \quad (11)$$

$$= \frac{\mu^j}{j!} \sum_{n=j}^{\infty} \frac{\mu^{n-j}}{(n-j)!} e^{-\mu} \lambda^{n-j} \quad (12)$$

tends to zero for μ tending to infinity except if $\lambda = 1$. If $\lambda = 1$ then there is no uniform upper bound on Φ^n except if the Jordan block is diagonal. Therefore $\Phi_\mu = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} e^{-\mu} \Phi^{on}$ converges to a map Φ_∞ that is diagonal with eigenvalues 0 and 1, i.e. a idempotent. Since Φ and Φ_∞ commute they have the same fix-points. \square

Proposition 5 Let Φ denote an affinity $\mathcal{K} \rightarrow \mathcal{L}$ and let Ψ denote an affinity $\mathcal{L} \rightarrow \mathcal{K}$. Then the set of fix-points of $\Psi \circ \Phi$ is a section of \mathcal{K} and the set of fix-points of $\Phi \circ \Psi$ is a section of \mathcal{L} . The affinities Φ and Ψ restricted to the fix-point sets are isomorphisms between these convex bodies.

3 Regret and Bregman Divergences

Consider a payoff function where the payoff may represent extracted energy or how much data can be compressed or some other quantity of interest. Our payoff depends both of the state of the system and of some choice that we can make. Let $F(\sigma)$ denote the maximal mean payoff when our knowledge is represented by σ . Then F is a convex function on the convex body.

The two most important examples are the squared Euclidean norm $F(\vec{v}) = \|\vec{v}\|_2^2$ defined on a vector space and minus the von Neuman entropy $F(\sigma) = \text{tr}[\sigma \ln(\sigma)]$. Note that Shannon entropy may be considered as a special case of von Neuman entropy when all operators commute. We may also consider $F(\sigma) = -S_\alpha(\sigma)$ where the Tsallis entropy of order $\alpha > 0$ is defined by

$$S_\alpha(\sigma) = -\text{tr}[\sigma \log_\alpha(\sigma)] \quad (13)$$

and where the logarithm of order $\alpha \neq 1$ is given by

$$\log_\alpha(x) = \frac{x^{\alpha-1} - 1}{\alpha - 1} \quad (14)$$

and $\log_1(x) = \ln(x)$. We will study such entropy functions via the corresponding regret functions that are defined by:

Definition 6 Let F denote a convex function defined on a convex body \mathcal{K} . For $\rho, \sigma \in \mathcal{K}$ we define the *regret function* D_F by

$$D_F(\rho, \sigma) = F(\rho) - \left(F(\sigma) + \lim_{t \rightarrow 0_+} \frac{F((1-t)\cdot\sigma + t\cdot\rho) - F(\sigma)}{t} \right). \quad (15)$$

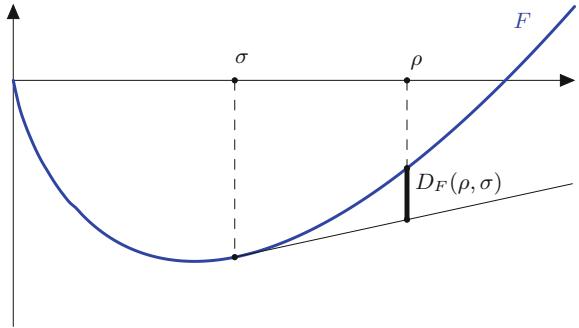
The regret function D_F is *strict* if $D_F(\rho, \sigma) = 0$ implies that $\rho = \sigma$. If F is differentiable the regret function is called a *Bregman divergence*.

The interpretation of the regret function is that $D_F(\rho, \sigma)$ tells how much more payoff one could have obtained if the state is ρ but one act as if the state was σ . This is illustrated in Fig. 4.

The two most important examples of Bregman divergences are squared Euclidean distance $\|\vec{v} - \vec{w}\|_2^2$ that is generated by the squared Euclidean norm and information divergence

$$D(\rho\|\sigma) = \text{tr}[\rho(\ln(\rho) - \ln(\sigma)) - \rho + \sigma] \quad (16)$$

Fig. 4 The regret equals the vertical distance between the curve and the tangent



that is generated by minus the von Neuman entropy. The Bregman divergence generated by $-S_\alpha$ is called the Bregman divergence of order α and is denoted $D_\alpha(\rho, \sigma)$. Various examples of payoff functions and corresponding regret functions are discussed in [8] where some basic properties of regret functions are also discussed. If F is differentiable the regret function is a Bregman divergence and the formula (15) reduces to

$$D_F(\rho, \sigma) = F(\rho) - (F(\sigma) + \langle \nabla F(\sigma) | \rho - \sigma \rangle). \quad (17)$$

Bregman divergences were introduced in [5], but they only gained popularity after their properties were investigated in great detail in [3]. A Bregman divergence satisfies the *Bregman equation*

$$\sum t_i \cdot D_F(\rho_i, \sigma) = \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma) \quad (18)$$

where (t_1, t_2, \dots) is a probability vector and $\bar{\rho} = \sum t_i \cdot \rho_i$.

Assume that D_F is a Bregman divergence on the convex body \mathcal{K} . If the state is not known exactly but we know that s is one of the states s_1, s_2, \dots, s_n then the *minimax regret* is defined as

$$C_F = \inf_{\sigma \in \mathcal{K}} \sup_{\rho \in \mathcal{K}} D_F(\rho, \sigma). \quad (19)$$

The point σ that achieves the minimax regret will be denoted by σ_{opt} .

Theorem 7 *If \mathcal{K} is a convex body with a Bregman divergence D_F and with a probability vector (t_1, t_2, \dots, t_n) on the points $\rho_1, \rho_2, \dots, \rho_n$ with $\bar{\rho} = \sum t_i \cdot \rho_i$ and σ_{opt} achieves the minimax regret then*

$$C_F \geq \sum t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma_{opt}). \quad (20)$$

Proof If σ_{opt} is optimal then

$$C_F = \sum_i t_i \cdot C_F \quad (21)$$

$$\geq \sum_i t_i \cdot D_F(\rho_i, \sigma_{opt}) \quad (22)$$

$$= \sum_i t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma_{opt}), \quad (23)$$

which proves Inequality (20). \square

One can formulate a minimax theorem for divergence, but we will prove a result that is stronger than a minimax theorem in the sense that it gives an upper bound on how close a specific strategy is to the optimal strategy. First we need the following lemma.

Lemma 8 *Let \mathcal{K} be a convex body with a Bregman divergence D_F that is lower semi-continuous. Let \mathcal{L} denote a closed convex subset of \mathcal{K} . For any $\sigma \in \mathcal{K}$ there exists a point $\sigma^* \in \mathcal{L}$ such that*

$$D_F(\rho, \sigma) \geq D_F(\rho, \sigma^*) + D_F(\sigma^*, \sigma) \quad (24)$$

for all $\rho \in \mathcal{L}$. In particular σ^* minimizes $D_F(\rho, \sigma)$ under the constraint that $\rho \in \mathcal{L}$.

Proof Using that \mathcal{L} is closed and lower semicontinuity of D_F we find a point $\sigma^* \in \mathcal{L}$ that minimizes $D_F(\rho, \sigma)$ under the constraint that $\rho \in \mathcal{L}$. Define

$$\rho_t = (1 - t) \cdot \sigma^* + t \cdot \rho. \quad (25)$$

Then according to the Bregman equation

$$\begin{aligned} (1 - t) \cdot D_F(\sigma^*, \sigma) + t \cdot D_F(\rho, \sigma) &= (1 - t) \cdot D_F(\sigma^*, \rho_t) + t \cdot D_F(\rho, \rho_t) + D_F(\rho_t, \sigma) \\ &\geq t \cdot D_F(\rho, \rho_t) + D_F(\sigma^*, \sigma). \end{aligned} \quad (26)$$

After reorganizing the terms and dividing by t we get

$$D_F(\rho, \sigma) \geq D_F(\rho, \rho_t) + D_F(\sigma^*, \sigma). \quad (27)$$

Inequality (24) is obtained by letting t tend to zero and using lower semi-continuity. \square

Theorem 9 *If \mathcal{K} is a convex body with a Bregman divergence D_F that is lower semi-continuous in both variables and such that F is continuously differentiable C^1 . Then*

$$C_F = \sup_{\vec{t}} \sum_i t_i \cdot D_F(\rho_i, \bar{\rho}) \quad (28)$$

where the supremum is taken over all probability vectors \vec{t} supported on \mathcal{K} . Further the following inequality holds

$$\sup_{\rho \in \mathcal{K}} D_F(\rho, \sigma) \geq C_F + D_F(\sigma_{opt}, \sigma) \quad (29)$$

for all σ .

Proof First we prove the theorem for a convex polytope $\mathcal{L} \subseteq \mathcal{K}$. Assume that $\rho_1, \rho_2, \dots, \rho_n$ are the extreme points of \mathcal{L} . Let $\sigma_{opt}(\mathcal{L})$ denote a point that minimizes $\sup_{\rho \in \mathcal{K}} D_F(\rho, \sigma)$. Let J denote the set of indices i for which

$$D_F(\rho_i, \sigma) = \sup_{\rho \in \mathcal{L}} D_F(\rho, \sigma). \quad (30)$$

Let \mathcal{M} denote the convex hull of $\rho_i, i \in J$. Let π denote the projection of σ_{opt} on \mathcal{M} . Then there exists a mixture such that $\sum_{i \in J} t_i \cdot \rho_i = \pi$. Then for any σ

$$\sup_{\rho \in \mathcal{L}} D_F(\rho, \sigma) \geq \sum_{i \in J} t_i \cdot D_F(\rho_i, \sigma) \quad (31)$$

$$= \sum_{i \in J} t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma). \quad (32)$$

Since all divergences $D_F(\rho_i, \sigma)$ where $i \in J$ can be decreased by moving σ from σ_{opt} towards π and the divergences $D_F(\rho_i, \sigma)$ where $i \notin J$ are below $C(\mathcal{L})$ as long as σ is only moved a little towards π we have that $\pi = \sigma_{opt}$ and that (28) holds. Inequality (29) follows from inequality (31) when $\bar{\rho} = \rho_{opt}$.

Let $\mathcal{L}_1 \subseteq \mathcal{L}_2 \subseteq \dots \subseteq \mathcal{K}$ denote an increasing sequence of polytopes such that the union contain the interior of \mathcal{K} . We have

$$C_F(\mathcal{L}_1) \leq C_F(\mathcal{L}_2) \leq \dots \leq C_F(\mathcal{K}) \quad (33)$$

Let $\sigma_{opt,i}$ denote a point that is optimal for \mathcal{L}_i . By compactness of \mathcal{K} we may assume that $\sigma_i \rightarrow \sigma_\infty$ for $i \rightarrow \infty$ for some point $\sigma_\infty \in \mathcal{K}$. Otherwise we just replace the sequence by a subsequence. For any $\rho \in \mathcal{L}_i$ we have

$$D_F(\rho, \sigma_i) \leq \liminf_{i \rightarrow \infty} D_F(\rho, \sigma_i) \quad (34)$$

$$\leq \lim_{i \rightarrow \infty} C_F(\mathcal{L}_i). \quad (35)$$

By lower semi-continuity

$$D_F(\rho, \sigma_\infty) \leq \lim_{i \rightarrow \infty} C_F(\mathcal{L}_i). \quad (36)$$

By taking the supremum over all interior points $\rho \in \mathcal{K}$ we obtain

$$C_F(\mathcal{K}) \leq \sup_{\rho \in \mathcal{K}} D_F(\rho, \sigma_\infty) \quad (37)$$

$$= \lim_{i \rightarrow \infty} C_F(\mathcal{L}_i) \quad (38)$$

$$= \sup_{\vec{t}} \sum_i t_i \cdot D_F(\rho_i, \bar{\rho}), \quad (39)$$

which in combination with (33) proves (28) and also proves that σ_∞ is optimal. We also have

$$\sup_{\rho \in \mathcal{K}} D_F(\rho, \sigma) = \lim_{i \rightarrow \infty} \sup_{\rho \in \mathcal{L}_i} D_F(\rho, \sigma) \quad (40)$$

$$\geq \liminf_{i \rightarrow \infty} (C_F(\mathcal{L}_i) + D_F(\sigma_i, \sigma)) \quad (41)$$

$$\geq C_F + D_F(\sigma_\infty, \sigma), \quad (42)$$

which proves Inequality (29). \square

4 Spectral Sets

Let \mathcal{K} denote a convex body of rank 2. Then $\sigma \in \mathcal{K}$ is said to have *unique spectrality* if all orthogonal decompositions $\sigma = (1-t) \cdot \sigma_0 + t \cdot \sigma_1$ have the same coefficients $\{1-t, t\}$ and the set $\{1-t, t\}$ is called the *spectrum* of σ . If all elements of \mathcal{K} have unique spectrality we say that \mathcal{K} is *spectral*. A convex body \mathcal{K} is said to be *centrally symmetric* with *center* c if for any point $\sigma \in \mathcal{K}$ there exists a *centrally inverted* point $\tilde{\sigma}$ in \mathcal{K} , i.e. a point $\tilde{\sigma} \in \mathcal{K}$ such that $\frac{1}{2}\sigma + \frac{1}{2}\tilde{\sigma} = c$.

Theorem 10 *A spectral set \mathcal{K} of rank 2 is centrally symmetric.*

Proof Let $S : [0, 1] \rightarrow \mathcal{K}$ denote a section. Let $\pi_0 \in \mathcal{K}$ denote an arbitrary extreme point and let π_1 denote a point on the boundary such that $(1-s) \cdot \pi_0 + s \cdot \pi_1 = S(1/2)$ where $0 \leq s \leq 1/2$. Then $S(1/2)$ can be written as a mixture $(1-t) \cdot \sigma_0 + t \cdot \sigma_1$ of points on the boundary such that t is minimal. As in the proof of Theorem 1 we see that σ_0 and σ_1 are orthogonal. Since \mathcal{K} is spectral we have $t = 1/2$. Since $t \leq s \leq 1/2$ we have $s = 1/2$ implying that \mathcal{K} is symmetric around $S(1/2)$. \square

Proposition 11 *Let $S : \mathcal{L} \rightarrow \mathcal{K}$ denote a section with retraction $R : \mathcal{K} \rightarrow \mathcal{L}$. If \mathcal{K} is a spectral set of rank 2 and \mathcal{L} is not a singleton then \mathcal{L} is also a spectral set of rank 2. If c is the center of \mathcal{K} then $R(c)$ is the center of \mathcal{L} and $S(R(c)) = c$, i.e. the section goes through the center of \mathcal{K} .*

Proof Let $\sigma \rightarrow \tilde{\sigma}$ denote reflection in the point $c \in \mathcal{K}$. If $\rho \in \mathcal{L}$ then

$$R(c) = R\left(\frac{1}{2} \cdot S(\rho) + \frac{1}{2} \cdot \widetilde{S(\rho)}\right) \quad (43)$$

$$= \frac{1}{2} \cdot R(S(\rho)) + \frac{1}{2} \cdot R(\widetilde{S(\rho)}) \quad (44)$$

$$= \frac{1}{2} \cdot \rho + \frac{1}{2} \cdot R(\widetilde{S(\rho)}) \quad (45)$$

so that \mathcal{L} is centrally symmetric around $R(c)$. If \mathcal{F} is a proper face of \mathcal{L} then $S(\mathcal{F})$ is a proper face of \mathcal{K} implying that $S(\mathcal{F})$ is a singleton. Therefore $\mathcal{F} = R(S(\mathcal{F}))$ is a singleton implying that \mathcal{L} has rank 2. If $\rho \in \mathcal{L}$ is an extreme point then $\tilde{\rho} = R(\widetilde{S(\rho)})$ is also an extreme point of \mathcal{L} . Now

$$S(R(c)) = S\left(\frac{1}{2} \cdot \rho + \frac{1}{2} \cdot R(\widetilde{S(\rho)})\right) \quad (46)$$

$$= \frac{1}{2} \cdot S(\rho) + \frac{1}{2} \cdot S(R(\widetilde{S(\rho)})). \quad (47)$$

Since $\tilde{\rho} \in \mathcal{L}$ is an extreme point and $\tilde{\rho} = R(\widetilde{S(\rho)})$ we have that $R^{-1}(\tilde{\rho})$ is a proper face of \mathcal{K} and thereby a singleton. Therefore $S(R(\widetilde{S(\rho)})) = \widetilde{S(\rho)}$ and $S(R(c)) = \frac{1}{2} \cdot S(\rho) + \frac{1}{2} \cdot \widetilde{S(\rho)} = c$. \square

Corollary 12 *If σ is an extreme point of a spectral set \mathcal{K} of rank 2 then there exists a unique element in \mathcal{K} that is orthogonal to σ .*

If a centrally symmetric set has a proper face that is not an extreme point then the set is not spectral as illustrated in Fig. 5.

Let $\rho = x \cdot \sigma_0 + y \cdot \sigma_1$ denote an orthogonal decomposition of an element of the vector space generated by a spectral set of rank 2. Then we may define

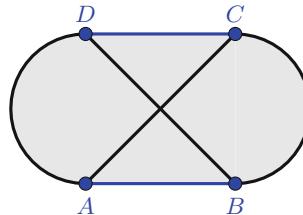


Fig. 5 A centrally symmetric convex body with non-trivial faces \overline{AB} and \overline{CD} . The Caratheodory number is 2, but the rank is 3. The points A, B, C , and D are orthogonal extreme points and any point in the interior of the square $\square ABCD$ has several orthogonal decompositions with different mixing coefficients with weights on A, B, C , and D . Points in the convex body but outside the triangles can be decomposed as a mixture of two orthogonal extreme points on the semi circles

$$f(\rho) = f(x) \cdot \sigma_0 + f(y) \cdot \sigma_1. \quad (48)$$

If $\rho = x \cdot \rho_0 + y \cdot \rho_1$ is another orthogonal decomposition then $x = y$ and

$$f(x) \cdot \sigma_0 + f(y) \cdot \sigma_1 = 2f(x) \cdot \frac{\sigma_0 + \sigma_1}{2} \quad (49)$$

and

$$f(x) \cdot \rho_0 + f(y) \cdot \rho_1 = 2f(x) \cdot \frac{\rho_0 + \rho_1}{2}. \quad (50)$$

Since

$$\frac{\sigma_0 + \sigma_1}{2} = \frac{\rho_0 + \rho_1}{2} = c \quad (51)$$

different orthogonal decompositions will result in the same value of $f(\rho)$. Note in particular that for the constant function $f(x) = 1/2$ we have $f(\rho) = c$. In this sense $c = 1/2$ and from now on we will use $\frac{1}{2}$ in bold face instead of c as notation for the center of a spectral set. If $\frac{1}{2} \cdot \rho + \frac{1}{2} \cdot \sigma = c$ then $\frac{1}{2} \cdot \rho + \frac{1}{2} \cdot \sigma = \frac{1}{2}$ so that $\rho + \sigma = \mathbf{1}$ so that the central inversion of ρ equals $\mathbf{1} - \rho$. We note that if $f(x) \geq 0$ for all x then $f(\rho)$ is element in the positive cone. Therefore $\sum_i \rho_i^2 = 0$ implies that $\rho_i = 0$ for all i , where ρ_i^2 is defined via Eq. (48). Note also that if Φ is an isomorphism then $\Phi(f(\rho)) = f(\Phi(\rho))$.

5 Sufficient Regret Functions

There are a number of equivalent ways of defining sufficiency, and the present definition of sufficiency is based on [18]. We refer to [14] where the notion of sufficiency is discussed in great detail.

Definition 13 Let $(\sigma_\theta)_\theta$ denote a family of points in a convex body \mathcal{K} and let Φ denote an affinity $\mathcal{K} \rightarrow \mathcal{L}$ where \mathcal{K} and \mathcal{L} denote convex bodies. Then Φ is said to be *sufficient* for $(\sigma_\theta)_\theta$ if there exists an affinity $\Psi : \mathcal{L} \rightarrow \mathcal{K}$ such that $\Psi(\Phi(\sigma_\theta)) = \sigma_\theta$, i.e. the states σ_θ are fix-points of $\Psi \circ \Phi$.

The notion of sufficiency as a property of general divergences was introduced in [11]. It was shown in [15] that a Bregman divergence on the simplex of distributions on an alphabet that is not binary determines the divergence up to a multiplicative factor. In [8] this result was extended to C^* -algebras. Here we are interested in the binary case and its generalization that is convex bodies of rank 2.

Definition 14 We say that the regret function D_F on the convex body \mathcal{K} satisfies *sufficiency* if

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma) \quad (52)$$

for any affinity $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ that is sufficient for (ρ, σ) .

Definition 15 A regret function is said to be *separable* if it is defined on a spectral convex body and it is generated by a function of the form

$$F(\sigma) = \text{tr}[f(\sigma)] \quad (53)$$

for some convex function $f : [0, 1] \rightarrow \mathbb{R}$.

Lemma 16 If a strict regret function on a convex body of rank 2 satisfies sufficiency, then the convex body is spectral and the regret function is separable.

Proof For $i = 1, 2$ assume that $S_i : [0, 1] \rightarrow \mathcal{K}$ are sections with retractions $R_i : \mathcal{K} \rightarrow [0, 1]$. Then $S_2 \circ R_1$ is sufficient for the pair $(S_1(t), S_1(1/2))$ with recovery map $S_1 \circ R_2$ implying that

$$D_F(S_1(t), S_1(1/2)) = D_F(S_2(t), S_2(1/2)). \quad (54)$$

Define $f(t) = D_F(S_1(t), S_1(1/2))$. Then $D_F(S_2(t), S_2(1/2)) = f(t)$ for any section S_2 , so this divergence is completely determined by the spectrum $(t, 1-t)$. In particular all orthogonal decompositions have the same spectrum so that the convex body is spectral.

Let \mathcal{K} denote a spectral convex set of rank 2 with center $\frac{1}{2}$. If the Bregman divergence D_F satisfies sufficiency then $D_F(\rho, \sigma) = D_F(\mathbf{1} - \rho, \mathbf{1} - \sigma)$ and

$$D_F(\rho, \sigma) = \frac{D_F(\rho, \sigma) + D_F(\mathbf{1} - \rho, \mathbf{1} - \sigma)}{2} \quad (55)$$

$$= \frac{D_F(\rho, \sigma) + D_{\tilde{F}}(\rho, \sigma)}{2} \quad (56)$$

$$= D_{\frac{F+\tilde{F}}{2}}(\rho, \sigma) \quad (57)$$

where $\tilde{F}(\sigma)$ is defined as $F(\mathbf{1} - \sigma)$. Now $\frac{F+\tilde{F}}{2}$ is convex and invariant under central inversion. Therefore a regret function on a spectral set of rank 2 is generated by a function that is invariant under central inversion.

Let F denote a convex function that is invariant under central inversion and assume that D_F satisfies sufficiency. If σ_0 and σ_1 are orthogonal we may define $f(t) = \frac{1}{2} \cdot F((1-t) \cdot \sigma_0 + t \cdot \sigma_1)$ for $t \in [0, 1]$. Then

$$\text{tr}[f(\sigma)] = \text{tr}[f(1-t) \cdot \sigma_0 + f(t) \cdot \sigma_1] \quad (58)$$

$$= f(1-t) \cdot \mathbf{1} + f(t) \cdot \mathbf{1} \quad (59)$$

$$= 2 \cdot f(t) \quad (60)$$

$$= 2 \cdot \frac{1}{2} \cdot F((1-t) \cdot \sigma_0 + t \cdot \sigma_1) \quad (61)$$

$$= F(\sigma), \quad (62)$$

which proves Eq. (53). \square

Proposition 17 Let \mathcal{K} denote a spectral convex set of rank 2. If $f : [0, 1] \rightarrow \mathbb{R}$ is convex then $F(\sigma) = \text{tr}[f(\sigma)]$ defines a convex function on \mathcal{K} and the regret function D_F satisfies sufficiency.

Proof Let $S_1 : [0, 1] \rightarrow \mathcal{K}$ denote a section and let R_1 denote a corresponding retraction. We will prove that F is decreasing under the idempotent affinity $S_1 \circ R_1$. Assume that $\sigma = S_2(t)$ for some section S_2 . The affinity $R_1 \circ S_2 : [0, 1] \rightarrow [0, 1]$ maps $1/2$ to $1/2$. Therefore $R_1 \circ S_2$ is a contraction around $1/2$ and it has the form $(R_1 \circ S_2)(t) = (1 - r) \cdot 1/2 + r \cdot t$ for some $r \in [0, 1]$. Therefore

$$(S_1 \circ R_1)(\sigma) = (S_1 \circ R_1)(S_2(t)) \quad (63)$$

$$= S_1((R_1 \circ S_2)(t)) \quad (64)$$

$$= S_1((1 - r) \cdot 1/2 + r \cdot t). \quad (65)$$

Then

$$F(\sigma) = f(1 - t) + f(t) \quad (66)$$

$$\leq f((1 - r) \cdot 1/2 + r \cdot (1 - t)) + f((1 - r) \cdot 1/2 + r \cdot t) \quad (67)$$

$$= F((S_1 \circ R_1)(\sigma)) \quad (68)$$

by convexity of $t \rightarrow f(1 - t) + f(t)$.

Let σ_0 and σ_1 denote points in \mathcal{K} . Assume that $S : [0, 1] \rightarrow \mathcal{K}$ is a section such that $S(t) = (1 - s) \cdot \sigma_0 + s \cdot \sigma_1$ and let $R : \mathcal{K} \rightarrow [0, 1]$ denote a retraction corresponding to S . Then

$$F((1 - s) \cdot \sigma_0 + s \cdot \sigma_1) = F(S(R((1 - s) \cdot \sigma_0 + s \cdot \sigma_1))) \quad (69)$$

$$= F((1 - s) \cdot S(R(\sigma_0)) + s \cdot S(R(\sigma_1))) \quad (70)$$

$$\leq (1 - s) \cdot F(S(R(\sigma_0))) + s \cdot F(S(R(\sigma_1))) \quad (71)$$

$$\leq (1 - s) \cdot F(\sigma_0) + s \cdot F(\sigma_1). \quad (72)$$

Now we will prove that D_F satisfies sufficiency. Let $\rho, \sigma \in \mathcal{K}$ denote two points and let $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ denote an affinity that is sufficient for ρ, σ with recovery map Ψ . Then $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are retractions and the fixpoint set of $\Psi \circ \Phi$ and $\Phi \circ \Psi$ are isomorphic convex bodies. According to Proposition 11 the center of \mathcal{K} a fixpoint under retractions and we see that a decomposition into orthogonal extreme point in a fixpoint set is also an orthogonal decomposition in \mathcal{K} . Therefore $\text{tr}[f(\sigma)]$ has the same value when the calculation is done within the fixpoint set of $\Psi \circ \Phi$, which proves the proposition. \square

Theorem 18 Let \mathcal{K} denote a convex body of rank 2 with a sufficient Bregman divergence D_F that is strict. Then the center of \mathcal{K} is the unique point that achieves the minimax regret.

Proof Let $S : [0, 1] \rightarrow \mathcal{K}$ denote a section. Then $S(1/2) = \frac{1}{2}$ and

$$C_F \geq \frac{1}{2} \cdot D_F(S(0), S(1/2)) + \frac{1}{2} \cdot D_F(S(1), S(1/2)) + D_F(S(1/2), \sigma_{opt}) \quad (73)$$

$$= D_F\left(S(1), \frac{1}{2}\right) + D_F\left(\frac{1}{2}, \sigma_{opt}\right). \quad (74)$$

Further we have

$$\sup_{\rho \in \mathcal{K}} D_F\left(\rho, \frac{1}{2}\right) \geq C_F + D_F\left(\sigma_{opt}, \frac{1}{2}\right). \quad (75)$$

Now $\rho = S_\rho(t)$ for some section S_ρ and some $t \in [0, 1]$. Therefore

$$D_F\left(\rho, \frac{1}{2}\right) = D_F(S_\rho(t), S(1/2)) \quad (76)$$

$$= D_F(S_\rho(t), S_\rho(1/2)) \quad (77)$$

$$= D_F(S(t), S(1/2)) \quad (78)$$

$$\leq D_F\left(S(1), \frac{1}{2}\right). \quad (79)$$

Therefore $C_F = D_F(S(1), \frac{1}{2})$ and $D_F(\sigma_{opt}, \frac{1}{2}) = 0$ implying $\sigma_{opt} = \frac{1}{2}$. \square

If the Bregman divergence is based on Shannon entropy then the minimax regret is called the capacity and the result is that a convex body of rank 2 has a capacity of 1 bit.

6 Spin Factors

We say that a convex body is a *Hilbert ball* if the convex body can be embedded as a unit ball in a d dimensional real Hilbert space \mathcal{H} with some inner product that will be denoted $\langle \cdot | \cdot \rangle$.

The direct sum $\mathcal{H} \oplus \mathbb{R}$ can be equipped a product \bullet by

$$(\vec{v}, s) \bullet (\vec{w}, t) = (t \cdot \vec{v} + s \cdot \vec{w}, \langle \vec{v} | \vec{w} \rangle + s \cdot t). \quad (80)$$

This product is distributive and $(\vec{v}, 1) \bullet (-\vec{v}, 1) = 0$. Therefore x^2 defined via (48) will be equal to $x \bullet x$ and $(\mathcal{H} \oplus \mathbb{R}, \bullet)$ becomes a *formally real Jordan algebra* of the type that is called a *spin factor* and is denoted $JSpin_d$. The unit of a spin factor is $(\vec{0}, 1)$ and will be denoted $\mathbf{1}$. See [16] for general results on Jordan algebras. The positive elements are the elements (\vec{v}, s) where $\|\vec{v}\|_2 \leq s$. The trace of the spin factor is $\text{tr}[(\vec{v}, s)] = 2s$.

Let $\mathcal{M}_n(\mathbb{F})$ denote $n \times n$ matrices over \mathbb{F} where \mathbb{F} may denote the real numbers \mathbb{R} or the complex numbers \mathbb{C} or the quaternions \mathbb{H} or the octonions \mathbb{O} . Let $(\mathcal{M}_n(\mathbb{F}))_h$ denote the set of self-adjoint matrices of $\mathcal{M}_n(\mathbb{F})$. Then $(\mathcal{M}_n(\mathbb{F}))_h$ is a formally real Jordan algebra with a Jordan product \bullet is given by

$$x \bullet y = \frac{1}{2} (xy + yx) \quad (81)$$

except for $\mathbb{F} = \mathbb{O}$ where one only get a Jordan algebra when $n \leq 3$. The self-adjoint 2×2 matrices with real, complex, quaternionic or octonionic entries can be identified with spin factors with dimension $d = 2, d = 3, d = 5$, or $d = 9$. The most important examples of spin factors are the bit $JSpin_1$ and the qubit $JSpin_3$.

We introduce the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (82)$$

and observe that $\sigma_1 \bullet \sigma_3 = \mathbf{0}$. Let v_1, v_2, \dots, v_d denote a basis of the Hilbert space \mathcal{H} . Let the function $S : JSpin_d \rightarrow (\mathcal{M}_2(\mathbb{R}))^{\otimes(d-1)}$ be defined by

$$S(\mathbf{1}) = \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1}, \quad (83)$$

$$S(v_1) = \sigma_1 \otimes \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1}, \quad (84)$$

$$S(v_2) = \sigma_3 \otimes \sigma_1 \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1}, \quad (85)$$

$$S(v_3) = \sigma_3 \otimes \sigma_3 \otimes \sigma_1 \otimes \cdots \otimes \mathbf{1}, \quad (86)$$

$$\vdots \quad (87)$$

$$S(v_{d-1}) = \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \cdots \otimes \sigma_1, \quad (88)$$

$$S(v_d) = \sigma_3 \otimes \sigma_3 \otimes \sigma_3 \otimes \cdots \otimes \sigma_3. \quad (89)$$

Then S can be linearly extended and one easily checks that

$$S(x \bullet y) = S(x) \bullet S(y). \quad (90)$$

Now $S(JSpin_d)$ is a linear subspace of the real Hilbert space $(\mathcal{M}_2(\mathbb{R}))^{\otimes(d-1)}$ so there exists a projection of $(\mathcal{M}_2(\mathbb{R}))^{\otimes(d-1)}$ onto $S(JSpin_d)$ and this projection maps symmetric matrices in $(\mathcal{M}_2(\mathbb{R}))^{\otimes(d-1)}$ into symmetric matrices. Therefore S is a section with a retraction generated by the projection. In this way $JSpin_d$ is a section of a Jordan algebra of symmetric matrices with real entries. The Jordan algebra $\mathcal{M}_n(\mathbb{R})_h$ is obviously a section of $\mathcal{M}_n(\mathbb{C})_h$ so $JSpin_d$ is a section of $\mathcal{M}_n(\mathbb{C})_h$. Note that the projection of $\mathcal{M}_n(\mathbb{C})_h$ on a spin factor is not necessarily completely positive.

Since the standard formalism of quantum theory represents states as density matrices in $\mathcal{M}_n(\mathbb{C})_h$ we see that spin factors appear as sections of state spaces of the usual formalism of quantum theory. Therefore the points in the Hilbert ball are called *states*

and the Hilbert ball is called the *state space* of the spin factor. The extreme points in the state space are called *pure states*.

The positive cone of a spin factor is self-dual in the sense that any positive functional $\phi : JSpin_d \rightarrow \mathbb{R}$ is given by $\phi(x) = \text{tr}[x \bullet y]$ for some uniquely determined positive element y . We recall the definition of the polar set of a convex body $\mathcal{K} \subseteq \mathbb{R}^d$

$$\mathcal{K}^\circ = \{y \in \mathbb{R}^d \mid \langle x, y \rangle \leq 1 \text{ for all } x \in \mathcal{K}\}. \quad (91)$$

Proposition 19 *Assume that the cone generated by a spectral convex body \mathcal{K} of rank 2 is self-dual. Then it can be represented as a spin factor.*

Proof If ϕ is a test on \mathcal{K} then $2 \cdot \phi - 1$ maps \mathcal{K} into $[0, 1]$, which is an element in the polar set of \mathcal{K} embedded in a Hilbert space with the center as the origin. Since the cone is assumed to be self-dual the set \mathcal{K} is self-polar and Hilbert balls are the only self-polar sets. The result follows because a Hilbert ball can be represented as the state space of a spin factor. \square

A convex body \mathcal{K} of rank 2 is said to have *symmetric transission probabilities* if for any extreme points σ_1 and σ_2 there exists retractions $R_1 : \mathcal{K} \rightarrow [-1, 1]$ and $R_2 : \mathcal{K} \rightarrow [-1, 1]$ such that $R_i(\sigma_i) = 1$, and $R_1(\sigma_2) = R_2(\sigma_1)$.

Theorem 20 *A spectral convex body \mathcal{K} of rank 2 with symmetric transmission probabilities can be represented by a spin factor.*

Proof For almost all extreme points σ of \mathcal{K} a retraction $R : \mathcal{K} \rightarrow [-1, 1]$ with $R(\sigma) = 1$ is uniquely determined. Let σ_1 and σ_2 be two extreme points that are not antipodal and with unique retractions R_1 and R_2 . Let \mathcal{L} denote the intersection of \mathcal{K} with the affine span of σ_1, σ_2 and the center. Embed \mathcal{L} in a 2-dimensional coordinate system with the center of \mathcal{L} as origin of the coordinate system. Let σ denote an extreme point with a unique retraction R . Then $R_1(\sigma) \cdot \sigma - \sigma_1$ is parallel with $R_2(\sigma) \cdot \sigma - \sigma_2$ because

$$R(R_i(\sigma) \cdot \sigma - \sigma_i) = R_i(\sigma) \cdot R(\sigma) - R(\sigma_i) \quad (92)$$

$$= R_i(\sigma) \cdot 1 - R_i(\sigma) \quad (93)$$

$$= 0. \quad (94)$$

Therefore the determinant of $R_1(\sigma) \cdot \sigma - \sigma_1$ and $R_2(\sigma) \cdot \sigma - \sigma_2$ is zero, but the determinant can be calculated as

$$\begin{aligned} & \det(R_1(\sigma) \cdot \sigma - \sigma_1, R_2(\sigma) \cdot \sigma - \sigma_2) \\ &= 0 - \det(R_1(\sigma) \cdot \sigma, \sigma_2) - \det(\sigma_1, R_2(\sigma) \cdot \sigma) + \det(\sigma_1, \sigma_2) \\ &= \det(\sigma, R_2(\sigma) \cdot \sigma_1 - R_1(\sigma) \cdot \sigma_2) - \det(\sigma_2, \sigma_1). \end{aligned} \quad (95)$$

This means that σ satisfies the following equation

$$\det(\sigma, R_2(\sigma) \cdot \sigma_1 - R_1(\sigma) \cdot \sigma_2) = \det(\sigma_2, \sigma_1). \quad (96)$$

This is a quadratic equation in the coordinates of σ , which implies that σ lies on a conic section. Since \mathcal{L} is bounded this conic section must be a circle or an ellipsoid. Almost all extreme points of \mathcal{L} have unique retractions. Therefore almost all extreme points lie on a circle or an ellipsoid which by convexity implies that all extreme points of \mathcal{L} lie on an ellipsoid or a circle. Since this holds for almost all pairs σ_1 and σ_2 the convex body \mathcal{K} must be an ellipsoid, which can be mapped into a ball. \square

Definition 21 Let $\mathcal{A} \subseteq JSpin_d$ denote a subalgebra of a spin factor. Then $\mathbb{E} : JSpin_d \rightarrow A$ is called a conditional expectation if $\mathbb{E}(1) = 1$ and $\mathbb{E}(a \bullet x) = a \bullet \mathbb{E}(x)$ for any $a \in \mathcal{A}$.

Theorem 22 Let \mathcal{K} denote the state space of a spin factor and assume that $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ is an idempotent that preserves the center. Then Φ is a conditional expectation of the spin factor into a sub-algebra of the spin factor.

Proof Assume that the spin factor is based on the Hilbert space $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ and that the idempotent Φ is the identity on \mathcal{H}_1 and maps \mathcal{H}_2 into the origin. Let $\mathbf{v}, \mathbf{w}_1 \in \mathcal{H}_1$ and $\mathbf{w}_2 \in \mathcal{H}_2$ and $s, t \in R$. Then

$$\Phi((\mathbf{v}, s) \bullet (\mathbf{w}_1 + \mathbf{w}_2, t)) = \Phi((\mathbf{v}, s) \bullet (\mathbf{w}_1, t)) + \Phi((\mathbf{v}, s) \bullet (\mathbf{w}_2, 0)) \quad (97)$$

$$= (\mathbf{v}, s) \bullet (\mathbf{w}_1, t) + \Phi(s \cdot \mathbf{w}_2, \langle \mathbf{v}, \mathbf{w}_2 \rangle) \quad (98)$$

$$= (\mathbf{v}, s) \bullet \Phi(\mathbf{w}_1 + \mathbf{w}_2, t), \quad (99)$$

which proves the theorem. \square

7 Monotonicity Under Dilations

Next we introduce the notion of monotonicity. In thermodynamics monotonicity is associated with decrease of free energy in a closed system and in information theory it is associated with the data processing inequality.

Definition 23 Let D_F denote a regret function on the convex body \mathcal{K} . Then D_F is said to be *monotone* if

$$D_F(\Phi(\rho), \Phi(\sigma)) \leq D_F(\rho, \sigma) \quad (100)$$

for any affinity $\Phi : \mathcal{K} \rightarrow \mathcal{K}$.

A simple example of a monotone regret function is squared Euclidean distance in a Hilbert ball, but later we shall see that there are many other examples. All monotone regret functions are Bregman divergences [8, Prop. 6] that satisfy sufficiency [8, Prop. 8]. We shall demonstrate that a convex body of rank 2 with a monotone Bregman divergence can be represented by a spin factor.

We will need to express the Bregman divergence as an integral involving a different type of divergence. Define

$$D^F(\sigma_0, \sigma_1) = \frac{d^2}{dt^2} F(\sigma_t)|_{t=0} \quad (101)$$

where $\sigma_t = (1-t) \cdot \sigma_0 + t \cdot \sigma_1$. If F is C^2 and $\mathbf{H}(y)$ is the Hesse matrix of F calculated in the point y then a Riemannian metric is defined by

$$D^F(\sigma_0, \sigma_1) = \langle \sigma_0 - \sigma_1 | \mathbf{H}(y) | \sigma_0 - \sigma_1 \rangle. \quad (102)$$

We will need the following lemma.

Lemma 24 *Let F denote a convex function defined on a convex body \mathcal{K} . Then for almost all $y \in \mathcal{K}$ we have*

$$\lim_{\rho \rightarrow \sigma} \frac{D_F(\rho, \sigma) - \frac{1}{2} D^F(\rho, \sigma)}{\|x - y\|^2} = 0. \quad (103)$$

Proof According to our definitions

$$\begin{aligned} & D_F(\rho, \sigma) - \frac{1}{2} D^F(\rho, \sigma) \\ &= F(\rho) - \left(F(\sigma) + \langle \nabla F(\sigma) | \rho - \sigma \rangle + \frac{1}{2} \langle \rho - \sigma | \mathbf{H}(\sigma) | \rho - \sigma \rangle \right) \end{aligned} \quad (104)$$

and we see that Lemma 24 states that a convex function is twice differentiable almost everywhere, which is exactly Alexandrov's theorem [1]. \square

It is easy to verify that

$$\frac{d}{ds} D_F(\sigma_0, \sigma_s) = \frac{D^F(\sigma_0, \sigma_s)}{s} \quad (105)$$

Proposition 25 *Let $F : [0, 1] \rightarrow \mathbb{R}$ denote a twice differentiable convex function. If $\sigma_s = (1-s) \cdot \sigma_0 + s \cdot \sigma_1$. Then*

$$D_F(\sigma_0, \sigma_1) = \int_0^1 \frac{D^F(\sigma_0, \sigma_s)}{s} ds, \quad (106)$$

where $D^F(\sigma, \sigma_t)$ is given by one of the Eqs. (101), (102), or (105).

A similar result appear in [12, Eq. 2.118]. In the context of complex matrices the result was proved in [19].

Since

$$F(\sigma_s) = F(\sigma_1) + \langle \nabla F(\sigma_1) \mid \sigma_s - \sigma_1 \rangle + D_F(\sigma_s, \sigma_1) \quad (107)$$

we also have

$$\frac{d}{ds} D_F(\sigma_s, \sigma_1)|_{s=1} = 0 \quad (108)$$

$$\frac{d^2}{ds^2} D_F(\sigma_s, \sigma_1)|_{s=1} = D^F(\sigma_0, \sigma_1). \quad (109)$$

Lemma 26 *If F is twice differentiable then D_F is a monotone Bregman divergence if and only if D^F is monotone.*

Proof Assume that D_F is monotone and that Φ is some affinity and that $\sigma_s = (1-s) \cdot \sigma_0 + s \cdot \sigma_1$. Then

$$D_F(\Phi(\sigma_s), \Phi(\sigma_1)) \leq D_F(\sigma_s, \sigma_1). \quad (110)$$

We also have

$$D_F(\sigma_1, \sigma_1) = 0, \quad (111)$$

$$\frac{d}{ds} D_F(\sigma_s, \sigma_1)|_{s=1} = 0, \quad (112)$$

$$D_F(\Phi(\sigma_1), \Phi(\sigma_1)) = 0, \quad (113)$$

$$\frac{d}{ds} D_F(\Phi(\sigma_s), \Phi(\sigma_1))|_{s=1} = 0. \quad (114)$$

Therefore we must have

$$D^F(\Phi(\sigma_0), \Phi(\sigma_1)) = \frac{d^2}{ds^2} D_F(\Phi(\sigma_s), \Phi(\sigma_1))|_{s=1} \quad (115)$$

$$\leq \frac{d^2}{ds^2} D_F(\sigma_s, \sigma_1)|_{s=1} \quad (116)$$

$$\leq D^F(\sigma_0, \sigma_1). \quad (117)$$

If D^F is monotone then Proposition 25 implies that D_F is monotone. \square

Theorem 27 *Let \mathcal{K} denote a convex body with a sufficient regret function D_F that is monotone under dilations. Then F is C^2 . In particular D_F is a Bregman divergence.*

Proof Since D_F is monotone under dilation we have that D^F is monotone under dilations whenever D^F is defined. Let y be a point where F is differentiable and let $0 < r < 1$ and z be a point such that F is differentiable in $(1-r) \cdot z + r \cdot y$. Then

$$D^F((1-r) \cdot z + r \cdot x, (1-r) \cdot z + r \cdot y) \leq D^F(x, y) \quad (118)$$

$$\langle r \cdot x - r \cdot y | \mathbf{H}((1-r) \cdot z + r \cdot y) | r \cdot x - r \cdot y \rangle \leq \langle x - y | \mathbf{H}(y) | x - y \rangle \quad (119)$$

$$r^2 \cdot \langle x - y | \mathbf{H}((1-r) \cdot z + r \cdot y) | x - y \rangle \leq \langle x - y | \mathbf{H}(y) | x - y \rangle \quad (120)$$

$$r^2 \cdot \mathbf{H}((1-r) \cdot z + r \cdot y) \leq \mathbf{H}(y). \quad (121)$$

Let $\mathcal{L}_1 \subseteq K$ denote a ball around y with radius R_1 and let \mathcal{L}_2 denote a ball around y with radius $R_2 < R_1$. Then for any $w \in \mathcal{L}_2$ there exists a $z \in \mathcal{L}_1$ such that $w = (1-r) \cdot z + r \cdot y$ where $r \geq 1 - \frac{R_2}{R_1}$ implying that

$$\left(1 - \frac{R_2}{R_1}\right)^2 \cdot \mathbf{H}(w) \leq \mathbf{H}(y). \quad (122)$$

There also exists a $\tilde{z} \in \mathcal{L}_1$ such that $y = (1-\tilde{r}) \cdot z + \tilde{r} \cdot w$ where $\tilde{r} \geq 1 - \frac{R_2}{R_1 + R_2}$ implying that

$$\left(1 - \frac{R_2}{R_1 + R_2}\right)^2 \cdot \mathbf{H}(y) \leq \mathbf{H}(w). \quad (123)$$

We see that if R_2 is small Then $y \rightarrow \mathbf{H}(y)$ is uniformly continuous on any compact subset of the interior of \mathcal{K} restricted to points where F is twice differentiable. Therefore \mathbf{H} has a unique continuous extension to \mathcal{K} and we can use the extension of \mathbf{H} to get an extension of D^F . The last thing we need to prove is that the unique extended function \mathbf{H} actually gives the Hesse matrix in any interior point in \mathcal{K} . Let $x, y \in \mathcal{K}$. Introduce $x_r = (1-r)z + r \cdot x$ and $y_r = (1-r)z + r \cdot y$. Then

$$\begin{aligned} D_F(x, y) - \frac{1}{2}D^F(x, y) &\geq D_F((1-r)z + r \cdot x, (1-r)z + r \cdot y) - \frac{1}{2}D^F(x, y) \\ &= D_F(x_r, y_r) - \frac{1}{2}D^F(x_r, y_r) + \frac{1}{2}D^F(x_r, y_r) - \frac{1}{2}D^F(x, y) - \frac{1}{2}D^F(x, y) \\ &\geq D_F(x_r, y_r) - \frac{1}{2}D^F(x_r, y_r) + \frac{1}{2}\langle x - y | r^2 \cdot \mathbf{H}(y_r) - \mathbf{H}(y) | x - y \rangle \\ &\geq D_F(x_r, y_r) - \frac{1}{2}D^F(x_r, y_r) - \frac{1}{2}\|r^2 \cdot \mathbf{H}((1-r)z + r \cdot y) - \mathbf{H}(y)\| \|x - y\|^2 \end{aligned} \quad (124)$$

Therefore

$$\begin{aligned} \frac{D_F(x, y) - \frac{1}{2}D^F(x, y)}{\|x - y\|^2} &\geq r^2 \frac{D_F(x_r, y_r) - \frac{1}{2}D^F(x_r, y_r)}{\|rx - ry\|^2} \\ &\quad - \frac{1}{2}\|r^2 \cdot \mathbf{H}((1-r)z + r \cdot y) - \mathbf{H}(y)\| \end{aligned} \quad (125)$$

and

$$\liminf_{x \rightarrow y} \frac{D_F(x, y) - \frac{1}{2} D^F(x, y)}{\|x - y\|^2} \geq -\frac{1}{2} \|r^2 \cdot \mathbf{H}((1-r)z + r \cdot y) - \mathbf{H}(y)\|. \quad (126)$$

Since this holds for all positive $r < 1$ we have

$$\liminf_{x \rightarrow y} \frac{D_F(x, y) - \frac{1}{2} D^F(x, y)}{\|x - y\|^2} \geq 0. \quad (127)$$

One can prove that \limsup is less than 0 in the same way. \square

Theorem 28 Assume that $f : [0, 1] \rightarrow \mathbb{R}$ is a convex symmetric function and that the function F is defined as $F(\sigma) = \text{tr}[f(\sigma)]$. If the Bregman divergence D_F is monotone under dilations then $y \rightarrow y^2 \cdot f''(y)$ is an increasing function.

Proof Assume that D_F is monotone under dilations. Let $S : [0, 1] \rightarrow \mathcal{K}$ denote a section. Then a dilation around $S(0)$ commutes with the retraction corresponding to the section S . Therefore D_F restricted to $S([0, 1])$ is monotone, so we may without loss of generality assume that the convex body is the interval $[0, 1]$.

Then F is C^2 and D^F is monotone.

$$D^F(r \cdot x, r \cdot y) = F''(r \cdot y) \cdot (r \cdot x - r \cdot y)^2 \quad (128)$$

$$= F''(r \cdot y) \cdot (r \cdot y)^2 \cdot \left(\frac{x}{y} - 1\right)^2. \quad (129)$$

Therefore $y^2 \cdot F''(y)$ and $y^2 \cdot f''(y)$ are increasing. \square

Theorem 29 Let \mathcal{K} denote a convex body of rank 2 with a sufficient and strict regret function D_F that is monotone under dilations. Then \mathcal{K} can be represented by a spin factor.

Proof First we note that \mathcal{K} is a spectral set with a center that we will denote c . We will embed \mathcal{K} in a vector space with c as the origin. If σ and ρ are points on the boundary and $\lambda \in [0, 1/2]$ then

$$D_F((1-\lambda)\sigma + \lambda \cdot c, c) = D_F((1-\lambda)\rho + \lambda \cdot c, c). \quad (130)$$

Therefore

$$D^F(\sigma, c) = k \quad (131)$$

for some constant k . Equation (131) can be written in terms of the Hesse matrix as

$$\langle \sigma - c | \mathbf{H}(c) | \sigma - c \rangle = k, \quad (132)$$

and this is the equation for an ellipsoid. The result follows because any ellipsoid is isomorphic to a ball. \square

One easily check that if $f : C^2([0, 1])$ then $F(\sigma) = \text{tr}[f(\sigma)]$ defines a C^2 -function on any spin factor.

Theorem 30 Assume that $f : C^3([0, 1])$ is a convex symmetric function and that the function F is defined as $F(\sigma) = \text{tr}[f(\sigma)]$ on a spin factor. If $y \rightarrow y^2 f(y)$ is an increasing function then the Bregman divergence D_F is monotone under dilations.

Proof Assume that $y \rightarrow y^2 f(y)$ is an increasing function. It is sufficient to prove that $D^F(x, y)$ is decreasing under dilations. Let $x \rightarrow (1 - r)z + rx$ denote a dilation around z by a factor of $r \in [0, 1]$. Then

$$D^F((1 - r)z + rx, (1 - r)z + ry) = r^2 \langle x - y | \mathbf{H}((1 - r)z + ry) | x - y \rangle. \quad (133)$$

so it is sufficient to prove that $r \rightarrow r^2 \mathbf{H}((1 - r)z + ry)$ is an increasing matrix function. Since f is C^3 we may differentiate with respect to r and we have to prove the inequality

$$2r\mathbf{H}((1 - r)z + ry) + r^2 \frac{d}{dr} \mathbf{H}((1 - r)z + ry) \geq 0. \quad (134)$$

Without loss of generality we may assume $r = 1$ so that we have to prove that

$$2\mathbf{H}(y) + \frac{d}{dr} \mathbf{H}((1 - r)z + ry)|_{r=1} \geq 0. \quad (135)$$

If $y = (y_1, y_2, \dots, y_d)$ and $\mathbf{H} = (\mathbf{H}_{i,j})$ then

$$\frac{d}{dr} \mathbf{H}((1 - r)z + ry)|_{r=1} = \left(\frac{d}{dr} \mathbf{H}_{i,j}((1 - r)z + ry) \right)_{|r=1} \quad (136)$$

$$= \langle \nabla \mathbf{H}_{i,j}((1 - r)z + ry) | y - z \rangle_{|r=1} \quad (137)$$

$$= \langle \nabla \mathbf{H}_{i,j}(y) | y - z \rangle. \quad (138)$$

Since inequality (135) is invariant under rotations that leave the center and y invariant the same must be the case for the inequality

$$2(\mathbf{H}_{i,j}) + \langle \nabla \mathbf{H}_{i,j}(y) | y - z \rangle \geq 0, \quad (139)$$

but this inequality is linear in z so we may take the mean under all rotated versions of this inequality. If \bar{z} denotes the mean of rotated versions of z we have to prove that

$$2(\mathbf{H}_{i,j}) + \langle \nabla \mathbf{H}_{i,j}(y) | y - \bar{z} \rangle \geq 0. \quad (140)$$

Since \bar{z} is collinear with the y and the center we have reduced the problem to dilations of a one-dimensional spin factor which is covered in Theorem 31. \square

For the Tsallis entropy of order α we have $F(x) = \frac{x^\alpha + (1-x)^\alpha - 1}{\alpha-1}$ so that $F''(x) = \alpha(x^{\alpha-2} + (1-x)^{\alpha-2})$ and

$$x^2 F''(x) = x^2 \alpha (x^{\alpha-2} + (1-x)^{\alpha-2}) \quad (141)$$

$$= \alpha (x^\alpha + x^2 (1-x)^{\alpha-2}). \quad (142)$$

The derivative is

$$\alpha (\alpha x^{\alpha-1} + 2x(1-x)^{\alpha-2} - x^2(\alpha-2)(1-x)^{\alpha-3}) \quad (143)$$

$$= \alpha (\alpha x^{\alpha-1} + (2x(1-x) - x^2(\alpha-2))(1-x)^{\alpha-3}) \quad (144)$$

$$= \alpha (\alpha x^{\alpha-1} + x(2-\alpha x)(1-x)^{\alpha-3}) \quad (145)$$

$$= \alpha x^{\alpha-1} \left(\alpha + \left(\frac{2}{x} - \alpha \right) \left(\frac{1}{x} - 1 \right)^{\alpha-3} \right). \quad (146)$$

Set $z = \frac{1}{x} - 1$ so that $x = \frac{1}{z+1}$ which gives

$$\alpha + \left(\frac{2}{x} - \alpha \right) \left(\frac{1}{x} - 1 \right)^{\alpha-3} = \alpha + (2z + 2 - \alpha) z^{\alpha-3}. \quad (147)$$

For $\alpha \leq 2$ the derivative is always positive. For $\alpha < 3$ and z tending to zero the derivative tends to $-\infty$ if $2 - \alpha$ is negative so we do not have monotonicity for $2 < \alpha < 3$.

For $\alpha \geq 3$ we calculate the derivative in order to determine the minimum.

$$2z^{\alpha-3} + (2z + 2 - \alpha)(\alpha - 3)z^{\alpha-4} = 0, \quad (148)$$

which has the solution $z = \frac{\alpha-3}{2}$. Plugging this solution the expression in Eq.(147) gives the value

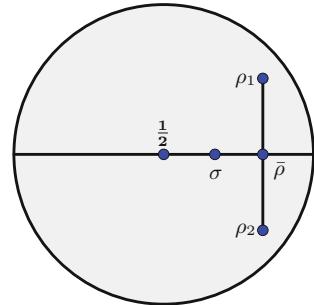
$$\alpha + \left(2 \cdot \frac{\alpha-3}{2} + 2 - \alpha \right) \left(\frac{\alpha-3}{2} \right)^{\alpha-3} = \alpha - \left(\frac{\alpha-3}{2} \right)^{\alpha-3}. \quad (149)$$

Numerical calculations show that this function is positive for values of α between 3 and 6.43779.

8 Monotonicity of Bregman Divergences on Spin Factors

We recall that a binary system can be represented as the spin factor $J Spin_1$ or as the interval $[0,1]$.

Fig. 6 Illustration of Δ and the relative position of the states mentioned in the proof of Lemma 32



Theorem 31 Let $F : [0, 1] \rightarrow \mathbb{R}$ denote a convex and symmetric function. Then D_F is monotone if and only if $F \in C^2([0, 1])$ and $y \rightarrow y^2 \cdot F''(y)$ is increasing.

Proof The convex body $[0, 1]$ has the identity and a reflection as the only isomorphisms. Any affinity can be decomposed into an isomorphism and two dilations where each dilation is a dilation around one of the extreme points $\{0, 1\}$. Therefore D_F is monotone if and only if it is monotone under dilations. \square

Next we will study monotonicity of Bregman divergences in spin factors $J Spin_d$ for $d \geq 2$.

Lemma 32 Let D_F denote a separable Bregman divergence on $J Spin_d$ where $d \geq 2$. If the restriction to $J Spin_2$ is monotone, then D_F is monotone on $J Spin_d$.

Proof Assume that D_F satisfies sufficiency and that the restriction of D_F to $J Spin_2$ is monotone. Let $\rho_1, \sigma \in J Spin_d$ and let $\Phi : J Spin_d \rightarrow J Spin_d$ denote a positive trace preserving affinity. Let Δ denote the disc spanned of ρ_1, σ and $\frac{1}{2}$. Then $\Phi(\rho_1), \Phi(\sigma)$ and $\Phi(\frac{1}{2})$ spans a disc $\tilde{\Delta}$ in $J Spin_d$. The restriction of Φ to Δ can be written as $\Phi|_{\Delta} = \Phi_2 \circ \Phi_1$ where Φ_1 is an affinity $\Delta \rightarrow \Delta$ and Φ_2 is an isomorphism $\Delta \rightarrow \tilde{\Delta}$. Essentially Φ_2 maps a great circle into a small circle where the great circle is the boundary of Δ and the small circle is the boundary of $\tilde{\Delta}$. According to our assumptions Φ_1 is monotone so it is sufficient to prove that Φ_2 is monotone (Fig. 6).

Let ρ_2 denote a state such that

$$D_F(\rho_2, \sigma) = D_F(\rho_1, \sigma) \quad (150)$$

$$D_F\left(\rho_2, \frac{1}{2}\right) = D_F\left(\rho_1, \frac{1}{2}\right). \quad (151)$$

Then

$$D_F(\rho_1, \sigma) = \frac{1}{2} \cdot D_F(\rho_1, \sigma) + \frac{1}{2} \cdot D_F(\rho_2, \sigma) \quad (152)$$

$$= \frac{1}{2} \cdot D_F(\rho_1, \bar{\rho}) + \frac{1}{2} \cdot D_F(\rho_2, \bar{\rho}) + D_F(\bar{\rho}, \sigma) \quad (153)$$

where $\bar{\rho} = \frac{1}{2} \cdot \rho_1 + \frac{1}{2} \cdot \rho_2$. Now $\bar{\rho}$, σ and $\frac{1}{2}$ are co-linear and so are $\Phi_2(\bar{\rho})$, $\Phi_2(\sigma)$, and $\Phi_2\left(\frac{1}{2}\right)$ so the restriction of Φ to the span of $\bar{\rho}$, σ and $\frac{1}{2}$ is an interval and the span of $\Phi_2(\bar{\rho})$, $\Phi_2(\sigma)$, $\Phi_2\left(\frac{1}{2}\right)$, and $\frac{1}{2}$ is a disc so by assumption the restriction is monotone implying that

$$D_F(\Phi_2(\bar{\rho}), \Phi_2(\sigma)) \leq D_F(\bar{\rho}, \sigma). \quad (154)$$

Let $\bar{\pi} \in \Delta$ denote a state that is colinear with $\bar{\rho}$ and $\frac{1}{2}$ and such that $D_F(\bar{\pi}, \frac{1}{2}) = D_F(\Phi_2(\bar{\rho}), \frac{1}{2})$. Then there exists an affinity $\Psi : \Delta \rightarrow \Delta$ such that $\Psi(\bar{\rho}) = \bar{\pi}$ and for $i = 1, 2$

$$D_F(\Phi(\rho_i), \Phi(\bar{\rho})) = D_F(\Psi(\rho_i), \Psi(\bar{\rho})). \quad (155)$$

Since Ψ is monotone

$$D_F(\Phi(\rho_1), \Phi(\bar{\rho})) = D_F(\Phi(\rho_2), \Phi(\bar{\rho})) \leq D_F(\rho_1, \bar{\rho}). \quad (156)$$

Therefore

$$\begin{aligned} D_F(\Phi(\rho_1), \Phi(\sigma)) &= \frac{1}{2} \cdot D_F(\Phi(\rho_1), \Phi(\sigma)) + \frac{1}{2} \cdot D_F(\Phi(\rho_2), \Phi(\sigma)) \\ &= \frac{1}{2} \cdot D_F(\Phi(\rho_1), \Phi(\bar{\rho})) + \frac{1}{2} \cdot D_F(\Phi(\rho_2), \Phi(\bar{\rho})) + D_F(\Phi(\bar{\rho}), \Phi(\sigma)) \\ &\leq \frac{1}{2} \cdot D_F(\rho_1, \bar{\rho}) + \frac{1}{2} \cdot D_F(\rho_2, \bar{\rho}) + D_F(\bar{\rho}, \sigma) = D_F(\rho_1, \sigma). \end{aligned} \quad (157)$$

□

Theorem 33 *Information divergence is monotone on spin factors.*

Proof According to Lemma 32 we just have to check monotonicity on spin factors of dimension 2, but these are sections of qubits. Müller-Hermes and Reeb [17] proved that quantum relative entropy is monotone on density matrices on complex Hilbert spaces. In particular quantum relative entropy is monotone on qubits. Therefore information divergence is monotone on any spin factor. □

We will need the following lemma.

Lemma 34 *Let $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ denote an affinity of a centrally symmetric set into itself. Let Ψ_r denote a dilation around the center c with a factor $r \in]0, 1]$. Then $\Psi_r \circ \Phi \circ \Psi_r^{-1}$ maps \mathcal{K} into itself.*

Proof Embed \mathcal{K} in a vector space V with origin in the center of \mathcal{K} . Then Φ is given by $\Phi(\vec{v}) = A\vec{v} + \vec{b}$ and $\Psi_r(\vec{v}) = r \cdot \vec{v}$. Then

$$(\Psi_r \circ \Phi \circ \Psi_r^{-1})(\vec{v}) = r \cdot \left(A \left(\frac{1}{r} \cdot \vec{v} \right) + \vec{b} \right) \quad (158)$$

$$= A\vec{v} + r \cdot \vec{b}. \quad (159)$$

Assume that $\vec{v} \in \mathcal{K}$. Then $\Phi(\vec{v}) \in \mathcal{K}$ and $-\Phi(-\vec{v}) \in \mathcal{K}$. Hence for $(1-t) \cdot \Phi(\vec{v}) + t \cdot (-\Phi(-\vec{v})) \in \mathcal{K}$. Now

$$(1-t) \cdot \Phi(\vec{v}) + t \cdot (-\Phi(-\vec{v})) = (1-t) \cdot \left(A\vec{v} + \vec{b} \right) + t \cdot \left(-\left(A(-\vec{v}) + \vec{b} \right) \right) \quad (160)$$

$$= A\vec{v} + (1-2t) \cdot \vec{b}. \quad (161)$$

For $t = \frac{1-r}{2}$ we get

$$(\Psi_r \circ \Phi \circ \Psi_r^{-1})(\vec{v}) = (1-t) \cdot \Phi(\vec{v}) + t \cdot (-\Phi(-\vec{v})) \in \mathcal{K}, \quad (162)$$

which completes the proof. \square

Theorem 35 *If D_F is a monotone Bregman divergence on a spin factor and $F_r(x) = F((1-r) \cdot \frac{1}{2} + r \cdot x)$ then the Bregman divergence D_{F_r} is also monotone.*

Proof We have

$$D_{F_r}(\rho, \sigma) = D_F\left((1-r) \cdot \frac{1}{2} + r \cdot \rho, (1-r) \cdot \frac{1}{2} + r \cdot \sigma\right) \quad (163)$$

$$= D_F(\Psi_r(\rho), \Psi_r(\sigma)) \quad (164)$$

where Ψ_r denotes a dilation around $\frac{1}{2}$ by a factor $r \in]0, 1]$. Let Φ denote an affinity of the state space into itself. Then according to Lemma 34

$$D_{F_r}(\Phi(\rho), \Phi(\sigma)) = D_F(\Psi_r(\Phi(\rho)), \Psi_r(\Phi(\sigma))) \quad (165)$$

$$= D_F((\Psi_r \circ \Phi)(\Psi_r^{-1} \circ \Psi_r(\rho)), (\Psi_r \circ \Phi)(\Psi_r^{-1} \circ \Psi_r(\sigma))) \quad (166)$$

$$= D_F((\Psi_r \circ \Phi \circ \Psi_r^{-1})(\Psi_r(\rho)), (\Psi_r \circ \Phi \circ \Psi_r^{-1})(\Psi_r(\sigma))) \quad (167)$$

$$\leq D_F(\Psi_r(\rho), \Psi_r(\sigma)) \quad (168)$$

$$= D_{F_r}(\rho, \sigma), \quad (169)$$

which proves the theorem. \square

In [19] joint convexity of Bregman divergences on complex density matrices was studied (see also [20]).

Theorem 36 *The separable Bregman divergence D_F given by $F(x) = \text{tr}[f(x)]$ is jointly convex if and only if f has the form*

$$f(x) = a(x) + \frac{\gamma}{2}q(x) + \int_0^\infty e_\lambda d\mu(\lambda) \quad (170)$$

where a is affine and

$$q(x) = x^2 \quad (171)$$

and

$$e_\lambda(x) = (\lambda + x) \ln(\lambda + x). \quad (172)$$

This result is related to the *matrix entropy class* introduced in [6] and further studied in [7]. The function q generates the Bregman divergence $D_q(\rho, \sigma) = \text{tr}[(\rho - \sigma)^2]$ and the function e_λ generates the Bregman divergence

$$D_{e_\lambda}(\rho, \sigma) = D(\rho + \lambda \parallel \sigma + \lambda). \quad (173)$$

We note that

$$\begin{aligned} & D(\rho + \lambda \parallel \sigma + \lambda) \\ &= (1 + 2\lambda) \cdot D\left(\frac{1}{1+2\lambda} \cdot \rho + \frac{2\lambda}{1+2\lambda} \cdot c \middle\| \frac{1}{1+2\lambda} \cdot \sigma + \frac{2\lambda}{1+2\lambda} \cdot c\right), \end{aligned} \quad (174)$$

which implies that $2\lambda(1 + 2\lambda) \cdot D_{e_\lambda}(\rho, \sigma) \rightarrow \text{tr}[(\rho - \sigma)^2]$ so the Bregman divergence D_2 may be considered as a limiting case. Now

$$D_f(\rho, \sigma) = \frac{\gamma}{2} \text{tr}[(\rho - \sigma)^2] + \int_0^\infty D(\rho + \lambda \parallel \sigma + \lambda) d\mu(\lambda). \quad (175)$$

Note that the Bregman divergence of order α can be written in this way for $\alpha \in [1, 2]$.

Theorem 37 *Any Bregman divergence based on a function of the form (170) is monotone on spin factors.*

Proof The result follows from Eqs.(174) and (175) in combination with Theorem 33. \square

9 Strict Monotonicity

Definition 38 We say that a regret function is *strictly monotone* if

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma) \quad (176)$$

implies that Φ is sufficient for ρ, σ .

In [10] it was proved that strict monotonicity implies monotonicity. As we shall see in Theorem 40 on convex bodies of rank 2 strictness and monotonicity is equivalent to strict monotonicity as long as the Bregman divergence is based on an analytic function.

Lemma 39 Let σ denote a point in a convex body \mathcal{K} with a monotone Bregman divergence D_F . If $\Phi : \mathcal{K} \rightarrow \mathcal{K}$ is an affinity then the set

$$C = \{\rho \in C \mid D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)\} \quad (177)$$

is a convex body that contains σ .

Proof Assume that $\rho_0, \rho_1 \in C$ and $t \in [0, 1]$ and $\bar{\rho} = (1-t) \cdot \rho_0 + t \cdot \rho_1$. Then according to the Bregman identity

$$\begin{aligned} & (1-t) \cdot D_F(\Phi(\rho_0), \Phi(\sigma)) + t \cdot D_F(\Phi(\rho_1), \Phi(\sigma)) \\ &= (1-t) \cdot D_F(\Phi(\rho_0), \Phi(\bar{\rho})) + t \cdot D_F(\Phi(\rho_1), \Phi(\bar{\rho})) + D_F(\Phi(\bar{\rho}), \Phi(\sigma)) \\ &\leq (1-t) \cdot D_F(\rho_0, \bar{\rho}) + t \cdot D_F(\rho_1, \bar{\rho}) + D_F(\bar{\rho}, \sigma) \\ &= (1-t) \cdot D_F(\rho_0, \sigma) + t \cdot D_F(\rho_1, \sigma). \end{aligned} \quad (178)$$

Therefore the inequality must hold with equality and

$$D_F(\Phi(\bar{\rho}), \Phi(\sigma)) = D_F(\bar{\rho}, \sigma), \quad (179)$$

which proves the lemma. \square

Theorem 40 Let D_F denote a monotone Bregman divergence that is strict on a spin factor based on an analytic function f . Then D_F is strictly monotone.

Proof Assume that D_F is monotone and that

$$D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma). \quad (180)$$

Let ρ_0 and ρ_1 denote extreme points such that ρ and σ lie on the line segment between ρ_0 and ρ_1 . Lemma 26 implies that

$$D^F(\Phi(\rho), \Phi(\sigma_t)) = D^F(\rho, \sigma_t) \quad (181)$$

for all $s \in [0, 1]$ where $\sigma_s = (1-s) \cdot \rho + s \cdot \sigma$. Since f is assumed to be analytic the identity (181) must hold for all t for which $(1-s) \cdot \rho + s \cdot \sigma \geq 0$ and this set of values of s coincides with set of values for which $(1-s) \cdot \Phi(\rho) + s \cdot \Phi(\sigma) \geq 0$. The identity (181) also holds if ρ is replaced by any point ρ' on the line segment between ρ_0 and ρ_1 because both sides of Eq. (181) are quadratic functions in the first variable. Using Proposition 25 we see that Eq. (180) can be extended to any pair of points on the line segment between ρ_0 and ρ_1 . In particular

$$D_F(\Phi(\rho_i), \Phi(\bar{\rho})) = D_F(\rho_i, \bar{\rho}) \quad (182)$$

for $i = 0, 1$ and $\bar{\rho} = \frac{1}{2} \cdot \rho_0 + \frac{1}{2} \cdot \rho_1$. Since both ρ_i and $\Phi(\rho_i)$ are extreme points we have

$$D_F \left(\Phi(\rho_i), \frac{1}{2} \right) = D_F \left(\rho_i, \frac{1}{2} \right) \quad (183)$$

we have $D_F \left(\bar{\rho}, \frac{1}{2} \right) = D_F \left(\Phi(\bar{\rho}), \frac{1}{2} \right)$. Therefore the points $\bar{\rho}$ and $\Phi(\bar{\rho})$ have the same distance to the center $\frac{1}{2}$, and there exists a rotation Ψ that maps $\Phi(\rho_i)$ into ρ_i . Since Ψ is a recovery map of the states ρ_i it is also a recovery map of ρ and σ . \square

An affinity in a Hilbert ball has a unique extension to a positive trace preserving map in the corresponding spin factor. Here we shall study such maps with respect to existence of recovery maps and with respect to monotonicity of Bregman divergences. Let Φ denote a positive trace preserving map of $JSpin_d$ into itself. The adjoint map Φ^* is defined by

$$\langle \Phi^*(x), y \rangle = \langle x, \Phi(y) \rangle. \quad (184)$$

If $\Phi(\sigma)$ is not singular then we may define

$$\Psi(\rho) = \sigma^{1/2} \Phi^* \left((\Phi(\sigma))^{-1/2} \rho (\Phi(\sigma))^{-1/2} \right) \sigma^{1/2}. \quad (185)$$

We observe that $\Psi(\Phi(\sigma)) = \sigma$. If Φ is an isomorphism then $\Phi(y) = O^* y O$ where O is an orthogonal map on $JSpin_d$ as a Hilbert space. Then

$$\langle \Phi^*(x), y \rangle = \langle x, \Phi(y) \rangle \quad (186)$$

$$= \text{tr} [x O^* y O] \quad (187)$$

$$= \text{tr} [O x O^* y] \quad (188)$$

$$= \langle O x O^*, y \rangle \quad (189)$$

so that $\Phi^*(x) = O x O^*$. Then

$$\Psi(\Phi(\rho)) = \sigma^{1/2} \Phi^* \left(\Phi \left(\sigma^{-1/2} \right) \Phi(\rho) \Phi \left(\sigma^{-1/2} \right) \right) \sigma^{1/2} \quad (190)$$

$$= \sigma^{1/2} O \left(\left(O^* \left(\sigma^{-1/2} \right) O \right) O^* \rho O \left(O^* \left(\sigma^{-1/2} \right) O \right) \right) O^* \sigma^{1/2} \quad (191)$$

$$= \rho. \quad (192)$$

Therefore Ψ is a recovery map. This formula extends to any ρ for which there exists a recovery map because Φ is an isomorphism between two sections of the state space that contain ρ and σ .

Acknowledgements I would like to thank Howard Barnum for pointing my attention to the notion of pairs of sections and retractions that proved to be very useful in stating and proving results on this topic. I would also like to thank two anonymous reviewers for their careful reading and their useful comments.

References

1. Alexandrov, A.D.: Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. Leningrad State Univ. Ann. [Uchenye Zapiski] **6**(335) (1939)
2. Alfsen, E.M., Shultz, F.W.: Geometry of State Spaces of Operator Algebras. Birkhäuser, Boston (2003)
3. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with Bregman divergences. *J. Mach. Learn. Res.* **6**, 1705–1749 (2005). <https://doi.org/10.1137/1.9781611972740.22>
4. Barnum, H., Barret, J., Krumm, M., Müller, M.P.: Entropy, majorization and thermodynamics in general probabilistic theories. In: Heunen, C., Selinger, P., Vicary, J. (eds.) Proceedings of the 12th International Workshop on Quantum Physics and Logic, Electronic Proceedings in Theoretical Computer Science, vol. 195, pp. 43–58 (2015). <https://arxiv.org/pdf/1508.03107.pdf>
5. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **7**, 200–217 (1967). Translated from Russian
6. Chen, R.Y., Tropp, J.: Subadditivity of matrix ϕ -entropy and concentration of random matrices. *Electron. J. Probab.* **19**(paper 27), 1–30 (2014). <https://doi.org/10.1214/EJP.v19-2964>
7. Hansen, F., Zhang, Z.: Characterisation of matrix entropies. *Lett. Math. Phys.* **105**(10), 1399–1411 (2015). <https://doi.org/10.1007/s11005-015-0784-8>
8. Harremoës, P.: Divergence and sufficiency for convex optimization. *Entropy* **19**(5) (2017). <https://doi.org/10.3390/e19050206>. Article no. 206
9. Harremoës, P.: Maximum entropy and sufficiency. *AIP Conf. Proc.* **1853**(1), 040001 (2017). <https://doi.org/10.1063/1.4985352>
10. Harremoës, P.: Quantum information on spectral sets. In: 2017 IEEE International Symposium on Information Theory, pp. 1549–1553 (2017). [https://doi.org/978-1-5090-4096-4/17/\\$31.00](https://doi.org/978-1-5090-4096-4/17/$31.00)
11. Harremoës, P., Tishby, N.: The information bottleneck revisited or how to choose a good distortion measure. In: 2007 IEEE International Symposium on Information Theory, pp. 566–570. IEEE Information Theory Society (2007). <https://doi.org/10.1109/ISIT.2007.4557285>
12. Hayashi, M.: Quantum Information Theory: Mathematical Foundation. Springer, Berlin (2016)
13. Holevo, A.S.: Probabilistic and Statistical Aspects of Quantum Theory. North-Holland Series in Statistics and Probability, vol. 1. North-Holland, Amsterdam (1982)
14. Jenčová, A., Petz, D.: Sufficiency in quantum statistical inference: a survey with examples. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **09**(03), 331–351 (2006). <https://doi.org/10.1142/S0219025706002408>
15. Jiao, J., Courtade, T., No, A., Venkat, K., Weissman, T.: Information measures: the curious case of the binary alphabet. *IEEE Trans. Inf. Theory* **60**(12), 7616–7626 (2014). <https://doi.org/10.1109/TIT.2014.2360184>
16. McCrimmon, K.: A Taste of Jordan Algebras. Springer, Berlin (2004)
17. Müller-Hermes, A., Reeb, D.: Monotonicity of the quantum relative entropy under positive maps. *Annales Henri Poincaré* **18**(5), 1777–1788 (2017). <https://doi.org/10.1007/s00023-017-0550-9>
18. Petz, D.: Sufficiency of channels over von Neumann algebras. *Q. J. Math. Oxford* **39**(1), 97–108 (1988). <https://doi.org/10.1093/qmath/39.1.97>
19. Pitrik, J., Virosztek, D.: On the joint convexity of the Bregman divergence of matrices. *Lett. Math. Phys.* **105**(5), 675–692 (2015). <https://doi.org/10.1007/s11005-015-0757-y>
20. Virosztek, D.: Jointly convex quantum Jensen divergences. *Linear Algebra and its Applications* (2018). <https://doi.org/10.1016/j.laa.2018.03.002>, <http://www.sciencedirect.com/science/article/pii/S0024379518301046>, arXiv:1712.05324

Information Geometry Associated with Generalized Means



Shinto Eguchi, Osamu Komori and Atsumi Ohara

Abstract This paper aims to provide a natural extension for the standard framework of information geometry. The main idea is to define a generalization of e-geodesic and m-geodesic, which associates with the canonical divergence and generalized expectation. The generalized e-geodesic is given by Kolmogorov–Nagumo means; the generalized m-geodesic is characterized to preserve the generalized expectation. The Pythagorean theorem with respect to the canonical divergence is shown in the space of all probability density functions employing the generalized e-geodesic and m-geodesic. As a result, we provide a wide class of generalization for the standard framework, in which there is still a dualistic structure associated with the two generalized geodesics as similar to the standard framework.

Keywords e-geodesic · KL divergence · m-geodesic Pythagorian theorem

1 Introduction

Information geometry is a research area to give essential understandings for decision-making based on phenomena with unknown uncertainty. This leads to the universal development connecting among a variety of disciplines in mathematical sciences.

S. Eguchi (✉)

Institute of Statistical Mathematics, Tachikawa 190-8562, Japan
e-mail: eguchi@ism.ac.jp

O. Komori

Department of Computer and Information Science, Seikei University, Musashino-shi,
Tokyo 180-8633, Japan

A. Ohara

Department of Electrical and Electronics Engineering, University of Fukui,
Fukui 910-8507, Japan

See for example, [1, 3, 5]. The key is to discuss geometric perspectives in a space

$$\mathcal{F} = \{f(x) : f(x) > 0, \int f(x)d\mathbb{P}(x) = 1\}, \quad (1)$$

where \mathbb{P} is a base probability measure. For example, \mathbb{P} is taken as a standard Gaussian probability distribution for a continuous case; \mathbb{P} is taken as a finite uniform distribution corresponding for a finite discrete case. A probability density function f of \mathcal{F} induces a probability measure $P_f(B) = \int_B f(x)d\mathbb{P}(x)$. All probabilistic statements with respect to P_f are described by the integration of $f(x)$.

Let us have a brief review for the original framework of information geometry in order to suggest a generalized framework in a subsequent discussion. For this, we focus on the e-geodesic and m-geodesic in \mathcal{F} . The pair of geodesics provides a dualistic view of \mathcal{F} as follows. Take three distinct density functions f, g and h in \mathcal{F} . Then the m-geodesic connecting between f and g is given as

$$C^{(m)} = \{f_\sigma^{(m)}(x) := (1 - \sigma)f(x) + \sigma g(x) : \sigma \in [0, 1]\}; \quad (2)$$

while the e-geodesic connecting between g and h is

$$C^{(e)} = \{h_\tau^{(e)}(x) := \exp((1 - \tau)\log g(x) + \tau \log h(x) - \kappa_\tau) : \tau \in [0, 1]\}, \quad (3)$$

where κ_τ is the normalizing factor, called the cumulant or free energy, defined by

$$\log \int \exp((1 - \tau)\log g + \tau \log h)d\mathbb{P}.$$

Henceforth, we assume that the domain of the e-geodesic $C^{(e)}$ is extended to an open interval containing the closed interval $[0, 1]$, see [29, 30] for rigorous discussions for \mathcal{F} in a framework of Orlicz space. If two geodesic curves $C^{(m)}$ and $C^{(e)}$ orthogonally intersect at g , then the Pythagorean identity

$$D_0(f, h) = D_0(f, g) + D_0(g, h) \quad (4)$$

holds, cf. [3], where the orthogonality is defined by the Fisher information and $D_0(f, g)$ is the Kullback–Leibler (KL) divergence defined by

$$D_0(f, g) = \int (\log f - \log g) f d\mathbb{P}. \quad (5)$$

Furthermore, it holds for any σ and τ of $[0, 1]$ that

$$D_0(f_\sigma^{(m)}, h_\tau^{(e)}) = D_0(f_\sigma^{(m)}, g) + D_0(g, h_\tau^{(e)}) \quad (6)$$

In this sense, the space \mathcal{F} associates with the right triangle connecting $f_\sigma^{(m)}$, g and $h_\tau^{(e)}$ on the m-geodesic $C^{(m)}$ and e-geodesic $C^{(e)}$. cf. [32]. We observe that the e-geodesic induces the KL divergence, or the canonical divergence in the sense that

$$-\left(\frac{d\kappa_\tau}{d\tau}\right)_{\tau=0} = D_0(g, h). \quad (7)$$

Thus, the two geodesics give such an intuitive understanding that plays an essential role on the foundation of information geometry.

The paper is organized as follows. We will extend this framework to a generalized framework in Sect. 2, considering the generalized analogues of $C^{(m)}$ and $C^{(e)}$. As similar to (7) we can define the canonical divergence, called generalized KL divergence from the generalized e-geodesic. Similarly, we will show that the generalized KL divergence satisfies the Pythagorean identity as in (4). For this the key idea is to build a generalized definition of expectation rather than the usual expectation, cf. [21]. In Sect. 3, we will discuss a generalized analogue of the Boltzmann–Gibbs–Shannon entropy, which is defined from the generalized KL divergence. Thus the maximum entropy distribution is derived from the generalized entropy, which is shown to form a generalized exponential model for a constrain of generalized expectation preserving space. In Sect. 4, we discuss the geometric characteristics of the generalized framework, and show the generalized Pythagorean theorem based on the divergence geometry. We finally give a brief discussion in Sect. 5.

2 Generalized m-Geodesic and e-Geodesic

Let $\phi : \mathbb{R} \rightarrow (0, \infty)$ be a surjective, strictly increasing and convex function, and so the inverse function ϕ^{-1} is defined on $(0, \infty)$, and strictly increasing and concave. We employ ϕ as a generator function for defining generalized m-geodesic and e-geodesic as follows. A ϕ -path connecting between f and g in \mathcal{F} is defined by

$$f_\tau^{(\phi)}(x) := \phi((1 - \tau)\phi^{-1}(f(x)) + \tau\phi^{-1}(g(x)) - \kappa_\tau^{(\phi)}), \quad (8)$$

where $\kappa_\tau^{(\phi)}$ is a normalizing factor satisfying

$$\int \phi((1 - \tau)\phi^{-1}(f(x)) + \tau\phi^{-1}(g(x)) - \kappa_\tau^{(\phi)}) d\mathbb{P}(x) = 1. \quad (9)$$

We call $C^{(\phi)} := \{f_\tau^{(\phi)}(x) : \tau \in [0, 1]\}$ the generalized e-geodesic connecting f and g embedded in \mathcal{F} noting that $C^{(\phi)}$ is reduced to the e-geodesic $C^{(e)}$ defined in (3) if $\phi(t) = \exp t$. See [13] for the existence of $\kappa_\tau^{(\phi)}$ for all $\tau \in [0, 1]$. Similarly, we assume that the domain of $C^{(\phi)}$ is extended to an open interval containing $[0, 1]$. The generalized e-geodesic is associated with Kolmogorov–Nagumo means for positive numbers a and b given as $\phi((1 - \tau)\phi^{-1}(a) + \tau\phi^{-1}(b))$, in which the generalized

mean is applied to two probability density functions $f(x)$ and $g(x)$ in place of positive numbers a and b . See [23–25] for the detailed discussion from statistical physics, and [27] for information bounds in statistical prediction.

The condition (9) leads to

$$\frac{\partial \kappa_{\tau}^{(\phi)}}{\partial \tau} = \frac{\int \{\phi^{-1}(g) - \phi^{-1}(f)\}\phi'(H_{\tau}(f, g))d\mathbb{P}}{\int \phi'(H_{\tau}(f, g))d\mathbb{P}} \quad (10)$$

and

$$\frac{\partial^2 \kappa_{\tau}^{(\phi)}}{\partial \tau^2} = \frac{\int \left\{ \phi^{-1}(g) - \phi^{-1}(f) - \frac{d}{d\tau} \kappa_{\tau}^{(\phi)} \right\}^2 \phi''(H_{\tau}(f, g))d\mathbb{P}}{\int \phi'(H_{\tau}(f, g))d\mathbb{P}} \quad (11)$$

if we assume that $f_{\tau}^{(\phi)}(x)$ is twice differentiable under the integral sign, where

$$H_{\tau}(f, g) = (1 - \tau)\phi^{-1}(f) + \tau\phi^{-1}(g) - \kappa_{\tau}^{(\phi)}.$$

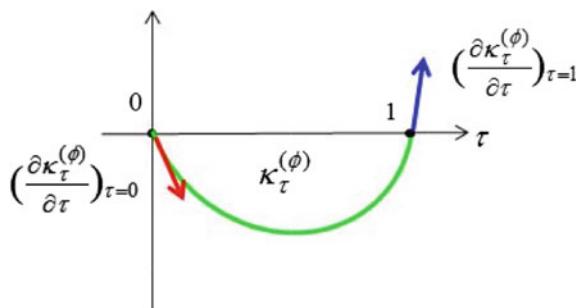
Henceforth, we assume the differentiability. We confirm that $\kappa_{\tau}^{(\phi)}$ is convex on $(0, 1)$ since the second order derivative (11) is positive because of the convexity for $\phi(s)$. Hence, we observe a property similar to (7) that

$$-\left(\frac{\partial \kappa_{\tau}^{(\phi)}}{\partial \tau}\right)_{\tau=0} = \frac{\int \{\phi^{-1}(f) - \phi^{-1}(g)\}\phi'(\phi^{-1}(f))d\mathbb{P}}{\int \phi'(\phi^{-1}(f))d\mathbb{P}} \quad (12)$$

which is equal to or larger than 0 with equality if and only if $f = g$ (a.e. \mathbb{P}). Because $\kappa_{\tau}^{(\phi)}$ is convex and nonpositive on $(0, 1)$ with $\kappa_0^{(\phi)} = 0$ and $\kappa_1^{(\phi)} = 0$ as seen in Fig. 1.

In accordance with this, the canonical divergence, called the generalized KL divergence can be defined by

Fig. 1 The derivative of the free energy function at $\tau = 0, 1$



$$D^{(\phi)}(f, g) = \int \{\phi^{-1}(f) - \phi^{-1}(g)\}\Phi(f)d\mathbb{P}, \quad (13)$$

where

$$\Phi(f(x)) = \frac{\phi'(\phi^{-1}(f(x)))}{\int \phi'(\phi^{-1}(f))d\mathbb{P}}. \quad (14)$$

See [31] for another derivation. Here we observe that Φ is bijective on \mathcal{F} under integrability conditions, that is

$$\Phi^{-1}(f(x)) = \phi((\phi')^{-1}(c_f f(x))), \quad (15)$$

where c_f is a normalizing constant depending on f to satisfy the total mass one. If $\phi(t) = \exp t$, then $D^{(\phi)}(f, g)$ is reduced to the Kullback–Leibler divergence. We note that

$$\left(\frac{\partial \kappa_\tau^{(\phi)}}{\partial \tau}\right)_{\tau=1} = \int \{\phi^{-1}(g) - \phi^{-1}(f)\}\Phi(g)d\mathbb{P}, \quad (16)$$

which is nothing but $D_\phi(g, f)$. The symmetrized form of the canonical divergence $D^{(\phi)}$ is given by

$$\int \{\phi^{-1}(f) - \phi^{-1}(g)\}\{\Phi(f) - \Phi(g)\}d\mathbb{P}. \quad (17)$$

Let \mathcal{S} and \mathcal{S}^* be linear hulls of $\{\phi^{-1}(f) : f \in \mathcal{F}\}$ and $\{\Phi(f) : f \in \mathcal{F}\}$, respectively. Then (17) associates with a canonical bilinear functional defined by

$$\langle s, s^* \rangle = \int s(x)s^*(x)d\mathbb{P}(x)$$

for s in \mathcal{S} and s^* in \mathcal{S}^* , see [26]. This bilinear form is essential to show the Pythagorean theorem which generalizes (4) employing $D^{(\phi)}$. We focus on the linear functional $F(s) := \langle s, s^* \rangle$. Thus, considering the functional $F(s)$, a generalized expectation for a random variable $t(X)$ is defined by

$$\mathbb{E}_f^{(\phi)}\{t(X)\} = \int t(x)\Phi(f(x))d\mathbb{P}(x). \quad (18)$$

Note that $\mathbb{E}^{(\phi)}$ is reduced to the usual expectation for a case when $\phi = \exp$. Thus, we can write

$$D^{(\phi)}(f, g) = \mathbb{E}_f^{(\phi)}\{\phi^{-1}(f(X)) - \phi^{-1}(g(X))\}. \quad (19)$$

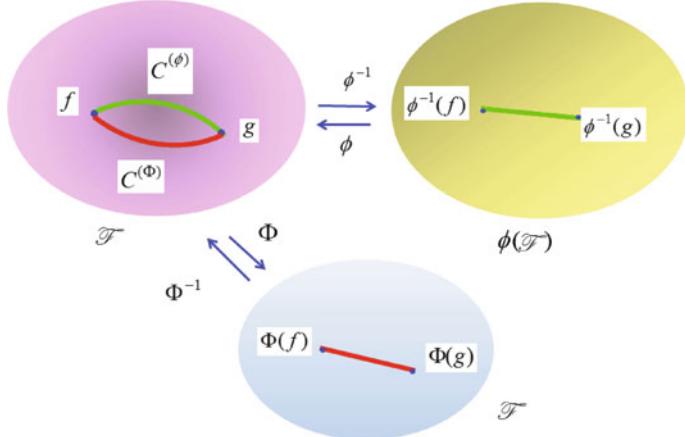


Fig. 2 The e-geodesic and m-geodesic in \mathcal{F} are injectively connected

Let us consider a generalized m-geodesic connecting between f and g of \mathcal{F} rather than the generalized e-geodesic in (8). The geodesic $\{f_{\sigma}^{(\phi)}\}$ is defined by

$$f_{\sigma}^{(\phi)}(x) = \Phi^{-1}\left((1 - \sigma)\Phi(f(x)) + \sigma\Phi(g(x))\right), \quad (20)$$

which we call $C^{(\phi)} := \{f_{\sigma}^{(\phi)}(x) : \sigma \in [0, 1]\}$ the generalized m-geodesic. Thus, we defined the generalized e-geodesic $C^{(\phi)} = \{f_{\tau}^{(\phi)} : \tau \in [0, 1]\}$ and m-geodesic $C^{(\phi)} = \{f_{\sigma}^{(\phi)} : \sigma \in [0, 1]\}$ connecting f with g in the space \mathcal{F} , in which $C^{(\phi)}$ is a line segment connecting $\phi^{-1}(f)$ and $\phi^{-1}(g)$ in the space $\phi^{-1}(\mathcal{F})$; while $C^{(\phi)}$ is a line segment connecting $\Phi(f)$ and $\Phi(g)$ in the space \mathcal{F} , cf. Fig. 2.

In fact, (20) is written as

$$\frac{1}{\int \frac{1}{(\phi^{-1})'(f_{\tau}^{(\phi*)}(x))} d\mathbb{P}} = (1 - \tau) \frac{1}{\int \frac{1}{(\phi^{-1})'(f(x))} d\mathbb{P}} + \tau \frac{1}{\int \frac{1}{(\phi^{-1})'(g(x))} d\mathbb{P}} \quad (21)$$

In accordance with this, the the generalized m-geodesic is closely related with the quasi-harmonic mean. If $\phi(f) = \exp f$, $C^{(\phi)}$ is reduced to the mixture geodesic $C^{(m)}$ in (2); if $\phi(f) = \frac{1}{2}t^{\frac{1}{2}}$, (21) is reduced to

$$\frac{1}{\int \frac{1}{f_{\tau}^{(\phi)}(x)} d\mathbb{P}} = (1 - \tau) \frac{1}{\int \frac{1}{f(x)} d\mathbb{P}} + \tau \frac{1}{\int \frac{1}{g(x)} d\mathbb{P}}, \quad (22)$$

Remark 1 We observe for a random vector $t(X)$ that

$$\mathbb{E}_{f_\sigma^{(\Phi)}}\{t(X)\} = (1 - \sigma)\mathbb{E}_f^{(\Phi)}\{t(X)\} + \sigma\mathbb{E}_g^{(\Phi)}\{t(X)\}. \quad (23)$$

The generalized m-geodesic preserves the Φ -mean in the following sense. If $\mathbb{E}_g^{(\Phi)}\{t(X)\} = \mathbb{E}_f^{(\Phi)}\{t(X)\}$, then $\mathbb{E}_{f_\sigma^{(\Phi)}}\{t(X)\} = \mathbb{E}_f^{(\Phi)}\{t(X)\}$ for all $\sigma \in [0, 1]$, that is, the generalized m-geodesic is in the Φ expectation preserving space

$$\mathcal{E}_f^{(\Phi)} = \{g : \mathbb{E}_f^{(\phi)}\{t(X)\} = \mathbb{E}_g^{(\Phi)}\{t(X)\}\}. \quad (24)$$

for arbitrarily fixed f of \mathcal{F} .

Here we discuss some examples of the generator function ϕ . The first example is a power exponential function $\phi_\beta(t) = (1 + \beta t)^{\frac{1}{\beta}}$, so $\phi_\beta^{-1}(s) = \frac{1}{\beta}(s^\beta - 1)$. See [6, 7, 12, 16] for the divergence by use of the generalized exponential function. The generalized e-geodesic is given by

$$f^{(\phi_\beta)}(x) = \{(1 - \tau)f(x)^\beta + \tau g(x)^\beta - \kappa_\tau^{(\phi_\beta)}\}^{\frac{1}{\beta}}; \quad (25)$$

the generalized m-geodesic is given by

$$f^{(\phi_\beta)}(x) = z_\tau^{(\phi_\beta)} \left\{ (1 - \tau) \frac{f(x)^{1-\beta}}{\int f^{1-\beta} dP} + \tau \frac{g(x)^{1-\beta}}{\int g^{1-\beta} dP} \right\}^{\frac{1}{1-\beta}}, \quad (26)$$

where $\kappa_\tau^{(\phi_\beta)}$ and $z_\tau^{(\phi_\beta)}$ are normalizing constants to satisfy the total mass one. The canonical divergence is written by

$$D_{\phi_\beta}(f, g) = \frac{1}{\beta \int f^{1-\beta} dP} \left\{ 1 - \int f^{1-\beta} g^\beta dP \right\}. \quad (27)$$

The second example is $\phi(t) = \exp(-t^{-1})$, so $\phi_\nu^{-1}(s) = -(\log s)^{-1}$. The generalized e-geodesic is given by

$$f^{(\phi)}(x) = \exp \left\{ \frac{1}{(1 - \tau) \frac{1}{\log f(x)} + \tau \frac{1}{\log g(x)}} - \kappa_\tau^{(\phi)} \right\}, \quad (28)$$

the generalized m-geodesic is not given in a closed form. The canonical divergence is written by

$$D_\phi(f, g) = \int \left\{ \frac{1}{\log g(x)} - \frac{1}{\log f(x)} \right\} \frac{f(\log f)^2}{\int f(\log f)^2 d\mathbb{P}} d\mathbb{P}. \quad (29)$$

3 Maximum Entropy Distribution

Let us consider an exponential model,

$$M^{(e)} = \{f_\theta^{(e)}(x) := \exp(\theta^\top t(x) - \kappa(\theta)) : \theta \in \Theta\}, \quad (30)$$

where $\kappa(\theta)$ is a normalizing factor and $\Theta = \{\theta : \kappa(\theta) < \infty\}$. Here $t(x)$ is called a canonical statistic; θ is called a canonical parameter. On the other hand, the m-geodesic associates with a expectation preserving space of the random vector $t(X)$ with a density function f of \mathcal{F} ,

$$\mathcal{E}_f = \{g \in \mathcal{F} : \mathbb{E}_g\{t(X)\} = \mathbb{E}_f\{t(X)\}\}, \quad (31)$$

where \mathbb{E}_f is a statistical expectation with respect to f . By definition, if $C^{(m)}$ is the m-geodesic between any two density functions of \mathbb{E} , $C^{(m)}$ is embedded in \mathcal{E} . In this sense, the m-geodesic preserves the expectation on \mathcal{E} . Let us consider the Boltzmann–Gibbs–Shannon entropy

$$H_0(f) = - \int f(x) \log f(x) d\mathbb{P}(x). \quad (32)$$

Then we observe that, if \mathcal{E} is the expectation preserving space of $t(X)$ with $f_\theta^{(e)}$, then

$$f_\theta^{(e)} = \underset{f \in \mathcal{E}}{\operatorname{argmax}} H_0(f). \quad (33)$$

Because, if f is in \mathcal{E} , then $H_0(f_\theta^{(e)}) - H_0(f) = D_0(f_\theta^{(e)}, f)$, which is always non-negative with equality if and only if $f = f_\theta^{(e)}$. Thus, the exponential model (30) is characterized by the maximum entropy, cf. [17]. Consider a minimization problem of $D_0(f, g)$ with constrain $g \in M^{(e)}$ for a fixed f outside $M^{(e)}$. Then the projection of f onto $M^{(e)}$ leads to the Pythagorean foliation such that the disjoint union $\bigcup_{g \in M^{(e)}} \mathcal{E}_g$ of $t(X)$ with g spans the total space \mathcal{F} on which the Pythagorean identity

$$D_0(f, h) = D_0(f, g) + D_0(g, h), \quad (34)$$

holds for any f of \mathcal{E}_g and any h of $M^{(e)}$.

We discuss the ϕ -cross entropy defined by

$$C^{(\phi)}(f, g) = -\mathbb{E}_f^{(\Phi)}\{\phi^{-1}(g(X))\}, \quad (35)$$

which expresses that $D^{(\phi)}(f, g) = C^{(\phi)}(f, g) - H^{(\phi)}(f)$, where the ϕ -entropy $H^{(\phi)}(f)$ is defined as $C^{(\phi)}(f, f)$, see [11] for the framework on the U -divergence and U -cross entropy similar to this argument. Hence there is an elementary inequality

$$C^{(\phi)}(f, g) \geq H^{(\phi)}(f) \quad (36)$$

with equality if and only if $f = g$ (a.e. \mathbb{P}). We next consider the maximum entropy density function associated with the ϕ -entropy $H^{(\phi)}(f)$. Here we suppose a mean constrain for a random vector $t(X)$ of d -dimension as

$$\mathcal{E}^{(\phi)} = \{g \in \mathcal{F} : \mathbb{E}_g^{(\Phi)}\{t(X)\} = \mathbb{E}_f^{(\Phi)}\{t(X)\}\} \quad (37)$$

rather than the usual mean constrain, where f is a fixed density function. The Euler–Lagrange function is written by

$$\mathcal{L}^{(\phi)}(f) = \int \{\phi^{-1}(f) - \theta^\top t + c_\theta\} \Phi(f) d\mathbb{P}, \quad (38)$$

where θ is the parameter of Lagrangian multipliers and $c_\theta = \mathbb{E}_f^{(\Phi)}\{\theta^\top t(X)\}$. Let f^{\max} be a maximum $H^{(\phi)}$ entropy density under the Φ mean constraint. Then, by the definition of f^{\max} , for any curve f_ϵ such that $f_\epsilon|_{\epsilon=0} = f^{\max}$ in $\mathcal{E}^{(\phi)}$

$$\frac{d}{d\epsilon} \mathcal{L}^{(\phi)}(f_\epsilon)|_{\epsilon=0} = 0, \quad (39)$$

which implies that

$$\int \frac{d}{d\epsilon} f_\epsilon \left[1 - \frac{\phi''(f_\epsilon)}{\phi'(f_\epsilon)^2} \{\phi^{-1}(f_\epsilon) - \theta^\top t - \text{const}\} d\mathbb{P} \right] \Big|_{\epsilon=0} = 0. \quad (40)$$

Hence $\phi^{-1}(f^{\max}(x)) = \theta^\top t(x) + \text{const}$, since (40) must hold for any f_ϵ , and $\int \frac{d}{d\epsilon} f_\epsilon d\mathbb{P}|_{\epsilon=0} = 0$. We conclude that

$$f_\theta^{(\phi)}(x) = \phi(\theta^\top t(x) - \kappa^{(\phi)}(\theta)) \quad (41)$$

as the maximizer of $\mathcal{L}^{(\phi)}(f)$, where $\kappa^{(\phi)}(\theta)$ is a normalizing factor. In fact, if f is in $\mathcal{E}_{f_\theta^{(\phi)}}^{(\phi)}$, then

$$C^{(\phi)}(f, f_\theta^{(\phi)}) = -\mathbb{E}_{f_\theta^{(\phi)}}^{(\Phi)}\{\phi(f_\theta^{(\phi)}(X))\} \quad (42)$$

which is exactly $H^{(\phi)}(f_\theta^{(\phi)})$. Therefore, it follows from (36) that $f_\theta^{(\phi)}$ is the maximizer of $H^{(\phi)}(f)$ in f under the constraints $\mathcal{E}_{f_\theta^{(\phi)}}^{(\phi)}$.

Definition 1 Let $t(X)$ be a random vector. A generalized exponential model is defined by

$$M^{(\phi)} = \{f_\theta^{(\phi)}(x) := \phi(\theta^\top t(x) - \kappa^{(\phi)}(\theta)) : \theta \in \Theta\}, \quad (43)$$

where $\kappa^{(\phi)}(\theta)$ is a normalizing factor and $\Theta = \{\theta : \kappa^{(\phi)}(\theta) < \infty\}$.

In this way, the generalized exponential model is derived by the maximum $H^{(\phi)}$ entropy distribution, see [14, 15] for the duality between the maximum entropy and minimum divergence, and [19] for deformed exponential families. We confirm that $H^{(\phi)}$ and $M^{(\phi)}$ are the Boltzmann entropy and the exponential model if $\phi(f) = \exp f$, in which $M^{(\phi)}$ is well known as the maximum Boltzmann entropy model.

4 Divergence Geometry

We discuss the divergence geometry for a model $M = \{f_\theta(x) : \theta \in \Theta\}$ associated with ϕ -divergence. If we apply the formula given in [8, 9] for $D^{(\phi)}$ to M , then the Riemannian metric $G^{(\phi)}$ is defined by

$$G_{ij}^{(\phi)}(\theta) = \int \frac{\partial}{\partial \theta_i} \phi^{-1}(f_\theta) \frac{\partial}{\partial \theta_j} \Phi(f_\theta) d\mathbb{P}, \quad (44)$$

and the pair of affine connections are

$$\Gamma_{ij,k}^{(\phi)}(\theta) = \int \frac{\partial}{\partial \theta_k} \phi^{-1}(f_\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Phi(f_\theta) d\mathbb{P} \quad (45)$$

and

$${}^* \Gamma_{ij,k}^{(\phi)}(\theta) = \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} \phi^{-1}(f_\theta) \frac{\partial}{\partial \theta_k} \Phi(f_\theta) d\mathbb{P}. \quad (46)$$

as the components with respect to θ .

Remark 2 We find that

$$\frac{\partial}{\partial \theta_k} G_{ij}^{(\phi)}(\theta) = \Gamma_{kj,i}^{(\phi)}(\theta) + {}^* \Gamma_{ki,j}^{(\phi)}(\theta), \quad (47)$$

in which Γ and ${}^* \Gamma$ are said to be conjugate with respect to $G^{(\phi)}$.

Remark 3 If we consider

$$\tilde{D}^{(\phi)}(f, g) = \int \{\phi^{-1}(f) - \phi^{-1}(g)\}\phi'(\phi^{-1}(f))d\mathbb{P} \quad (48)$$

as a divergence measure, then the geometry associated with $\tilde{D}^{(\phi)}$ is a conformal correspondence to that with $D^{(\phi)}$, in which the conformal factor is given by $(\int \phi'(\phi^{-1}(f))d\mathbb{P})^{-1}$, cf. [18]. Here we confirm from the convexity of ϕ that

$$\{\phi^{-1}(f(x)) - \phi^{-1}(g(x))\}\phi'(\phi^{-1}(f(x))) \geq \phi(\phi^{-1}(f(x))) - \phi(\phi^{-1}(g(x))), \quad (49)$$

in which the integration of both sides in \mathbb{P} establishes the nonnegativity of $\tilde{D}^{(\phi)}(f, g)$.

We next review the divergence geometry by U -divergence, cf. [10, 11].

Remark 4 Let $U(s)$ be a strictly and convex function on \mathbb{R} . Then the U -divergence is defined by

$$D_U(f, g) = \int \{\xi(f) - \xi(g)\}f d\mathbb{P} - \int \{U(\xi(f)) - U(\xi(g))\}d\mathbb{P} \quad (50)$$

where $\xi(t)$ is the inverse function of the derivative of $U(s)$, cf. [11]. The divergence geometry with D_U on the model $M = \{f_\theta : \theta \in \Theta\}$ is given by

$$G_{ij}^{(U)}(\theta) = \int \frac{\partial}{\partial \theta_i} \xi(f_\theta) \frac{\partial}{\partial \theta_j} f_\theta d\mathbb{P}, \quad (51)$$

The pair of affine connections are

$$\Gamma_{ij,k}^{(U)}(\theta) = \int \frac{\partial}{\partial \theta_k} \xi(f_\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta d\mathbb{P} \quad (52)$$

and

$${}^* \Gamma_{ij,k}^{(U)}(\theta) = \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} \xi(f_\theta) \frac{\partial}{\partial \theta_k} f_\theta d\mathbb{P}. \quad (53)$$

as the components with respect to θ . Therefore, the affine connection $\Gamma^{(U)}$ is essentially the same as the mixture connection. If we assume that $\xi = \phi^{-1}$, then U is a primitive function of ϕ , in which the U -divergence is closely related with the ϕ -divergence.

Get back to the generalized exponential model $M^{(\phi)} = \{f_\theta^{(\phi)}(x) : \theta \in \Theta\}$ defined in (43). Then the canonical parameter θ leads to

$$\left(G_{ij}^{(\phi)}(\theta), \Gamma_{ij,k}^{(\phi)}(\theta), {}^* \Gamma_{ij,k}^{(\phi)}(\theta) \right) = \left(\frac{\partial^2 \kappa^{(\phi)}(\theta)}{\partial \theta_i \partial \theta_j}, \frac{\partial^3 \kappa^{(\phi)}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, 0 \right). \quad (54)$$

On the other hand, if we consider the ϕ mean parameter $\eta = \mathbb{E}_{f_\theta^{(\phi)}}^{\langle\phi\rangle}\{t(X)\}$, then

$$\left(G_{ij}^{(\phi)}(\eta), \Gamma_{ij,k}^{(\phi)}(\eta), {}^*\Gamma_{ij,k}^{(\phi)}(\eta)\right) = \left(\frac{\partial^{2*}\kappa^{(\phi)}(\eta)}{\partial\eta_i\partial\eta_j}, 0, \frac{\partial^{3*}\kappa^{(\phi)}(\eta)}{\partial\eta_i\partial\eta_j\partial\eta_k}\right), \quad (55)$$

where ${}^*\kappa^{(\phi)}(\eta)$ is a conjugate convex function of $\kappa^{(\phi)}(\theta)$, that is,

$${}^*\kappa^{(\phi)}(\eta) = \sup_{\theta \in \Theta}\{\theta^\top \eta - \kappa^{(\phi)}(\theta)\}. \quad (56)$$

Thus we observe that θ is ${}^*\Gamma$ -affine parameter; η is Γ -affine parameter. The duality between parameters θ and η is deeply explored in [2, 4, 18, 20, 21].

Let f and g be in \mathcal{F} . We consider a geometric property of a generalized e-geodesic connecting between f and g as defined in (8). Let

$$\mathcal{T}^{(\phi)} = \{t(x) : \int \phi(t(x))d\mathbb{P}(x) < \infty\}. \quad (57)$$

By definition, $\mathcal{T}^{(\phi)}$ is a convex set, that is, if $t(x)$ and $s(x)$ are in $\mathcal{T}^{(\phi)}$, then $(1 - \tau)s(x) + \tau t(x)$ is in $\mathcal{T}^{(\phi)}$ for $\tau, 0 < \tau < 1$. Because ϕ is a convex function from assumption. We assume that there are K distinct functions $t_k(x)$ for $k = 1, \dots, K$ in $\mathcal{T}^{(\phi)}$ such that

$$f(x) = \phi \left(\sum_{k=1}^K \theta_{f_k} t_k(x) - \kappa^{(\phi)}(\theta_f) \right) \quad (58)$$

and

$$g(x) = \phi \left(\sum_{k=1}^K \theta_{g_k} t_k(x) - \kappa^{(\phi)}(\theta_g) \right). \quad (59)$$

Consider the ϕ model generated by $t(x) = (t_1(x), \dots, t_K(x))^\top$, that is,

$$M^{(\phi)} = \{\phi(\theta^\top t(x) - \kappa^{(\phi)}(\theta)) : \theta \in \Theta\}, \quad (60)$$

where $\Theta = \{\theta : \kappa^{(\phi)}(\theta) < \infty\}$. Note that f and g are both in $M^{(\phi)}$. Then the ${}^*\Gamma$ -geodesic connecting between f and g in $M^{(\phi)}$ is given by

$$\theta(\tau) = (1 - \tau)\theta_f + \tau\theta_g \quad (61)$$

in Θ because the canonical parameter θ is an affine parameter with respect to ${}^*\Gamma$ as seen in (54). So the ${}^*\Gamma$ -geodesic is written as

$$f_\tau(x) = \phi((1 - \tau)\theta_f^\top t(x) + \tau\theta_g^\top t(x) - \kappa^{(\phi)}(\theta(\tau))) \quad (62)$$

in $M^{(\phi)}$, where $c_\tau = \kappa(\theta(\tau))$. It follows from expressions (58) and (59) that

$$f_\tau(x) = \phi((1 - \tau)\phi(g(x)) + \tau\phi(f(x)) - \kappa_\tau^{(\phi)}) \quad (63)$$

where $\kappa_\tau^{(\phi)} = (1 - \tau)\kappa^{(\phi)}(\theta_f) + \tau\kappa^{(\phi)}(\theta_g) + \kappa^{(\phi)}(\theta(\tau))$. Therefore the ${}^*\Gamma$ -geodesic is nothing but the generalized e-geodesic.

Remark 5 Assumptions (58) and (59) are not very restrictive because we can take arbitrarily a large set $\{t_k(x) : 1 \leq k \leq K\}$. In fact, the model $M^{(\phi)}$ can well approximate \mathcal{F} for sufficiently a large K using a standard function approximation theorem if we assume several analytic assumptions for \mathcal{F} including a condition of compact support and so forth.

We next consider a geometric property of a ϕ^* -path connecting between f and g as defined in (20). Let us take K density functions $h_k(x)$ for $k = 1, \dots, K$ such that $h_1 = f$ and $h_2 = g$. Then we consider a parametric model $M^{(\phi)} = \{f_\eta^{(\phi)}(x) : \eta \in H\}$ such that

$$\Phi(f_\eta^{(\phi)}(x)) = \sum_{k=1}^K \eta_k \Phi(h_k(x)) \quad (64)$$

where $H = \{\eta : \kappa^*(\eta) < \infty\}$ and $\sum_{k=1}^K \eta_k = 1$ and $\eta_k \geq 0$ for $k, 1 \leq k \leq K$. Thus, we note that $f_{\eta^{(1)}}^{(\phi^*)} = f$ and $f_{\eta^{(2)}}^{(\phi^*)} = g$ for $\eta^{(1)} = (1, 0, 0, \dots, 0)^\top$ and $\eta^{(2)} = (0, 1, 0, \dots, 0)^\top$, respectively. Then, the parameter η is a Γ affine parameter from the definition in (52) because

$$\frac{\partial^2}{\partial \eta_i \partial \eta_j} \Phi(f_\eta^{(\phi)}(x)) = 0. \quad (65)$$

Hence, the Γ -geodesic is given by $\eta(\tau) = (1 - \tau)\eta^{(1)} + \tau\eta^{(2)}$ in H . In $M^{(\phi)}$ the Γ -geodesic $f_\tau(x)$ satisfies

$$\Phi(f_\tau(x)) = (1 - \tau)\Phi(f(x)) + \tau\Phi(g(x)), \quad (66)$$

which leads that the Γ -geodesic in $M^{(\phi)}$ is equal to the generalized m-geodesic defined in (20). In fact, we note that $M^{(\phi)}$ can also well approximate \mathcal{F} if K is taken sufficiently large.

We consider a triangle in \mathcal{F} . Let us take distinct density functions f , g and h in \mathcal{F} . Then consider the generalized e-geodesic $\{f_t\}_{0 \leq t \leq 1}$ and generalized m-geodesic $\{h_s\}_{0 \leq s \leq 1}$ as follows:

$$\Phi(f_t) = (1 - t)\Phi(g) + t\Phi(f) \quad (67)$$

and

$$\phi^{-1}(h_s) = (1 - s)\phi^{-1}(g) + s\phi^{-1}(h) - \kappa_s^{(\phi)}, \quad (68)$$

where $\kappa_s^{(\phi)}$ is the normalizing constant satisfying $\int h_s d\mathbb{P} = 1$. Note that

$$\mathbb{E}_{f_t}^{(\phi)}\{s(X)\} = (1-t)\mathbb{E}_g^{(\phi)}\{s(X)\} + t\mathbb{E}_f^{(\phi)}\{s(X)\} \quad (69)$$

for any statistic $s(X)$.

Theorem 1 Let $\{f_t\}_{0 \leq t \leq 1}$ and $\{h_s\}_{0 \leq s \leq 1}$ be the two geodesic curves as defined in (67) and (68), respectively. Then,

$$D^{(\phi)}(f, h) = D^{(\phi)}(f, g) + D^{(\phi)}(g, h) \quad (70)$$

if and only if

$$G^{(\phi)}\left(\frac{\partial}{\partial t} f_t, \frac{\partial}{\partial s} h_s\right) \Big|_{t=0, s=0} = 0. \quad (71)$$

Proof By definition,

$$G^{(\phi)}\left(\frac{\partial}{\partial t} f_t, \frac{\partial}{\partial s} h_s\right) \Big|_{t=0, s=0} = \int \frac{\partial}{\partial s} \phi(h_s) \frac{\partial}{\partial t} \Phi(f_t), \quad (72)$$

which is written by

$$\begin{aligned} & \int \{\phi(g) - \phi(h) - \dot{\kappa}_s^{(\phi)}\}\{\Phi(f) - \Phi(g)\}d\mathbb{P} \\ &= \int \{\phi(g) - \phi(h)\}\{\Phi(f) - \Phi(g)\}d\mathbb{P}, \end{aligned} \quad (73)$$

which is equal to $D^{(\phi)}(f, h) - \{D^{(\phi)}(f, g) + D^{(\phi)}(g, h)\}$. This completes the proof.

We have a look at a foliation structure of the full space \mathcal{F} as seen in Sect. 3. Let $M^{(\phi)}$ be a generalized exponential model, that is

$$M^{(\phi)} = \{\phi(\theta^\top t(x) - \kappa^{(\phi)}(\theta)) : \theta \in \Theta\} \quad (74)$$

where $\Theta = \{\theta \in \mathbb{R}^d : \kappa^{(\phi)}(\theta) < \infty\}$. Let

$$\mathcal{E}_g^{(\phi)} = \{f \in \mathcal{F} : \mathbb{E}_f^{(\phi)}\{t(X)\} = \mathbb{E}_g^{(\phi)}\{t(X)\}\} \quad (75)$$

for a fixed g in \mathcal{F} . Then from Theorem 1 we see that, if g is in $M^{(\phi)}$,

$$D^{(\phi)}(f, h) = D^{(\phi)}(f, g) + D^{(\phi)}(g, h) \quad (76)$$

for any f of $\mathcal{E}_g^{(\phi)}$ and any h of $M^{(\phi)}$. Thus the Pythagorean foliation

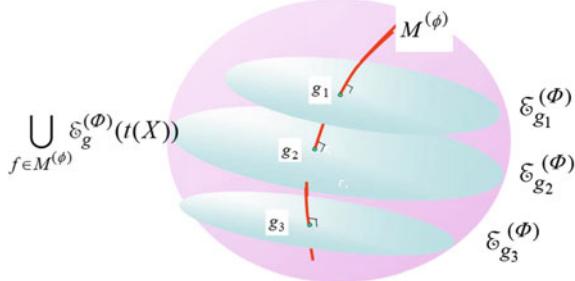


Fig. 3 Pythagorean foliation of \mathcal{F}

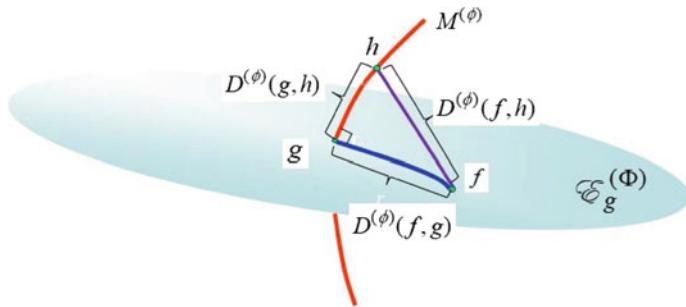


Fig. 4 The orthogonal transversality of the generalized exponential model and the generalized expectation preserving space

$$\bigcup_{g \in M^{(\phi)}} \mathcal{E}_g^{(\Phi)}(t(X)) \quad (77)$$

holds with leaves $\mathcal{E}_g^{(\Phi)}$ on the ray $M^{(\phi)}$, see Fig. 3. We remark that this is associated with the minimization problem $\min_{h \in M^{(\phi)}} D^{(\phi)}(f, h)$, see Fig. 4.

5 Discussion

We have discussed the extension of the standard framework of information geometry. However, we pose a strong assumption on the e-geodesic and generalized e-geodesic. There still remains incomplete discussion for the nonparametric framework of the the generalized geometry, see [22] for the related topics.

The key of this paper is that the generator function ϕ defined on the set of positive values leads to an operation of $\phi\text{-sum-}\phi^{-1}$, which defines the generalized e-geodesic, extending the e-geodesic led to the geometric mean, or exp-sum-log. Thus, any element of \mathcal{F} is connected by other elements in a form of $\phi\text{-sum-}\phi^{-1}$. The elegant property of the foliation structure of \mathcal{F} with an exponential model as a Rey and the

expectation preserving space as a leaf is naturally extended to that of the generalized exponential model and the generalized mean space. Here is a dualistic idea of the operation of ϕ -sum- ϕ^{-1} . Let us consider a space of real-valued functions, say $\mathcal{S} = \{s(x)\}$. Then, we consider

$$s_\tau^{(\phi)}(x) = \phi^{-1}((1 - \tau)\phi(s_1(x)) + \tau\phi(s_2(x))) \quad (78)$$

as an operation of ϕ^{-1} -sum- ϕ . Typically, the log-sum-exp gives a nonlinear connection for functions $s_1(x)$ and $s_2(x)$. In a context of pattern recognition such operations are discussed in [28], in which the prediction performance is shown to be more powerful than a linear connection. In principle, it is closely related with the deep leaning in a context of machine learning, however, the geometric considerations are not completely given. In a near future the dualistic understanding will be offered in a framework of information geometry.

Acknowledgements We are grateful to two referees' constructive comments, which are greatly helpful for the revision of paper.

References

1. Amari, S.: Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics, vol. 28. Springer, New York (1985)
2. Amari, S.: Information geometry of positive measures and positive-definite matrices: decomposable dually flat structure. *Entropy* **16**, 2131–2145 (2014)
3. Amari, S., Nagaoka, H.: Methods of Information Geometry. Oxford University Press, Oxford (2000)
4. Amari, S., Ohara, A., Matsuzoe, H.: Geometry of deformed exponential families: invariant, dually-flat and conformal geometries. *Phys. A Stat. Mech. Appl.* **391**(18), 4308–4319 (2012)
5. Ay, N., Jost, J., Van Le, H., Schwachhofer, L.: Information Geometry, vol. 64. Springer, Berlin (2017)
6. Cichocki, A., Amari, S.I.: Families of alpha-beta-and gamma-divergences: flexible and robust measures of similarities. *Entropy* **12**, 1532–1568 (2010)
7. Cichocki, A., Cruces, S., Amari, S.: Generalized alpha-beta divergences and their application to Robust nonnegative matrix factorization. *Entropy* **13**, 134–170 (2011)
8. Eguchi, S.: Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Stat.* **11**, 793–803 (1983)
9. Eguchi, S.: Geometry of minimum contrast. *Hiroshima Math. J.* **22**, 631–647 (1992)
10. Eguchi, S.: Information geometry and statistical pattern recognition. *Sugaku Expositions*, vol. 19, pp. 197–216. American Mathematical Society (2006)
11. Eguchi, S.: Information divergence geometry and the application to statistical machine learning. In: Emmert-Streib, F., Dehmer, M. (eds.) *Information Theory and Statistical Learning*, pp. 309–332. Springer, Berlin (2008)
12. Eguchi, S., Kato, S.: Entropy and divergence associated with power function and the statistical application. *Entropy* **12**, 262–274 (2010)
13. Eguchi, S., Komori, O.: Path connectedness on a space of probability density functions. *Geometric Science of Information*, pp. 615–624. Springer International Publishing, Berlin (2015)
14. Eguchi, S., Komori, O., Kato, S.: Projective power entropy and maximum Tsallis entropy distributions. *Entropy* **13**, 1746–1764 (2011)

15. Eguchi, S., Komori, O., Ohara, A.: Duality of maximum entropy and minimum divergence. *Entropy* **16**(7), 3552–3572 (2014)
16. Fujisawa, H., Eguchi, S.: Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* **99**, 2053–2081 (2008)
17. Grunwald, P.D., Dawid, A.P.: Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Ann. Stat.* 1367–1433 (2004)
18. Matsuzoe, H.: Hessian structures on deformed exponential families and their conformal structures. *Differ. Geom. Appl.* **35**, 323–333 (2014)
19. Matsuzoe, H., Wada, T.: Deformed algebras and generalizations of independence on deformed exponential families. *Entropy* **17**, 5729–5751 (2015); Naudts, J.: Generalised Thermostatistics. Springer, London (2011)
20. Matsuzoe, H., Henmi, M.: Hessian structures on deformed exponential families. *Geometric Science of Information*, pp. 275–282. Springer, Berlin (2013)
21. Matsuzoe, H., Henmi, M.: Hessian structures and divergence functions on deformed exponential families. *Geometric Theory of Information*, pp. 57–80. Springer International Publishing, Berlin (2014)
22. Montrucchio, L., Pistone, G.: Deformed exponential bundle: the linear growth case. In: *International Conference on Geometric Science of Information*, pp. 239–246. Springer, Cham (2017)
23. Naudts, J.: Generalized exponential families and associated entropy functions. *Entropy* **10**, 131–149 (2008)
24. Naudts, J.: The q -exponential family in statistical physics. *Central Eur. J. Phys.* **7**, 405–413 (2009)
25. Naudts, J.: Generalized Thermostatistics. Springer Science & Business Media, Berlin (2011)
26. Newton, N.: An infinite dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **263**, 1661–1681 (2012)
27. Nielsen, F.: Generalized Bhattacharyya and Chernoff upper bounds on Bayes error using quasi-arithmetic means. *Pattern Recognit. Lett.* **42**, 25–34 (2014)
28. Omae, K., Komori, O., Eguchi, S.: Quasi-linear score for capturing heterogeneous structure in biomarkers. *BMC Bioinform.* **18**(1), 308 (2017)
29. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* 1543–1561 (1995)
30. Santacroce, M., Siri, P., Trivellato, B.: New results on mixture and exponential models by Orlicz spaces. *Bernoulli* **22**, 1431–1447 (2016)
31. Vigelis, R., David, C., Charles, C.: New metric and connections in statistical manifolds. *Geometric Science of Information*, pp. 222–229. Springer International Publishing, Berlin (2015)
32. Zhang, J.: Nonparametric information geometry: from divergence function to referential-representational biduality on statistical manifolds. *Entropy* **15**, 5384–5418 (2013)

Information Geometry with (Para-)Kähler Structures



Jun Zhang and Teng Fei

Abstract We investigate conditions under which a statistical manifold \mathfrak{M} (with a Riemannian metric g and a pair of torsion-free conjugate connections ∇, ∇^*) can be enhanced to a (para-)Kähler structure. Assuming there exists an almost (para-)complex structure L compatible with g on a statistical manifold \mathfrak{M} (of even dimension), then we show $(\mathfrak{M}, g, L, \nabla)$ is (para-)Kähler if ∇ and L are Codazzi coupled. Other equivalent characterizations involve a symplectic form $\omega \equiv g(L \cdot, \cdot)$. In terms of the compatible triple (g, ω, L) , we show that (i) each object in the triple induces a conjugate transformation on ∇ and becomes an element of an (Abelian) Klein group; (ii) the compatibility of any two objects in the triple with ∇ leads to the compatible quadruple (g, ω, L, ∇) in which any pair of objects are mutually compatible. This is what we call *Codazzi-(para-)Kähler manifold* [8] which admits the family of torsion-free α -connections (convex mixture of ∇, ∇^*) compatible with (g, ω, L) . Finally, we discuss the properties of divergence functions on $\mathfrak{M} \times \mathfrak{M}$ that lead to Kähler (when $L = J, J^2 = -id$) and para-Kähler (when $L = K, K^2 = id$) structures.

Keywords Statistical manifold · Torsion · Codazzi coupling · Conjugation of connection · Kähler structure · Para-Kähler structure · Codazzi-(para-)Kähler Compatible triple · Compatible quadruple

1 Introduction

Let \mathfrak{M} be a smooth (real) manifold of *even dimension* and ∇ be an affine connection on it. In this paper, we would investigate the interaction of ∇ with three geometric structures on \mathfrak{M} , namely, a pseudo-Riemannian metric g , a nondegenerate 2-form

J. Zhang (✉)
University of Michigan, Ann Arbor, MI 48109, USA
e-mail: junz@umich.edu

T. Fei
Columbia University, New York, NY 10027, USA
e-mail: tfei@math.columbia.edu

ω , and a tangent bundle isomorphism $L : T\mathfrak{M} \rightarrow T\mathfrak{M}$, often forming a “compatible triple” together. The interaction of the compatible triple (g, ω, L) with ∇ , in terms of parallelism, is well understood, leading to integrability of L and of ω , and turning almost (para-)Hermitian structure of \mathfrak{M} to (para-)Kähler structure on \mathfrak{M} . Here, we investigate the interaction of ∇ with the compatible triple (g, ω, L) in terms of Codazzi coupling, a relaxation of parallelism.

We start by recalling that the statistical structure $(\mathfrak{M}, g, \nabla)$ can be defined either as (i) a manifold $(\mathfrak{M}, g, \nabla, \nabla^*)$ with a pair, ∇ and ∇^* , of torsion-free g -conjugate connections (Lauritzen’s definition [21]); or (ii) a manifold $(\mathfrak{M}, g, \nabla)$ with a torsion-free connection ∇ that is Codazzi coupled to g (Kurose’s definition [20]). Though the two definitions can be shown to be equivalent, they represent two different perspectives of generalizing Levi-Civita connection which is, by definition, parallel to g . Section 2.1 aims at clarifying the distinction and the link between (i) the concept of h -conjugate transformation of connection ∇ ; and (ii) the concept of Codazzi coupling associated to the pair (∇, h) , where h is an arbitrary $(0, 2)$ -tensor. The special cases of $h = g$ (symmetric) and $h = \omega$ (skew-symmetric) are highlighted, because both g -conjugation and ω -conjugation are involutive operations. Codazzi coupling ∇ with h then, is the precise characterization of the condition for such conjugate operations on a connection to preserve its torsion.

In Sect. 2.2, we investigate Codazzi coupling of ∇ with a tangent bundle isomorphism L , in particular the cases of $L = J$, $J^2 = -id$ (almost complex structure) and $L = K$, $K^2 = id$ (almost para-complex structure, with same multiplicity for ± 1 eigenvalues). Such coupling is shown to lead to integrability of L .

In Sects. 2.3 and 2.4, the interaction of ∇ with the compatible triple (g, ω, L) is studied. We follow the same approach of Sect. 2.1 in distinguishing (i) the conjugation transformation of ∇ by, and (ii) the Codazzi coupling of ∇ with respect to each of the (g, ω, L) . In the former case (Sect. 2.3), we show that g -conjugate, ω -conjugate, and L -gauge transformation (together with identity transform) form a Klein group of transformations of connections. In the latter case (Sect. 2.4), we show that Codazzi couplings of ∇ with any two of the compatible (g, ω, L) lead to its coupling with the third (and hence turning the compatible triple into a compatible quadruple (g, ω, L, ∇)). After studying the implications of the existence of such couplings (Sect. 2.5), this then leads to the definition of *Codazzi-(para-)Kähler* structure (Sect. 2.6); its relations with various other geometric structures (Hermitian, symplectic, etc) are also discussed there.

Section 3 investigates how (para-)Kähler structures on $\mathfrak{M} \times \mathfrak{M}$ may arise from divergence functions. After a brief review how divergence functions induce a statistical structure (Sect. 3.1), we study how they may induce a symplectic structure on $\mathfrak{M} \times \mathfrak{M}$ (Sect. 3.2). We then show constraints on divergence functions if the induced structures on $\mathfrak{M} \times \mathfrak{M}$ are further para-Kähler (Sect. 3.3) or Kähler (Sect. 3.4). As an exercise, we relate our construction of Kähler structure to Calabi’s diastatic function approach (Sect. 3.5).

In this paper, we investigate integrability of L and of ω while g and L are not necessarily covariantly constant (i.e., parallel) with respect to ∇ , but instead are Codazzi coupled to it. The results were known in the parallel case: the exis-

tence of a torsion-free connection ∇ on \mathfrak{M} such that $\nabla g = 0$ (metric-compatible) and $\nabla L = 0$ (complex connection) implies that (\mathfrak{M}, g, L) is (para-)Kähler. When Codazzi coupling replaces parallelism, our results show that (para-)Kähler manifolds may still admit a pair of conjugate connections ∇ and ∇^* , much like statistical manifolds do. In recent work [15], we showed that such pair of connections are in fact both (para-)holomorphic for the (para-)Kähler manifolds; general conditions for (para-)holomorphicity of g -conjugate and L -gauge transformations of connections for (para-)Hermitian manifolds are also studied there.

As most materials in Sect. 2 has already appeared in [8, 31], we only provide summary of results while omitting proofs. A small improvement to earlier results is showing the entire family of α -connections for the Codazzi-(para-)Kähler manifold. Section 3 contains results unpublished before. All materials of this paper were first presented at the fourth international conference on Information Geometry and Its Applications (IGAIA4).

2 Enhancing Statistical Structure to (Para-)Kähler Structures

In this Section, we investigate Codazzi couplings of an affine connection ∇ on a real manifold \mathfrak{M} with a pseudo-Riemannian metric g , a symplectic form ω , and a tangent bundle isomorphism $L : T\mathfrak{M} \rightarrow T\mathfrak{M}$. We prove that the Codazzi coupling between a torsion-free ∇ and a quadratic operator L leads to transversal foliations, and that the Codazzi coupling of any two of (g, ω, L) leads to the Codazzi coupling of the remaining third. Mirroring the study of these Codazzi couplings is the study of the transformations of ∇ by g -conjugation, by ω -conjugate, and by L -gauge, and of how their torsions are affected including when they are preserved. As a highlight, we show that these transformations generically are non-trivial elements of a four-element Klein group. This motivates the notions of *compatible quadruple* and *Codazzi-(para-)Kähler* manifolds.

2.1 Conjugate Transformation and Codazzi Coupling Associated to (h, ∇)

The simplest form of “coupling” relation between ∇ and h is that of “parallelism”: $\nabla h = 0$. In other words, covariant derivative of h under ∇ is zero. There are two ways of generalizing this notion of parallelism: the first involves introducing the notion of a h -conjugate transformation ∇^h of ∇ such that $\nabla^h = \nabla$ recovers $\nabla h = 0$, the second involves requiring ∇h to have some symmetry for which $\nabla h = 0$ is a special case. Below, we discuss them in turn.

Conjugation of a connection by h If h is any non-degenerate $(0, 2)$ -tensor, i.e., bilinear form, it induces isomorphisms $h(X, -)$ and $h(-, X)$ from vector fields X to one-forms. When h is not symmetric, these two isomorphisms are different. Given

an affine connection ∇ , we can take the covariant derivative of the one-form $h(X, -)$ with respect to Z , and obtain a non-tensorial object θ such that, when fixing X ,

$$\theta_Z(Y) = Z(h(X, Y)) - h(X, \nabla_Z Y).$$

Similarly, we can take the covariant derivative of the one-form $h(-, Y)$ with respect to Z , and obtain a corresponding object $\tilde{\theta}$ such that, when fixing Y ,

$$\tilde{\theta}_Z(X) = Z(h(X, Y)) - h(\nabla_Z X, Y).$$

Since h is non-degenerate, there exists a U and V such that $\theta_Z = h(U, -)$ and $\tilde{\theta}_Z = h(-, V)$ as one-forms, so that

$$\begin{aligned} Z(h(X, Y)) &= h(U(Z, X), Y) + h(X, \nabla_Z Y), \\ Z(h(X, Y)) &= h(\nabla_Z X, Y) + h(X, V(Z, Y)). \end{aligned}$$

Proposition 1 ([31], Proposition 7) *Let $\nabla_Z^{\text{left}} X := U(Z, X)$ and $\nabla_Z^{\text{right}} X := V(Z, X)$. Then ∇^{left} and ∇^{right} are both affine connections as induced from ∇ .*

The ∇^{left} and ∇^{right} are called, respectively, *left-h-conjugate* and *right-h-conjugate* of ∇ ; neither is involutive in general. From their definitions, it is easy to see that

$$(\nabla^{\text{left}})^{\text{right}} = (\nabla^{\text{right}})^{\text{left}} = \nabla.$$

Reference [31] provided the conditions under which left- and right-conjugate of h are the same.

Proposition 2 ([31], Proposition 15) *When the non-degenerate bilinear form h is either symmetric, $h(X, Y) = h(Y, X)$, or skew-symmetric, $h(X, Y) = -h(Y, X)$, then*

$$\nabla^{\text{left}} = \nabla^{\text{right}}.$$

The two special cases of h : symmetric or skew-symmetric bilinear form, are denoted as g and ω , respectively. Since the left- and right-conjugates with respect to such h are equal, we use ∇^* to denote g -conjugate and ∇^\dagger to denote ω -conjugate of an arbitrary affine connection ∇ . Note that both g -conjugation and ω -conjugation operations are involutive: $(\nabla^*)^* = \nabla$, $(\nabla^\dagger)^\dagger = \nabla$.

In information geometry, it is standard to consider α -connections for $\alpha \in \mathbb{R}$

$$\nabla_g^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^*, \quad \text{with } (\nabla_g^{(\alpha)})^* = \nabla_g^{(-\alpha)}.$$

Likewise, we can introduce

$$\nabla_{\omega}^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^{\dagger}, \quad \text{with } (\nabla_{\omega}^{(\alpha)})^{\dagger} = \nabla_{\omega}^{(-\alpha)}.$$

Remark 1 Despite of the skew-symmetric nature of ω , ω -conjugation is one and the same whether defined with respect to the first or second slot of ω :

$$Z\omega(X, Y) = \omega(\nabla_Z^{\dagger}X, Y) + \omega(X, \nabla_ZY) = \omega(\nabla_ZX, Y) + \omega(X, \nabla_Z^{\dagger}Y).$$

Codazzi coupling of ∇ and h We introduce the (0,3)-tensor C defined by:

$$C_h(X, Y, Z) \equiv (\nabla_Zh)(X, Y) = Z(h(X, Y)) - h(\nabla_ZX, Y) - h(X, \nabla_ZY). \quad (1)$$

The tensor C_h is called the *cubic form* associated with (∇, h) pair. Rewriting the above

$$C_h(X, Y, Z) = h((\nabla^{\text{left}} - \nabla)_ZX, Y) = h(X, (\nabla^{\text{right}} - \nabla)_ZY), \quad (2)$$

we see that

$$\nabla = \nabla^{\text{left}} = \nabla^{\text{right}}$$

if and only if $C_h = 0$. In this case, we say that ∇ is *parallel* to the bilinear form h , i.e.,

$$Z(h(X, Y)) = h(\nabla_ZX, Y) + h(X, \nabla_ZY).$$

In general, the cubic forms associated with $(\nabla^{\text{left}}, h)$ pair and with $(\nabla^{\text{right}}, h)$ pair are

$$(\nabla_Z^{\text{left}}h)(X, Y) = (\nabla_Z^{\text{right}}h)(X, Y) = -C_h(X, Y, Z) = -(\nabla_Zh)(X, Y).$$

From (2), we can derive

$$\begin{aligned} C_h(X, Y, Z) - C_h(Z, Y, X) &= h(T^{\nabla^{\text{left}}}(Z, X) - T^{\nabla}(Z, X), Y) \\ &= h(X, T^{\nabla^{\text{right}}}(Z, Y) - T^{\nabla}(Z, Y)). \end{aligned}$$

So $C_h(X, Y, Z) = C_h(Z, Y, X)$ if and only if the torsions of ∇ , ∇^{left} , ∇^{right} are all equal

$$T(X, Y) = T^{\nabla^{\text{left}}}(X, Y) = T^{\nabla^{\text{right}}}(X, Y).$$

Definition 1 Let h be a non-degenerate bilinear form, and ∇ an affine connection. Then (∇, h) is called a *Codazzi pair*, and ∇ and h are said to be *Codazzi coupled*, if

$$C_h(X, Y, Z) = C_h(Z, Y, X) \quad (3)$$

or explicitly

$$(\nabla_Z h)(X, Y) = (\nabla_X h)(Z, Y).$$

Now, let us investigate Codazzi coupling of ∇ with g (symmetric case) or ω (skew-symmetric case); in both cases $\nabla^{\text{left}} = \nabla^{\text{right}}$.

- For $h = g$: symmetry of g implies $C_g(X, Y, Z) = C_g(Y, X, Z)$. This, combined with (3), leads to

$$C_g(Z, Y, X) = C_g(X, Y, Z) = C_g(Y, X, Z) = C_g(Z, X, Y) = C_g(X, Z, Y) = C_g(Y, Z, X),$$

so $C_g(X, Y, Z) \equiv \nabla g$ is totally symmetric in its three slots.

- For $h = \omega$: skew-symmetry of ω implies $C_\omega(X, Y, Z) = -C_\omega(Y, X, Z)$. This, combined with (3), leads to

$$\begin{aligned} C_\omega(X, Y, Z) &= C_\omega(Z, Y, X) = -C_\omega(Y, Z, X) = -C_\omega(X, Z, Y) \\ &= C_\omega(Z, X, Y) = C_\omega(Y, X, Z) = -C_\omega(X, Y, Z), \end{aligned}$$

hence $C_\omega(X, Y, Z) \equiv \nabla\omega = 0$.

We therefore conclude

Proposition 3 Let ∇^* and ∇^\dagger denote the g -conjugate and ω -conjugate of an arbitrary connection ∇ with respect to g and ω , respectively.

1. The following are equivalent:

- (i) (∇, g) is Codazzi-coupled;
- (ii) (∇^*, g) is Codazzi-coupled;
- (iii) ∇g is totally symmetric;
- (iv) $\nabla^* g$ is totally symmetric;
- (v) $T^\nabla = T^{\nabla^*}$.

2. The following are equivalent:

- (i) $\nabla\omega = 0$;
- (ii) $\nabla^\dagger\omega = 0$;
- (iii) $\nabla = \nabla^\dagger$;
- (iv) $T^\nabla = T^{\nabla^\dagger}$.

2.2 Tangent Bundle Isomorphism

Codazzi coupling of ∇ with L For a smooth manifold \mathfrak{M} , an isomorphism L of the tangent bundle $T\mathfrak{M}$ is a smooth section of the bundle $\text{End } T\mathfrak{M}$ such that it is

invertible everywhere. Starting from a (not necessarily torsion-free) connection ∇ on \mathfrak{M} , an *L-gauge transformation* of a connection ∇ is a new connection ∇^L defined by

$$\nabla_X^L Y = L^{-1}(\nabla_X(LY))$$

for any vector fields X and Y . It can be verified that indeed ∇^L is an affine connection. Note that gauge transformations of a connection form a group, with operator composition as group multiplication.

Definition 2 L and ∇ are said to be *Codazzi coupled* if the following identity holds

$$(\nabla_X L)Y = (\nabla_Y L)X, \quad (4)$$

where

$$(\nabla_X L)Y \equiv \nabla_X(LY) - L(\nabla_X Y).$$

We have the following characterization of Codazzi coupling of ∇ with L

Lemma 1 (e.g., [27]) *Let ∇ be an affine connection, and let L be a tangent bundle isomorphism. Then the following statements are equivalent:*

- (i) (∇, L) is Codazzi-coupled.
- (ii) (∇^L, L^{-1}) is Codazzi-coupled.
- (iii) $T^\nabla = T^{\nabla^L}$.

Integrability of L A tangent bundle isomorphism $L : T\mathfrak{M} \rightarrow T\mathfrak{M}$ is said to be a *quadratic operator* if it satisfies a real coefficient quadratic polynomial equation with distinct roots, i.e., there exists $\alpha \neq \beta \in \mathbb{C}$ such that $\alpha + \beta$, $\alpha\beta$ are real numbers and

$$L^2 - (\alpha + \beta)L + \alpha\beta \cdot \text{id} = 0.$$

Note that L is an isomorphism, so $\alpha\beta \neq 0$.

Let E_α and E_β be the eigenbundles of L corresponding to the eigenvalues α and β respectively, i.e., at each point $x \in \mathfrak{M}$, the fiber is defined by

$$E_\lambda(x) := \{X \in T_x\mathfrak{M} : L_x(X) = \lambda X\} \text{ for } \lambda = \alpha, \beta.$$

As subbundles of the tangent bundle $T\mathfrak{M}$, E_α and E_β are distributions. We call $E_\alpha(E_\beta)$ a foliation if for any vector fields X, Y with value in $E_\alpha(E_\beta)$, so is their Lie bracket $[X, Y]$.

The Nijenhuis tensor N_L associated with L is defined as

$$N_L(X, Y) = -L^2[X, Y] + L[X, LY] + L[LX, Y] - [LX, LY].$$

When $N_L = 0$, the operator L is said to be integrable. In this case, the eigen-bundles of L form foliations, i.e., subbundles that are closed with respect to Lie bracket operation $[., .]$.

One can derive (see [8]) that, when a quadratic operator L is Codazzi-coupled to an affine connection ∇ , then the Nijenhuis tensor N_L has the expression

$$N_L(X, Y) = L^2 T^\nabla(X, Y) - LT^\nabla(X, LY) - LT^\nabla(LX, Y) + T^\nabla(LX, LY).$$

An immediate consequence is that $N_L = 0$ vanishes when ∇ is torsion-free $T^\nabla = 0$. That is,

Proposition 4 *A quadratic operator L is integrable if it is Codazzi-coupled to a torsion-free connection ∇ .*

Combining Proposition 4 with Lemma 1 yields

Corollary 1 *A quadratic operator L is integrable if there exists a torsion-free connection ∇ such that ∇^L is torsion-free.*

Almost-complex J and almost-para-complex K operator The most important examples of the bundle isomorphism L are almost complex structures and almost para-complex structures. By definition, L is called an *almost complex structure* if $L^2 = -\text{id}$. Analogously, L is known as an *almost para-complex structure* if $L^2 = \text{id}$ and the multiplicities of the eigenvalues ± 1 are equal. We will use J and K to denote almost complex structures and almost para-complex structures, respectively, and use L when these two structures can be treated in a unified way. It is clear from our definitions that such structures exist only when \mathfrak{M} is of even dimension.

The following results follow readily from Lemma 1 for the special case of $L^2 = \pm \text{id}$.

Corollary 2 *When $L = J$ or $L = K$,*

1. $\nabla^L = \nabla^{L^{-1}}$, i.e., L -gauge transformation is involutive, $(\nabla^L)^L = \nabla$.
2. (∇, L) is Codazzi-coupled if and only if (∇^L, L) is Codazzi-coupled.

Compatible triple (g, ω, L) The compatibility condition between a metric g and an almost (para-)complex structure $J(K)$ is well-known, where $J^2 = -\text{id}$ and $K^2 = \text{id}$. We say that g is compatible with J if J is orthogonal, i.e.

$$g(JX, JY) = g(X, Y) \tag{5}$$

holds for any vector fields X and Y . Similarly we say that g is compatible with K if

$$g(KX, KY) = -g(X, Y) \tag{6}$$

is always satisfied, which implies that g must be of split signature. When expressed using L , (5) and (6) have the same form

$$g(X, LY) + g(LX, Y) = 0. \quad (7)$$

Hence a two-form ω can be defined

$$\omega(X, Y) = g(LX, Y), \quad (8)$$

and turns out to satisfy

$$\omega(X, LY) + \omega(LX, Y) = 0. \quad (9)$$

Of course, one can also start with ω and define $g(X, Y) = \omega(L^{-1}X, Y)$, then show that imposing compatibility of ω and L via (9) leads to the desired symmetry of g . Finally, given the knowledge of both g and ω , the bundle isomorphism L defined by (8) is uniquely determined, which satisfies (7), (9) and $L^2 = \pm id$. Whether L takes the form of J or K depends on whether (5) as opposed to (6) is to be satisfied.

In any case, the three objects g , ω and L with $L^2 = \pm id$ form a *compatible triple* such that given any two, the third one is rigidly “interlocked”.

2.3 Klein Group of Transformations on ∇

We now show a key relationship between the three transformations of a connection ∇ : its g -conjugate ∇^* , its ω -conjugate ∇^\dagger , and its L -gauge transform ∇^L .

Theorem 1 ([8], Theorem 2.13) *Let (g, ω, L) be a compatible triple, and ∇^* , ∇^\dagger , and ∇^L denote, respectively, g -conjugation, ω -conjugation, and L -gauge transformation of an arbitrary connection ∇ . Then, $(id, *, \dagger, L)$ realizes a 4-element Klein group action on the space of affine connections:*

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

Theorem 1 and Proposition 3, part (2) immediately lead to

Corollary 3 *Given a compatible triple (g, ω, L) , $\nabla\omega = 0$ if and only if*

$$\nabla^* = \nabla^L.$$

Explicitly written,

$$\nabla_Z^* X = \nabla_Z X + L^{-1}((\nabla_Z L)X) = \nabla_Z X + L((\nabla_Z L^{-1})X). \quad (10)$$

Remark 2 Note that, in both Theorem 1 and Corollary 3, there is no requirement of ∇ to be torsion-free nor is there any assumption about its Codazzi coupling with L or with g . In particular, Corollary 3 says that, when viewing $\omega(X, Y) = g(LX, Y)$, $\nabla\omega = 0$ if and only if the torsions introduced by $*$ and by L are cancelled.

There have been confusing statements about (10), even for the special case of $L = J$, the almost complex structure. In Ref. [11, Proposition 2.5(2)], (10) was shown after assuming (g, ∇) to be a statistical structure. On the other hand, [24, Lemma 4.2] claimed the converse, also under the assumption of $(\mathfrak{M}, g, \nabla)$ being statistical. As Corollary 3 shows, the Codazzi coupling of ∇ and g is not relevant for (10) to hold; (10) is entirely a consequence of $\nabla\omega = 0$. Corollary 3 is a special case of a more general theorem ([31], Theorem 21).

2.4 Compatible Quadruple (g, ω, L, ∇)

We now consider simultaneous Codazzi couplings by the same ∇ with a compatible triple (g, ω, L) . We first have the following result.

Theorem 2 Let ∇ be a torsion-free connection on \mathfrak{M} , and $L = J, K$. Consider the following three statements regarding any compatible triple (g, ω, L)

- (i) (∇, g) is Codazzi-coupled;
- (ii) (∇, L) is Codazzi-coupled;
- (iii) $\nabla\omega = 0$.

Then

1. Given (iii), then (i) and (ii) imply each other;
2. Assume ∇ is torsion-free, then (i) and (ii) imply (iii).

Proof First, assuming (iii), we show that (i) and (ii) imply each other. This is because by Theorem 1, (iii) amounts to $\nabla = \nabla^\dagger$. Therefore, $\nabla^* = \nabla^L$. Hence: $T^{\nabla^*} = T^\nabla$ iff $T^{\nabla^L} = T^\nabla$. By Proposition 3 part (1), $T^{\nabla^*} = T^\nabla$ is equivalent to (g, ∇) being Codazzi coupled. By Lemma 1, $T^{\nabla^L} = T^\nabla$ is equivalent to (L, ∇) being Codazzi coupled. Hence, we proved that (i) and (ii) imply each other.

Next, assuming (i) and (ii), (iii) holds under the condition that ∇ is torsion-free. The proof is much involved, see the proof of Theorem 3.4 of [8].

In [8], we propose the notion of ‘‘compatible quadruple’’ to describe the compatibility between the four objects g, ω, L , and ∇ on a manifold \mathfrak{M} .

Definition 3 ([8], Definition 3.9) A *compatible quadruple* on \mathfrak{M} is a quadruple (g, ω, L, ∇) , where g and ω are symmetric and skew-symmetric non-degenerate $(0,2)$ -tensors respectively, L is either an almost complex or almost para-complex structure, and ∇ is a torsion-free connection, that satisfy the following relations:

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

for any vector fields X, Y, Z on \mathfrak{M} .

As a consequence of Theorem 2, we have the following proposition regarding compatible quadruple.

Proposition 5 *Given a torsion-free connection ∇ , (g, ω, L, ∇) forms a compatible quadruple if any of the following conditions holds:*

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).

In other words, compatibility of ∇ with any two objects of the compatible triple makes a compatible quadruple, i.e., satisfying the three conditions as specified by either (1), (2), or (3) will lead to the satisfaction of all conditions (i)–(vi) of Definition 3.

2.5 Role of Connection ∇

A manifold \mathfrak{M} admitting a compatible quadruple (g, ω, L, ∇) , when ∇ is furthermore torsion-free, is in fact a (para-)Kähler manifold. This is because:

1. Codazzi coupling of L with a torsion-free ∇ ensures that L is integrable;
2. $\nabla\omega = 0$ with ∇ torsion-free ensures that $d\omega = 0$ (see Lemma 3.1 of [8]).

So the existence of a torsion-free connection ∇ on \mathfrak{M} that is Codazzi couple to the compatible triple (g, ω, L) on \mathfrak{M} gives rise to a (para-)Kähler structure on \mathfrak{M} .

Let us recall definitions of various types of structures on a manifold. A manifold (\mathfrak{M}, g, L) where g is a Riemannian metric is said to be *almost (para-)Hermitian* if g and L are compatible; when furthermore L is integrable, then (\mathfrak{M}, g, L) is called a *(para-)Hermitian* manifold. On the other hand, a manifold (\mathfrak{M}, ω) with a nondegenerate 2-form ω is said to be *symplectic* if we require ω to be closed, $d\omega = 0$. Amending (\mathfrak{M}, ω) with a (non-necessarily integrable) L turns $(\mathfrak{M}, \omega, L)$ into an *almost (para-)Kähler manifold* when L and ω are compatible. If furthermore we require both (i) an integrable L and (ii) a closed ω , then what we have on \mathfrak{M} is a *(para-)Kähler structure*.

Note that in the definitions of (para-)Hermitian, symplectic, and (para-)Kähler structures, no affine connections are explicitly involved. In particular, (para-)Kähler manifold is defined by the integrability conditions of L and closedness of ω , which

are related to topological properties of \mathfrak{M} . However, it is well-known in (para-)Kähler geometry that (\mathfrak{M}, g, L) is (para-)Kähler if and only if L is parallel under the Levi-Civita connection of g , i.e., if there exists a torsion-free connection ∇ such that

$$\nabla g = 0, \quad \nabla L = 0.$$

So the existence of a “nice enough” connection on \mathfrak{M} will enable a (para-)Kähler structure on it.

One the other hand, a *symplectic connection* ∇ is a connection that is both torsion-free and parallel to ω : $\nabla\omega = 0$. A symplectic manifold (\mathfrak{M}, ω) , where $d\omega = 0$, equipped with a symplectic connection is known as a *Fedosov manifold* [14]. Since the parallelism of L with respect to any torsion-free ∇ implies that L is integrable, a symplectic manifold (\mathfrak{M}, ω) can be enhanced to a (para-)Kähler manifold if any symplectic connection on \mathfrak{M} also renders L parallel:

$$\nabla\omega = 0, \quad \nabla L = 0.$$

Again, it is the existence of a “nice enough” connection that enhances the symplectic manifold to a (para-)Kähler manifold.

The contribution of our work is to extend the involvement of a connection ∇ from “parallelism” to “Codazzi coupling”; this is how statistical manifolds extend Riemannian manifolds. To this end, Theorem 2 says that for an arbitrary statistical manifold $(\mathfrak{M}, g, \nabla)$, if there exists an almost (para-)complex structure L compatible with g such that (the necessarily torsion-free, by definition of a statistical manifold) ∇ and L are Codazzi-coupled, then what we have of $(\mathfrak{M}, g, L, \nabla)$ is a (para-)Kähler manifold.

Theorem 2 also says that, for any Fedosov manifold $(\mathfrak{M}, \omega, \nabla)$, if there exists an almost (para-)complex structure L compatible with ω such that (the necessarily torsion-free, by definition of symplectic connection of a Fedosov manifold) ∇ and L are Codazzi-coupled, then $(\mathfrak{M}, \omega, L, \nabla)$ is a (para-)Kähler manifold. In other words, Codazzi coupling of ∇ with L turns a statistical manifold or a Fedosov manifold into a (para-)Kähler manifold, which is then both statistical and symplectic.

Proposition 6 *Given compatible triple (g, ω, L) on a manifold \mathfrak{M} , then any two of the following three statements imply the third, meanwhile turning \mathfrak{M} into a (para-)Kähler manifold:*

- (i) $(\mathfrak{M}, g, \nabla)$ is a statistical manifold;
- (ii) $(\mathfrak{M}, \omega, \nabla)$ is a Fedosov manifold;
- (iii) (∇, L) is Codazzi coupled.

2.6 Codazzi-(Para-)Kähler Structure

Insofar as a compatible quadruple (g, ω, L, ∇) gives rise to a special kind of (para-)Kähler manifold, where the torsion-free ∇ is integrated snugly into the compatible triple (g, ω, L) , we can call such a manifold *Codazzi-(para-)Kähler manifold*.

More generally, since as seen from Proposition 4, integrability of L may result from the existence of an affine connection $\bar{\nabla}$ that is Codazzi coupled to L under the condition that ∇ is torsion-free, we can have the following definition.

Definition 4 ([8], Definition 3.8) An almost Codazzi-(para-)Kähler manifold \mathfrak{M} is by definition an almost (para-)Hermitian manifold (\mathfrak{M}, g, L) with an affine connection ∇ (not necessarily torsion-free) which is Codazzi-coupled to both g and L . If ∇ is torsion-free, then L is automatically integrable and ω is parallel, so in this case we will call $(\mathfrak{M}, g, L, \nabla)$ a Codazzi-(para-)Kähler manifold instead.

So an almost Codazzi-(para-)Kähler manifold is an almost (para-)Hermitian manifold with a specified nice affine connection. Such structure exists on all almost (para-)Hermitian manifolds (\mathfrak{M}, g, L) . In particular, one can take ∇ to be any (para-)Hermitian connection [12, 18], which satisfies

$$\nabla g = 0 \text{ and } \nabla L = 0.$$

In the like manner, any (para-)Kähler manifold is trivially Codazzi-(para-)Kähler, because one can always take its Levi-Civita connection to be the desired ∇ , turning the compatible triple into a compatible quadruple.

In a Codazzi-(para-)Kähler manifold, because of $\nabla\omega = 0$ which leads to $\nabla = \nabla^\dagger$ (Theorem 1), so $\nabla^* = \nabla^L$. Therefore, any Codazzi-(para-)Kähler manifold admits a pair (∇, ∇^C) of torsion-free connections, where ∇^C is called the *Codazzi dual* of ∇ :

$$\nabla^C = \nabla^* = \nabla^L.$$

Proposition 7 For any Codazzi-(para-)Kähler manifold, its Codazzi dual connection ∇^C satisfies:

- (i) $(\nabla_X^C L)Y = (\nabla_Y^C L)X$;
- (ii) $(\nabla_X^C g)(Y, Z) = (\nabla_Y^C g)(X, Z)$;
- (iii) $(\nabla_X^C \omega)(Y, Z) = 0$.

Introducing a family of α -connections for $\alpha \in \mathbb{R}$

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1+\alpha}{2}\nabla^C, \quad \text{with} \quad (\nabla^{(\alpha)})^C = \nabla^{(-\alpha)}.$$

Then, we can easily show

- (i) $(\nabla_X^{(\alpha)} L)Y = (\nabla_Y^{(\alpha)} L)X$;
- (ii) $(\nabla_X^{(\alpha)} g)(Y, Z) = (\nabla_Y^{(\alpha)} g)(X, Z)$;
- (iii) $(\nabla_X^{(\alpha)} \omega)(Y, Z) = 0$.

Remark 3 When $\alpha = 0$, this is the familiar case of Levi-Civita connection (which is also the Chern connection) on the (para-)Kähler manifold. We can see here that the entire family of α -connections are compatible with the same Codazzi-(para-)Kähler structure.

Let us now investigate how to enhance a statistical structure to Codazzi-(para-)Kähler structure. To this end, we have:

Theorem 3 *Let ∇ be a torsion-free connection on \mathfrak{M} , and $(\nabla^*, \nabla^\dagger, \nabla^L)$ are the transformations of ∇ induced by the compatible triple (g, ω, L) . Then (g, ω, L, ∇) forms a compatible quadruple if any two of the following three statements are true:*

- (i) ∇^* is torsion-free;
- (ii) ∇^\dagger is torsion-free;
- (iii) ∇^L is torsion-free.

The proof is rather straight-forward, invoking Proposition 3 and Lemma 1 which link Codazzi coupling condition to torsion preservation in conjugate and gauge transformations.

In this case, i.e., when any two of the above three statements are true, \mathfrak{M} is Codazzi-(para-)Kähler. Hence, Theorem 3 can be viewed as a characterization theorem for Codazzi-(para-)Kähler structure, in the same way that condition (i) above alone characterizes statistical structure (a la Lauritzen [21]). This provides the affirmative answer to the key question posed by our paper: A statistical structure $(\mathfrak{M}, g, \nabla)$ can be “enhanced” to a Codazzi-(para-)Kähler structure $(\mathfrak{M}, g, \omega, L, \nabla)$ by

1. supplying it with an L that is compatible with g and that is Codazzi coupled with ∇ ;
2. supplying it with an L that is compatible with g and such that ∇^L is torsion-free; or
3. supplying it with an ω such that $\nabla\omega = 0$.

To summarize, in relation to more familiar types of manifolds, a Codazzi-(para-)Kähler manifold is a (para-)Kähler manifold which is at the same time statistical; it is also a Fedosov (hence symplectic) manifold which is at the same time statistical.

3 Divergence Functions and (Para-)Kähler Structures

Roughly speaking, a divergence function provides a measure of “directed distance” between two probability distributions in a family parameterized by a manifold \mathfrak{M} . Starting from a (local) divergence function on \mathfrak{M} , there are standard techniques to

generate a statistical structure on the diagonal $\mathfrak{M}_\Delta := \{(x, y) \in \mathfrak{M} \times \mathfrak{M} : x = y\} \subset \mathfrak{M} \times \mathfrak{M}$ as well as a symplectic structure on $\mathfrak{M} \times \mathfrak{M}$. We will first review these techniques and then show that para-Kähler structures on $\mathfrak{M} \times \mathfrak{M}$ arise naturally in this setting. Kähler structures will also be discussed. In the end, we study the case where \mathfrak{M} is Kähler and the local divergence function is taken to be Calabi's diastatic function. Its very rich geometric structures can be built in this scenario.

3.1 Classical Divergence Functions and Statistical Structures

Definition 5 (*Classical divergence function*) Let \mathfrak{M} be a smooth manifold of dimension n . A *classical divergence function* is a non-negative smooth function $\mathcal{D} : \mathfrak{M} \times \mathfrak{M} \rightarrow \mathbb{R}_{\geq 0}$ satisfying the following conditions:

- (i) $\mathcal{D}(x, y) \geq 0$ for any $(x, y) \in \mathfrak{M} \times \mathfrak{M}$, with equality holds if and only if $x = y$;
- (ii) The diagonal $\mathfrak{M}_\Delta \subset \mathfrak{M} \times \mathfrak{M}$ is a critical submanifold of \mathfrak{M} with respect to \mathcal{D} , in other words, $\mathcal{D}_i(x, x) = \mathcal{D}_{j,i}(x, x) = 0$ for any $1 \leq i, j \leq n$;
- (iii) $-\mathcal{D}_{i,j}(x, x)$ is positive definite at any $(x, x) \in \mathfrak{M}_\Delta$.

Here $\mathcal{D}_i(x, y) = \partial_{x^i} \mathcal{D}(x, y)$, $\mathcal{D}_{j,i}(x, y) = \partial_{y^j} \mathcal{D}(x, y)$, and $\mathcal{D}_{i,j}(x, y) = \partial_{x^i} \partial_{y^j} \mathcal{D}(x, y)$ and so on, where $\{x^i\}_{i=1}^n$ and $\{y^j\}_{j=1}^n$ are local coordinates of \mathfrak{M} near x and y respectively. When $x = y$, we further require that $\{x^i\}_{i=1}^n$ and $\{y^j\}_{j=1}^n$ give the same coordinates on \mathfrak{M} . Under such assumption, one can easily check that properties (i), (ii) and (iii) are independent of the choice of local coordinates. Note that \mathcal{D} does not have to satisfy $\mathcal{D}(x, y) = \mathcal{D}(y, x)$.

A standard example of classical divergence function is the Bregman divergence [3]. Given any smooth and strictly convex function $\Phi : \Omega \rightarrow \mathbb{R}$ on a closed convex set Ω , the Bregman divergence $\mathcal{B}_\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$ is defined by

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle x - y, \nabla \Phi(y) \rangle \quad (11)$$

where $\nabla \Phi$ is the usual gradient of Φ and $\langle \cdot, \cdot \rangle$ denotes the standard inner product on \mathbb{R}^n . Generalizing Bregman's divergence, we have the following Φ -divergence for all $\alpha \in \mathbb{R}$ [35]:

$$\mathcal{D}_\Phi^{(\alpha)}(x, y) = \frac{4}{1 - \alpha^2} \left(\frac{1 - \alpha}{2} \Phi(x) + \frac{1 + \alpha}{2} \Phi(y) - \Phi \left(\frac{1 - \alpha}{2} x + \frac{1 + \alpha}{2} y \right) \right). \quad (12)$$

It is known [7] that a statistical structure on \mathfrak{M}_Δ can be induced from a classical divergence function \mathcal{D} . Consider the Taylor expansion of \mathcal{D} along \mathfrak{M}_Δ , we obtain:

- (i) (2nd order): a Riemannian metric g

$$g_{ij}(x) = -\mathcal{D}_{i,j}(x, x) = \mathcal{D}_{ij}(x, x) = -\mathcal{D}_{j,i}(x, x).$$

(ii) (3rd order): a pair of conjugate connections

$$\Gamma_{ij,k}(x) = -\mathcal{D}_{ij,k}(x, x), \quad \Gamma_{ij,k}^*(x) = -\mathcal{D}_{k,ij}(x, x).$$

One can verify that the definitions of g , ∇ and ∇^* are independent of the choice of coordinates and indeed ∇^* is the g -conjugate of ∇ , i.e.,

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}^*.$$

Moreover, ∇ is torsion-free and (∇, g) is Codazzi-coupled, so we obtain a statistical structure on \mathfrak{M}_Δ .

As an example, from the Φ -divergence $\mathcal{D}_\phi^{(\alpha)}(x, y)$, we get the α -Hessian structure on \mathfrak{M} (see [35]) consisting of

$$g_{ij}(x) = \Phi_{ij}(x)$$

and

$$\Gamma_{ij,k}^{(\alpha)}(x) = \frac{1-\alpha}{2} \Phi_{ijk}(x), \quad \Gamma_{ij,k}^{*(\alpha)}(x) = \frac{1+\alpha}{2} \Phi_{ijk}(x).$$

3.2 Generalized Divergence Functions and Symplectic Structures

In this subsection, we will use a slightly different notion of divergence functions.

Definition 6 (*Generalized divergence function*) Let \mathfrak{M} be a smooth manifold of dimension n . A *generalized divergence function* is a smooth function $\mathcal{D} : \mathfrak{M} \times \mathfrak{M} \rightarrow \mathbb{R}$ satisfying the following conditions:

- (i) The diagonal $\mathfrak{M}_\Delta \subset \mathfrak{M} \times \mathfrak{M}$ is a critical submanifold of \mathfrak{M} with respect to \mathcal{D} ; in other words, $\mathcal{D}_i(x, x) = \mathcal{D}_{,j}(x, x) = 0$ for any $1 \leq i, j \leq n$;
- (ii) $\mathcal{D}_{i,j}(x, y)$ is a nondegenerate matrix at any point $(x, y) \in \mathfrak{M} \times \mathfrak{M}$.

Once again, $\{x^i\}_{i=1}^n$ and $\{y^j\}_{j=1}^n$ are arbitrary local coordinates of \mathfrak{M} near x and y respectively. It is obvious that this definition does not rely on the choice of local coordinates. Using the same ingredient as before, we can cook up a metric and a pair of conjugate torsion-free connections on \mathfrak{M}_Δ . However this metric may be indefinite.

Barndorff-Nielsen and Jupp [2] associated a symplectic form on $\mathfrak{M} \times \mathfrak{M}$ with \mathcal{D} (called “yoke” there), defined as (apart from a minus sign added here)

$$\omega_{\mathcal{D}}(x, y) = -\mathcal{D}_{i,j}(x, y) dx^i \wedge dy^j. \quad (13)$$

In particular, Bregman divergence \mathcal{B}_Φ (which fulfills the definition of a generalized divergence function) induces the symplectic form $\sum \Phi_{ij} dx^i \wedge dy^j$.

Such a construction essentially treated the divergence function as the Type II generating function of the symplectic structure on $\mathfrak{M} \times \mathfrak{M}$, see [22]. Let us consider the map $L_{\mathcal{D}} : \mathfrak{M} \times \mathfrak{M} \rightarrow T^*\mathfrak{M}$ given by

$$(x, y) \mapsto (x, d(\mathcal{D}(\cdot, y))(x)) = \left(x, \sum_i \mathcal{D}_i(x, y) dx^i \right).$$

Given y , we think of $\mathcal{D}(\cdot, y)$ as a smooth function of $x \in \mathfrak{M}$ and $d(\mathcal{D}(\cdot, y))(x)$ is nothing but the value of its differential at point x .

Recall that $T^*\mathfrak{M}$ admits a canonical symplectic form ω_{can} . A local calculation shows that

$$\omega_{\mathcal{D}} = -L_{\mathcal{D}}^* \omega_{\text{can}}.$$

In addition, it is not hard to see that condition (ii) in Definition 6 of a \mathcal{D} is equivalent to that $L_{\mathcal{D}}$ is a local diffeomorphism. Therefore $\omega_{\mathcal{D}}$ is indeed a symplectic form on $\mathfrak{M} \times \mathfrak{M}$. Similarly we can consider the map $R_{\mathcal{D}} : \mathfrak{M} \times \mathfrak{M} \rightarrow T^*\mathfrak{M}$ given by

$$(x, y) \mapsto (y, d(\mathcal{D}(x, \cdot))(y)) = \left(y, \sum_j \mathcal{D}_{,j}(x, y) dy^j \right).$$

In the same manner, we see that

$$\omega_{\mathcal{D}} = R_{\mathcal{D}}^* \omega_{\text{can}}.$$

Let $\mathfrak{M}_x = \{x\} \times \mathfrak{M}$ and $\mathfrak{M}_y = \mathfrak{M} \times \{y\}$. From the expression (13), we see immediately that \mathfrak{M}_x , \mathfrak{M}_y and \mathfrak{M}_{Δ} are Lagrangian submanifolds of $(\mathfrak{M} \times \mathfrak{M}, \omega_{\mathcal{D}})$.

3.3 Para-Kähler Structure on $\mathfrak{M} \times \mathfrak{M}$

Let M be a smooth manifold and $\mathcal{D} : \mathfrak{M} \times \mathfrak{M} \rightarrow \mathbb{R}$ be a generalized divergence function per Definition 6. From (13), \mathcal{D} induces a symplectic form $\omega_{\mathcal{D}}$ on $\mathfrak{M} \times \mathfrak{M}$. Actually, this symplectic form comes from a natural para-Kähler structure on $\mathfrak{M} \times \mathfrak{M}$ as we show below.

Let $(x, y) \in \mathfrak{M} \times \mathfrak{M}$ be an arbitrary point. Using the canonical identification

$$T_{(x,y)}(\mathfrak{M} \times \mathfrak{M}) = T_x\mathfrak{M} \oplus T_y\mathfrak{M},$$

we can produce an almost para-complex structure K on $\mathfrak{M} \times \mathfrak{M}$ by assigning

$$K_{(x,y)} = \text{id} \text{ on } T_x\mathfrak{M} \oplus 0 \quad \text{and} \quad K_{(x,y)} = -\text{id} \text{ on } 0 \oplus T_y\mathfrak{M}.$$

It is clear from this definition that K is integrable. Moreover, it can be checked that K and $\omega_{\mathcal{D}}$ are compatible in the sense that

$$\omega_{\mathcal{D}}(KX, KY) = -\omega_{\mathcal{D}}(X, Y). \quad (14)$$

Therefore the associated metric $g(X, Y) = \omega(KX, Y)$ is also compatible with K and we get a para-Kähler structure on $\mathfrak{M} \times \mathfrak{M}$.

Now let E_1 and E_{-1} be the eigen-distributions of K of eigenvalues 1 and -1 respectively. We see instantly that they are Lagrangian foliations with leaves \mathfrak{M}_x 's and \mathfrak{M}_y 's respectively.

Note that (14) does not impose any restriction on the form of the generalized divergence function \mathcal{D} . So we have the following structure theorem of manifolds admitting a generalized divergence function.

Theorem 4 *Let \mathfrak{M} be a smooth manifold admitting a generalized divergence function \mathcal{D} . Then \mathfrak{M} must be orientable, non-compact and parallelizable. Moreover, M supports an affine structure, i.e., there exists a torsion-free flat connection on \mathfrak{M} .*

Proof Assuming the existence of \mathcal{D} , we can produce a symplectic form $\omega_{\mathcal{D}}$ on $\mathfrak{M} \times \mathfrak{M}$ as in the last subsection. Therefore $\mathfrak{M} \times \mathfrak{M}$ is orientable and so is \mathfrak{M} . Moreover, $\omega_{\mathcal{D}}$ is an exact symplectic form since it is pull-back of an exact symplectic form (the canonical symplectic form on a cotangent bundle), therefore M cannot be compact.

As $E_{\pm 1}$ are Lagrangian foliations, it follows from Weinstein's result [34] that \mathfrak{M} , diffeomorphic to a leaf of a Lagrangian foliation, is affine. Indeed, such torsion-free flat connections can be construct explicitly on \mathfrak{M} . Let ∇^{LC} be the Levi-Civita connection associated to the para-Kähler structure $(\mathfrak{M} \times \mathfrak{M}, K, g)$. A straightforward calculation (see [16] and [32, Proposition 3.2]) shows that the connections induced by ∇^{LC} on leaves of $E_{\pm 1}$ are flat. Therefore, by identifying \mathfrak{M} with $\mathfrak{M}_x(\mathfrak{M}_y)$ for varying $x(y)$, we actually obtain two families of affine structure on \mathfrak{M} parameterized by \mathfrak{M} itself.

Finally, as $T_x \mathfrak{M}$ and $T_y \mathfrak{M}$ are Lagrangian subspaces of $(T_{(x,y)}(\mathfrak{M} \times \mathfrak{M}), \omega_{\mathcal{D}})$, we obtain an isomorphism

$$T_x \mathfrak{M} \cong (T_y \mathfrak{M})^*$$

using $\omega_{\mathcal{D}}$. If we fix $y = y_0$, then we get a smooth identification

$$T_x \mathfrak{M} \cong (T_{y_0} \mathfrak{M})^* \cong T_{x'} \mathfrak{M}$$

for any $x, x' \in \mathfrak{M}$, which parallelize $T \mathfrak{M}$.

Remark 4 The signature (n, n) of the pseudo-Riemannian metric g on $\mathfrak{M} \times \mathfrak{M}$ can be written down explicitly as

$$g = -\mathcal{D}_{i,j} dx^i \otimes dy^j.$$

Therefore the induced metric on \mathfrak{M}_Δ agrees with the metric constructed by 2nd order expansion of \mathcal{D} in Sect. 3.1. However in general, the pair of conjugate connections ∇ and ∇^* on \mathfrak{M}_Δ constructed from 3rd order expansion are distinct, therefore they do not coincide with ∇^{LC} associated to g .

In fact, we can give a full characterization of manifold with generalized divergence functions.

Theorem 5 *An n -dimensional manifold \mathfrak{M} admits a generalized divergence function \mathcal{D} if and only if M can be immersed into \mathbb{R}^n .*

Proof Fix a point $y_0 \in \mathfrak{M}$ and linear independent tangent vectors $v_1, \dots, v_n \in T_{y_0}\mathfrak{M}$. If \mathfrak{M} admits a generalized divergence function \mathcal{D} , we can consider the map $f : \mathfrak{M} \rightarrow \mathbb{R}^n$ given by

$$f(x) = (v_1\mathcal{D}(x, y_0), \dots, v_n\mathcal{D}(x, y_0)).$$

Then by the nondegeneracy condition of \mathcal{D} , we know that f has invertible Jacobian, hence it is an immersion.

On the other hand, $\mathcal{D}_0 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ given by $\mathcal{D}_0(x, y) = x \cdot y$ is a generalized divergence function on \mathbb{R}^n . If \mathfrak{M} can be immersed into \mathbb{R}^n , then we can pull-back \mathcal{D}_0 to get a generalized divergence function on \mathfrak{M} .

Remark 5 All the results of Theorem 4 follow trivially from Theorem 5. However, we state it independently because the constructions in the proof of Theorem 4 are canonical. It should also be noted that the condition that \mathfrak{M} can be immersed into \mathbb{R}^n is a much weaker than that \mathfrak{M} can be imbedded as an open subset of \mathbb{R}^n . For example, if we let $\mathfrak{M} = (S^2 \times S^1) \setminus \{\text{pt}\}$, then \mathfrak{M} can be immersed into \mathbb{R}^3 but not imbedded into it.

Para-complex manifolds have very rich geometric structures. For instance, one can recognize various Dirac structures [5] on them. Let \mathfrak{N} be a smooth manifold of dimension n . Following Courant, we define a Dirac structure on \mathfrak{N} as a rank n subbundle of $T\mathfrak{N} \oplus T^*\mathfrak{N}$ which is closed under the Courant bracket $[\cdot, \cdot]_C$:

$$[X \oplus \xi, Y \oplus \eta]_C = [X, Y] \oplus (\mathcal{L}_X \eta - \mathcal{L}_Y \xi - \frac{1}{2}d(\iota_X \eta - \iota_Y \xi))$$

for any smooth vector fields X, Y and 1-forms ξ, η , where ι is the interior product and \mathcal{L} is the Lie derivative. If (\mathfrak{N}, K) is a para-complex manifold, then we have the decomposition

$$T\mathfrak{N} = E_1 \oplus E_{-1},$$

where $E_{\pm 1}$ are eigen-distributions of eigenvalue 1 and -1 with respect to K . This splitting induces the decomposition for cotangent bundle:

$$T^*\mathfrak{N} = E_1^* \oplus E_{-1}^*.$$

It is not hard to check that $D_{\pm 1} = E_{\pm 1} \oplus E_{\mp 1}^*$ define two transversal Dirac structures on \mathfrak{N} . In particular, we obtain such structures on $\mathfrak{M} \times \mathfrak{M}$ if \mathfrak{M} admits a generalized divergence function. It is of great interest to understand the statistical interpretation of Courant bracket, as well as Dirac structures. We also refer to [33] for a general discussion of para-complex manifolds and Dirac structures.

3.4 Local Divergence Functions and Kähler Structures

A natural question to ask is whether one can construct a Kähler structure on $\mathfrak{M} \times \mathfrak{M}$ from a divergence function on \mathfrak{M} . The first problem one has to solve is to construct a complex structure J on $\mathfrak{M} \times \mathfrak{M}$. Unlike the para-complex case, there seems to be no canonical choice of J . So instead, we only consider a local version of this problem, i.e., constructing a Kähler structure on a neighborhood of \mathfrak{M}_Δ inside $\mathfrak{M} \times \mathfrak{M}$.

Definition 7 (*Local divergence function*) Let \mathfrak{M} be a smooth manifold of dimension n . A *local divergence function* is a nonnegative smooth function \mathcal{D} defined on an open neighborhood U of \mathfrak{M}_Δ inside $\mathfrak{M} \times \mathfrak{M}$ such that

- (i) $\mathcal{D}(x, y) \geq 0$ for any $(x, y) \in U$, with equality holds if and only if $x = y$;
- (ii) The diagonal \mathfrak{M}_Δ is a critical submanifold of \mathfrak{M} with respect to \mathcal{D} , in other words, $\mathcal{D}_i(x, x) = \mathcal{D}_{,j}(x, x) = 0$ for any $1 \leq i, j \leq n$;
- (iii) $-\mathcal{D}_{i,j}(x, x)$ is positive definite at any $(x, x) \in \mathfrak{M}_\Delta$.

It is obvious from this definition that classical divergence functions are local divergence functions. On the other hand, by a partition of unity argument, one can always extend a local divergence function to a classical divergence function. And moreover, local divergence is indeed a local version of divergence function we defined in Sect. 3.2.

To define a complex structure on a neighborhood of \mathfrak{M}_Δ , let us assume that \mathfrak{M} is an affine manifold, i.e., there exists a coordinate cover of \mathfrak{M} such that coordinate transformations are affine transformations. Let $\{U_\alpha\}_\alpha$ be the set of affine coordinate charts on \mathfrak{M} . Then

$$U = \bigcup_{\alpha} U_\alpha \times U_\alpha \subset \mathfrak{M} \times \mathfrak{M}$$

is an open neighborhood of \mathfrak{M}_Δ . We can define a complex structure J on U as follows. For any point $(x, y) \in U_\alpha \times U_\alpha \subset U$, we define J by assigning

$$J \frac{\partial}{\partial x^i} = \frac{\partial}{\partial y^i} \quad \text{and} \quad J \frac{\partial}{\partial y^i} = -\frac{\partial}{\partial x^i}$$

for $1 \leq i \leq n$, here $\{x^i\}_{i=1}^n, \{y^i\}_{i=1}^n$ are two copies of the same coordinates on U_α . As a consequence of \mathfrak{M} being affine, J does not depend on the choice of U_α . Furthermore, J is integrable since we may use $z^j = x^j + iy^j$ as holomorphic coordinates. However in general, J cannot be extended to a complex structure on $\mathfrak{M} \times \mathfrak{M}$.

Analogous to the para-Kähler case (14), we would like to have the compatibility condition between $\omega_{\mathcal{D}}$ and J

$$\omega_{\mathcal{D}}(JX, JY) = \omega_{\mathcal{D}}(X, Y).$$

As $\omega_{\mathcal{D}}$ is induced by the generalized divergence function \mathcal{D} via (13), the above condition does impose a restriction on the generalized divergence function \mathcal{D}

$$\mathcal{D}_{i,j} = \mathcal{D}_{j,i}$$

or explicitly

$$\frac{\partial^2 \mathcal{D}}{\partial x^i \partial y^j} = \frac{\partial^2 \mathcal{D}}{\partial y^i \partial x^j}.$$

We call such divergence functions “proper”. This condition was first derived in Zhang and Li [37]. As an example, the Φ -divergence given in (12) satisfies this condition of properness.

Now let us take the local proper divergence function \mathcal{D} into account. Using \mathcal{D} as a Kähler potential, we obtain

$$i\partial\bar{\partial}\mathcal{D} = \frac{i}{4}(\mathcal{D}_{jk} + \mathcal{D}_{,jk} + i\mathcal{D}_{j,k} - i\mathcal{D}_{k,j})dz^j \wedge d\bar{z}^k = \frac{i}{4}(\mathcal{D}_{jk} + \mathcal{D}_{,jk})dz^j \wedge d\bar{z}^k.$$

When restricting to \mathfrak{M}_Δ , we see that

$$\mathcal{D}_{jk}(x, x) + \mathcal{D}_{,jk}(x, x) = -2\mathcal{D}_{j,k}(x, x)$$

form a positive definite matrix. Therefore in a sufficiently small open neighborhood U , the (1,1)-form $i\partial\bar{\partial}\mathcal{D}$ is Kähler and we obtain a Kähler structure on U whose restriction on \mathfrak{M}_Δ agrees with the original metric on \mathfrak{M} up to a scalar.

3.5 An Example: The Case of Analytic Kähler Manifold

When \mathfrak{M} itself is an analytic Kähler manifold, we have a canonical choice of local divergence function: the diastatic function defined by Calabi [4].

Let $(\mathfrak{M}, I_0, \Omega_0)$ be an analytic Kähler manifold, that is, \mathfrak{M} is a Kähler manifold with complex structure I_0 such that the Kähler metric Ω_0 is real analytic with respect to the natural analytic structure on \mathfrak{M} . Let $\bar{\mathfrak{M}}$ be the conjugate manifold of \mathfrak{M} . By this,

we mean a complex manifold related to \mathfrak{M} by a diffeomorphism mapping $p \in \mathfrak{M}$ onto a point $\bar{p} \in \bar{\mathfrak{M}}$, such that for each local holomorphic coordinate $\{z^1, \dots, z^n\}$ in a neighborhood V of p , there exists a local holomorphic coordinate $\{w^1, \dots, w^n\}$ in the image \bar{V} of V , satisfying

$$w^j(\bar{q}) = \overline{z^j(q)}, \text{ for } j = 1, \dots, k.$$

Abstractly, $\bar{\mathfrak{M}}$ is the complex manifold $(\mathfrak{M}, -I_0)$ with local holomorphic coordinates of specified as above.

Let Ψ be a Kähler potential of Ω_0 , that is, Ψ is a locally defined real-valued function such that $i\partial\bar{\partial}\Psi = \Omega_0$. In local coordinates on V , we have

$$\Omega_0 = i \frac{\partial^2 \Psi(z, \bar{z})}{\partial z^j \partial \bar{z}^k} dz^j \wedge d\bar{z}^k.$$

As by our assumption Ω_0 is real analytic, so is Ψ , therefore in a small enough neighborhood Ψ can be written as a convergent power series of z and \bar{z} . Think of \bar{z} as coordinates on $\bar{\mathfrak{M}}$, then using this power series expansion, Ψ is a local holomorphic function on $\mathfrak{M} \times \bar{\mathfrak{M}} \cong \mathfrak{M} \times \mathfrak{M}$ defined in a neighborhood U of diagonal \mathfrak{M}_Δ .

Calabi defined the diastatic function $\mathcal{D}_d : U \rightarrow \mathbb{R}$ by

$$\mathcal{D}_d(p, \bar{q}) = \Psi(p, \bar{p}) + \Psi(q, \bar{q}) - \Psi(p, \bar{q}) - \Psi(q, \bar{p}). \quad (15)$$

Using our language, Calabi essentially proved the following theorem:

Theorem 6 ([4, Proposition 1–5]) *The diastatic function \mathcal{D}_d defined by (15) does not depend on the choice of local holomorphic coordinate.*

In other words, \mathcal{D}_d is a local divergence function. Now we use \mathcal{D}_d to perform the constructions in previous sections.

In local coordinates, write $z^j = x^j + iy^j$ and $w^j = u^j - iv^j$. Due to the complex conjugation we need to identify $\bar{\mathfrak{M}}$ with $(\mathfrak{M}, -I_0)$, we have that $\{x^j, y^k\}_{j,k=1}^n$ and $\{u^j, v^k\}_{j,k=1}^n$ form two copies of identical coordinates. As

$$\mathcal{D}_d(x, y, u, v) = \Psi(z, \bar{z}) + \Psi(\bar{w}, w) - \Psi(z, w) - \Psi(\bar{w}, \bar{z}),$$

we can compute directly that

$$\begin{aligned} \frac{\partial^2 \mathcal{D}_d}{\partial x^j \partial u^k} &= -\frac{\partial^2 \Psi}{\partial z^j \partial \bar{z}^k}(z, w) - \frac{\partial^2 \Psi}{\partial z^k \partial \bar{z}^j}(\bar{w}, \bar{z}) = \frac{\partial^2 \mathcal{D}_d}{\partial y^j \partial v^k}, \\ \frac{\partial^2 \mathcal{D}_d}{\partial x^j \partial v^k} &= i \frac{\partial^2 \Psi}{\partial z^j \partial \bar{z}^k}(z, w) - i \frac{\partial^2 \Psi}{\partial z^k \partial \bar{z}^j}(\bar{w}, \bar{z}) = -\frac{\partial^2 \mathcal{D}_d}{\partial y^j \partial u^k}. \end{aligned}$$

Therefore the holomorphic symplectic form, as induced via (13), is given by

$$\begin{aligned}
\Omega &= (\Psi_{j\bar{k}}(z, w) + \Psi_{k\bar{j}}(\bar{w}, \bar{z}))(\mathrm{d}x^j \wedge \mathrm{d}u^k + \mathrm{d}y^j \wedge \mathrm{d}v^k) - i(\Psi_{j\bar{k}}(z, w) \\
&\quad - \Psi_{k\bar{j}}(\bar{w}, \bar{z}))(\mathrm{d}x^j \wedge \mathrm{d}v^k - \mathrm{d}y^j \wedge \mathrm{d}u^k) \\
&= \Psi_{j\bar{k}}(z, w)\mathrm{d}z^j \wedge \mathrm{d}w^k + \Psi_{k\bar{j}}(\bar{w}, \bar{z})\mathrm{d}\bar{z}^j \wedge \mathrm{d}\bar{w}^k \\
&= \Omega_{\mathbb{C}} + \overline{\Omega_{\mathbb{C}}},
\end{aligned}$$

where $\Omega_{\mathbb{C}} = \Psi_{j\bar{k}}(z, w)\mathrm{d}z^j \wedge \mathrm{d}w^k$ is a well-defined complex-valued 2-form.

There are two natural complex structures on $\mathfrak{M} \times \mathfrak{M} \cong \mathfrak{M} \times \overline{\mathfrak{M}}$, i.e., $J^+ := (I_0, I_0)$ and $J^- := (I_0, -I_0)$, whose holomorphic coordinates are given by $\{z^j, \bar{w}^k\}_{j,k=1}^n$ and $\{\bar{z}^j, w^k\}_{j,k=1}^n$ respectively.

It is clear from the above expression of Ω that Ω is a (1,1)-form with respect to J^+ , therefore (U, J^+, Ω) is a pseudo-Kähler manifold such that \mathfrak{M}_Δ is a Lagrangian submanifold. On the other hand, with respect to J^- , we see that $\Omega_{\mathbb{C}}$ is a holomorphic symplectic form whose restriction on \mathfrak{M}_Δ is the Kähler form Ω_0 on \mathfrak{M} up to a purely imaginary scalar. It was proved years ago independently by Kaledin [19] and Feix [9], using different methods, that U actually admits a hyperkähler metric.

Notice that J^+ commutes with J^- and $-J^+J^- = K$ is the para-complex structure we specified in Sect. 3.3. A manifold with such structures was used by [13] and many other places in string theory as “modified Calabi–Yau manifolds”, see [26] for more details.

If we further assume that \mathfrak{M} is also affine in the sense that the holomorphic coordinates on \mathfrak{M} change by affine transformations, then we can use the recipe in Sect. 3.4 to construct a complex structure J on U with Kähler metric $i\partial\bar{\partial}\mathcal{D}_d$. To be specific, $\{x^j + iu^j, y^k + iv^k\}_{j,k=1}^n$ gives local holomorphic coordinates on U with respect to J . It is straightforward to see that J^+ commutes with J while J^- anticommutes with J , which leads to a modified Calabi–Yau structure and a hypercomplex structure on U , respectively.

4 Discussions

Codazzi coupling is the cornerstone of affine differential geometry (e.g., [25, 30]), and in particular so for information geometry. In information geometry, the Riemannian metric g and a pair of torsion-free g -conjugate affine connections ∇, ∇^* are naturally induced by the so-called divergence (or “contrast”) function on a manifold \mathfrak{M} (see [1]). While a statistical structure is naturally induced on \mathfrak{M} , the divergence function will additionally induce a symplectic structure ω on the product manifold $\mathfrak{M} \times \mathfrak{M}$, see [2, 37]. Reference [31] appears to be the first to extend the definition of conjugate connection with respect to g to that with respect to ω . And [8] proved that the g -conjugate, ω -conjugate, L -gauge transformations of ∇ form a Klein group. Based on these, it is shown that Codazzi coupling of torsion-free ∇ with any two of the compatible triple (g, ω, L) implies its coupling with the remaining third, turning (g, ω, L, ∇) into a compatible quadruple and hence the manifold \mathfrak{M} into

a (para-)Kähler one. Therefore, our results here provide precise conditions under which a statistical manifold could be “enhanced” to a Kähler and/or para-Kähler manifold, and clarify some confusions in the literature regarding the roles of Codazzi coupling of ∇ with g and with L in the interactions between statistical structure (as generalized Riemannian structure), symplectic structure, and (para-)complex structure.

Codazzi-(para-)Kähler manifolds are generalizations of special Kähler manifolds by removing the requirement of ∇ to be (dually) flat in the latter. Special Kähler manifolds are first mathematically formulated by Freed [10], and they have been extensively studied in physics literature since 1980s. For example, special Kähler structures are found on the base of algebraic integrable systems [6] and moduli space of complex Lagrangian submanifolds in a hyperkähler manifold [17]. From the above discussions, we can view special Kähler manifolds as “enhanced” from the class of dually-flat statistical manifold, namely, Hessian manifolds [29]. In information geometry, non-flat affine connections are abundant – the family of $\nabla^{(\alpha)}$ connections associated with a pair of dually-flat connections ∇, ∇^* are non-flat except $\alpha = \pm 1$ [36]. So our generalization of special Kähler geometry to Codazzi-Kähler geometry, which shifts attention from curvature to torsion, may be meaningful for the investigation of bidualistic geometric structures in statistical and information sciences [35].

Acknowledgements This collaborative research started while the first author (J.Z.) was on sabbatical visit at the Center for Mathematical Sciences and Applications at Harvard University in the Fall of 2014 under the auspices of Prof. S.-T. Yau. The writing of this paper is supported by DARPA/ARO Grant W911NF-16-1-0383 (PI: Jun Zhang).

References

1. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, vol. 191. AMS, Providence (2000)
2. Barndorff-Nielsen, O.E., Jupp, P.E.: Yokes and symplectic structures. *J. Stat. Plan. Inference* **63**(2), 133–146 (1997)
3. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **7**(3), 200–217 (1967)
4. Calabi, E.: Isometric imbedding of complex manifolds. *Ann. Math.* **58**(1), 1–23 (1953)
5. Courant, T.J.: Dirac manifolds. *Trans. Am. Math. Soc.* **319**(2), 631–661 (1990)
6. Donagi, R., Witten, E.: Supersymmetric Yang–Mills theory and integrable systems. *Nucl. Phys. B* **460**(2), 299–334 (1996)
7. Eguchi, S.: Geometry of minimum contrast. *Hiroshima Math. J.* **22**(3), 631–47 (1992)
8. Fei, T., Zhang, J.: Interaction of Codazzi coupling and (para)-Kähler geometry. *Results in Mathematics* (2017). (in print)
9. Feix, B.: Hyperkähler metrics on cotangent bundles. *J. für die reine und angewandte Math.* **532**, 33–46 (2001)
10. Freed, D.S.: Special Kähler manifolds. *Commun. Math. Phys.* **203**(1), 31–52 (1999)
11. Furuhata, H.: Hypersurfaces in statistical manifolds. *Differ. Geom. Appl.* **27**(3), 420–429 (2009)

12. Gauduchon, P.: Hermitian connections and Dirac operators. *Bollettino della Unione Matematica Italiana-B* **11**(2, Suppl.), 257–288 (1997)
13. Gates, S.J., Hull, C.M., Rocek, M.: Twisted multiplets and new supersymmetric non-linear σ -models. *Nucl. Phys. B* **248**(1), 157–186 (1984)
14. Gelfand, I., Retakh, V., Shubin, M.: Fedosov manifolds. *Adv. Math.* **136**(1), 104–140 (1998)
15. Grigorian, S., Zhang, J.: (Para-)Holomorphic connections for information geometry. *Geometric Science of Information*. Springer International Publishing, Berlin
16. Hess, H.: Connections on symplectic manifolds and geometric quantization. *Differential Geometrical Methods in Mathematical Physics*. Lecture Notes in Mathematics, vol. 836, pp. 153–166. Springer, Berlin (1980)
17. Hitchin, N.J.: The moduli space of complex Lagrangian submanifolds. *Surveys in Differential Geometry VII*, pp. 327–345. International Press, Somerville (2000)
18. Ivanov, S., Zamkovoy, S.: Parahermitian and paraquaternionic manifolds. *Differ. Geom. Appl.* **23**(2), 205–234 (2005)
19. Kaledin, D.B.: Hyperkähler metrics on total spaces of cotangent bundles. *Quaternionic Structures in Mathematics and Physics* (Rome, 1999), pp. 195–230. International Press, Somerville (1999)
20. Kurose, T.: Dual connections and affine geometry. *Math. Z.* **203**(1), 115–121 (1990)
21. Lauritzen, S.L.: Statistical manifolds. *Differential Geometry in Statistical Inference*. IMS Lecture Notes Monograph Series, vol. 10, pp. 163–216. Institute of Mathematical Statistics, Hayward (1987)
22. Leok, M., Zhang, J.: Connecting information geometry and geometric mechanics. *Entropy* **19**(10), 518 (2017)
23. Moroianu, A.: Lectures on Kähler Geometry. London Mathematical Society Student Texts, vol. 69. Cambridge University Press, Cambridge (2007)
24. Noda, T.: Symplectic structures on statistical manifolds. *J. Aust. Math. Soc.* **90**(3), 371–384 (2011)
25. Nomizu, K., Sasaki, T.: *Affine Differential Geometry: Geometry of Affine Immersions*. Cambridge Tracts in Mathematics, vol. 111. Cambridge University Press, Cambridge (1994)
26. Rocek, M.: Modified Calabi–Yau manifolds with torsion. *Essays on Mirror Manifolds*, pp. 480–488. International Press, Hong Kong (1992)
27. Schwenk-Schellschmidt, A., Simon, U.: Codazzi-equivalent affine connections. *Results Math.* **56**(1–4), 211–229 (2009)
28. Schwenk-Schellschmidt, A., Simon, U., Wiehe, M.: Generating higher order Codazzi tensors by functions. *TU Fachbereich Mathematik* **3** (1998)
29. Shima, H.: On certain locally flat homogeneous manifolds of solvable Lie groups. *Osaka J. Math.* **13**(2), 213–229 (1976)
30. Simon, U.: Affine differential geometry. *Handbook of Differential Geometry*, vol. 1, pp. 905–961. North-Holland, Amsterdam (2000)
31. Tao, J., Zhang, J.: Transformations and coupling relations for affine connections. *Differ. Geom. Appl.* **49**, 111–130
32. Vaisman, I.: Symplectic curvature tensors. *Monatshefte für Math.* **100**(4), 299–327 (1985)
33. Wade, A.: Dirac structures and paracomplex manifolds. *Comptes Rendus Math.* **338**(11), 889–894 (2004)
34. Weinstein, A.D.: Symplectic manifolds and their Lagrangian submanifolds. *Adv. Math.* **6**(3), 329–346 (1971)
35. Zhang, J.: Divergence function, duality, and convex analysis. *Neural Comput.* **16**(1), 159–195 (2004)
36. Zhang, J.: A note on curvature of α -connections of a statistical manifold. *Ann. Inst. Stat. Math.* **59**(1), 161–170 (2007)
37. Zhang, J., Li, F.-B.: Symplectic and Kähler structures on statistical manifolds induced from divergence functions. *Geometric Science of Information*. Lecture Notes in Computer Science, vol. 8085, pp. 595–603. Springer, Berlin (2013)

Doubly Autoparallel Structure on the Probability Simplex



Atsumi Ohara and Hideyuki Ishi

Abstract On the probability simplex, we can consider the standard information geometric structure with the e- and m-affine connections mutually dual with respect to the Fisher metric. The geometry naturally defines submanifolds simultaneously autoparallel for the both affine connections, which we call *doubly autoparallel submanifolds*. In this note we discuss their several interesting common properties. Further, we algebraically characterize doubly autoparallel submanifolds on the probability simplex and give their classification.

Keywords Statistical manifold · Dual affine connections · Doubly autoparallel submanifolds · Mutation of Hadamard product

1 Introduction

Let us consider information geometric structure [1] (g, ∇, ∇^*) on a manifold \mathcal{M} , where g , ∇ , ∇^* are, respectively, a Riemannian metric and a pair of torsion-free affine connections satisfying

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z), \quad \forall X, Y, Z \in \mathcal{X}(\mathcal{M}).$$

Here, $\mathcal{X}(\mathcal{M})$ denotes the set of all vector fields on \mathcal{M} . Such a manifold with the structure (g, ∇, ∇^*) is called a *statistical manifold* and we say ∇ and ∇^* are *mutually dual* with respect to g . When curvature tensors of ∇ and ∇^* vanish, the statistical manifold

A. Ohara (✉)
University of Fukui, Fukui 910-8507, Japan
e-mail: ohara@fuee.u-fukui.ac.jp

H. Ishi
Nagoya University, Furo-cho, Nagoya 464-8602, Japan
e-mail: hideyuki@math.nagoya-u.ac.jp

H. Ishi
JST, PRESTO, 4-1-8, Honcho, Kawaguchi 332-0012, Japan

is said *dually flat*. For a statistical manifold, we can introduce a one-parameter family of affine connections called α -connection:

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^*, \quad \alpha \in \mathbf{R}.$$

It is seen that $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are mutually dual with respect to g .

In a statistical manifold, we can naturally define a submanifold \mathcal{N} that is simultaneously autoparallel with respect to both ∇ and ∇^* .

Definition 1 Let $(\mathcal{M}, g, \nabla, \nabla^*)$ be a statistical manifold and \mathcal{N} be its submanifold. We call \mathcal{N} *doubly autoparallel* in \mathcal{M} when the followings hold:

$$\nabla_X Y \in \mathcal{X}(\mathcal{N}), \quad \nabla_X^* Y \in \mathcal{X}(\mathcal{N}), \quad \forall X, Y \in \mathcal{X}(\mathcal{N}).$$

We immediately see that doubly autoparallel submanifolds \mathcal{N} possess the following properties. (Note that the statement (5) holds if $(\mathcal{M}, g, \nabla, \nabla^*)$ is dually flat.) They are somehow complemented in Appendix A.

Proposition 1 *The following statements are equivalent:*

- (1) *a submanifold \mathcal{N} is doubly autoparallel (DA),*
- (2) *a submanifold \mathcal{N} is autoparallel w.r.t. to $\nabla^{(\alpha)}$ for two different α 's,*
- (3) *a submanifold \mathcal{N} is autoparallel w.r.t. to $\nabla^{(\alpha)}$ for all α 's,*
- (4) *every α -geodesic curve passing through an arbitrary point p in \mathcal{N} that is tangent to \mathcal{N} at p lies in \mathcal{N} for all α 's,*
- (5) *a submanifold \mathcal{N} is affinely constrained in both ∇ - and ∇^* -affine coordinates of \mathcal{M} .*

In particular, let \mathcal{M} be a parametric statistical model that is dually flat w.r.t. the Fisher metric g , the exponential connection $\nabla = \nabla^{(e)}$ and the mixture connection $\nabla^ = \nabla^{(m)}$ [1]. If \mathcal{N} is DA, the α -projections to \mathcal{N} are unique for all α 's.*

The concept of doubly autoparallelism has sometimes appeared but played important roles in several applications of information geometry [2–5]. For example, optimization problems on the positive semidefinite matrices, which is called semidefinite program, can be proved explicitly solvable if feasible regions for the problems are doubly autoparallel [2]. However, the literature mostly treat doubly autoparallel submanifolds in symmetric cones, and cases for statistical models have not been exploited yet.

In this note, we consider doubly autoparallel structure on the probability simplex, which can be identified with probability distributions on discrete and finite sample spaces. As a result, we give an algebraic characterization and classification of doubly autoparallel submanifolds in the probability simplex. It should be remarked that in several aspects of the approach here Hadamard product naturally appears, and hence, we can employ technique similar to the case of symmetric cones [5] where Jordan subalgebras play crucial roles.

Doubly autoparallel submanifolds commonly possess the above interesting properties. Hence, the obtained results might be expected to give a useful insight into constructing statistical models for wide area of applications in information science, mathematics and statistical physics and so on [6–8]. Actually, Nagaoka has recently reported the significance of this concept in study of statistical models [9]. He has characterized models statistically equivalent to the probability simplex by doubly autoparallelism. Further, it should be also mentioned that Matúš and Ay have studied the corresponding statistical models based on motivations from learning theory [7]. They have reached characterizations different from ours, using partitions of the sample space [10].

We demonstrate results for the case of the probability simplex in this note and a study for general statistical manifolds is left in the future work.

2 Preliminaries

2.1 Information geometry of \mathcal{S}^n and \mathbf{R}_+^{n+1}

Let us represent an element $p \in \mathbf{R}^{n+1}$ with its components p_i , $i = 1, \dots, n + 1$ as $p = (p_i) \in \mathbf{R}^{n+1}$. Denote, respectively, the positive orthant by

$$\mathbf{R}_+^{n+1} := \{p = (p_i) \in \mathbf{R}^{n+1} | p_i > 0, i = 1, \dots, n + 1\},$$

and the relative interior of the probability simplex by

$$\mathcal{S}^n := \left\{ p \in \mathbf{R}_+^{n+1} \left| \sum_{i=1}^{n+1} p_i = 1 \right. \right\}.$$

For a subset $\mathcal{Q} \subset \mathbf{R}_+^{n+1}$ and an element $p \in \mathcal{Q}$, we simply write

$$\log \mathcal{Q} := \{\log p | p \in \mathcal{Q}\}, \quad \log p := (\log p_i) \in \mathbf{R}^{n+1}.$$

Each element p in the closure of \mathcal{S}^n denoted by $\text{cl}\mathcal{S}^n$ can be identified with a discrete probability distribution for the sample space $\Omega = \{1, 2, \dots, n, n + 1\}$. However, we only consider distributions $p(X)$ with positive probabilities, i.e., $p(i) = p_i > 0$, $i = 1, \dots, n + 1$, defined by

$$p(X) = \sum_{i=1}^{n+1} p_i \delta_i(X), \quad \delta_i(j) = \delta_i^j \text{ (the Kronecker's delta)},$$

which is identified with \mathcal{S}^n .

A statistical model in \mathcal{S}^n is represented with parameters $\xi = (\xi_j)$, $j = 1, \dots, d \leq n$ by

$$p(X; \xi) = \sum_{i=1}^{n+1} p_i(\xi) \delta_i(X),$$

where each p_i is a function of ξ . For example, $p_i = \xi_i$, $i = 1, \dots, n$ with the condition $\sum_{i=1}^n \xi_i < 1$ is the full model, i.e.,

$$p(X; \xi) = \sum_{i=1}^n \xi_i \delta_i(X) + \left(1 - \sum_{i=1}^n \xi_i\right) \delta_{n+1}(X)$$

For the submodel, ξ^j , $j = 1, \dots, d < n$ can be also regarded as coordinates of the corresponding submanifold in \mathcal{S}^n .

The standard information geometric structure on \mathcal{S}^n [1] denoted by $(g, \nabla^{(e)}, \nabla^{(m)})$ are composed of the pair of flat affine connections $\nabla^{(e)}$ and $\nabla^{(m)}$. The affine connections $\nabla^{(e)} = \nabla^{(1)}$ and $\nabla^{(m)} = \nabla^{(-1)}$ are respectively called the *exponential connection* and the *mixture connection*. They are mutually dual with respect to the Fisher metric g .

By writing $\partial_i := \partial/\partial\xi_i$, $i = 1, \dots, n$, they are explicitly represented as follows:

$$g_{ij}(p) = \sum_{X \in \Omega} p(X)(\partial_i \log p(X))(\partial_j \log p(X)), \quad i, j = 1, \dots, n,$$

$$\Gamma_{ij,k}^{(m)}(p) = \sum_{X \in \Omega} p(X)(\partial_i \partial_j p(X))(\partial_k \log p(X)) \quad i, j, k = 1, \dots, n, \quad (1)$$

$$\Gamma_{ij,k}^{(e)}(p) = \sum_{X \in \Omega} p(X)(\partial_i \partial_j \log p(X))(\partial_k p(X)), \quad i, j, k = 1, \dots, n. \quad (2)$$

There exist two special coordinate systems. The one is the *expectation coordinate* $\eta_i := \sum_{X \in \Omega} p(X)\delta_i(X) = p_i$, $i = 1, \dots, n$, which is $\nabla^{(m)}$ -affine from (1). It implies that if each η_i is an affine function of all the model parameters ξ_i 's, then the statistical model is $\nabla^{(m)}$ -*autoparallel* (or sometimes called *m-flat*).

The other is the *canonical coordinate* θ^i , which is defined by

$$\theta^i := \log \left(\frac{p_i}{1 - \sum_{i=1}^n p_i} \right), \quad i = 1, \dots, n. \quad (3)$$

Since θ^i 's satisfy

$$p(X) = \exp \left\{ \sum_{i=1}^n \theta^i \delta_i(X) - \psi(\theta) \right\}, \quad \psi(\theta) := \log \left(1 + \sum_{i=1}^n \exp \theta^i \right),$$

they are $\nabla^{(e)}$ -affine from (2). Hence, it implies that if each θ^i is an affine function of all the model parameters ξ_i 's, then the statistical model is $\nabla^{(e)}$ -*autoparallel* (or sometimes called *e-flat*).

Note that from the property of the expectation coordinates, a $\nabla^{(e)}$ -autoparallel submanifold in \mathcal{S}^n , denoted by M , should be represented by $M = W \cap \mathcal{S}^n$ for a certain subspace $W \subset \mathbf{R}_+^{n+1}$. This fact is used later.

Finally, we introduce information geometric structure $(\tilde{g}, \tilde{\nabla}^{(e)}, \tilde{\nabla}^{(m)})$ on \mathbf{R}_+^{n+1} . The structure $(g, \nabla^{(e)}, \nabla^{(m)})$ on \mathcal{S}^n is a submanifold geometry induced from this ambient structure. For arbitrary coordinates $\tilde{\xi}_i$, $i = 1, \dots, n+1$ of \mathbf{R}_+^{n+1} , let us take $\tilde{\partial}_i := \partial/\partial \tilde{\xi}_i$. Then their components are given by

$$\tilde{g}_{ij}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \log p(X))(\tilde{\partial}_j \log p(X)), \quad i, j = 1, \dots, n+1,$$

$$\tilde{\Gamma}_{ij,k}^{(m)}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \tilde{\partial}_j p(X))(\tilde{\partial}_k \log p(X)), \quad i, j, k = 1, \dots, n+1, \quad (4)$$

$$\tilde{\Gamma}_{ij,k}^{(e)}(p) = \sum_{X \in \Omega} p(X)(\tilde{\partial}_i \tilde{\partial}_j \log p(X))(\tilde{\partial}_k p(X)), \quad i, j, k = 1, \dots, n+1. \quad (5)$$

Thus, we find that p_i 's are $\tilde{\nabla}^{(m)}$ -affine coordinates and $\log p_i$'s are $\tilde{\nabla}^{(e)}$ -affine coordinates, respectively, from (4), (5) and $\log p(X) = \sum_{X \in \Omega} (\log p_i) \delta_i(X)$.

2.2 An Example

Example Let $v^{(k)} = (\delta_i^k) \in \mathbf{R}^{n+1}$, $k = 1, \dots, n+1$ represent vertices on $\text{cl}\mathcal{S}^n$. Take a vector $v^{(0)} \in \mathbf{R}_+^{n+1}$ that is linearly independent of $\{v^{(k)}\}_{k=1}^d$ ($d < n$) and define a subspace of dimension $d+1$ by $W = \text{span}\{v^{(0)}, v^{(1)}, \dots, v^{(d)}\}$. Then $M = \mathcal{S}^n \cap W$ is doubly autoparallel.

We show this for the case $d = 2$ but similar arguments hold for general d . For the simplicity we take the following $v^{(i)}$, $i = 0, 1, 2$:

$$v^{(0)} = (0 \ 0 \ p_3 \ \dots \ p_{n+1})^T, \quad \sum_{i=3}^{n+1} p_i = 1, \quad p_i > 0, \quad i = 3, \dots, n+1,$$

$$v^{(1)} = (1 \ 0 \ \dots \ 0)^T, \quad v^{(2)} = (0 \ 1 \ 0 \ \dots \ 0)^T, \quad (\cdot^T \text{ denotes the transpose}).$$

Since for $p \in M$ we have a convex combination by parameters ξ_i as

$$p = \xi_1 v^{(1)} + \xi_2 v^{(2)} + (1 - \xi_1 - \xi_2) v^{(0)},$$

the expectation coordinates η_i 's are

$$\begin{aligned}\eta_1 &= \xi_1, \quad \eta_2 = \xi_2, \quad \eta_i = (1 - \xi_1 - \xi_2)p_i, \quad i = 3, \dots, n+1, \\ (\xi_1 &> 0, \quad \xi_2 > 0, \quad \xi_1 + \xi_2 < 1).\end{aligned}$$

Thus, each η_i is affine in ξ_i , $i = 1, 2$.

On the other hand, the canonical coordinates θ^i 's are

$$\begin{aligned}\theta^1 &= \zeta_1, \quad \theta^2 = \zeta_2, \quad \theta^i = \log p_i + c, \quad i = 3, \dots, n+1, \\ (\zeta_i &= \log\{\xi_i/(1 - \xi_1 - \xi_2)\}, \quad i = 1, 2, \quad c = -\log p_{n+1}).\end{aligned}$$

Thus, each θ^i is affine in parameters ζ_i , $i = 1, 2$. Hence, M is doubly autoparallel.

2.3 Denormalization

Definition 2 Let M be a submanifold in S^n . The submanifold \tilde{M} in \mathbf{R}_+^{n+1} defined by

$$\tilde{M} = \{\tau p \in \mathbf{R}_+^{n+1} \mid p \in M, \tau > 0\}$$

is called a *denormalization* of M [1].

Lemma 1 A submanifold M is $\nabla^{(\pm 1)}$ -autoparallel in S^n if and only if the denormalization \tilde{M} is $\tilde{\nabla}^{(\pm 1)}$ -autoparallel in \mathbf{R}_+^{n+1} .

A key observation derived from the above lemma is as follows:

Since p_i , $i = 1, \dots, n+1$ are $\nabla^{(m)}$ -affine coordinates for S^n , a submanifold $M \subset S^n$ is $\nabla^{(m)}$ -autoparallel if and only if it is represented as $M = W \cap S^n$ for a subspace $W \subset \mathbf{R}^{n+1}$. Hence, by definition \tilde{M} is nothing but

$$\tilde{M} = W \cap \mathbf{R}_+^{n+1}. \quad (6)$$

On the other hand, since the coordinates $\log p_i$, $i = 1, \dots, n+1$ for \mathbf{R}^{n+1} are $\tilde{\nabla}^{(e)}$ -affine, \tilde{M} is $\tilde{\nabla}^{(e)}$ -autoparallel if and only if there exist a subspace $V \subset \mathbf{R}^{n+1}$ and a constant element $b \in \mathbf{R}^{n+1}$ satisfying

$$\log \tilde{M} = b + V, \quad (7)$$

where $\dim W = \dim V$. If so, M is also $\nabla^{(e)}$ -autoparallel from Lemma 1.

Thus, we study conditions for the denormalization \tilde{M} to have simultaneously dualistic representations (6) and (7), which is equivalent to doubly autoparallelism of M .

3 Main Results

First we introduce an algebra $(\mathbf{R}^{n+1}, \circ)$ via the Hadamard product \circ , i.e.,

$$x \circ y = (x_i) \circ (y_i) := (x_i y_i), \quad x, y \in \mathbf{R}^{n+1},$$

where the identity element e and an inverse x^{-1} are

$$e = \mathbf{1}, \quad x^{-1} = \left(\frac{1}{x_i} \right),$$

respectively. Here, $\mathbf{1} \in \mathbf{R}_+^{n+1}$ is the element all the components of which are one. Note that the set of invertible elements

$$\mathcal{I} := \{x = (x_i) \in \mathbf{R}^{n+1} | x_i \neq 0, i = 1, \dots, n+1\}$$

contains \mathbf{R}_+^{n+1} . We simply write x^k for the powers recursively defined by $x^k = x \circ x^{k-1}$.

For an arbitrarily fixed $a \in \mathcal{I}$ the algebra $(\mathbf{R}^{n+1}, \circ)$ induces another algebra called a *mutation* $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$, the product of which is defined by

$$x \circ_{a^{-1}} y := x \circ a^{-1} \circ y = (x_i y_i / a_i), \quad x, y \in \mathbf{R}^{n+1},$$

with its identity element a . We write $x^{(\circ a^{-1})k}$ for the powers by $\circ_{a^{-1}}$.

We give a basic result in terms of $(\mathbf{R}^{n+1}, \circ)$.

Theorem 1 Assume that $a \in \tilde{M} = W \cap \mathbf{R}_+^{n+1}$. Then, there exists a subspace V satisfying

$$\log \tilde{M} = \log \{(a + W) \cap \mathbf{R}_+^{n+1}\} = \log a + V$$

if and only if the following two conditions hold:

$$(1) V = a^{-1} \circ W, \quad (2) \forall u, w \in W, \quad u \circ a^{-1} \circ w \in W.$$

Proof (“only if” part): For all $w \in W$ and small $t \in \mathbf{R}_+$, we have $\log(a + tw) \in \log a + V$. Hence, it holds that

$$\left. \frac{d}{dt} \log(a + tw) \right|_{t=0} = a^{-1} \circ w \in V.$$

Thus, the condition (1) holds.

Similarly, for all $u, w \in W$ and small $t \in \mathbf{R}_+$ and $s \in \mathbf{R}_+$, we have $\log(a + su + tw) \in \log a + V$ and obtain

$$\frac{\partial}{\partial s} \left(\frac{\partial}{\partial t} \log(a + su + tw) \Big|_{t=0} \right) \Big|_{s=0} = -a^{-1} \circ u \circ a^{-1} \circ w \in V.$$

Hence, we see that the condition (2) holds, using the condition (1).

(“if” part): For $w = (w_i) \in W$ satisfying $a + w \in \mathbf{R}_+^{n+1}$, take $t \in \mathbf{R}_+$ be larger than $(1 + \max_i \{w_i/a_i\})/2$. Then there exists $u = (u_i) \in W$ satisfying

$$a + w = ta + u, \quad ta_i > |u_i|, \quad i = 1, \dots, n+1. \quad (8)$$

Hence, we have

$$\begin{aligned} \log(a + w) &= \log(ta + u) = \log\{(ta) \circ (e + (ta)^{-1} \circ u)\} \\ &= (\log t)e + \log a + \log\{e + (ta)^{-1} \circ u\}. \end{aligned} \quad (9)$$

Using the inequalities in (8) and the Taylor series

$$\log(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k, \quad |x| < 1,$$

we expand the right-hand side of (9) as

$$(\log t)e + \log a + a^{-1} \circ \left(\frac{1}{t} \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{-1}{t} \right)^{k-1} u^{(\circ a^{-1})k} \right).$$

Since each $u^{(\circ a^{-1})k}$ belongs to W from the condition (2), the third term is in V by the condition (1). Further the condition (1) implies that $e \in V$, so is the first term. This completes the proof.

- Remark 1* (i) The condition (2) claims that W is a *subalgebra* of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$.
(ii) The affine subspace $\log a + V$ is independent of the choice of $a \in \tilde{M} = W \cap \mathbf{R}_+^{n+1}$. This follows from the proof of “if” part by taking $a' = a + w$.

The following algebraic characterization of doubly autoparallel submanifold in \mathcal{S}^n is immediate from the above theorem and Lemma 1 in Sect. 2.

Corollary 1 A $\nabla^{(m)}$ -autoparallel submanifold $M = W \cap \mathcal{S}^n$ is doubly autoparallel if and only if the subspace W is a subalgebra of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$ with $a \in \tilde{M}$.

Finally, in order to answer a natural question what structure is necessary and sufficient for W , we classify subalgebras of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$. Let q and r be integers that meet $q \geq 0$, $r > 0$ and $q + r = \dim W < n + 1$. Define integers n_l , $l = 1, \dots, r$ satisfying

$$q + \sum_{l=1}^r n_l = n + 1, \quad 2 \leq n_1 \leq \dots \leq n_r.$$

Constructing subvectors $a_l \in \mathbf{R}_+^{n_l}$, $l = 1, \dots, r$ with components arbitrarily extracted from $a \in W \cap \mathbf{R}_+^{n+1}$ without duplications, we denote by Π the permutation matrix that meets

$$(a_0^T \ a_1^T \ \dots \ a_r^T)^T = \Pi a, \quad (10)$$

where the subvector $a_0 \in \mathbf{R}^q$ is composed of the remaining components in a . We give the classification via the canonical form for $W_0 = \Pi W = \{w' \in \mathbf{R}^{n+1} | w' = \Pi w, w \in W\}$ based on this partition instead of the original form for W .

Theorem 2 *For the above setup, W is a subalgebra of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$ with $a \in \tilde{M}$ if and only if W is isomorphic to $\mathbf{R}^q \times \mathbf{R}a_1 \times \dots \times \mathbf{R}a_r$ and represented by $\Pi^{-1}W_0$, where*

$$W_0 = \{(y^T \ t_1 a_1^T \ \dots \ t_r a_r^T)^T \in \mathbf{R}^{n+1} | \forall y \in \mathbf{R}^q, \ a_l \in \mathbf{R}_+^{n_l}, \ \forall t_l \in \mathbf{R}, \ l = 1, \dots, r\}.$$

Proof (“only if” part) Let V be a subspace in \mathbf{R}^{n+1} defined by $V = a^{-1} \circ W$. Then it is straightforward that W is a subalgebra of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$ if and only if V is a subalgebra of $(\mathbf{R}^{n+1}, \circ)$. Using this equivalence, we consider the necessity condition.

Since e and x^k are in V for any $x = (x_i) \in V$ and positive integer k , the square matrix Ξ defined by

$$\Xi := (e \ x \ \dots \ x^n) = \begin{pmatrix} 1 & x_1 & \dots & x_1^n \\ 1 & x_2 & \dots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n+1} & \dots & x_{n+1}^n \end{pmatrix}$$

is singular. The determinant of the Vandermonde’s matrix Ξ is calculated using the well-known formula, as

$$\det \Xi = (-1)^{(n+1)n/2} \left(\prod_{i < j} (x_i - x_j) \right).$$

Hence, it is necessary for x to belong to V that

$$\exists(i, j), \ x_i = x_j. \quad (11)$$

Denoting basis vectors of V by $v^{(k)} = (v_i^{(k)}) \in \mathbf{R}^{n+1}$, $k = 1, \dots, q+r$ ($= \dim V$), we can represent any x as $x = \sum_{k=1}^{q+r} \alpha_k v^{(k)}$ using a coefficient vector $(\alpha_k) \in \mathbf{R}^{q+r}$. Hence, the necessary condition (11) is equivalent to

$$\forall(\alpha_k) \in \mathbf{R}^{q+r}, \ \exists(i, j), \ \sum_{k=1}^{q+r} \alpha_k (v_i^{(k)} - v_j^{(k)}) = 0. \quad (12)$$

It is easy to see, by contradiction, that (12) implies the following condition:

$$\exists(i, j), \forall k, \quad v_i^{(k)} = v_j^{(k)}. \quad (13)$$

By normalization $v_i^{(k)} = v_j^{(k)} = 1$ for (i, j) satisfying (13) and a proper permutation of $i = 1, \dots, n+1$, we find that possible canonical form of subspace V , which we denote by V_0 , is restricted to

$$V_0 = \{(z^T \ t_1 \mathbf{1}^T \ \cdots \ t_r \mathbf{1}^T)^T \in \mathbf{R}^{n+1} \mid \forall z \in \mathbf{R}^q, \ t_l \mathbf{1} \in \mathbf{R}^{n_l}, \ \forall t_l \in \mathbf{R}, \ l = 1, \dots, r\}$$

for q, r and $n_l, l = 1, \dots, r$ given in the setup. Using the above permutation as Π in the setup, i.e., $V = (\Pi^{-1} V_0)$, we have an isomorphic relation $W = a \circ (\Pi^{-1} V_0)$. Thus, this means that $W_0 = \Pi W = (\Pi a) \circ V_0$.

(“if” part) Conversely it is easy to confirm V_0 is a subalgebra of $(\mathbf{R}^{n+1}, \circ)$. We show that any other proper subspaces in V_0 cannot be a subalgebra with e , except for the trivial cases where several t_l ’s or components of $z = (z_i)$ are fixed to be zeros.¹ or equal to each other²

Consider a subspace $V' \subset V_0$ with nontrivial linear constraints between t_l ’s and z_i ’s. If V' is a subalgebra, then for all $x \in V'$ and integer m we have

$$V' \ni x^m = ((z^m)^T \ t_1^m \mathbf{1}^T \ \cdots \ t_r^m \mathbf{1}^T)^T, \quad z^m = (z_i^m) \in \mathbf{R}^q,$$

where t_l^m ’s and z_i^m ’s should satisfy the same linear constraints. We, however, find this is impossible by the similar arguments with the Vandermonde’s matrix in the “only if” part. This completes the proof.

Example (continued from Sect. 2.2) As $a \in \tilde{M} = W \cap \mathbf{R}_+^{n+1}$ we set

$$a = (1 \ 2 \ p_3 \ \cdots \ p_{n+1})^T, \quad a_0 = (1 \ 2)^T, \quad a_1 = (p_3 \ \cdots \ p_{n+1})^T.$$

Then we have $q = 2$, $r = 1$, $n_1 = n - 1$ and need no permutation, i.e., $W = W_0$. Since every element in W can be represented by

$$w = (\xi_1 \ \xi_2 \ tp_3 \ \cdots \ tp_{n+1})^T, \quad \xi_1, \ \xi_2, \ t \in \mathbf{R}$$

we can confirm W is a subalgebra of $(\mathbf{R}^{n+1}, \circ_{a^{-1}})$ and

$$V_0 = V = a^{-1} \circ W = \{(z^T \ t \mathbf{1}^T)^T \in \mathbf{R}^{n+1} \mid \forall z \in \mathbf{R}^2, \ t \mathbf{1} \in \mathbf{R}^{n-1}, \ \forall t \in \mathbf{R}\}.$$

¹These cases contradict the fact that $e \in V_0$.

²These cases correspond to choosing smaller q or r in the setup.

4 Concluding Remarks

We have studied doubly autoparallel structure of statistical models in the family of probability distributions on discrete and finite sample space. Identifying it by the probability simplex and using the mutation of Hadamard product, we give an algebraic characterization of doubly autoparallel submanifolds and their classification.

As is described in Sect. 1, several characterizations are also proved by different approaches in [9, 10]. However, in the literature, there seems to be no clear and simple result on classification of the structure, which we consider very important to distinguish or construct doubly autoparallel statistical models in applications. Because of the simplicity our result in Theorem 2 may be useful for such purposes.

Acknowledgements The authors are grateful to a reviewer who has pointed out the Ref. [10] and given comments to improve the original manuscript. A. O. is partially supported by JSPS Grant-in-Aid (C) 15K04997 and H. I. is partially supported by JST PRESTO and JSPS Grant-in Aid (C) 16K05174.

A Complements for Proposition 1

We shortly summarize basic notions of the relation between submanifolds and affine connections and give an outline of the proof for Proposition 1.

For an n -dimensional manifold \mathcal{M} equipped with an affine connection ∇ , we say that a submanifold \mathcal{N} is *autoparallel* with respect to ∇ when it holds that $\nabla_X Y \in \mathcal{X}(\mathcal{N})$ for arbitrary $X, Y \in \mathcal{X}(\mathcal{N})$. Since $\nabla^{(\alpha)}$ is nothing but an affine combination of $\nabla = \nabla^{(1)}$ and $\nabla^* = \nabla^{(-1)}$, we see that \mathcal{N} is actually DA if \mathcal{N} is autoparallel for arbitrary two connections $\nabla^{(\alpha)}$ and $\nabla^{(\alpha')}$ with $\alpha \neq \alpha'$. Hence, the equivalence of the statements (1), (2) and (3) is straightforward.

If every ∇ -geodesic curve passing through a point $p \in \mathcal{N}$ that is tangent to \mathcal{N} at p lies in \mathcal{N} , then \mathcal{N} is called *totally geodesic* at p . If \mathcal{N} is totally geodesic at every point of \mathcal{N} , then we say that \mathcal{N} is a *totally geodesic submanifold* in \mathcal{M} . When ∇ is torsion-free, a submanifold is totally geodesic if and only if it is autoparallel [11]. Since every $\nabla^{(\alpha)}$ is torsion-free by the assumption, the equivalence of (3) and (4) holds.

When ∇ is flat, there exists a coordinate system (x^1, \dots, x^n) of \mathcal{M} satisfying $\nabla_{\partial/\partial x^i} \partial/\partial x^j = 0$ for all i, j , which we call a ∇ -affine coordinate system. It is well known ([1, Theorem 1.1], for example) that a submanifold is ∇ -autoparallel if and only if it is expressed as an affine subspace with respect to ∇ -affine coordinate system. Then the equivalence of (1) and (5) immediately follows.

For the final statement a proof is given in [12] when \mathcal{M} is the probability simplex. Since the proof for this general case is similar, it is omitted.

References

1. Amari, S.-I., Nagaoka, H.: Methods of Information Geometry. Translations of mathematical monographs, vol. 191, American Mathematical Society and Oxford University Press (2000)
2. Ohara, A.: Information Geometric Analysis of an Interior Point Method for Semidefinite Programming. In: Barndorff-Nielsen, O., Jensen, V. (eds.) Proceedings of Geometry in Present Day Science, pp. 49–74. World Scientific (1999)
3. Ohara, A.: Geodesics for dual connections and means on symmetric cones. *Integral Equ. Oper. Theory* **50**, 537–548 (2004)
4. Ohara, A., Wada, T.: Information geometry of q-gaussian densities and behaviours of solutions to related diffusion equations. *J. Phys. A: Math. Theor.* **43**, 035002 (18pp.) (2010)
5. Uohashi, K., Ohara, A.: Jordan algebras and dual affine connections on symmetric cones. *Positivity* **8**(4), 369–378 (2004)
6. Ikeda, S., Tanaka, T., Amari, S.-I.: Stochastic reasoning, free energy, and information geometry. *Neural Comput.* **16**, 1779–1810 (2004)
7. Matúš, F., Ay, N.: On maximization of the information divergence from an exponential family. In: Proceedings of WUPES'03, pp. 199–204 (2003)
8. Montúfar, G.F.: Mixture decompositions of exponential families using a decomposition of their sample spaces. *Kybernetika* **49**(1), 23–39 (2013)
9. Nagaoka, H.: Information-geometrical characterization of statistical models which are statistically equivalent to probability simplex (2017). [arXiv:1701.07736v2](https://arxiv.org/abs/1701.07736v2)
10. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry. Springer, Berlin (2017)
11. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry II. Interscience Publishers (1969)
12. Ohara, A.: Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **370**, 184–193 (2007)

Complementing Chentsov's Characterization



Akio Fujiwara

On the occasion of Professor Amari's 80th birthday

Abstract It is shown that Markov invariant tensor fields on the manifold of probability distributions are closed under the operations of raising and lowering indices with respect to the Fisher metric. As a result, every (r, s) -type Markov invariant tensor field can be obtained by raising indices of some $(0, r + s)$ -type Markov invariant tensor field.

Keywords Information geometry · Probability simplex · Chentsov's characterization · Markov invariant tensor fields · Fisher metric · Statistical isomorphism

1 Introduction

In his seminal book [4], Chentsov characterized several covariant tensor fields on the manifold of probability distributions that fulfil certain invariance property, now referred to as the Markov invariance. Since Markov invariant $(0, 2)$ - and $(0, 3)$ -type tensor fields play essential roles in introducing a metric and affine connections on the manifold of probability distributions, Chentsov's theorem is regarded as one of the most fundamental achievements in information geometry [2].

Let, for each $n \in \mathbb{N}$,

A. Fujiwara (✉)

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

$$\mathcal{S}_{n-1} := \left\{ p : \Omega_n \rightarrow \mathbb{R}_{++} \mid \sum_{\omega \in \Omega_n} p(\omega) = 1 \right\}$$

be the manifold of probability distributions on a finite set $\Omega_n = \{1, 2, \dots, n\}$, where \mathbb{R}_{++} denotes the set of strictly positive real numbers. In what follows, each point $p \in \mathcal{S}_{n-1}$ is identified with the vector $(p(1), p(2), \dots, p(n)) \in \mathbb{R}_{++}^n$.

Given natural numbers n and ℓ satisfying $2 \leq n \leq \ell$, let

$$\Omega_\ell = \bigsqcup_{i=1}^n C_{(i)} \quad (1)$$

be a direct sum decomposition of the index set $\Omega_\ell = \{1, \dots, \ell\}$ into n mutually disjoint nonempty subsets $C_{(1)}, \dots, C_{(n)}$. We put labels on elements of the i th subset $C_{(i)}$ as follows:

$$C_{(i)} = \{i_1, \dots, i_{r_i}\},$$

where r_i is the number of elements in $C_{(i)}$. A map

$$f : \mathcal{S}_{n-1} \longrightarrow \mathcal{S}_{\ell-1} : (p_1, \dots, p_n) \longmapsto (q_1, \dots, q_\ell)$$

is called a *Markov embedding* associated with the partition (1) if it takes the form

$$q_{i_s} := \lambda_{i_s} p_i \quad \left(\lambda_{i_s} > 0, \quad \sum_{s=1}^{r_i} \lambda_{i_s} = 1 \right) \quad (2)$$

for each $i = 1, \dots, n$ and $s = 1, \dots, r_i$. A simple example of a Markov embedding is illustrated in Fig. 1, where $n = 2$ and $\ell = 3$.

A series $\{F^{[n]}\}_{n \in \mathbb{N}}$ of $(0, s)$ -type tensor fields, each on \mathcal{S}_{n-1} , is said to be *Markov invariant* if

$$F_p^{[n]}(X_1, \dots, X_s) = F_{f(p)}^{[\ell]}(f_* X_1, \dots, f_* X_s)$$

holds for all Markov embeddings $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ with $2 \leq n \leq \ell$, points $p \in \mathcal{S}_{n-1}$, and tangent vectors $X_1, \dots, X_s \in T_p \mathcal{S}_{n-1}$. When no confusion arises, we simply use an abridged notation F for $F^{[n]}$.

Now, the Chentsov Theorem [4] (cf., [3, 5]) asserts that the only Markov invariant tensor fields of type $(0, s)$, with $s \in \{1, 2, 3\}$, on \mathcal{S}_{n-1} are given, up to scaling, by

$$T_p(X) = E_p[(X \log p)] \quad (= 0), \quad (3)$$

$$g_p(X, Y) = E_p[(X \log p)(Y \log p)], \quad (4)$$

$$S_p(X, Y, Z) = E_p[(X \log p)(Y \log p)(Z \log p)], \quad (5)$$

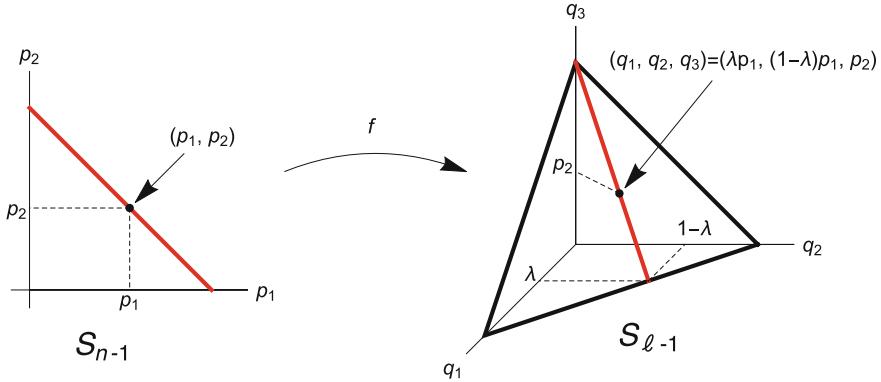


Fig. 1 A Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ for $n = 2$ and $\ell = 3$ associated with the partition $\Omega_3 = C_{(1)} \sqcup C_{(2)}$, where $C_{(1)} = \{1, 2\}$ and $C_{(2)} = \{3\}$

where $p \in \mathcal{S}_{n-1}$, and $E_p[\cdot]$ denotes the expectation with respect to p . In particular, the $(0, 2)$ -type tensor field g is nothing but the Fisher metric, and the $(0, 3)$ -type tensor field S yields the α -connection $\nabla^{(\alpha)}$ through the relation

$$g(\nabla_X^{(\alpha)} Y, Z) := g(\bar{\nabla}_X Y, Z) - \frac{\alpha}{2} S(X, Y, Z),$$

where $\bar{\nabla}$ is the Levi-Civita connection with respect to the Fisher metric g . Chentsov's theorem is thus a cornerstone of information geometry.

Despite this fact, it is curious that the above-mentioned formulation only concerns characterization of covariant tensor fields. Put differently, discussing the Markov invariance of contravariant and/or mixed-type tensor fields is beyond the scope. To the best of the author's knowledge, however, there have been no attempts toward such generalization. The objective of the present paper is to extend Chentsov's characterization to generic tensor fields.

2 Main Results

Associated with each Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ is a unique affine map

$$\varphi_f : \mathcal{S}_{\ell-1} \longrightarrow \mathcal{S}_{n-1} : (q_1, \dots, q_\ell) \longmapsto (p_1, \dots, p_n)$$

that satisfies

$$\varphi_f \circ f = \text{id}.$$

In fact, it is explicitly given by the following relations

$$p_i = \sum_{j \in C_{(i)}} q_j \quad (i = 1, \dots, n)$$

that allocate each event $C_{(i)}$ ($\subset \Omega_\ell$) to the singleton $\{i\}$ ($\subset \Omega_n$), (cf., Appendix A). We shall call the map φ_f the *coarse-graining* associated with a Markov embedding f . Note that the coarse-graining φ_f is determined only by the partition (1), and is independent of the internal ratios $\{\lambda_{i_s}\}_{i,s}$ that specifies f as (2).

For example, let us consider a Markov embedding

$$f : \mathcal{S}_1 \longrightarrow \mathcal{S}_3 : (p_1, p_2) \longmapsto (\lambda p_1, (1 - \lambda)p_1, \mu p_2, (1 - \mu)p_2), \quad (0 < \lambda, \mu < 1)$$

associated with the partition $\Omega_4 = C_{(1)} \sqcup C_{(2)}$, where

$$C_{(1)} = \{1, 2\}, \quad C_{(2)} = \{3, 4\}.$$

The coarse-graining $\varphi_f : \mathcal{S}_3 \rightarrow \mathcal{S}_1$ associated with f is given by

$$\varphi_f : (q_1, q_2, q_3, q_4) \longmapsto (q_1 + q_2, q_3 + q_4).$$

There are of course other affine maps $\bar{\varphi}_f : \mathcal{S}_3 \rightarrow \mathbb{R}_{++}^2$ that satisfy the relation $\bar{\varphi}_f \circ f = \text{id}$ on \mathcal{S}_1 : for example,

$$\bar{\varphi}_f : (q_1, q_2, q_3, q_4) \longmapsto \left(\frac{q_1}{\lambda}, \frac{q_3}{\mu} \right).$$

However, this is not a map of the form $\varphi_f : \mathcal{S}_3 \rightarrow \mathcal{S}_1$.

Now we introduce a generalized Markov invariance. A series $\{F^{[n]}\}_{n \in \mathbb{N}}$ of (r, s) -type tensor fields, each on \mathcal{S}_{n-1} , is said to be *Markov invariant* if

$$F_p^{[n]}(\omega^1, \dots, \omega^r, X_1, \dots, X_s) = F_{f(p)}^{[\ell]}(\varphi_f^* \omega^1, \dots, \varphi_f^* \omega^r, f_* X_1, \dots, f_* X_s)$$

holds for all Markov embeddings $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ with $2 \leq n \leq \ell$, points $p \in \mathcal{S}_{n-1}$, cotangent vectors $\omega^1, \dots, \omega^r \in T_p^* \mathcal{S}_{n-1}$, and tangent vectors $X_1, \dots, X_s \in T_p \mathcal{S}_{n-1}$. When no confusion arises, we simply use an abridged notation F for $F^{[n]}$.

The main result of the present paper is the following.

Theorem 1 *Markov invariant tensor fields are closed under the operations of raising and lowering indices with respect to the Fisher metric g .*

Theorem 1 has a remarkable consequence: every (r, s) -type Markov invariant tensor field can be obtained by raising indices of some $(0, r + s)$ -type Markov invariant tensor field. This fact could be paraphrased by saying that Chentsov's original approach was universal.

3 Proof of Theorem 1

We first prove that raising indices with respect to the Fisher metric preserves Markov invariance, and then prove that lowering indices also preserves Markov invariance.

3.1 Raising Indices Preserves Markov Invariance

Suppose we want to know whether the $(1, 2)$ -type tensor field $F^i_{jk} := g^{im} S_{mjk}$ is Markov invariant, where S is the Markov invariant $(0, 3)$ -type tensor field defined by (5). Put differently, we want to investigate if, for some (then any) local coordinate system (x^a) of \mathcal{S}_{n-1} , the $(1, 2)$ -type tensor field F defined by $F \left(dx^a, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c} \right) := g^{ae} S_{ebc}$ exhibits

$$F_p \left(dx^a, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c} \right) = F_{f(p)} \left(\varphi_f^* dx^a, f_* \frac{\partial}{\partial x^b}, f_* \frac{\partial}{\partial x^c} \right). \quad (6)$$

In order to handle such a relation, it is useful to identify the Fisher metric g on the manifold \mathcal{S}_{n-1} and its inverse g^{-1} with the following linear maps:

$$\begin{aligned} g &: T\mathcal{S}_{n-1} \longrightarrow T^*\mathcal{S}_{n-1}: \frac{\partial}{\partial x^a} \longmapsto g_{ab} dx^b, \\ g^{-1} &: T^*\mathcal{S}_{n-1} \longrightarrow T\mathcal{S}_{n-1}: dx^a \longmapsto g^{ab} \frac{\partial}{\partial x^b}. \end{aligned}$$

Note that these maps do not depend on the choice of a local coordinate system (x^a) of \mathcal{S}_{n-1} .

Now, observe that

$$\begin{aligned}\text{LHS of (6)} &= S_p \circ (g_p^{-1} \otimes I \otimes I) \left(dx^a, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c} \right) \\ &= S_p \left(g_p^{ae} \frac{\partial}{\partial x^e}, \frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c} \right)\end{aligned}$$

and

$$\begin{aligned}\text{RHS of (6)} &= S_{f(p)} \circ (g_{f(p)}^{-1} \otimes I \otimes I) \left(\varphi_f^* dx^a, f_* \frac{\partial}{\partial x^b}, f_* \frac{\partial}{\partial x^c} \right) \\ &= S_{f(p)} \left(g_{f(p)}^{-1} (\varphi_f^* dx^a), f_* \frac{\partial}{\partial x^b}, f_* \frac{\partial}{\partial x^c} \right).\end{aligned}$$

Since the $(0, 3)$ -type tensor field S is Markov invariant, the following Lemma establishes (6).

Lemma 2 *For any Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$, it holds that*

$$f_* \left(g_p^{ae} \frac{\partial}{\partial x^e} \right) = g_{f(p)}^{-1} (\varphi_f^* dx^a). \quad (7)$$

In other words, the diagram

$$\begin{array}{ccc} T_p^* \mathcal{S}_{n-1} & \xrightarrow{\varphi_f^*} & T_{f(p)}^* \mathcal{S}_{\ell-1} \\ g^{-1} \downarrow & & \downarrow g^{-1} \\ T_p \mathcal{S}_{n-1} & \xrightarrow{f_*} & T_{f(p)} \mathcal{S}_{\ell-1} \end{array}$$

is commutative.

Proof In view of a smooth link with the expression (2) of a Markov embedding, we make use of the $\nabla^{(m)}$ -affine coordinate system

$$\hat{\eta}_i := p_i \quad (i = 1, \dots, n-1)$$

as a coordinate system of \mathcal{S}_{n-1} , and the $\nabla^{(m)}$ -affine coordinate system

$$\eta_{i_s} := q_{i_s} \quad (i = 1, \dots, n-1; s = 1, \dots, r_i \text{ and } i = n; s = 1, \dots, r_n - 1)$$

as a coordinate system of $\mathcal{S}_{\ell-1}$, given a Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$. Note that the component $q_{n_{r_n}}$ that corresponds to the last element n_{r_n} of $C_{(n)}$ is excluded in this coordinate system because of the normalisation.

We shall prove (7) by showing the identity

$$\hat{h}_{im} f_* \frac{\partial}{\partial \hat{\eta}_m} = g_{f(p)}^{-1}(\varphi_f^* d\hat{\eta}_i), \quad (8)$$

where $\hat{h}_{im} := h_p(d\hat{\eta}_i, d\hat{\eta}_m)$, with h being the $(2, 0)$ -type tensor field defined by

$$h(dx^a, dx^b) := g^{ab}.$$

Note that, due to the duality of the η - and θ -coordinate systems, \hat{h}_{im} is identical to the component \hat{g}_{im} of the Fisher metric g with respect to the θ -coordinate system $(\hat{\theta}^i)$ of \mathcal{S}_{n-1} , and is explicitly given by

$$\hat{g}_{im} = \hat{\eta}_i \delta_{im} - \hat{\eta}_i \hat{\eta}_m.$$

Similarly, the components of h with respect to the η -coordinate system (η_j) of $\mathcal{S}_{\ell-1}$ is simply denoted by $h_{ij} := h_{f(p)}(d\eta_i, d\eta_j)$, and is identical to $g_{ij} = \eta_i \delta_{ij} - \eta_i \eta_j$.

Due to the choice of the coordinate systems $(\hat{\eta}_i)_i$ and $(\eta_{i_s})_{i,s}$, we have

$$\hat{\eta}_i = \sum_{j \in C(i)} \eta_j = \sum_{s=1}^{r_i} \eta_{i_s} \quad (i = 1, \dots, n-1),$$

so that

$$\varphi_f^* d\hat{\eta}_i = \sum_j \frac{\partial \hat{\eta}_i}{\partial \eta_j} d\eta_j = \sum_{s=1}^{r_i} d\eta_{i_s}.$$

Thus

$$\text{RHS of (8)} = g_{f(p)}^{-1} \left(\sum_{s=1}^{r_i} d\eta_{i_s} \right) = \sum_{s=1}^{r_i} \left(\sum_j h_{i_s, j} \frac{\partial}{\partial \eta_j} \right). \quad (9)$$

On the other hand, since

$$\begin{cases} \eta_{i_s} = \lambda_{i_s} \hat{\eta}_i & (i = 1, \dots, n-1; s = 1, \dots, r_i) \\ \eta_{n_s} = \lambda_{n_s} \left(1 - \sum_{i=1}^{n-1} \hat{\eta}_i \right) & (s = 1, \dots, r_n - 1) \end{cases},$$

we see that, for each $m = 1, \dots, n-1$,

$$\begin{aligned} f_* \frac{\partial}{\partial \hat{\eta}_m} &= \sum_{i=1}^{n-1} \sum_{s=1}^{r_i} \frac{\partial \eta_{i_s}}{\partial \hat{\eta}_m} \frac{\partial}{\partial \eta_{i_s}} + \sum_{s=1}^{r_n-1} \frac{\partial \eta_{n_s}}{\partial \hat{\eta}_m} \frac{\partial}{\partial \eta_{n_s}} \\ &= \sum_{s=1}^{r_m} \lambda_{m_s} \frac{\partial}{\partial \eta_{m_s}} - \sum_{s=1}^{r_n-1} \lambda_{n_s} \frac{\partial}{\partial \eta_{n_s}}. \end{aligned}$$

Consequently,

$$\text{LHS of (8)} = \sum_{m=1}^{n-1} \hat{h}_{im} \sum_{s=1}^{r_m} \lambda_{m_s} \frac{\partial}{\partial \eta_{m_s}} - \left(\sum_{m=1}^{n-1} \hat{h}_{im} \right) \left(\sum_{s=1}^{r_n-1} \lambda_{n_s} \frac{\partial}{\partial \eta_{n_s}} \right). \quad (10)$$

To prove (8), let us compare, for each j , the coefficients of $\frac{\partial}{\partial \eta_j}$ in (9) and (10). The index j runs through

$$\left(\bigsqcup_{k=1}^{n-1} C_{(k)} \right) \sqcup \{n_s\}_{s=1}^{r_n-1}.$$

So suppose that $j = k_u$, the u th element of $C_{(k)}$, where $1 \leq k \leq n$. Then

$$\text{coefficient of } \frac{\partial}{\partial \eta_{k_u}} \text{ in (9)} = \sum_{s=1}^{r_i} h_{i_s, k_u}. \quad (11)$$

On the other hand,

$$\text{coefficient of } \frac{\partial}{\partial \eta_{k_u}} \text{ in (10)} = \begin{cases} \hat{h}_{ik} \lambda_{k_u} & (1 \leq k \leq n-1) \\ - \left(\sum_{m=1}^{n-1} \hat{h}_{im} \right) \lambda_{n_u} & (k = n) \end{cases}. \quad (12)$$

We show that (11) equals (12) for all indices i, k , and u .

When $1 \leq k \leq n-1$,

$$\begin{aligned} (11) &= \sum_{s=1}^{r_i} (\eta_{i_s} \delta_{i_s, k_u} - \eta_{i_s} \eta_{k_u}) \\ &= \delta_{ik} \eta_{k_u} - \hat{\eta}_i \eta_{k_u} \\ &= \lambda_{k_u} (\delta_{ik} \hat{\eta}_k - \hat{\eta}_i \hat{\eta}_k) \\ &= (12). \end{aligned}$$

When $k = n$, on the other hand,

$$\begin{aligned}
(11) &= \sum_{s=1}^{r_i} (-\eta_{i_s} \eta_{n_u}) \\
&= -\hat{\eta}_i \lambda_{n_u} \left(1 - \sum_{m=1}^{n-1} \hat{\eta}_m \right) \\
&= -\lambda_{n_u} \sum_{m=1}^{n-1} (\hat{\eta}_i \delta_{im} - \hat{\eta}_i \hat{\eta}_m) \\
&= (12).
\end{aligned}$$

This proves the identity (8). \square

Now that Lemma 2 is established, a repeated use of the line of argument that precedes Lemma 2 leads us to the following general assertion: raising indices with respect to the Fisher metric preserves Markov invariance.

3.2 Lowering Indices Preserves Markov Invariance

Suppose that, given a Markov invariant $(3, 0)$ -type tensor field T , we want to know whether the $(2, 1)$ -type tensor field F defined by

$$F \left(\frac{\partial}{\partial x^a}, dx^b, dx^c \right) := g_{ae} T^{ebc}$$

satisfies Markov invariance:

$$F_p \left(\frac{\partial}{\partial x^a}, dx^b, dx^c \right) = F_{f(p)} \left(f_* \frac{\partial}{\partial x^a}, \varphi_f^* dx^b, \varphi_f^* dx^c \right)$$

or equivalently

$$T_p (g_{ae} dx^e, dx^b, dx^c) = T_{f(p)} \left(g_{f(p)} \left(f_* \frac{\partial}{\partial x^a} \right), \varphi_f^* dx^b, \varphi_f^* dx^c \right).$$

This question is resolved affirmatively by the following

Lemma 3 *For any Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$, it holds that*

$$\varphi_f^* ((g_p)_{ae} dx^e) = g_{f(p)} \left(f_* \frac{\partial}{\partial x^a} \right).$$

In other words, the diagram

$$\begin{array}{ccc}
T_p^* \mathcal{S}_{n-1} & \xrightarrow{\varphi_f^*} & T_{f(p)}^* \mathcal{S}_{\ell-1} \\
g \uparrow & & g \uparrow \\
T_p \mathcal{S}_{n-1} & \xrightarrow{f_*} & T_{f(p)} \mathcal{S}_{\ell-1}
\end{array}$$

is commutative.

Proof Since g is an isomorphism, this is a straightforward consequence of Lemma 2. \square

Lemma 3 has the following implication: lowering indices with respect to the Fisher metric preserves Markov invariance.

Theorem 1 is now an immediate consequence of Lemmas 2 and 3.

4 Concluding Remarks

We have proved that raising and lowering indices with respect to the Fisher metric preserve Markov invariance of tensor fields on the manifold of probability distributions. For example, g^{ij} is, up to scaling, the only $(2, 0)$ -type Markov invariant tensor field. It may be worthwhile to mention that not every operation in tensor calculus preserves Markov invariance. The following example is due to Amari [1].

With the $\nabla^{(e)}$ -affine coordinate system $\theta = (\theta^1, \dots, \theta^{n-1})$ of \mathcal{S}_{n-1} defined by

$$\log p(\omega) = \sum_{i=1}^{n-1} \theta^i \delta_i(\omega) - \log \left(1 + \sum_{k=1}^{n-1} \exp \theta^k \right),$$

the $(0, 3)$ -type tensor field (5) has the following components:

$$S_{ijk} = \begin{cases} \eta_i(1 - \eta_i)(1 - 2\eta_i), & (i = j = k) \\ -\eta_i(1 - 2\eta_i)\eta_k, & (i = j \neq k) \\ -\eta_j(1 - 2\eta_j)\eta_i, & (j = k \neq i) \\ -\eta_k(1 - 2\eta_k)\eta_j, & (k = i \neq j) \\ 2\eta_i\eta_j\eta_k, & (i \neq j \neq k \neq i) \end{cases}.$$

Here, $\eta = (\eta_1, \dots, \eta_{n-1})$ is the $\nabla^{(m)}$ -affine coordinate system of \mathcal{S}_{n-1} that is dual to θ . By using the formula

$$g^{ij} = \frac{1}{\eta_0} + \frac{\delta^{ij}}{\eta_i}, \quad \left(\eta_0 := 1 - \sum_{i=1}^{n-1} \eta_i \right),$$

the $(1, 2)$ -type tensor field $F^i_{jk} := g^{im} S_{mjk}$ is readily calculated as

$$F^i_{jk} = \begin{cases} 1 - 2\eta_i, & (i = j = k) \\ -\eta_k, & (i = j \neq k) \\ -\eta_j, & (i = k \neq j) \\ 0, & (i \neq j, i \neq k) \end{cases}.$$

We know from Theorem 1 that F is Markov invariant. However, the following contracted $(0, 1)$ -type tensor field

$$\tilde{T}_k := F^i_{ik} = 1 - n\eta_k$$

is non-zero, and hence is not Markov invariant; see (3). This demonstrates that the contraction, which is a standard operation in tensor calculus, does not always preserve Markov invariance.

Chentsov's idea of imposing the invariance of geometrical structures under Markov embeddings $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ is based on the fact that \mathcal{S}_{n-1} is statistically isomorphic to $f(\mathcal{S}_{n-1})$. Put differently, the invariance only involves direct comparison between \mathcal{S}_{n-1} and its image $f(\mathcal{S}_{n-1})$, and is nothing to do with the complement of $f(\mathcal{S}_{n-1})$ in the ambient space $\mathcal{S}_{\ell-1}$. On the other hand, the partial trace operation $F^i_{jk} \mapsto F^i_{ik}$ on $\mathcal{S}_{\ell-1}$ (more precisely, on $T_{f(p)}\mathcal{S}_{\ell-1} \otimes T_{f(p)}^*\mathcal{S}_{\ell-1}$) makes the output F^i_{ik} "contaminated" with information from outside the submanifold $f(\mathcal{S}_{n-1})$. It is thus no wonder such an influx of extra information manifests itself as the non-preservation of Markov invariance. In this respect, a distinctive characteristic of Lemmas 2 and 3 lies in the fact that raising and lowering indices preserve Markov invariance although they are represented in the forms of contraction such as $g^{i\ell} S_{mjk} \mapsto g^{im} S_{mjk}$ or $g_{i\ell} T^{mjk} \mapsto g_{im} T^{mjk}$.

Acknowledgements The author would like to express his sincere gratitude to Professor Amari for all his encouragement and inspiring discussions. He is also grateful to Professor Nagaoka for many insightful comments, and to the anonymous referee for constructive criticism and helpful suggestions. The present study was supported by JSPS KAKENHI Grants No. JP22340019, No. JP15K13459, and No. JP17H02861.

Appendix

A Uniqueness of φ_f

A Markov embedding $f : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_{\ell-1}$ defined by (1) and (2) uniquely extends to a linear map $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}^\ell : (x_1, \dots, x_n) \mapsto (y_1, \dots, y_\ell)$ as

$$y_{i_s} := \lambda_{i_s} x_i \quad \left(\lambda_{i_s} > 0, \quad \sum_{s=1}^{r_i} \lambda_{i_s} = 1 \right).$$

Similarly, let $\tilde{\varphi}_f : \mathbb{R}^\ell \rightarrow \mathbb{R}^n$ be the unique linear extension of $\varphi_f : \mathcal{S}_{\ell-1} \rightarrow \mathcal{S}_{n-1}$.

Since

$$\left(\sum_{k=1}^{\ell} (\tilde{\varphi}_f)_{ik} q_k \right)_{i=1,\dots,n} \in \mathcal{S}_{n-1} \quad (13)$$

for all $q = (q_1, \dots, q_\ell) \in \mathcal{S}_{\ell-1}$, we see that the matrix elements of $\tilde{\varphi}_f$ are nonnegative:

$$(\tilde{\varphi}_f)_{ik} \geq 0 \quad (\forall i = 1, \dots, n; \forall k = 1, \dots, \ell). \quad (14)$$

The relation (13) also entails that, for all $q \in \mathcal{S}_{\ell-1}$,

$$\sum_{k=1}^{\ell} q_k = 1 = \sum_{i=1}^n \left(\sum_{k=1}^{\ell} (\tilde{\varphi}_f)_{ik} q_k \right),$$

so that

$$\sum_{k=1}^{\ell} \left(\sum_{i=1}^n (\tilde{\varphi}_f)_{ik} - 1 \right) q_k = 0.$$

Consequently,

$$\sum_{i=1}^n (\tilde{\varphi}_f)_{ik} = 1 \quad (\forall k = 1, \dots, \ell). \quad (15)$$

It then follows from (14) and (15) that

$$0 \leq (\tilde{\varphi}_f)_{ik} \leq 1 \quad (\forall i = 1, \dots, n; \forall k = 1, \dots, \ell). \quad (16)$$

Now, since $\tilde{\varphi}_f$ is a left inverse of \tilde{f} ,

$$\delta_{ij} = \sum_{k=1}^{\ell} (\tilde{\varphi}_f)_{ik} (\tilde{f})_{kj} = \sum_{k \in C(j)} (\tilde{\varphi}_f)_{ik} \lambda_k = \sum_{s=1}^{r_j} (\tilde{\varphi}_f)_{i,j_s} \lambda_{j_s}.$$

Because of (16), we have

$$(\tilde{\varphi}_f)_{i,j_s} = \delta_{ij} \quad (\forall s = 1, \dots, r_j).$$

This proves that φ_f is unique and is given by the coarse-graining associated with f .

References

1. Amari, S.-I.: private communication (14 January 2015)
2. Amari, S.-I., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, vol. 191. AMS and Oxford, Providence (2000)
3. Campbell, L.L.: An extended Čencov characterization of the information metric. Proc. Am. Math. Soc. **98**, 135–141 (1986)
4. Čencov, N.N.: Statistical Decision Rules and Optimal Inference. Translations of Mathematical Monographs, vol. 53. AMS, Providence (1982)
5. Fujiwara, A.: Foundations of Information Geometry. Makino Shoten, Tokyo (2015). (in Japanese)

Constant Curvature Connections On Statistical Models



Alexander Rylov 

Abstract We discuss the constant curvature geometry on some 2-dimensional examples: the normal, logistic, Pareto and Weibull statistical manifolds with connections of constant α -curvature. The Pareto two-dimensional statistical model has such a structure, each of its α -connection has the constant curvature $(-2\alpha - 2)$. It is known that if the statistical manifold has some α -connection of constant curvature then it is a conjugate symmetric manifold. The Weibull two-dimensional statistical model has such a structure and its 1-connection has the constant curvature. We compare this model with the known statistical models such as normal and logistic.

Keywords Statistical manifold · Constant α -curvature · Conjugate symmetric manifold · Pareto statistical model · Weibull statistical model

1 Introduction

We consider statistical manifolds with connections of constant α -curvature and more generally conjugate statistical manifolds [1, 2]. If the statistical manifold has some α -connection of constant curvature then it is a conjugate symmetric manifold. But the converse is not true [3].

Statistical models on families of probability distributions provide important examples of above-mentioned connections. Ivanova [4] compared the geometric properties of metric, or 0-connections on four statistical parameter spaces constituted by the normal, the exponential, the logistic, and the Weibull distribution, respectively. She gave the values of its 0-curvatures, which are constant.

It is known, see Arwini and Dodson [5], that normal statistical models have the constant α -curvatures

A. Rylov (✉)

Financial University under the Government of the Russian Federation,
125993 Leningradskiy Ave., 49, Moscow, Russia

e-mail: alexander_rylov@mail.ru

URL: <http://www.fa.ru/university/persons/Pages/view.aspx?ProfileId=9503>

$$k^{(\alpha)} = \frac{\alpha^2 - 1}{2} \text{ for any parameter } \alpha.$$

We obtain that the Pareto two-dimensional statistical model, see also Peng, Sun and Jiu [6], has such a structure: each of its α -connection has the constant curvature

$$k^{(\alpha)} = -2\alpha - 2.$$

The logistic two-dimensional statistical model has the constant 2-curvature

$$k^{(2)} = -\frac{162}{(\pi^2 + 3)^2},$$

thus, model is a conjugate symmetric manifold, see also our paper [7]. We consider the Weibull two-dimensional statistical model: its 1-connection has the constant curvature

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

where γ is Euler–Mascheroni constant. Thus, the Weibull model is a conjugate symmetric.

The paper is organized as follows. In Sect. 2 we recall the main concepts of a statistical manifold, an α -connection, a conjugate symmetric manifold and a statistical manifold of a constant α -curvature. Then we provide a relationship between these concepts. In Sect. 3 we review the concepts of a statistical model, Amari–Chentsov connections and illustrate them by well-known examples of the normal and the logistic models. Sections 4–5 is devoted to the calculation of statistical structure components for the Pareto and the Weibull model. Both models turned out to be a conjugate symmetric. In Sect. 6 we compare the constant values of α -curvatures for these different statistical models.

2 Statistical Manifolds of a Constant α -Curvature

Let M be a smooth manifold; $\dim M = n$; $\langle \cdot, \cdot \rangle = g$ be a Riemannian metrics; K be a $(2, 1)$ -tensor such that

$$K(X, Y) = K(Y, X), \tag{1}$$

$$\langle K(X, Y), Z \rangle = \langle Y, K(X, Z) \rangle, \tag{2}$$

where X, Y, Z are vector fields on M . Then a triple (M, g, K) is a *statistical manifold*, see Lauritzen [1] and Amari [2]. If D is the metric connection, i.e.

$$Dg = 0, \quad (3)$$

α is a real-valued parameter, then the linear connections of the 1-parameter family

$$\nabla^\alpha = D + \alpha \cdot K, \quad (4)$$

are called α -connections.

Denote R_{XY}^α a curvature operator of ∇^α , Ric^α a Ricci tensor of ∇^α , ω_g a volume element associated to the metrics. We call (M, g, K) a *conjugate symmetric manifold*, when

$$R_{XY}^\alpha g = 0 \text{ for any parameter } \alpha. \quad (5)$$

This is equivalent to the equality $R^\alpha = R^{-\alpha}$ for any pairs of the dual α - and $(-\alpha)$ -connections ($\alpha \neq 0$), see Lauritzen [1] and Noguchi [3].

Theorem 1 ([7]) *If some α -connection ($\alpha \neq 0$) has the constant α -curvature $k^{(\alpha)}$, i.e.*

$$R^\alpha(X, Y)Z := k^{(\alpha)}(\langle Y, Z \rangle X - \langle X, Z \rangle Y), \quad (6)$$

then (M, g, K) is a conjugate symmetric manifold.

The converse of the statement in Theorem 1 is not true, see the example by Noguchi [3]. Therefore, the sufficient condition for a constant α -curvature requires additional assumptions.

Theorem 2 ([8]) *If (M, g, K) is a conjugate symmetric manifold with connections which are*

- *equiprojective α -connections, i.e.*

$$R^\alpha(X, Y)Z = \frac{1}{n-1}(Ric^\alpha(X, Z)Y - Ric^\alpha(Y, Z)X), \quad (7)$$

- *strongly compatible α -connections with the metrics g i.e.*

$$(\nabla_X^\alpha g)(Y, Z) = (\nabla_Y^\alpha g)(X, Z); \quad (8)$$

$$\nabla^\alpha \omega_g = 0, \quad (9)$$

then (M, g, K) is a statistical manifold of a constant α -curvature.

Remark Some properties of the dual α -connections have been described by the author in earlier papers [9, 10].

3 Two-Dimensional Statistical Models

Let $S = \{P_{\theta^i} \mid i = 1, \dots, n\}$ be a family of probability distributions on a measurable space with a probability measure P , where each probability density function $p := p(x \mid \theta^i)$ of a continuous random variable x may be parameterized using the set of real-valued n variables $\theta = (\theta^i)$. Then S is called an *n -dimensional statistical model* [2]. We denote

$$\partial_i := \frac{\partial}{\partial \theta^i},$$

$\log p(x \mid \theta^i)$:= natural logarithm of the likelihood function.

The components of the Fisher information matrix I

$$I_{ij}(\theta) := \int_{-\infty}^{+\infty} \partial_i \log p \cdot \partial_j \log p \cdot p \cdot dP \quad (i, j = 1, \dots, n) \quad (10)$$

and the covariant components of the tensor K

$$K_{ijk}(\theta) := -\frac{1}{2} \int_{-\infty}^{+\infty} \partial_i \log p \cdot \partial_j \log p \cdot \partial_k \log p \cdot p \cdot dP \quad (i, j, k = 1, \dots, n) \quad (11)$$

give the structure of the statistical manifold (I, K) on S . For real-valued parameter α the covariant components of α -connections named *Amari–Chentsov connections* are

$$\Gamma_{ijk}(\theta) := \int_{-\infty}^{+\infty} \left(\partial_i \partial_j \log p \cdot \partial_k \log p + \frac{1-\alpha}{2} \partial_i \log p \cdot \partial_j \log p \cdot \partial_k \log p \right) \cdot p \cdot dP. \quad (12)$$

We provide two well-known examples of two-dimensional statistical models with its statistical structures.

3.1 Normal Model

The probability density function of 1-dimensional normal distribution is

$$p(x \mid \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}; \quad \mu = \theta^1, \sigma = \theta^2 > 0, \quad (13)$$

where x is a random variable, $x \in R$, μ is the mean value, σ is the standard deviation, and σ^2 is the variance. Therefore, we consider a family of normal distributions as the *normal* two-dimensional statistical model. From (10)–(12) we determine the

components of the Fisher information matrix

$$I = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix} \quad (14)$$

and the non-zero components of the α -curvatures tensor [5]

$${}^{(\alpha)}R_{1212} = -{}^{(\alpha)}R_{1221} = {}^{(\alpha)}R_{2121} = -{}^{(\alpha)}R_{2112} = \frac{1 - \alpha^2}{2\sigma^4}. \quad (15)$$

Therefore, normal statistical models have the constant α -curvatures

$$k^{(\alpha)} = \frac{\alpha^2 - 1}{2}. \quad (16)$$

3.2 Logistic Model

The logistic density function is defined as follows

$$p(x | a, b) = \frac{a \cdot \exp(-ax - b)}{(1 + \exp(-ax - b))^2}; \quad a = \theta^1 > 0, b = \theta^2, \quad (17)$$

where x is a variable, $x \in R$, a is the positive scale parameter, b is the location parameter. The distribution function is

$$F(x | a, b) = \frac{1}{1 + \exp(-ax - b)}, \quad (18)$$

the natural logarithm of the likelihood function is

$$\log p = \log a - ax - b - 2 \log(1 + \exp(-ax - b)), \quad (19)$$

and the partial derivatives are

$$\begin{aligned} \partial_1 \log p &:= \frac{\partial \log p}{\partial a} = \frac{1}{a} - x + \frac{2x \cdot \exp(-ax - b)}{1 + \exp(-ax - b)}; \\ \partial_2 \log p &:= \frac{\partial \log p}{\partial b} = \frac{-1 + \exp(-ax - b)}{1 + \exp(-ax - b)}. \end{aligned} \quad (20)$$

Then, using (10) and (20) we can obtain the components of the Fisher information matrix

$$I = \frac{1}{3a^2} \begin{pmatrix} b^2 + \frac{\pi^2}{3} + 1 & -ab \\ -ab & a^2 \end{pmatrix} \quad (21)$$

From (11) and (20) we determine the covariant components of the tensor K

$$\begin{aligned} K_{111} &= \frac{1}{6a^3} \cdot (\pi^2 + 3b^2), & K_{122} = K_{212} = K_{221} &= \frac{1}{6a}, \\ K_{211} = K_{112} = K_{121} &= -\frac{b}{3a^2}, & K_{222} &= 0. \end{aligned} \quad (22)$$

The components of Amari–Chentsov α -connections on the logistic model have long analytical expressions, see also [7], so we consider the case $\alpha = 2$. We have the non-zero components of the 2-connection as follows

$${}^{(2)}\Gamma_{11}^1 = \frac{2\pi^2 - 3}{(\pi^2 + 3)a}, \quad {}^{(2)}\Gamma_{11}^2 = -\frac{9b}{(\pi^2 + 3)a^2}, \quad {}^{(2)}\Gamma_{12}^2 = {}^{(2)}\Gamma_{21}^2 = \frac{1}{a}. \quad (23)$$

and the non-zero covariant components of the 2-curvature tensor

$${}^{(2)}R_{1212} = -{}^{(2)}R_{1221} = {}^{(2)}R_{2121} = -{}^{(2)}R_{2112} = \frac{3}{(\pi^2 + 3)a^2}. \quad (24)$$

Hence, the logistic model has the constant 2-curvature

$$k^{(2)} = -\frac{162}{(\pi^2 + 3)^2}. \quad (25)$$

Thus, by the Theorem 1 it is a conjugate symmetric manifold.

4 Pareto Statistical Model

Let us consider a family of probability distributions having power distribution function

$$F(x | a, \rho) = 1 - \left(\frac{a}{x}\right)^\rho; \quad a = \theta^1 > 0, \rho = \theta^2 > 0, \quad (26)$$

where x is a random variable on the domain $x \in [a; +\infty]$, a is the scale parameter, b is the shape parameter. The probability density functions

$$p(x | a, \rho) := \frac{dF(x | a, \rho)}{dx} = \rho \cdot a^\rho x^{-\rho-1}, \quad (27)$$

called the Pareto functions, form the two-dimensional *Pareto* statistical model. We have the natural logarithm of the likelihood function as

$$\log p = \log \rho + \rho \cdot \log a - (\rho + 1) \log x, \quad (28)$$

and its partial derivatives:

$$\begin{aligned}\partial_1 \log p &:= \frac{\partial \log p}{\partial a} = \frac{\rho}{a}; \\ \partial_2 \log p &:= \frac{\partial \log p}{\partial \rho} = \frac{1}{\rho} + \log a - \log x.\end{aligned}\quad (29)$$

To obtain the components of the Pareto statistical structure we use auxiliary improper integrals:

$$\begin{aligned}\int_a^{+\infty} x^{-\rho-1} dx &= \frac{1}{\rho a^\rho}, \\ \int_a^{+\infty} \log x \cdot x^{-\rho-1} dx &= \frac{1 + \rho \log a}{\rho^2 a^\rho}, \\ \int_a^{+\infty} \log^2 x \cdot x^{-\rho-1} dx &= \frac{(1 + \rho \log a)^2 + 1}{\rho^3 a^\rho}, \\ \int_a^{+\infty} \log^3 x \cdot x^{-\rho-1} dx &= \frac{(1 + \rho \log a)^3 + 3(1 + \rho \log a) + 2}{\rho^4 a^\rho}.\end{aligned}$$

Now we determine the components (10) of the Fisher matrix:

$$I_{11} = \int_a^{+\infty} \left(\frac{\rho}{a}\right)^2 \cdot \rho \cdot a^\rho x^{-\rho-1} dx = \left(\frac{\rho}{a}\right)^2, \quad (30)$$

$$I_{12} = I_{21} = \int_a^{+\infty} \frac{\rho}{a} \left(\frac{1}{\rho} + \log a - \log x\right) \cdot \rho \cdot a^\rho x^{-\rho-1} dx = 0, \quad (31)$$

$$I_{22} = \int_a^{+\infty} \left(\frac{1}{\rho} + \log a - \log x\right)^2 \cdot \rho \cdot a^\rho x^{-\rho-1} dx = \frac{1}{\rho^2}. \quad (32)$$

Thus, we obtain the Fisher information matrix as follows

$$I = \begin{pmatrix} \left(\frac{\rho}{a}\right)^2 & 0 \\ 0 & \frac{1}{\rho^2} \end{pmatrix} \quad (33)$$

Similarly, we calculate the covariant components (11) of the tensor K :

$$K_{111} = -\frac{1}{2} \int_a^{+\infty} \left(\frac{\rho}{a}\right)^3 \cdot \rho \cdot a^\rho x^{-\rho-1} dx = -\frac{1}{2} \cdot \left(\frac{\rho}{a}\right)^3, \quad (34)$$

$$\begin{aligned}K_{112} &= K_{121} = K_{211} \\ &= -\frac{1}{2} \int_a^{+\infty} \left(\frac{\rho}{a}\right)^2 \cdot \left(\frac{1}{\rho} + \log a - \log x\right) \cdot \rho \cdot a^\rho x^{-\rho-1} dx = 0,\end{aligned}\quad (35)$$

$$K_{122} = K_{212} = K_{221}$$

$$= -\frac{1}{2} \int_a^{+\infty} \frac{\rho}{a} \cdot \left(\frac{1}{\rho} + \log a - \log x \right)^2 \cdot \rho \cdot a^\rho x^{-\rho-1} dx = -\frac{1}{2a\rho}, \quad (36)$$

$$K_{222} = -\frac{1}{2} \int_a^{+\infty} \left(\frac{1}{\rho} + \log a - \log x \right)^3 \cdot \rho \cdot a^\rho x^{-\rho-1} dx = \frac{1}{\rho^3}. \quad (37)$$

Using the Christoffel symbols, i.e. the components of 0-connections, and the components of the tensor K we can find the components of Amari–Chentsov α -connections by the formula (4):

$${}^{(\alpha)}\Gamma_{11}^1 := {}^{(0)}\Gamma_{11}^1 + \alpha \cdot K_{11}^1 = -\frac{1}{a} + \alpha \cdot \left(-\frac{\rho}{2a} \right) = -\frac{2 + \alpha \cdot \rho}{2a}, \quad (38)$$

$${}^{(\alpha)}\Gamma_{11}^2 := {}^{(0)}\Gamma_{11}^2 + \alpha \cdot K_{11}^2 = -\frac{\rho^3}{a^2} + \alpha \cdot 0 = -\frac{\rho^3}{a^2}, \quad (39)$$

$${}^{(\alpha)}\Gamma_{12}^1 = {}^{(\alpha)}\Gamma_{21}^1 := {}^{(0)}\Gamma_{12}^1 + \alpha \cdot K_{12}^1 = \frac{1}{\rho} + \alpha \cdot 0 = \frac{1}{\rho}, \quad (40)$$

$${}^{(\alpha)}\Gamma_{12}^2 = {}^{(\alpha)}\Gamma_{21}^2 := {}^{(0)}\Gamma_{12}^2 + \alpha \cdot K_{12}^2 = 0 + \alpha \cdot \left(-\frac{\rho}{2a} \right) = -\frac{\alpha \cdot \rho}{2a}, \quad (41)$$

$${}^{(\alpha)}\Gamma_{22}^1 := {}^{(0)}\Gamma_{22}^1 + \alpha \cdot K_{22}^1 = 0 + \alpha \cdot \left(-\frac{a}{2\rho^3} \right) = -\frac{\alpha \cdot a}{2\rho^3}, \quad (42)$$

$${}^{(\alpha)}\Gamma_{22}^2 := {}^{(0)}\Gamma_{22}^2 + \alpha \cdot K_{22}^2 = -\frac{1}{\rho} + \alpha \cdot \frac{1}{\rho} = -\frac{1 - \alpha}{\rho}. \quad (43)$$

Theorem 3 *The Pareto two-dimensional statistical model is a statistical manifold of the constant α -curvature*

$$k^{(\alpha)} = -2\alpha - 2. \quad (44)$$

In particular,

$$k^{(0)} = -2, \quad (45)$$

i.e. this model is the Riemannian manifold of constant negative curvature.

Proof We show that every Amari–Chentsov α -connection of the Pareto model has the constant curvature $(-2\alpha - 2)$. The components of the α -curvature tensor take the form:

$$\begin{aligned} {}^{(\alpha)}R_{1212} &= -{}^{(\alpha)}R_{1221} = {}^{(\alpha)}R_{2121} = -{}^{(\alpha)}R_{2112} \\ &= {}^{(\alpha)}R_{121}^1 \cdot g_{12} + {}^{(\alpha)}R_{121}^2 \cdot g_{22} = (\alpha + 1) \left(\frac{\rho}{a} \right)^2 \cdot \frac{1}{\rho^2} = \frac{\alpha + 1}{a^2}. \end{aligned} \quad (46)$$

Therefore, the curvature of an arbitrary α -connection has the constant value

$$k^{(\alpha)} = -2 \cdot \det I^{-1} \cdot {}^{(\alpha)}R_{1212} = -2 \cdot a^2 \cdot \frac{\alpha + 1}{a^2} = -2\alpha - 2.$$

For $\alpha = 0$ we obtain the metric connection of constant negative curvature $k^{(0)} = -2$. \square

5 Weibull Statistical Model

The Weibull distribution is a generalization of the previously considered normal and logistic distributions in a certain sense [4]. The Weibull density function is defined as follows

$$p(x | \lambda, k) = \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{k-1} \cdot \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}; \quad \lambda = \theta^1 > 0, k = \theta^2 > 0, \quad (47)$$

where x is a variable on the domain $x \in (0; +\infty)$, λ is the scale parameter, k is the shape parameter. The distribution function is

$$F(x | \lambda, k) = 1 - \exp\left\{-\left(\frac{x}{\lambda}\right)^k\right\}, \quad (48)$$

the natural logarithm of the likelihood function is

$$\log p = \log\left(\frac{k}{\lambda}\right) + (k-1) \cdot \log\left(\frac{x}{\lambda}\right) - \left(\frac{x}{\lambda}\right)^k, \quad (49)$$

and the partial derivatives are

$$\begin{aligned} \partial_1 \log p &:= \frac{\partial \log p}{\partial \lambda} = -\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k; \\ \partial_2 \log p &:= \frac{\partial \log p}{\partial k} = \frac{1}{k} + \left(1 - \left(\frac{x}{\lambda}\right)^k\right) \cdot \log\left(\frac{x}{\lambda}\right). \end{aligned} \quad (50)$$

To obtain the components of the Weibull statistical structure we need useful improper integrals, for natural n and for real $k > 0$:

$$\begin{aligned} \int_0^{+\infty} x^{nk-1} \cdot \exp(-x^k) \cdot dx &= \frac{(n-1)!}{k}; \\ \int_0^{+\infty} \log x \cdot x^{k-1} \cdot \exp(-x^k) \cdot dx &= -\frac{\gamma}{k^2}; \\ \int_0^{+\infty} \log^2 x \cdot x^{k-1} \cdot \exp(-x^k) \cdot dx &= \frac{\pi^2 + 6\gamma^2}{6k^3}; \\ \int_0^{+\infty} \log^3 x \cdot x^{k-1} \cdot \exp(-x^k) \cdot dx &= -\frac{4\zeta(3) + \gamma\pi^2 + 2\gamma^3}{2k^4}, \end{aligned}$$

where

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{m=1}^{\infty} \frac{1}{m} - \log n \right) \approx 0,577 \text{ is Euler–Mascheroni constant,} \quad (51)$$

$$\zeta(3) = \sum_{m=1}^{\infty} \frac{1}{m^3} \approx 1,202 \text{ is Apery's constant.} \quad (52)$$

We determine the components (10) of the Fisher matrix:

$$I_{11} = \int_0^{+\infty} \left(-\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k \right)^2 \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ -\left(\frac{x}{\lambda} \right)^k \right\} \cdot dx = \left(\frac{k}{\lambda} \right)^2, \quad (53)$$

$$\begin{aligned} I_{12} &= I_{21} \\ &= \int_0^{+\infty} \left(-\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k \right) \cdot \left(\frac{1}{k} + \left(1 - \left(\frac{x}{\lambda} \right)^k \right) \cdot \log \left(\frac{x}{\lambda} \right) \right) \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ -\left(\frac{x}{\lambda} \right)^k \right\} \cdot dx \\ &= \frac{\gamma - 1}{\lambda}, \end{aligned} \quad (54)$$

$$\begin{aligned} I_{22} &= \int_0^{+\infty} \left(\frac{1}{k} + \left(1 - \frac{x}{\lambda} \right)^k \right) \cdot \log \left(\frac{x}{\lambda} \right)^2 \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ -\left(\frac{x}{\lambda} \right)^k \right\} \cdot dx \\ &= \frac{\pi^2 + 6\gamma^2 - 12\gamma + 6}{6k^2}. \end{aligned} \quad (55)$$

Therefore, we have the Fisher information matrix and its determinant:

$$I = \begin{pmatrix} \left(\frac{k}{\lambda} \right)^2 & \frac{\gamma - 1}{\lambda} \\ \frac{\gamma - 1}{\lambda} & \frac{\pi^2 + 6(\gamma - 1)^2}{6k^2} \end{pmatrix}; \quad (56)$$

$$\det I = \frac{\pi^2}{6\lambda^2}. \quad (57)$$

Similarly, we calculate the covariant components (11) of the tensor K :

$$K_{111} = -\frac{1}{2} \int_0^{+\infty} \left(-\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k \right)^3 \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ -\left(\frac{x}{\lambda} \right)^k \right\} \cdot dx = -\left(\frac{k}{\lambda} \right)^3, \quad (58)$$

$$\begin{aligned} K_{112} &= K_{121} = K_{211} \\ &= -\frac{1}{2} \int_0^{+\infty} \left(-\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k \right)^2 \left(\frac{1}{k} + \left(1 - \left(\frac{x}{\lambda} \right)^k \right) \cdot \log \left(\frac{x}{\lambda} \right) \right) \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ -\left(\frac{x}{\lambda} \right)^k \right\} \cdot dx \\ &= \frac{(2 - \gamma)k}{\lambda^2}, \end{aligned} \quad (59)$$

$$\begin{aligned}
K_{122} &= K_{212} = K_{221} \\
&= -\frac{1}{2} \int_0^{+\infty} \left(-\frac{k}{\lambda} + \frac{k}{\lambda^{k+1}} x^k \right) \left(\frac{1}{k} + \left(1 - \left(\frac{x}{\lambda} \right)^k \right) \cdot \log \left(\frac{x}{\lambda} \right) \right) \cdot \frac{2k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ - \left(\frac{x}{\lambda} \right)^k \right\} \cdot dx \\
&= -\frac{\pi^2 + 6\gamma^2 - 24\gamma + 12}{6\lambda k},
\end{aligned} \tag{60}$$

$$\begin{aligned}
K_{222} &= -\frac{1}{2} \int_0^{+\infty} \left(\frac{1}{k} + \left(1 - \left(\frac{x}{\lambda} \right)^k \right) \cdot \log \left(\frac{x}{\lambda} \right) \right)^3 \cdot \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda} \right)^{k-1} \cdot \exp \left\{ - \left(\frac{x}{\lambda} \right)^k \right\} \cdot dx \\
&= \frac{\pi^2(2-\gamma) - 2\gamma^3 + 12\gamma^2 - 12\gamma + 2 - 4\zeta(3)}{2k^3}.
\end{aligned} \tag{61}$$

The components of Amari–Chentsov α -connections on the Weibull model have long analytical expressions as for logistic model in Sect. 3.2. Let us consider the case $\alpha = 1$.

Theorem 4 *The Weibull two-dimensional statistical model is a conjugate symmetric statistical manifold. Its 1-connection has the constant curvature*

$$k^{(1)} = \frac{12\pi^2\gamma - 144\gamma + 72}{\pi^4}, \tag{62}$$

where γ is Euler–Mascheroni constant.

Proof We show that the Amari–Chentsov 1-connection of the Weibull model has the constant curvature. Then, by Theorem 1, it will follow that the model is conjugate symmetric. We provide the components of the 1-connection according (4):

$${}^{(1)}\Gamma_{11}^1 := {}^{(0)}\Gamma_{11}^1 + 1 \cdot K_{11}^1 = -\frac{\pi^2 - 6(\gamma - 1)k}{\pi^2 \lambda} - k \cdot \frac{\pi^2 + 6(\gamma - 1)}{\pi^2 \lambda} = -\frac{k + 1}{\lambda}, \tag{63}$$

$${}^{(1)}\Gamma_{11}^2 := {}^{(0)}\Gamma_{11}^2 + 1 \cdot K_{11}^2 = -\frac{6k^3}{\pi^2 \lambda^2} + \frac{6k^3}{\pi^2 \lambda^2} = 0, \tag{64}$$

$${}^{(1)}\Gamma_{12}^1 = {}^{(1)}\Gamma_{21}^1 := {}^{(0)}\Gamma_{12}^1 + 1 \cdot K_{12}^1 = \frac{\pi^2 + 6(\gamma - 1)^2}{\pi^2 k} + \frac{\pi^2 - 6\gamma(\gamma - 1)}{\pi^2 k} = \frac{2\pi^2 - 6\gamma + 6}{\pi^2 k}, \tag{65}$$

$${}^{(1)}\Gamma_{12}^2 = {}^{(1)}\Gamma_{21}^2 := {}^{(0)}\Gamma_{12}^2 + 1 \cdot K_{12}^2 = -\frac{6(\gamma - 1)k}{\pi^2 \lambda} - k \cdot \frac{\pi^2 - 6\gamma}{\pi^2 \lambda} = -\frac{(\pi^2 - 6)k}{\pi^2 \lambda}, \tag{66}$$

$$\begin{aligned}
{}^{(1)}\Gamma_{22}^1 &:= {}^{(0)}\Gamma_{22}^1 + 1 \cdot K_{22}^1 \\
&= \lambda \cdot \frac{(\pi^2 + 6(\gamma-1)^2)(\gamma-1)}{\pi^2 k^3} - \lambda \cdot \frac{\pi^2(\pi^2 + 12\gamma - 18) - 6(\gamma-1)(\pi^2\gamma + 12\zeta(3) - 6(\gamma^2 - 1))}{6\pi^2 k^3} \\
&= -\lambda \cdot \frac{\pi^4 - 12\pi^2 + 6\pi^2\gamma(2-\gamma) + 72(\gamma-1)^2 - 72\zeta(3)(\gamma-1)}{6\pi^2 k^3}, \tag{67}
\end{aligned}$$

$$\begin{aligned}
{}^{(1)}\Gamma_{22}^2 &:= {}^{(0)}\Gamma_{22}^2 + 1 \cdot K_{22}^2 \\
&= \frac{\pi^2 + 6(\gamma-1)^2}{\pi^2 k} + \frac{\pi^2(5-2\gamma) + 6(\gamma^2 - 1) - 12\zeta(3)}{\pi^2 k} = \frac{2\pi^2(2-\gamma) + 12(\gamma-1) - 12\zeta(3)}{\pi^2 k}. \tag{68}
\end{aligned}$$

The components of the 1-curvature tensor take the form:

$$\begin{aligned}
{}^{(1)}R_{1212} &= -{}^{(1)}R_{1221} = {}^{(1)}R_{2121} = -{}^{(1)}R_{2112} = {}^{(1)}R_{121}^1 g_{12} + {}^{(1)}R_{121}^2 g_{22} \\
&= \frac{-\pi^4 + 6\pi^2(\gamma+1) - 36(\gamma-1)}{\pi^4 \lambda} \cdot \frac{\gamma-1}{\lambda} + \frac{6k^2(6-\pi^2)}{\pi^4 \lambda^2} \cdot \frac{\pi^2 + 6(\gamma-1)^2}{6k^2} \\
&= \frac{-\pi^2\gamma + 12\gamma - 6}{\pi^2 \lambda^2}. \tag{69}
\end{aligned}$$

Therefore, the curvature of 1-connection has a constant value

$$k^{(1)} = -2 \cdot \det I^{-1} \cdot {}^{(1)}R_{1212} = -2 \cdot \frac{6\lambda^2}{\pi^2} \cdot \frac{-\pi^2\gamma + 12\gamma - 6}{\pi^2 \lambda^2} = \frac{12\pi^2\gamma - 144\gamma + 72}{\pi^4}. \tag{70}$$

□

6 Comparison of Curvatures

Finally, we combine the results obtained above into a summary Table 1.

Table 1 α -curvatures of two-dimensional statistical models ($\alpha = 0; 1; 2$)

∇^α	Normal	Logistic	Pareto	Weibull
$\alpha = 0$	$-0, 5$	$\approx -1, 4$	-2	$\approx -1, 22$
$\alpha = 1$	0		-4	$\approx 0, 59$
$\alpha = 2$	$1, 5$	$\approx -0, 98$	-6	

References

1. Lauritzen, S.: Conjugate connections in statistical theory. In: Dodson, C.T.J. (ed.) Geometrization of Statistical Theory, pp. 33–51. ULDM Publications, Lancaster (1987)
2. Amari, S.: Information Geometry and Its Applications. Applied Mathematical Sciences, vol. 194. Springer, Japan (2016). <https://doi.org/10.1007/978-4-431-55978-8>
3. Noguchi, M.: Geometry of statistical manifolds. *Diff. Geom. Appl.* **2**, 197–222 (1992). [https://doi.org/10.1016/0926-2245\(92\)90011-B](https://doi.org/10.1016/0926-2245(92)90011-B)
4. Ivanova, R.: A geometric observation on four statistical parameter spaces. *Tensor, N.S.* **72**, 188–195 (2010)
5. Arwini, K., Dodson, C.T.J.: Information Geometry: Near Randomness and Near Independence. Lecture Notes in Mathematics, vol. 1953. Springer, Berlin (2008). <https://doi.org/10.1007/978-3-540-69393-2>
6. Peng, L., Sun, H., Jiu, L.: The geometric structure of the Pareto distribution. *Bol. Asoc. Math. Venez.* **XIV**(1–2), 5–13 (2007)
7. Rylov, A.: Amari–Chentsov connections on the logistic model. *Izv. PGPU Belinskogo* **26**, 195–206 (2011). (in Russian)
8. Rylov, A.: Constant curvature connections on the Pareto statistical model. *Izv. PGPU Belinskogo* **30**, 155–163 (2012). (in Russian)
9. Rylov, A.: Connections compatible with a metric and statistical manifolds. *Russ. Math.* **36**(12), 47–56 (1992)
10. Rylov, A.: Connections compatible with a Riemannian metric in the theory of statistical manifolds. *Russ. Math.* **38**(3), 60–62 (1994)

Relation Between the Kantorovich–Wasserstein Metric and the Kullback–Leibler Divergence



Roman V. Belavkin

Abstract We discuss a relation between the Kantorovich–Wasserstein (KW) metric and the Kullback–Leibler (KL) divergence. The former is defined using the optimal transport problem (OTP) in the Kantorovich formulation. The latter is used to define entropy and mutual information, which appear in variational problems to find optimal channel (OCP) from the rate distortion and the value of information theories. We show that OTP is equivalent to OCP with one additional constraint fixing the output measure, and therefore OCP with constraints on the KL-divergence gives a lower bound on the KW-metric. The dual formulation of OTP allows us to explore the relation between the KL-divergence and the KW-metric using decomposition of the former based on the law of cosines. This way we show the link between two divergences using the variational and geometric principles.

Keywords Kantorovich metric · Wasserstein metric · Kullback–Leibler divergence · Optimal transport · Rate distortion · Value of information

1 Introduction

The study of the optimal transport problem (OTP), initiated by Gaspar Monge [10], was advanced greatly when Leonid Kantorovich reformulated the problem in the language of probability theory [8]. Let X and Y be two measurable sets, and let $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ be the sets of all probability measures on X and Y respectively, and let $\mathcal{P}(X \times Y)$ be the set of all joint probability measures on $X \times Y$. Let $c : X \times Y \rightarrow \mathbb{R}$ be a non-negative measurable function, which we shall refer to as the *cost function*. Often one takes $X \equiv Y$ and $c(x, y)$ to be a metric. We remind that when X is a complete and separable metric space (or if it is a homeomorphic image of it), then all probability measures on X are Radon (i.e. inner regular).

R. V. Belavkin (✉)
Middlesex University, London NW4 4BT, UK
e-mail: R.Belavkin@mdx.ac.uk

The expected cost with respect to probability measure $w \in \mathcal{P}(X \times Y)$ is the integral¹:

$$\mathbb{E}_w\{c\} := \int_{X \times Y} c(x, y) dw(x, y)$$

It is often assumed that the cost function is such that the above integral is lower semicontinuous or closed functional of w (i.e. the set $\{w : \mathbb{E}_w\{c\} \leq v\}$ is closed for all $v \in \mathbb{R}$). In particular, this is the case when $c(w) := \mathbb{E}_w\{c\}$ is a continuous linear functional.

Given two probability measures $q \in \mathcal{P}(X)$ and $p \in \mathcal{P}(Y)$, we denote by $\Gamma[q, p]$ the set of all joint probability measures $w \in \mathcal{P}(X \times Y)$ such that their marginal measures are $\pi_X w = q$ and $\pi_Y w = p$:

$$\Gamma[q, p] := \{w \in \mathcal{P}(X \times Y) : \pi_X w = q, \pi_Y w = p\}$$

Kantorovich's formulation of OTP is to find optimal joint probability measure in $\Gamma[q, p]$ minimizing the expected cost $\mathbb{E}_w\{c\}$. The optimal joint probability measure $w \in \mathcal{P}(X \times Y)$ (or the corresponding conditional probability measure $dw(y | x)$) is called the *optimal transportation plan*. The corresponding optimal value is often denoted

$$K_c[p, q] := \inf \{\mathbb{E}_w\{c\} : w \in \Gamma[q, p]\} \quad (1)$$

The non-negative value above allows one to compare probability measures, and when the cost function $c(x, y)$ is a metric on $X \equiv Y$, then $K_c[p, q]$ is a metric on the set $\mathcal{P}(X)$ of all probability measures on X , and it is often called the *Wasserstein metric* due to a paper by Dobrushin [7, 15], even though it was introduced much earlier by Kantorovich [8]. Thus, we shall refer to it as the Kantorovich–Wasserstein (KW) metric. It is known that the KW-metric (or the related to it Kantorovich–Rubinstein metric) induces a topology on $\mathcal{P}(X)$ equivalent to the weak topology [6].

Another important functional used to compare probability measures is the Kullback–Leibler divergence [9]:

$$D[p, q] := \int_X \left[\ln \frac{dp(x)}{dq(x)} \right] dp(x) \quad (2)$$

where it is assumed that p is absolutely continuous with respect to q (otherwise the divergence can be infinite). It is not a metric, because it does not satisfy the symmetry and the triangle axioms, but it is non-negative, $D[p, q] \geq 0$, and $D[p, q] = 0$ if and only if $p = q$. The KL-divergence has a number of useful and sometimes unique to it properties (e.g. see [4] for an overview), and it plays an important role in physics and

¹We use the integral notation as it is more common in the optimal transport literature, and it can cover both countable and uncountable cases (indeed, a sum is an integral with respect to the counting measure). The spaces are infinite dimensional in general.

information theory, because entropy and Shannon's information are defined using the KL-divergence.

The main question that we discuss in this paper is whether these two, seemingly unrelated divergences have anything in common. In the next section, we recall some definitions and properties of the KL-divergence. Then we show that the optimal transport problem (OTP) has an implicit constraint, which allows us to relate OTP to variational problems of finding an optimal channel (OCP) that were studied in the rate distortion and the information value theories [12, 13]. Using the fact that OCP has fewer constraints than OTP, we show that OCP defines a lower bound on the Kantorovich metric, and it depends on the KL-divergence. Then we consider the dual formulation of the OTP and introduce an additional constraint, which allows us to define another lower bound on the Kantorovich metric. We then show that the KL-divergence can be decomposed into a sum, one element of which is this lower bound on the Kantorovich metric.

2 Entropy, Information and the Optimal Channel Problem

Entropy and Shannon's mutual information are defined using the KL-divergence. In particular, *entropy* of probability measure $p \in \mathcal{P}(X)$ relative to a reference measure r is defined as follows:

$$\begin{aligned} H[p/r] &:= - \int_X \left[\ln \frac{dp(x)}{dr(x)} \right] dp(x) \\ &= \ln r(X) - D[p, r/r(X)] \end{aligned}$$

where the second line is written assuming that the reference measure is finite $r(X) < \infty$. It shows that entropy is equivalent up to a constant $\ln r(X)$ to negative KL-divergence from a normalized reference measure. The entropy is usually defined with respect to some Haar measure as a reference, such as the counting measure (i.e. $r(E) = |E|$ for $E \subseteq X$ or $dr(x) = 1$). We shall often write $H[p]$ instead of $H[p/r]$, if the choice of a reference measure is clear (e.g. $dr(x) = 1$ or $dr(x) = dx$). We shall also use notation $H_p(x)$ and $H_p(x | y)$ to distinguish between the prior and conditional entropies.

Shannon's mutual information between two random variables $x \in X$ and $y \in Y$ is defined as the KL-divergence of a joint probability measure $w \in \mathcal{P}(X \times Y)$ from a product $q \otimes p \in \mathcal{P}(X \times Y)$ of the marginal measures $\pi_X w = q$ and $\pi_Y w = p$:

$$\begin{aligned} I(x, y) &:= D[w, q \otimes p] = \int_{X \times Y} \left[\ln \frac{dw(x, y)}{dq(x) dp(y)} \right] dw(x, y) \\ &= H_q(x) - H_q(x | y) = H_p(y) - H_p(y | x) \end{aligned}$$

The second line shows that mutual information can be represented by the differences of entropies and the corresponding conditional entropies (i.e. computed respectively using the marginal $dp(y)$ and conditional probability measures $dp(y | x)$). If both unconditional and conditional entropies are non-negative (this is always possible with a proper choice of a reference measure), then we have inequalities $H_q(x | y) \leq H_q(x)$ and $H_p(y | x) \leq H_p(y)$, because their differences (i.e. mutual information $I(x, y)$) is non-negative. In this case, mutual information satisfies Shannon's inequality:

$$0 \leq I(x, y) \leq \min[H_q(x), H_p(y)]$$

Thus, entropy can be defined as the supremum of information or as self-information [5]:

$$\sup_w \{I(x, y) : \pi_X w = q\} = I(x, x) = H_q(x)$$

Here, we assume that $H_q(x | x) = 0$ for the entropy of elementary conditional probability measure $q(E | x) = \delta_x(E)$, $E \subseteq X$. Let us now consider the following variational problem.

Given probability measure $q \in \mathcal{P}(X)$ and cost function $c : X \times Y \rightarrow \mathbb{R}$, find optimal joint probability measure $w = w(\cdot | x) \otimes q \in \mathcal{P}(X \times Y)$ minimizing the expected cost $\mathbb{E}_w\{c\}$ subject to the constraint on mutual information $I(x, y) \leq \lambda$. Because the marginal measure $\pi_X w = q$ is fixed, this problem is really to find an optimal conditional probability $dw(y | x)$, which we refer to as the *optimal channel*. We shall denote the corresponding optimal value as follows:

$$R_c[q](\lambda) := \inf \{\mathbb{E}_w\{c\} : I(x, y) \leq \lambda, \pi_X w = q\} \quad (3)$$

This problem was originally studied in the rate distortion theory [12] and later in the value of information theory [13]. The value of Shannon's mutual information is defined simply as the difference:

$$V(\lambda) := R_c[q](0) - R_c[q](\lambda)$$

It represents the maximum gain (in terms of reducing the expected cost) that is possible due to obtaining λ amount of mutual information.

The optimal solution to OCP has the following general form (see [3, 14] for details):

$$dw_{OCP}(x, y) = dq(x) dp(y) e^{-\beta c(x, y) - \kappa(\beta, x)} \quad (4)$$

where measures q and p are the marginals of w_{OCP} (not necessarily coinciding with the marginals of w_{OTP}), and the exponent β , sometimes called the *inverse temperature*, is the inverse of the Lagrange multiplier β^{-1} defined from the information constraint by the equality $I(x, y) = \lambda$. In fact, one can show that $\beta^{-1} = dV(\lambda)/d\lambda$, where $V(\lambda)$ is the value of information.

The normalizing function $\kappa(\beta, x) = \ln \int_Y e^{-\beta c(x, y)} dp(y)$ depends on x , and the solution (4) in general depends on the marginal measure $q \in \mathcal{P}(X)$. One can show, however, that if the cost function is translation invariant (i.e. $c(x + a, y + a) = c(x, y)$), then the function $dq(x) e^{-\kappa(\beta, x)} =: e^{-\kappa_0(\beta)}$ does not depend on x , which gives a simplified expression:

$$dw_{OCP}(x, y) = dp(y) e^{-\beta c(x, y) - \kappa_0(\beta)}$$

The measure above does not depend on the input marginal measure $q \in \mathcal{P}(X)$ explicitly, but only via its influence on the output measure $p \in \mathcal{P}(Y)$. Note, however, that although it is always possible to choose the same input measure $q \in \mathcal{P}(X)$ in OCP and OTP, because the output measure $p \in \mathcal{P}(Y)$ in OCP depends on the input measure, the output measures in OCP and OTP are usually different and $w_{OCP} \neq w_{OTP}$.

Let us compare the optimal values (1) of the Kantorovich’s OTP and (3) of the OCP. On one hand, the OCP problem has only one marginal constraint $\pi_X w = q$, while the OTP has two constraints $\pi_X w = q$ and $\pi_Y w = p$. On the other hand, the OCP has an information constraint $I(x, y) \leq \lambda$. Notice, however, that because fixing marginal measures q and p also fixes the values of their entropies $H_q(x)$ and $H_p(y)$, the OTP has information constraint implicitly, because mutual information is bounded above $I(x, y) \leq \min[H_q(x), H_p(y)]$ by the entropies. Therefore, in reality the OTP differs from OCP only by one extra constraint — fixing the output measure $\pi_Y w = p$. Let us define the following extended version of the OTP by introducing the information constraint explicitly:

$$K_c[p, q](\lambda) := \inf \{\mathbb{E}_w\{c\} : I(x, y) \leq \lambda, \pi_X w = q, \pi_Y w = p\}$$

For $\lambda = \min[H_q(x), H_p(y)]$ one recovers the original value $K_c[p, q]$ defined in (1). It is also clear that the following inequality holds for any λ :

$$R_c[q](\lambda) \leq K_c[p, q](\lambda)$$

In fact, the equality holds if and only if the solutions to both problems coincide.

Theorem 1 *Let w_{OCP} and $w_{OTP} \in \mathcal{P}(X \times Y)$ be optimal solutions to OCP and OTP respectively with the same information constraint $I(x, y) \leq \lambda$. Then $R_c[q](\lambda) = K_c[p, q](\lambda)$ if and only if $w_{OCP} = w_{OTP} \in \Gamma[p, q]$.*

Proof Measure w_{OCP} is a solution to OCP if and only if it is an element $w_{OCP} \in \partial D^*[-\beta c, q \otimes p]$ of subdifferential² of a closed convex functional

$$D^*[u, q \otimes p] = \ln \int_{X \times Y} e^{u(x, y)} dq(x) dp(y)$$

²The set of all elements of a dual space defining support hyperplanes of a functional at a point. The elements of subdifferential are called *subgradients*. If subdifferential contains only one element defining a unique support hyperplane, then it is called the *gradient* [11].

evaluated at function $u(x, y) = -\beta c(x, y)$. This can be shown using the standard method of Lagrange multipliers (e.g. see [3, 14]). The functional $D^*[u, q \otimes p]$ is the Legendre–Fenchel transform of the KL-divergence $D[w, q \otimes p]$ considered as a functional in the first variable $w \in \mathcal{P}(X \times Y)$:

$$D^*[u, q \otimes p] = \sup\{\mathbb{E}_w\{u\} - D[w, q \otimes p]\}$$

If there is another optimal measure w_{OTP} achieving the same optimal value, then it also must be an element of the subdifferential $\partial D^*[-\beta c, q \otimes p]$, as well as any convex combination $(1-t)w_{OCP} + tw_{OTP}$, $t \in [0, 1]$, because subdifferential is a convex set. But this means that the KL-divergence $D[w, q \otimes p]$, the dual of $D^*[u, q \otimes p]$, is not strictly convex, which is false. \square

The equality $w_{OTP} = w_{OCP}$ means that the solution to OTP has special form (4), and, as discussed earlier, this is not typical, because the output measure $p \in \mathcal{P}(Y)$ in OCP (and generally the solution w_{OCP}) depends on the input measure $q \in \mathcal{P}(X)$. Thus, in most cases $w_{OCP} \neq w_{OTP}$ implying strict inequality $R[q](\lambda) < K[p, q](\lambda)$ between the optimal values, and from game-theoretic point of view the optimal channels should be preferred to optimal transportation plans.

Also, $w_{OCP} \neq w_{OTP}$ means that w_{OTP} is not on the boundary of the set $\{w : \pi_X w = q, I(x, y) \leq \lambda\}$, so that the amount of mutual information $I(x, y)$ for w_{OTP} is strictly less than the constraint λ . In this case, conditional entropies $H_q(x | y)$ and $H_p(y | x)$ are not minimized, and the optimal transportation plan is mixed (i.e. w_{OTP} does not correspond to any function $f : X \rightarrow Y$ pushing measure $q \in \mathcal{P}(X)$ to $p \in \mathcal{P}(Y)$).

If the constraint $I(x, y) \leq \lambda$ is relaxed to the supremum of mutual information $\lambda = I(x, x) = H_q(x)$, then the solution w_{OCP} is defined as the limit of the exponential form (4) for $\beta \rightarrow \infty$. The optimal channel in this case corresponds to a function $f : X \rightarrow Y$ minimizing the cost function $c(x, f(x))$ for each $x \in X$, and the optimal value is the infimum

$$J_c[q](\lambda = H_q(x)) = \inf_{f:X \rightarrow Y} \int_X c(x, f(x)) dq(x)$$

If the cost function $c(x, y)$ is a metric, then $J_c[q](\lambda = H_q(x)) = 0$ corresponding to $f(x) = x$ and $p = q$. Note that $K_c[p, q] = 0$ if and only if $p = q$ for the KW-metric. From this point of view, the KW-metric represents the amount by which the output measure of the optimal channel w_{OCP} differs from the output measure of the transportation plan w_{OTP} (i.e. $\pi_Y w_{OTP} \neq \pi_Y w_{OCP}$). The optimal channel, therefore, can potentially be useful in the analysis of the optimal transportation.

Finally, let us point out in this section that the KL-divergence $D[p, q]$ between the marginal measures $p, q \in \mathcal{P}(X)$ can be related to mutual information via *cross-information*:

$$D[w, q \otimes q] = \underbrace{D[w, q \otimes p]}_{I(x, y)} + D[p, q] \quad (5)$$

The term cross-information for the KL-divergence $D[w, q \otimes q]$ (notice the difference from mutual information $D[w, q \otimes p]$) was introduced in [5] by analogy with cross-entropy. The expression (5) is a special case of Pythagorean theorem for the KL-divergence. As was shown in [2], a joint probability measure $w \in \mathcal{P}(X \times Y)$ together with its marginals $\pi_X w = q$ and $\pi_Y w = p$ defines a triangle $(w, q \otimes p, q \otimes q)$ in $\mathcal{P}(X \times Y)$, which is always a right triangle (and the same holds for triangle $(w, q \otimes p, p \otimes p)$). This means that the KL-divergence between the marginal measures q and p can be expressed as the difference:

$$D[p, q] = D[w, q \otimes q] - I(x, y)$$

Therefore, the constraint $I(x, y) \leq \lambda$ on mutual information can be expressed using the KL-divergence between the marginal measures and cross-information, and the inequality $K_c[p, q](\lambda) \geq R_c[q](\lambda)$ can be written as follows

$$K_c[p, q](\lambda) \geq \inf \{\mathbb{E}_w\{c\} : D[\pi_Y w, q] \geq D[w, q \otimes q] - \lambda, \pi_X w = q\}$$

The right-hand-side is the value $R_c[q](\lambda)$ of the OCP. Note that the marginal measures on the right are for the joint measure w_{OCP} , which may not coincide with the solution w_{OTP} defining the value $K_c[p, q](\lambda)$. In the case when OCP and OTP have the same solution $w \in \Gamma[p, q]$, then $p = \pi_Y w_{OCP}$, and the inequality above becomes equality relating the KW-metric and the KL-divergence in one expression.

3 Dual Formulation of the Optimal Transport Problem

Kantorovich's great achievement was the dual formulation of the optimal transport problem way before the development of convex analysis and the duality theory. Given a cost function $c : X \times Y \rightarrow \mathbb{R}$ consider real functions $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$ satisfying the condition: $f(x) - g(y) \leq c(x, y)$ for all $(x, y) \in X \times Y$. Then the dual formulation is the following maximization over all such functions:

$$J_c[p, q] := \sup \{\mathbb{E}_p\{f\} - \mathbb{E}_q\{g\} : f(x) - g(y) \leq c(x, y)\} \quad (6)$$

where we assumed $X \equiv Y$. It is clear that the following inequality holds:

$$J_c[p, q] \leq K_c[p, q]$$

We shall attempt to use this dual formulation to find another relation between the KL-divergence and the KW-metric. First, consider the following decomposition of the KL-divergence:

$$D[p, q] = D[p, r] + D[r, q] - \int_X \ln \frac{dq(x)}{dr(x)} [dp(x) - dr(x)] \quad (7)$$

$$= D[p, r] - D[q, r] - \int_X \ln \frac{dq(x)}{dr(x)} [dp(x) - dq(x)] \quad (8)$$

Equation (7) is the *law of cosines* for the KL-divergence (e.g. see [2]). It can be proved either by second order Taylor expansion in the first argument or directly by substitution. Equation (8) can be proved by using the formula:

$$D[q, r] + D[r, q] = \int_X \ln \frac{dq(x)}{dr(x)} [dq(x) - dr(x)]$$

We now consider functions $f(x) - g(y) \leq c(x, y)$ satisfying additional constraints:

$$\begin{aligned} \beta f(x) &= \nabla_p D[p, r] = \ln \frac{dp(x)}{dr(x)}, & \beta \geq 0 \\ \alpha g(x) &= \nabla_q D[q, r] = \ln \frac{dq(x)}{dr(x)}, & \alpha \geq 0 \end{aligned}$$

where by $\nabla D[\cdot, r]$ we denoted the gradient³ of the KL-divergence considered as a functional of the first variable. Thus, βf and αg are the gradients of divergences $D[p, r]$ and $D[q, r]$ respectively, and this means that probability measures $p, q \in \mathcal{P}(X)$ have the following exponential representations:

$$\begin{aligned} dp(x) &= e^{\beta f(x) - \kappa[\beta f]} dr(x) \\ dq(x) &= e^{\alpha g(x) - \kappa[\alpha g]} dr(x) \end{aligned}$$

where $\kappa[(\cdot)] = \ln \int_X e^{(\cdot)} dr(x)$ is the normalizing constant (the value of the cumulant generating function). One can show that

$$\begin{aligned} \frac{d}{d\beta} \kappa[\beta f] &= \mathbb{E}_p\{f\}, & D[p, r] &= \beta \mathbb{E}_p\{f\} - \kappa[\beta f] \\ \frac{d}{d\alpha} \kappa[\alpha g] &= \mathbb{E}_q\{g\}, & D[q, r] &= \alpha \mathbb{E}_q\{g\} - \kappa[\alpha g] \end{aligned}$$

Substituting these formulae into (8) we obtain

$$D[p, q] = \beta \mathbb{E}_p\{f\} - \alpha \mathbb{E}_q\{g\} - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int_X g(x) [dp(x) - dq(x)]$$

Let us define the following value:

³Gradient here is understood in the usual sense of convex analysis [11] as an element of the dual space defining the unique support hyperplane (i.e. it is the unique element of subdifferential).

$$J_{c,\varepsilon}[p, q] := \frac{1}{\varepsilon} [\beta \mathbb{E}_p\{f\} - \alpha \mathbb{E}_q\{g\}]$$

where $\varepsilon = \inf\{\epsilon \geq 0 : \beta f(x) - \alpha g(y) \leq \epsilon c(x, y)\}$. The value above reminds the value $J_c[p, q]$ of the dual problem to OTP, defined in (6). However, because we also require that functions f and g to satisfy additional constraints (the gradient conditions), we have the following inequality:

$$J_{c,\varepsilon}[p, q] \leq J_c[p, q] \leq K_c[p, q]$$

Using these inequalities, we can rewrite Eq. (8) as follows:

$$D[p, q] \leq \varepsilon K_c[p, q] - (\kappa[\beta f] - \kappa[\alpha g]) - \alpha \int g(x) [dp(x) - dq(x)]$$

Theorem 2 *Let the pair of functions (f, g) be the solution to the dual OTP (6). If there exists a reference measure $r \in \mathcal{P}(X)$ such that $f = \nabla_p D[p, r]$ and $g = \nabla_q D[q, r]$, then*

$$D[p, q] = K_c[p, q] - (\kappa[f] - \kappa[g]) - \int g(x) [dp(x) - dq(x)]$$

Proof The assumptions $f = \nabla_p D[p, r]$ and $g = \nabla_q D[q, r]$ mean that the Lagrange multipliers are $\alpha = \beta = 1$, and probability measures have the form $p = \exp(f - \kappa[f])r$ and $q = \exp(g - \kappa[g])r$. Substituting these expressions into Eq. (8) will result in the expression containing the difference of expectations $\mathbb{E}_p\{f\} - \mathbb{E}_q\{g\}$, which equals to $J_c[p, q] = K_c[p, q]$. \square

The right-hand-side of the equation in the above theorem uses the value $K_c[p, q]$ of the KW-metric that depends on the cost function $c(x, y)$, but the left-hand-side is the KL-divergence $D[p, q]$, which does not depend on the cost function explicitly. However, the condition of the theorem effectively requires that measures p and q are expressed as exponential families using functions f and g satisfying the condition $f(x) - g(y) \leq c(x, y)$, and therefore the cost function appears implicitly in the left-hand-side.

Related Work and Discussion

After this manuscript had been submitted, another article on a closely related topic considering the Wasserstein distance and the KL-divergence was published [1]. In this work, the authors considered a relaxed version of OTP, where both marginals were fixed $q = \pi_X w$, $p = \pi_Y w$, but with an additional constraint on entropy $H_w(x, y)$ of the joint distribution $w \in \Gamma[p, q]$. We note that the value of this relaxed OTP is equivalent to the value $K_c[p, q](\lambda)$ of a generalized OTP, defined in this paper.

Indeed, mutual information can be expressed by the difference $I(x, y) = H_q(x) + H_p(y) - H_w(x, y)$. Thus, because fixing both marginals also fixes their entropies $H_q(x)$ and $H_p(y)$, varying the entropy $H_w(x, y)$ of joint distribution is equivalent to varying mutual information. The main focus in [1] is on the relation between the KL-divergence $D[w, q \otimes p]$ for joint distributions w and $q \otimes p$ and the Wasserstein metric $K_c[p, q]$ between the marginals $q = \pi_X w$ and $p = \pi_Y w$. On the other hand, here we considered both the KL-divergence $D[p, q]$ and the Wasserstein metric $K_c[p, q]$ between the marginal measures.

We have demonstrated that by relaxing one constraint, namely fixing the output measure, the optimal transport problem becomes mathematically equivalent to the optimal channel problem in information theory, which uses a constraint on the KL-divergence between the joint and the product of marginal measures (i.e. on mutual information). This way, an optimal channel defines a lower bound on the KW-metric. One could argue that for this reason optimal channels should be preferred to optimal transportation plans purely from a game-theoretic point of view. Applying Pythagorean theorem for joint and product of marginal measures allowed us to relate the constraint on mutual information to the constraint on the KL-divergence between the marginal measures of the optimal channel.

In addition to this variational approach, we have considered a geometric idea based on the law of cosines for the KL-divergence to decompose the divergence between two probability measures into a sum that includes divergences from a third reference measure. We have shown then that a component of this decomposition can be related to the dual formulation of the optimal transport problem. Generally, the relations presented have a form of inequalities. Additional conditions have been derived in Theorems 1 and 2 for the cases when the relations hold with equalities.

Acknowledgements This work is dedicated to the anniversary of Professor Shun-ichi Amari, one of the founders of information geometry. The work was supported in part by the Biotechnology and Biological Sciences Research Council [grant numbers BB/L009579/1, BB/M021106/1].

References

1. Amari, S.I., Karakida, R., Oizumi, M.: Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem. *Information Geometry* (2018)
2. Belavkin, R.V.: Law of cosines and Shannon-Pythagorean theorem for quantum information. In: Nielsen, F., Barbaresco, F. (eds.) *Geometric Science of Information*. Lecture Notes in Computer Science, vol. 8085, pp. 369–376. Springer, Heidelberg (2013)
3. Belavkin, R.V.: Optimal measures and Markov transition kernels. *J. Glob. Optim.* **55**, 387–416 (2013)
4. Belavkin, R.V.: Asymmetric topologies on statistical manifolds. In: Nielsen, F., Barbaresco, F. (eds.) *Geometric Science of Information*. Lecture Notes in Computer Science, vol. 9389, pp. 203–210. Springer International Publishing, Berlin (2015)
5. Belavkin, R.V.: On variational definition of quantum entropy. In: A. Mohammad-Djafari, F. Barbaresco (eds.) *AIP Conference Proceedings of Bayesian Inference and Maximum Entropy*

- Methods in Science and Engineering (MAXENT 2014), Clos Lucé, Amboise, France, vol. 1641, p. 197 (2015)
- 6. Bogachev, V.I.: Measure theory, vol. I, II, Chap. xviii, xiv, pp. 500, 575. Springer, Berlin (2007)
 - 7. Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions. Theory Probab. Appl. **15**(3), 458–486 (1970)
 - 8. Kantorovich, L.V.: On translocation of masses. USSR AS Doklady **37**(7–8), 227–229 (1942). (in Russian). English translation. J. Math. Sci. **133**(4), 1381–1382 (2006)
 - 9. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
 - 10. Monge, G.: Mémoire sur la théorie des déblais et de remblais. Histoire de l’Academie Royale des Sciences avec les Mémoires de Mathématique & de Physique, Paris (1781)
 - 11. Rockafellar, R.T.: Conjugate Duality and Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 16. Society for Industrial and Applied Mathematics, PA (1974)
 - 12. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
 - 13. Stratonovich, R.L.: On value of information. Izv. USSR Acad. Sci. Techn. Cybern. **5**, 3–12 (1965). (In Russian)
 - 14. Stratonovich, R.L.: Information Theory. Sovetskoe Radio, Moscow (1975). (In Russian)
 - 15. Vasershtein, L.N.: Markov processes over denumerable products of spaces describing large system of automata. Probl. Inform. Transm. **5**(3), 47–52 (1969)

Part IV

Quantum Information Geometry

Some Inequalities for Wigner–Yanase Skew Information



Shunlong Luo and Yuan Sun

Abstract The purpose of this work is twofold: On one hand, we review concisely some important features of the Wigner–Yanase skew information and its intrinsic relations with other information quantities such as the Fisher information and entropy. We focus on several significant and widely used inequalities related to convexity, superadditivity, monotonicity, and uncertainty relations. On the other hand, we derive some novel inequalities concerning monotonicity of the skew information under quantum measurements. Some applications and implications to quantum information theory, including quantum coherence, quantum uncertainty, and quantum correlations, are discussed.

Keywords Fisher information · Skew information · Convexity · Monotonicity · Quantum coherence · Quantum correlations

1 Introduction

In 1950s, Wigner initiated the study of restrictions to quantum measurements imposed by the presence of conserved quantities, and discovered that observables which do not commute with the conserved quantities are more difficult to measure than commutative ones [1, 2]. Subsequently, Araki and Yanase proved that observables not commuting with the conserved quantities cannot be measured precisely [3], and Yanase further established a trade-off relation between the measurement accuracy and measuring apparatus size [4]. These constitute the contents of the

S. Luo (✉) · Y. Sun
Academy of Mathematics and Systems Science, Chinese Academy of Sciences,
Beijing 100190, China
e-mail: luosl@amt.ac.cn

S. Luo · Y. Sun
School of Mathematical Sciences, University of Chinese Academy of Sciences,
Beijing 100049, China

celebrated Wigner–Araki–Yanase theorem, which is a fundamental result in quantum measurement theory, and is further explored and exploited by many authors [5–29].

In 1963, Wigner and Yanase introduced a quantity, which they called the skew information, to address the quantum measurement problem with conserved quantities in a more quantitative way [30]. Many remarkable properties of the skew information are then revealed [31–34], various generalizations and extensive applications are studied [35–75].

The present work is devoted to providing a concise review of various inequalities related to the skew information, establishing some novel ones, and elucidating several applications of the skew information in quantum information theory. The paper is organized as follows. In Sect. 2, we illuminate the formal analogy between Fisher information and skew information, which indicates why the skew information, as a particular and significant version of quantum Fisher information, has so many nice information theoretic properties. In Sect. 3, we review various inequalities for the skew information, including convexity, superadditivity, uncertainty relations, and monotonicity under quantum operations. Several new inequalities concern monotonicity are also established. In Sect. 4, we review some usages of the skew information in quantum information theory, including the quantification of quantum coherence, quantum uncertainties, and quantum correlations. Finally, we summarize in Sect. 5.

2 Skew Information as Quantum Fisher Information

Recall that skew information was introduced by Wigner and Yanase in seeking a reasonable information measure with the following desiderata in mind [30].

- (i) *Convexity*: If two different states are mixed probabilistically, then the information content of the resulting mixture should be smaller than the average information content of the components.
- (ii) *Additivity*: The information content of the union of two systems should be the sum of the information contents of the components.
- (iii) *Invariance*: The information content of an isolated system should be independent of time.
- (iv) *Superadditivity*: When a joint system is separated into two parts, the information content should, in general, decrease.
- (v) *Monotonicity*: One should investigate the changes on the information content as a result of measurements.

As an analogy of minus von Neumann entropy, Wigner and Yanase defined the information content of a state ρ with respect to observables not commuting with (i.e., skew to) a conserved quantity H (an Hermitian operator) as

$$I(\rho, H) = -\frac{1}{2} \text{tr}[\sqrt{\rho}, H]^2, \quad (1)$$

where $[\cdot, \cdot]$ denotes commutator between two operators. This notion has played an important role in quantum information theory [30–75]. In order to put this quantity in perspective and to gain an intuitive understanding of it, we first recall the notion of classical Fisher information, and then elucidate that the skew information is a natural quantum analogue of the classical Fisher information [38, 39].

Parameter estimation is one of the most basic tasks in statistics and information theory [76–80]. Fisher information, first introduced by Fisher in 1925 [76], is an intrinsic measure of the amount of parameter information encoded in probability densities. It is now the central concept in the theory of statistical estimation. The celebrated Cramér–Rao inequality and the asymptotic normality of maximum likelihood estimation are both phrased in terms of the Fisher information.

Consider the following parameter estimation problem: suppose that $\{p_\theta : \theta \in \mathbb{R}\}$ is a parametric family of probability densities on \mathbb{R} , and we have observed independent and identically distributed samples x_1, x_2, \dots, x_N of p_θ . Our task is to estimate this θ as precisely as possible by use of the data observed. In this context, the Fisher information defined as [76, 77]

$$I_F(p_\theta) = \int_{\mathbb{R}} \left(\frac{\partial \log p_\theta(x)}{\partial \theta} \right)^2 p_\theta(x) dx \quad (2)$$

$$= 4 \int_{\mathbb{R}} \left(\frac{\partial \sqrt{p_\theta(x)}}{\partial \theta} \right)^2 dx \quad (3)$$

plays a crucial role. In particular, when $p_\theta(x) = p(x - \theta)$, i.e., θ is a translation parameter, then

$$I_F(p_\theta) = \int_{\mathbb{R}} \left(\frac{\partial \log p(x)}{\partial x} \right)^2 p(x) dx \quad (4)$$

$$= 4 \int_{\mathbb{R}} \left(\frac{\partial \sqrt{p(x)}}{\partial x} \right)^2 dx \quad (5)$$

is independent of θ , which may be regarded as the Fisher information of p . In this instance, we denote $I_F(p_\theta)$ by $I_F(p)$, which is the Fisher information of p with respect to the location parameter. Suppose that $\hat{\theta}$ is an unbiased estimator of θ , i.e., $E(\hat{\theta}) = \int_{\mathbb{R}} \hat{\theta}(x) p_\theta(x) dx = \theta$, then the celebrated Cramér–Rao inequality states that

$$\text{Var}(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2 \geq \frac{1}{I_F(p_\theta)}. \quad (6)$$

If n independent observations are carried out, then the Cramér–Rao inequality can be expressed as $\text{Var}(\hat{\theta}) \geq \frac{1}{n I_F(p_\theta)}$. This shows that Fisher information puts a fundamental

limit to the precision of parameter estimation. It is remarkable that the notion of Fisher information has enjoyed increasing popularity in an informational approach to physics [79–89].

The Fisher information is the unique monotone metric in the classical context, as established in Ref. [90]. However, when passing to the quantum scenario, there are many natural quantum extensions of the classical Fisher information [91], among which there are two distinguished ones. First, one can formally generalize Eq. (2). Motivated by

$$\frac{\partial}{\partial \theta} p_\theta = \frac{1}{2} \left(\frac{\partial \log p_\theta}{\partial \theta} p_\theta + p_\theta \frac{\partial \log p_\theta}{\partial \theta} \right), \quad (7)$$

one may formally introduce the quantum analogue of $\frac{\partial \log p_\theta}{\partial \theta}$ as L_θ (symmetric logarithmic derivative) determined by

$$\frac{\partial}{\partial \theta} \rho_\theta = \frac{1}{2} (L_\theta \rho_\theta + \rho_\theta L_\theta), \quad (8)$$

and define a quantum analogue of the classical Fisher information as

$$I_F(\rho_\theta) = \frac{1}{4} \text{tr} \rho_\theta L_\theta^2, \quad (9)$$

where the constant 1/4 is for simplicity and consistence with the definition in Ref. [39]. This generalization was first done by Helstrom [79], and plays an important role in quantum estimation. In particular, when $\rho_\theta = e^{-i\theta H} \rho e^{i\theta H}$, one has

$$I_F(\rho_\theta) = \frac{1}{4} \text{tr} \rho L^2 \quad (10)$$

which is independent of θ , and will be denoted by $I_F(\rho, H)$. Here L is determined by $i[\rho, H] = \frac{1}{2}(L\rho + \rho L)$.

Second, one can formally generalize Eq. (3). Replacing the integration by trace, the parametric probabilities p_θ by a parametric density operators ρ_θ on a Hilbert space, then one is naturally led to

$$I_W(\rho_\theta) = \text{tr} \left(\frac{\partial \sqrt{\rho_\theta}}{\partial \theta} \right)^2, \quad (11)$$

which may be regarded as a version of quantum Fisher information. In particular, when $\rho_\theta = e^{-i\theta H} \rho e^{i\theta H}$, then

$$I_W(\rho_\theta) = 2I(\rho, H) \quad (12)$$

is independent of θ , and is essentially the Wigner–Yanase skew information. It is known that [39]

$$I(\rho, H) \leq I_F(\rho, H) \leq 2I(\rho, H). \quad (13)$$

In the following, we mainly consider the skew information. There are several closely related interpretations of the skew information.

- (i) As the information content of ρ with respect to observables not commuting with (i.e., skew to) the conserved quantity H .
- (ii) As a measure quantifying the noncommutativity between ρ and H .
- (iii) As a version of quantum Fisher information with respect to the time parameter encoded in the evolution driven by the conserved quantity (Hamiltonian) H .
- (iv) As a measure of quantum uncertainty of H in the state ρ which is dual to (i).
- (v) As coherence of ρ with respect to H .
- (vi) As asymmetry of ρ relative to H .

The infinitesimal change of the Bures distance is the quantum Fisher information determined by the symmetric logarithmic derivative, while the infinitesimal change of the Hellinger distance is the quantum Fisher information determined by the commutator derivative, which happens to be the Wigner–Yanase skew information [40]: Let $D(\rho, \sigma) = \text{tr}(\sqrt{\rho} - \sqrt{\sigma})^2 = 2 - 2\text{tr}\sqrt{\rho}\sqrt{\sigma}$ be the quantum Hellinger distance and $D_b(\rho, \sigma) = 2 - 2\text{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}$ be the quantum Bures distance. Suppose $\rho_\theta = e^{-i\theta H}\rho e^{i\theta H}$ satisfies the Landau-von Neumann equation $i\frac{\partial\rho_\theta}{\partial\theta} = [H, \rho_\theta]$, $\rho_0 = \rho$, then (noting that here I_F is defined via Eq. (10), which differs from that in Ref. [40] by a factor 1/4)

- (i) $\frac{\partial^2 D(\rho_\theta, \rho_\xi)}{\partial\theta\partial\xi}|_{\theta=0, \xi=0} = 4I(\rho, H)$.
- (ii) $\frac{\partial^2 D(\rho_\theta, \rho)}{\partial\theta^2}|_{\theta=0} = 4I(\rho, H)$.
- (iii) $\frac{\partial^2 D_b(\rho_\theta, \rho_\xi)}{\partial\theta\partial\xi}|_{\theta=0, \xi=0} = 2I_F(\rho, H)$.
- (iv) $\frac{\partial^2 D_b(\rho_\theta, \rho)}{\partial\theta^2}|_{\theta=0} = 2I_F(\rho, H)$.

As mentioned by Wigner and Yanase [30], Dyson suggested the following more general quantity

$$I_\alpha(\rho, H) = -\frac{1}{2}\text{tr}[\rho^\alpha, H][\rho^{1-\alpha}, H], \quad 0 < \alpha < 1, \quad (14)$$

which is now termed the Wigner–Yanase–Dyson information, as a measure of information content of ρ with respect to H . This quantity plays a crucial role in the first proof of the strong subadditivity of the von Neumann entropy [32].

A remarkable generalization of the skew information, the metric adjusted skew information, was introduced by Hansen [53]. Let c be a Morozova–Chentsov function in the sense that

$$c(x, y) = \frac{1}{yf(xy^{-1})}, \quad x, y > 0, \quad (15)$$

where f is a positive operator monotone function defined on the positive half-axis satisfying the functional equation $f(t) = tf(t^{-1})$, $t > 0$. Let L_ρ and R_ρ be the left and right multiplication operators by ρ , respectively. Let

$$K_\rho^c(A, B) = \text{tr} A^* c(L_\rho, R_\rho) B, \quad \hat{c}(x, y) = (x - y)^2 c(x, y), \quad m(c) = \lim_{t \rightarrow 0} c(t, 1)^{-1} > 0,$$

then the metric adjusted skew information is defined as

$$\begin{aligned} I^c(\rho, A) &= \frac{m(c)}{2} K_\rho^c(i[\rho, A], i[\rho, A]) = \frac{m(c)}{2} \text{tr} i[\rho, A] c(L_\rho, R_\rho) i[\rho, A] \\ &= \frac{m(c)}{2} \text{tr} A \hat{c}(L_\rho, R_\rho) A. \end{aligned} \quad (16)$$

In particular, when the Morozova–Chentsov function c takes the form

$$c^{\text{WY}}(x, y) = \frac{4}{(\sqrt{x} + \sqrt{y})^2} \quad x, y > 0,$$

then the corresponding metric adjusted skew information reduces to the skew information; when c takes the form

$$c^{\text{WYD}}(x, y) = \frac{1}{\alpha(1-\alpha)} \cdot \frac{(x^\alpha - y^\alpha)(x^{1-\alpha} - y^{1-\alpha})}{(x - y)^2} \quad 0 < \alpha < 1,$$

then the metric adjusted skew information reduces to the Wigner–Yanase–Dyson information. Consequently, $I^c(\rho, A)$ is a generalization of the skew information. Hansen further showed that [53]

$$I^c(\rho, A) = \frac{m(c)}{2} \int_0^1 I^{c_\lambda}(\rho, A) \frac{(1+\lambda)^2}{\lambda} d\mu_c(\lambda), \quad (17)$$

where $c_\lambda(x, y) = \frac{1+\lambda}{2} \left(\frac{1}{x+\lambda y} + \frac{1}{\lambda x+y} \right)$, $\lambda \in [0, 1]$, μ_c is the representing measure of c . It is also proved that $0 \leq I^c(\rho, A) \leq V(\rho, A)$, and $I^c(\rho, A)$ is convex in ρ . It is obvious that $m(c) = f(0)$ and the Morozova–Chentsov function is determined by the function f . Furthermore, Gibilisco et al. generalized the inequality relation (13) to any metric adjusted skew information as [61]

$$I^c(\rho, A) \leq I_F(\rho, A) \leq \frac{1}{2f(0)} I^c(\rho, A), \quad (18)$$

and showed that the constant $1/(2f(0))$ is optimal.

When quantifying quantum uncertainty of quantum channels with respect to quantum states, Luo and Sun studied the quantity [92]

$$I(\rho, \Phi) = \sum_j I(\rho, E_j) \quad (19)$$

with

$$I(\rho, E_j) = \text{tr}[\sqrt{\rho}, E_j][\sqrt{\rho}, E_j]^\dagger. \quad (20)$$

Here $\Phi(\sigma) = \sum_j E_j \sigma E_j^\dagger$ is the Kraus representation of the channel, and E_j are in general not self-adjoint. The above quantity may be regarded as an extension of the skew information and is independent of the Kraus representations. It enjoys several remarkable properties and has interesting applications in characterizing channel-state coupling. In particular, $I(\rho, \Phi)$ is a bona fide measure quantifying the coherence of the state ρ with respect to the channel Φ [92].

3 Skew Information Inequalities

In this section, we review and investigate various inequalities in the desiderata of Wigner and Yanase for a measure of information content. More precisely, we are concerned with convexity, superadditivity, and monotonicity. We will also review some uncertainty inequalities involving the skew information.

3.1 Convexity

The convexity of the Wigner–Yanase–Dyson information,

$$I_\alpha(\lambda_1 \rho_1 + \lambda_2 \rho_2, H) \leq \lambda_1 I_\alpha(\rho_1, H) + \lambda_2 I_\alpha(\rho_2, H), \quad (21)$$

was firstly proved by Lieb [31]. Here $\lambda_i \geq 0$, $\lambda_1 + \lambda_2 = 1$ and ρ_1, ρ_2 are density operators. It is remarkable that the first proof of the monotonicity of quantum relative entropy and the strong subadditivity of quantum entropy is based on the convexity of the Wigner–Yanase–Dyson information [32, 33].

A quantity naturally generalizing the Wigner–Yanase–Dyson information is [51]

$$I_{\alpha,\beta}(\rho, H) = -\frac{1}{2} \text{tr}[\rho^\alpha, H][\rho^\beta, H] \rho^{1-\alpha-\beta} \quad (22)$$

where $\alpha, \beta \in [0, 1]$ are two fixed constants satisfying $0 \leq \alpha + \beta \leq 1$. It is interesting to note that the convexity of this generalized Wigner–Yanase–Dyson information depends on the parameters α and β [57]: If α, β satisfy $\alpha + 2\beta \leq 1$ and $2\alpha + \beta \leq 1$, then $I_{\alpha,\beta}(\rho, H)$ is convex in ρ . Moreover, suppose $0 < \alpha + \beta < 1$ and $I_{\alpha,\beta}(\rho, H)$ is convex in ρ , then $\alpha + 2\beta \leq 1$ or $2\alpha + \beta \leq 1$. In particular, $I_{\alpha,\alpha}(\rho, H)$ is convex in ρ if and only if $\alpha \in [0, \frac{1}{3}] \cup \{\frac{1}{2}\}$.

3.2 Superadditivity

Superadditivity refers to the conjecture that information content should decrease when a joint system is separated into two parts. This is true when the system is classical and the information content is quantified by the classical Fisher informa-

tion, i.e., classical Fisher information is superadditive [83]. In the quantum scenario, Wigner and Yanase conjectured that the skew information is superadditive in the sense that

$$I(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b) \geq I(\rho^a, H^a) + I(\rho^b, H^b) \quad (23)$$

where ρ^{ab} is a bipartite state shared by two parties a and b , $\rho^a = \text{tr}_b \rho^{ab}$ and $\rho^b = \text{tr}_a \rho^{ab}$ are the reduced states, H^a and H^b are observables, on parties a and b respectively. Wigner and Yanase have shown that the above superadditivity inequality is correct at least for pure states and product states [30], and Lieb further emphasized the significance of superadditivity [31]. Unfortunately, Hansen disproved the superadditivity conjecture by a counterexample [52]. Although superadditivity is not true in general, it is still desirable to identify the states satisfying superadditivity [31, 55, 56], and to study weak forms of superadditivity [58].

The Wigner–Yanase–Dyson information is superadditive in the following special instances [55, 56]: (i) $\rho^{ab} = |\phi^{ab}\rangle\langle\phi^{ab}|$ is pure; (ii) $\rho^{ab} = \rho^a \otimes \rho^b$ is a product state; (iii) ρ is diagonal in the representation diagonalizing H^a and H^b ; (iv) $[\rho^{ab}, H^a \otimes \mathbf{1}^b] = 0$ or $[\rho^{ab}, \mathbf{1}^a \otimes H^b] = 0$; (v) ρ is a classical-quantum state in the representation diagonalizing H^a .

The skew information is weak superadditive in the sense that [58]

$$I(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b) \geq \max\{I(\rho^a, H^a), I(\rho^b, H^b)\} \geq \frac{1}{2}(I(\rho^a, H^a) + I(\rho^b, H^b)) \quad (24)$$

and

$$I(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b) + I(\rho^{ab}, H^a \otimes \mathbf{1}^b - \mathbf{1}^a \otimes H^b) \geq 2(I(\rho^a, H^a) + I(\rho^b, H^b)). \quad (25)$$

It was further conjectured that [58]

$$\inf \frac{I(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b)}{I(\rho^a, H^a) + I(\rho^b, H^b)} = \frac{1}{2} \quad (26)$$

where the inf is over all states ρ^{ab} and observables H^a and H^b .

The weak forms of superadditivity of the skew information were extended to any metric adjusted skew information with monotone metric in Ref. [59]. Let \mathcal{F}_{op} be the set of functions $f : (0, \infty) \rightarrow (0, \infty)$ satisfying (i) $f(1) = 1$, (ii) $tf(t^{-1}) = f(t)$, (iii) f is operator monotone, then for any regular function f in \mathcal{F}_{op} , let c be the corresponding Morozova–Chentsov function defined by Eq. (15) and denote the metric adjusted skew information I^c defined by Eq. (16) as I_f , then

$$I_f(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b) \geq \frac{1}{2}(I_f(\rho^a, H^a) + I_f(\rho^b, H^b)) \quad (27)$$

and

$$I_f(\rho^{ab}, H^a \otimes \mathbf{1}^b + \mathbf{1}^a \otimes H^b) + I_f(\rho^{ab}, H^a \otimes \mathbf{1}^b - \mathbf{1}^a \otimes H^b) \geq 2(I_f(\rho^a, H^a) + I_f(\rho^b, H^b)). \quad (28)$$

3.3 Uncertainty Inequalities

Conventionally, the uncertainty of an observable A in a quantum state ρ was described by variance $V(\rho, A) = \text{tr}\rho A^2 - (\text{tr}\rho A)^2$, and the standard textbook form of the Heisenberg uncertainty relation is expressed as

$$V(\rho, A)V(\rho, B) \geq \frac{1}{4}|\text{tr}\rho[A, B]|^2. \quad (29)$$

In terms of the skew information, the above inequality can be refined as [45]

$$U(\rho, A)U(\rho, B) \geq \frac{1}{4}|\text{tr}\rho[A, B]|^2 \quad (30)$$

where

$$U(\rho, A) = \sqrt{V^2(\rho, A) - (V(\rho, A) - I(\rho, A))^2}. \quad (31)$$

It is obvious that $0 \leq I(\rho, A) \leq U(\rho, A) \leq V(\rho, A)$. In fact, the following inequalities hold

$$I(\rho, A)J(\rho, B) \geq \frac{1}{4}|\text{tr}\rho[A, B]|^2, \quad J(\rho, A)I(\rho, B) \geq \frac{1}{4}|\text{tr}\rho[A, B]|^2 \quad (32)$$

where $J(\rho, A) := \frac{1}{2}\text{tr}\{\rho^{1/2}, A_0\}^2$ with $A_0 = A - (\text{tr}\rho A)\mathbf{1}$, $\{A, B\} = AB + BA$ is the anticommutator.

Many further and remarkable results generalizing the above inequalities are obtained [60–69]. In particular [62], for any $f \in \mathcal{F}_{\text{op}}$, $f(0) \neq 0$,

$$U_f(\rho, A)U_f(\rho, B) \geq f(0)^2|\text{tr}\rho[A, B]|^2. \quad (33)$$

where

$$U_f(\rho, A) = \sqrt{V(\rho, A)^2 - (V(\rho, A) - I_f(\rho, A))^2}, \quad (34)$$

and $I_f(\rho, A)$ is the metric adjusted skew information. It is clear that

$$0 \leq I_f(\rho, A) \leq U_f(\rho, A) \leq V(\rho, A). \quad (35)$$

Furthermore, Yanagi showed that [67]

$$U_f(\rho, A)U_f(\rho, B) \geq f(0)|\text{tr}\rho[A, B]|^2 \quad (36)$$

for any $f \in \mathcal{F}_{\text{op}} \cap \{f : f(0) \neq 0\}$ satisfying $\frac{x+1}{2} + \hat{f}(x) \geq 2f(x)$, where $\hat{f}(x) = \frac{1}{2}\left((x+1) - (x-1)^2 \frac{f(0)}{f(x)}\right)$, $x > 0$. This can be seen as a refinement of the result of Gibilisco and Isola [62].

3.4 Monotonicity Under Operations

Since quantum operation usually disturbs a system and destroys the intrinsic information, it is natural to expect that skew information decreases under certain quantum operations. However, Du and Bai showed that skew information can increase under phase sensitive operations [74]. Luo and Zhang gave a sufficient condition for the decrease of the skew information [54]. Consider a quantum operation Φ , $\Phi(\rho) = \sum_j E_j \rho E_j^\dagger$ where the Kraus operators satisfy $\sum_j E_j^\dagger E_j = \mathbf{1}$. From Ref. [54], we know that $\Phi^\dagger(H) = H$ and $\Phi^\dagger(H^2) = H^2$ are satisfied if and only if $[E_j, H] = 0$, for any j . Moreover, if the quantum measurement Φ does not disturb the conserved observable H in the sense that $\Phi^\dagger(H) = H$ and $\Phi^\dagger(H^2) = H^2$ (or, equivalently, all E_j commute with H), then $I(\Phi(\rho), H) \leq I(\rho, H)$.

Now we further study the relationship between no-disturbance and commutativity, and disprove a conjecture in Ref. [54] by a counterexample. We will discuss the relation between $\Phi^\dagger(H) = H$ and $[E_j, H] = 0$ for any j in different dimensions. We will also identify some conditions such that the skew information decreases under quantum measurements.

Proposition 1 *If the system Hilbert space is of dimension 2, then the following three statements are equivalent:*

- (1). $\Phi^\dagger(H) = H$;
- (2). $\Phi^\dagger(H^2) = H^2$;
- (3). $[E_j, H] = 0$ for any j , where E_j are Kraus operators of Φ .

To establish this result, noting that H is hermitian, by spectral decomposition, we have $H = U \Lambda U^\dagger$, where U is a unitary operator and Λ is a real diagonal matrix. On one hand, for any j ,

$$\Phi^\dagger(H) = H \Leftrightarrow \sum_j E_j^\dagger H E_j = H \Leftrightarrow \sum_j (U^\dagger E_j U)^\dagger \Lambda (U^\dagger E_j U) = \Lambda.$$

On the other hand, $[E_j, H] = [E_j, U \Lambda U^\dagger] = U[U^\dagger E_j U, \Lambda]U^\dagger$, for any j . Therefore, $[E_j, H] = 0 \Leftrightarrow [U^\dagger E_j U, \Lambda] = 0$. Consequently, in order to prove the equivalence between items (1) and (3), we only need to prove it for any diagonal real matrix H . Recall Theorem 8.3 in Ref. [93], we know that all quantum operations on a two-dimensional Hilbert space can be generated by an operator-sum representation containing at most 4 elements. We first show the implication item (1) \Rightarrow item (3). Suppose $\Phi^\dagger(H) = H$, i.e., $\sum_{j=1}^4 E_j^\dagger H E_j = H$. Let

$$E_1 = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}, \quad E_2 = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix}, \quad E_3 = \begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix}, \quad E_4 = \begin{pmatrix} d_1 & d_2 \\ d_3 & d_4 \end{pmatrix}$$

satisfy $\sum_{j=1}^4 E_j^\dagger E_j = \mathbf{1}$, then

$$\begin{aligned} |a_1|^2 + |b_1|^2 + |c_1|^2 + |d_1|^2 + |a_3|^2 + |b_3|^2 + |c_3|^2 + |d_3|^2 &= 1, \\ |a_2|^2 + |b_2|^2 + |c_2|^2 + |d_2|^2 + |a_4|^2 + |b_4|^2 + |c_4|^2 + |d_4|^2 &= 1, \end{aligned}$$

where $a_i, b_i, c_i, d_i \in \mathbb{C}$, $i = 1, 2, 3, 4$. Let

$$H = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}.$$

When $\lambda_1 = \lambda_2$, $H = \lambda \mathbf{1}$ is a scalar matrix, then apparently, $[E_j, H] = 0$ for any $j = 1, 2, 3, 4$. When $\lambda_1 \neq \lambda_2$,

$$\Phi^\dagger(H) = \begin{pmatrix} \lambda_1(|a_1|^2 + |b_1|^2 + |c_1|^2 + |d_1|^2) & \lambda_1(a_1^*a_2 + b_1^*b_2 + c_1^*c_2 + d_1^*d_2) \\ +\lambda_2(|a_3|^2 + |b_3|^2 + |c_3|^2 + |d_3|^2) & +\lambda_2(a_3^*a_4 + b_3^*b_4 + c_3^*c_4 + d_3^*d_4) \\ \lambda_1(a_2^*a_1 + b_2^*b_1 + c_2^*c_1 + d_2^*d_1) + & \lambda_1(|a_2|^2 + |b_2|^2 + |c_2|^2 + |d_2|^2) + \\ +\lambda_2(a_4^*a_3 + b_4^*b_3 + c_4^*c_3 + d_4^*d_3) & +\lambda_2(|a_4|^2 + |b_4|^2 + |c_4|^2 + |d_4|^2) \end{pmatrix}$$

From $\Phi^\dagger(H) = H$ and previous equations, we obtain $a_i = b_i = c_i = d_i = 0$ for $i = 2, 3$, i.e., E_j is diagonal for $j = 1, 2, 3, 4$. Consequently, $[E_j, H] = 0$ for $j = 1, 2, 3, 4$.

The proof of the reversed implication item (3) \Rightarrow item (1) is straightforward. Next, the equivalence between items (2) and (3) can also be established similarly by replacing H with H^2 .

Since we have proved the equivalence between $\Phi^\dagger(H) = H$ and $[E_j, H] = 0$ for any j in two-dimensional case, it is natural to ask whether the equivalence is still true in higher dimensions. From the following examples, we see that the answer is negative.

Example 1 Let

$$\begin{aligned} H &= \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & E_1 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix}, \\ E_2 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{\sqrt{8}} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix}, & E_3 &= \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{\sqrt{3}} \end{pmatrix}, \end{aligned}$$

then $\sum_{j=1}^3 E_j^\dagger E_j = \mathbf{1}$ and $[E_j, H] \neq 0$ for any $j = 1, 2, 3$. Direct calculation yields $\Phi^\dagger(H) = \sum_{j=1}^3 E_j^\dagger H E_j = H$, and

$$\Phi^\dagger(H^2) = \begin{pmatrix} 9 & 0 & 0 \\ 0 & \frac{19}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq H^2$$

Example 2 Take

$$H = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix},$$

$$E_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{\sqrt{8}} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad E_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{\sqrt{3}} \end{pmatrix},$$

then, $\sum_{j=1}^3 E_j^\dagger E_j = \mathbf{1}$ and simple calculations indicate that

$$\Phi^\dagger(H) = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & \frac{3\sqrt{3}+2\sqrt{2}+3}{8} & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq H.$$

However, $\Phi^\dagger(H^2) = H^2$ and $[E_j, H] \neq 0$ for $j = 1, 2, 3$.

The above examples show that any two items in Proposition 1 are not equivalent in three-dimensional case. Luo and Zhang have proved that items (1) and (2) combined together is the necessary and sufficient conditions of item (3) in any finite-dimensional case [54]. We can also illustrate the conclusions from the following simple situation. Suppose $H = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, $E_j = (a_j^{kl})$, where a_j^{kl} is the (k, l) -matrix element of E_j , and $\sum_{j=1}^3 E_j^\dagger E_j = \mathbf{1}$. Without loss of generality, we assume $\lambda_1 > \lambda_2 > \lambda_3$. From $\Phi^\dagger(H) = H$, we conclude that E_j are of the forms

$$E_j = \begin{pmatrix} a_j^{11} & a_j^{12} & 0 \\ 0 & a_j^{22} & 0 \\ 0 & a_j^{32} & a_j^{33} \end{pmatrix},$$

which satisfy $\sum_{j=1}^3 E_j^\dagger E_j = \mathbf{1}$. In fact, $[E_j, H] = 0 \Leftrightarrow E_j$ is a diagonal matrix. Therefore, $[E_j, H] = 0$ for any j neither can be derived from the condition $\Phi^\dagger(H) = H$ nor can be obtained from the condition $\Phi^\dagger(H^2) = H^2$.

Combining Proposition 1 and the results in Ref. [54], we conclude that $\Phi^\dagger(H) = H$ is sufficient for the decreasing of skew information under measurements in two-dimensional spaces.

Proposition 2 Suppose that the system Hilbert space is two-dimensional, and $\Phi^\dagger(H) = H$, then $I(\Phi(\rho), H) \leq I(\rho, H)$.

The above statement is incorrect for the case of higher dimensional spaces, as the following example shows.

Example 3 Let

$$H = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$E_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad E_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{4} & 0 \\ 0 & \frac{1}{\sqrt{8}} & 0 \\ 0 & \frac{1}{4} & \frac{1}{\sqrt{3}} \end{pmatrix}, \quad E_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

By direct calculation, we have

$$\Phi^\dagger(H) = H, \quad \Phi^\dagger(H^2) = \begin{pmatrix} 9 & 0 & 0 \\ 0 & \frac{19}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \neq H^2$$

and

$$I(\rho, H) \approx 0.0637, \quad I(\Phi(\rho), H) \approx 0.1272$$

Consequently, $I(\rho, H) < I(\Phi(\rho), H)$. This indicates that $\Phi^\dagger(H) = H$ is not a sufficient condition for monotonicity.

4 Applications to Quantum Information

In this section, we review some applications of skew information inequalities in quantum information processing.

4.1 Coherence

In recent years, inspired by the seminal work of Baumgratz et al. [94], there is increasing interest in axiomatic and quantitative studies of coherence [94–113]. Girolami proposed to use the skew information $I(\rho, K) = -\frac{1}{2}\text{tr}[\sqrt{\rho}, K]^2$ as a quantifier of coherence of ρ with respect to K , and called it the K -coherence [95]. This is an intuitive measure satisfying all desirable properties of a coherence measure

as postulated by Baumgratz et al. [94], with the exception of monotonicity under incoherent operations [74, 101]. In order to obtain a monotone measure of coherence, Marvian et al. suggested to restrict the incoherent states to translation invariant states and the incoherent operations to translation invariant operations in the following sense [101, 102]: Consider $\mathcal{I}_H = \{\rho : e^{-iHt}\rho e^{iHt} = \rho, \forall t \in \mathbb{R}\}$ as the set of free states and $\mathcal{M}_H = \{\Phi : e^{-iHt}\Phi(\rho)e^{iHt} = \Phi(e^{-iHt}\rho e^{iHt}), \forall t \in \mathbb{R}\}$ as the set of free operations. In this resource theoretic framework, Marvian et al. showed that $I(\Phi_{\text{TI}}(\rho), H) \leq I(\rho, H)$ for any translation invariant operation $\Phi_{\text{TI}} \in \mathcal{M}_H$ [101, 102]. The monotonicity problem can also be remedied by defining coherence in terms of average skew information [109, 110].

The diversity of the notions of incoherent operations complicates the resource theoretic approach to coherence. A direct approach to coherence qualification via quantum uncertainty, which in turn is quantified via average quantum Fisher information, is proposed in Ref. [111]. Let $M = \{M_i : i = 1, 2, \dots, m\}$ be a positive operator valued measure (POVM) with $M_i \geq 0, \sum_{i=1}^m M_i = \mathbf{1}$, then a measure of the coherence of ρ with respect to the measurement M may be defined as

$$Q(\rho, M) = \sum_i I(\rho, M_i). \quad (37)$$

This is indeed a bona fide measure of coherence. In particular, it satisfies

(i) Decreasing under partial trace in the sense that

$$Q(\rho, M^a \otimes \mathbf{1}^b) \geq Q(\rho^a, M^a), \quad (38)$$

where M^a is a POVM on party a , $\rho^a = \text{tr}_b \rho$ is the reduced state on party a , $\mathbf{1}^b$ is the identity operator on party b .

(ii) Decreasing under nondisturbing operations in the sense that

$$Q(\rho, M) \geq Q(\Phi(\rho), M) \quad (39)$$

for any quantum operation Φ satisfying $\Phi^\dagger(\sqrt{M_i}) = \sqrt{M_i}, \Phi^\dagger(M_i) = M_i$ for any i .

(iii) Convexity in ρ , that is,

$$Q\left(\sum_j \lambda_j \rho_j, M\right) \leq \sum_j \lambda_j Q(\rho_j, M) \quad (40)$$

where $\lambda_j \geq 0, \sum_j \lambda_j = 1$ and ρ_j are states. When $\Pi = \{\Pi_i = |i\rangle\langle i| : i = 1, \dots, m\}$ is a von Neumann measurement, the coherence measure is reduced to the measure $C(\rho) = \sum_i I(\rho, |i\rangle\langle i|)$ introduced by Yu [108].

4.2 Quantum Uncertainty

Uncertainty is often measured by variance or entropies, and usually consists of both classical and quantum parts. The skew information has been interpreted as a measure of quantum uncertainty [45, 46]. In order to relate the skew information with entropy and to get an intrinsic quantity capturing the information content of the state ρ without reference to other quantity, one may take average of the skew information $Q(\rho) = \sum_i I(\rho, H_i)$ where $\{H_i : i = 1, 2, \dots, n^2\}$ is an orthonormal basis of the observables on the n -dimensional system Hilbert space with the Hilbert–Schmidt inner product $\langle A|B\rangle = \text{tr}A^\dagger B$. It is shown that this quantity is independent of the choice of the basis, and turns out to be [48]

$$Q(\rho) = \sum_{j < k} \left(\sqrt{\lambda_j} - \sqrt{\lambda_k} \right)^2 = n - (\text{tr}\sqrt{\rho})^2. \quad (41)$$

Therefore, this quantum uncertainty measure is intimately related to the notions of generalized entropies of Rényi and Tsallis [114, 115]. In contrast, if one defines $V(\rho) = \sum_i V(\rho, H_i)$, then $V(\rho) = n - \text{tr}\rho^2$ [116].

As a generalization of the quantum uncertainty measure based on the skew information, Li et al. introduced the following quantum uncertainty measure based on the Wigner–Yanase–Dyson information [69]

$$Q_\alpha(\rho) = \sum_i I_\alpha(\rho, H_i), \quad (42)$$

which turns out to be $Q_\alpha(\rho) = n - \text{tr}\rho^\alpha \text{tr}\rho^{1-\alpha}$.

Following some conjectures in Ref. [117], Cai studied quantum uncertainty in terms of any metric adjusted skew information by taking average [118], which was a further generalization of the quantum uncertainty based on the Wigner–Yanase–Dyson information.

4.3 Correlations

Quantum correlations have many manifestations such as nonlocality, steering, entanglement, quantum discord, etc. By use of the skew information, several measures of correlations have been introduced. Consider a bipartite state ρ^{ab} , one may define [70]

$$Q_a(\rho^{ab}) = \sum_i I(\rho^{ab}, X_i \otimes \mathbf{1}^b) \quad (43)$$

as the global information content of ρ^{ab} with respect to local observables $\{X_i\}_{i=1}^{m^2}$ which is an operator orthonormal basis of observables of party a . $Q_a(\rho^{ab})$ is inde-

pendent of the choice of local operator orthonormal basis $\{X_i\}_{i=1}^{m^2}$. A measure of correlations in ρ^{ab} may be defined as the information difference between ρ^{ab} and $\rho^a \otimes \rho^b$

$$F(\rho^{ab}) = Q_a(\rho^{ab}) - Q_a(\rho^a \otimes \rho^b) = Q_a(\rho^{ab}) - Q(\rho^a). \quad (44)$$

This measure captures the correlations in ρ^{ab} that can be probed by local observables of party a .

In a quite different approach, Girolami et al. defined local quantum uncertainty [72]

$$U_\Lambda^a(\rho^{ab}) = \min I(\rho^{ab}, K^a \otimes \mathbf{1}^b) \quad (45)$$

as a measure of quantum correlations. Here the min is over all local observables K^a on party a with fixed and non-degenerated spectrum Λ . The local quantum uncertainty $U_\Lambda^a(\rho^{ab})$ is a reasonable measure of non-classical correlations under the criteria for discord-like correlation quantifier [119].

Li et al. defined the following quantity [120]

$$N_s(\rho^{ab}) = \max_{\Pi^a} \sum_i I(\rho^{ab}, \Pi_i^a \otimes \mathbf{1}^b) \quad (46)$$

as a measure of the global effect caused by local invariant von Neumann measurements. Here the max is over all local von Neumann measurements $\Pi^a = \{\Pi_i^a\}$ on party a which do not disturb $\rho^a = \text{tr}_b \rho^{ab}$ locally. This is a measure similar to the measurement-induced nonlocality [121].

Let $\Pi^a = \{\Pi_i^a\}$ be any local von Neumann measurement on party a , and $\Pi = \{\Pi_i^a \otimes \mathbf{1}^b\}$ the corresponding Lüders measurement on the combined system ab . Let

$$C_H(\rho^{ab}|\Pi) = \|\sqrt{\rho^{ab}} - \Pi(\sqrt{\rho^{ab}})\|^2, \quad (47)$$

which is a bona fide measure for coherence [112], then it turns out that

$$C_H(\rho^{ab}|\Pi) = \sum_i I(\rho^{ab}, \Pi_i^a \otimes \mathbf{1}^b). \quad (48)$$

Define $\Delta_H(\rho^{ab}|\Pi) = C_H(\rho^{ab}|\Pi) - C_H(\rho^a|\Pi^a)$ as the coherence difference between ρ^{ab} and $\rho^a = \text{tr}_b \rho^{ab}$. Then in terms of $C_H(\rho^{ab}|\Pi)$ and $\Delta_H(\rho^{ab}|\Pi)$, one may introduce the following four quantities.

- (1) minimal coherence $C_{\min}(\rho^{ab}) = \min_{\Pi} C_H(\rho^{ab}|\Pi)$;
- (2) maximal coherence $C_{\max}(\rho^{ab}) = \max_{\Pi} C_H(\rho^{ab}|\Pi)$;
- (3) minimal coherence difference $D_{\min}(\rho^{ab}) = \min_{\Pi} \Delta_H(\rho^{ab}|\Pi)$;
- (4) maximal coherence difference $D_{\max}(\rho^{ab}) = \max_{\Pi} \Delta_H(\rho^{ab}|\Pi)$.

In particular, $C_{\min}(\rho^{ab})$ coincides with the modified geometric discord $D_H(\rho)$ introduced in Ref. [122]. Furthermore, we have [112]

$$C_{\max}(\rho^{ab}) \geq D_{\max}(\rho^{ab}) \geq C_{\min}(\rho^{ab}) \geq D_{\min}(\rho^{ab}). \quad (49)$$

In particular, for any pure state $\rho^{ab} = |\Psi\rangle\langle\Psi|$ with Schmidt decomposition $|\Psi\rangle = \sum_{i=1}^n \sqrt{s_i} |i\rangle_a |i\rangle_b$, where $\{|i\rangle_a\}$ and $\{|i\rangle_b\}$ are orthonormal bases for parties a and b , respectively, then

$$\begin{aligned} C_{\min}(\rho^{ab}) &= 1 - \sum_j s_j^2, \quad C_{\max}(\rho^{ab}) = 1 - \frac{1}{n}, \\ D_{\min}(\rho^{ab}) &= \frac{1}{n} \left(\left(\sum_j \sqrt{s_j} \right)^2 - 1 \right), \quad D_{\max}(\rho^{ab}) = 1 - \sum_j s_j^2, \end{aligned}$$

which implies that for pure state ρ^{ab} ,

$$C_{\max}(\rho^{ab}) \geq D_{\max}(\rho^{ab}) = C_{\min}(\rho^{ab}) \geq D_{\min}(\rho^{ab}). \quad (50)$$

By the definitions of $N_s(\rho^{ab})$, $C_{\min}(\rho^{ab})$ and $D_{\max}(\rho^{ab})$, we know that

$$D_{\max}(\rho^{ab}) \geq N_s(\rho^{ab}) \geq C_{\min}(\rho^{ab}). \quad (51)$$

and $D_{\max}(\rho^{ab}) = N_s(\rho^{ab}) = C_{\min}(\rho^{ab})$ if ρ^{ab} is a pure state.

The interferometric power [123]

$$P(\rho^{ab}) = \min_{H^a} I_F(\rho^{ab}, H^a \otimes \mathbf{1}^b) \quad (52)$$

was introduced as a measure of correlations motivated by interferometry. Here I_F is the quantum Fisher information defined by the symmetric logarithm derivative, and the min is over all observables H^a with given spectrum [123]. By the analytic expression of quantum Fisher information $I_F(\rho^{ab}, H^a \otimes \mathbf{1}^b)$ and the inequality $I_F(\rho^{ab}, H) \leq 2I(\rho^{ab}, H)$ [39], we have [112]

$$P(\rho^{ab}) \leq 2C_{\min}(\rho^{ab}). \quad (53)$$

5 Summary

We have reviewed the origin and significance of the skew information from a historical perspective and highlighted various fundamental inequalities for the skew information which have important implications in quantum information processing. The skew information may be intrinsically regarded as a version of quantum Fisher information based on square roots of density operators. Some novel inequalities concerning monotonicity of the skew information under quantum operations are established, and various applications of the skew information are indicated. It is expected

that the skew information will play an increasingly inspiring and instrumental role in quantum information theory.

Acknowledgements This work was supported by the National Natural Science Foundation of China, Grant No. 11375259, the National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Grant No. Y029152K51, the Key Laboratory of Random Complex Structures and Data Science, Chinese Academy of Sciences, Grant No. 2008DP173182.

References

1. Wigner, E.P.: Die messung quantenmechanischer operatoren. *Zeit. Phys.* **133**, 101 (1952)
2. Salecker, H., Wigner, E.P.: Quantum limitations of the measurement of space-time distances. *Phys. Rev.* **109**, 571 (1958)
3. Araki, H., Yanase, M.M.: Measurement of quantum mechanical operators. *Phys. Rev.* **120**, 622 (1960)
4. Yanase, M.M.: Optimal measuring apparatus. *Phys. Rev.* **123**, 666 (1961)
5. Wigner, E.P.: The problem of measurement. *Am. J. Phys.* **31**, 6 (1963)
6. Albertson, J.: Quantum-mechanical measurement operator. *Phys. Rev.* **129**, 940 (1963)
7. Jauch, M., Wigner, E.P., Yanase, M.M.: Some comments concerning measurements in quantum mechanics. *Nuovo Cimento* **48**, 144 (1967)
8. Fine, A.I.: On the general quantum theory of measurement. *Math. Proc. Camb. Philos. Soc.* **65**, 111 (1969)
9. Wigner, E.P., Yanase, M.M.: Analysis of the quantum mechanical measurement process. *Ann. Jpn. Assoc. Philos. Sci.* **4**, 171 (1973)
10. Ghirardi, G.C., Miglietta, F., Rimini, A., Weber, T.: Limitations on quantum measurements. I. determination of the minimal amount of nonideality and identification of the optimal measuring apparatuses. *Phys. Rev. D* **24**, 347 (1981)
11. Ghirardi, G.C., Miglietta, F., Rimini, A., Weber, T.: Limitations on quantum measurements. II. analysis of a model example. *Phys. Rev. D* **24**, 353 (1981)
12. Ghirardi, G.C., Rimini, A., Weber, T.: Quantum evolution in the presence of additive conservation laws and the quantum theory of measurement. *J. Math. Phys.* **23**, 1792 (1982)
13. Ozawa, M.: Quantum measuring processes of continuous observables. *J. Math. Phys.* **25**, 79 (1984)
14. Kudaka, S., Kakazu, K.: The Wigner–Araki–Yanase theorem and its extension in quantum measurement with generalized coherent states. *Prog. Theor. Phys.* **87**, 61 (1992)
15. Matsumoto, S.: A reexamination of the Wigner and Araki–Yanase theorem. *Prog. Theor. Phys.* **90**, 35 (1993)
16. Kakazu, K., Pascazio, S.: Alternative formulation of the Wigner–Araki–Yanase theorem. *Phys. Rev. A* **51**, 3469 (1995)
17. Nielsen, M.A.: Computable functions, quantum measurements, and quantum dynamics. *Phys. Rev. Lett.* **79**, 2915 (1997)
18. Ozawa, M.: Conservation laws, uncertainty relations, and quantum limits of measurements. *Phys. Rev. Lett.* **88**, 050402 (2002)
19. Miyadera, T., Imai, H.: Wigner–Araki–Yanase theorem on distinguishability. *Phys. Rev. A* **74**, 024101 (2006)
20. Bartlett, S.D., Rudolph, T., Spekkens, R.W.: Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.* **79**, 555 (2007)
21. Kimura, G., Meister, B.K., Ozawa, M.: Quantum limits of measurements induced by multiplicative conservation laws: extension of the Wigner–Araki–Yanase theorem. *Phys. Rev. A* **78**, 032106 (2008)

22. Loveridge, L., Busch, P.: ‘Measurement of quantum mechanical operators’ revisited. *Eur. Phys. J. D* **62**, 297 (2011)
23. Loveridge, L., Busch, P.: Position measurements obeying momentum conservation. *Phys. Rev. Lett.* **106**, 110406 (2011)
24. Ahmadi, M., Jennings, D., Rudolph, T.: The Wigner–Araki–Yanase theorem and the quantum resource theory of asymmetry. *New J. Phys.* **15**, 013057 (2013)
25. Marvian, I., Spekkens, R.W.: Modes of asymmetry: the application of harmonic analysis to symmetric quantum dynamics and quantum reference frames. *Phys. Rev. A* **90**, 062110 (2014)
26. Navascués, M., Popescu, S.: How energy conservation limits our measurements. *Phys. Rev. Lett.* **112**, 140502 (2014)
27. Miyadera, T., Loveridge, L., Busch, P.: Approximating relational observables by absolute quantities: a quantum accuracy-size trade-off. *J. Phys. A* **49**, 185301 (2016)
28. Tukiainen, M.: Wigner–Araki–Yanase theorem beyond conservation laws. *Phys. Rev. A* **95**, 012127 (2017)
29. Loveridge, L., Miyadera, T., Busch, P.: Symmetry, reference frames, and relational quantities in quantum mechanics (2017). [arXiv:1703.10434](https://arxiv.org/abs/1703.10434)
30. Wigner, E.P., Yanase, M.M.: Information contents of distribution. *Proc. Natl. Acad. Sci. U.S.A.* **49**, 910 (1963)
31. Lieb, E.H.: Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.* **11**, 267 (1973)
32. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum mechanical entropy. *Phys. Rev. Lett.* **30**, 434 (1973)
33. Uhlmann, A.: Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Commun. Math. Phys.* **54**, 21 (1977)
34. Wehrl, A.: General properties of entropy. *Rev. Mod. Phys.* **50**, 221 (1978)
35. Hasegawa, H.: α -Divergence of the non-commutative information geometry. *Rep. Math. Phys.* **33**, 87 (1993)
36. Gibilisco, P., Isola, T.: A characterisation of Wigner–Yanase skew information among statistically monotone metrics. *Inf. Dim. Anal. Quantum Probab. Rel. Top.* **4**, 553 (2001)
37. Gibilisco, P., Isola, T.: Wigner–Yanase information on quantum state space: the geometric approach. *J. Math. Phys.* **44**, 3752 (2003)
38. Luo, S.: Wigner–Yanase skew information and uncertainty relations. *Phys. Rev. Lett.* **91**, 180403 (2003)
39. Luo, S.: Wigner–Yanase skew information versus quantum Fisher information. *Proc. Am. Math. Soc.* **132**, 885 (2003)
40. Luo, S., Zhang, Q.: Informational distance on quantum-state space. *Phys. Rev. A* **69**, 032106 (2004)
41. Luo, S., Zhang, Z.: An information characterization of Schrödinger uncertainty relations. *J. Stat. Phys.* **114**, 1557 (2004)
42. Luo, S., Zhang, Q.: On skew information. *IEEE Trans. Inf. Theory* **50**, 1778 (2004); **51**, 4432 (2005)
43. Yanagi, K., Furuichi, S., Kuriyama, K.: A generalized skew information and uncertainty relation. *IEEE Trans. Inf. Theory* **51**, 4401 (2005)
44. Kosaki, H.: Matrix trace inequality related to uncertainty principle. *Int. J. Math.* **16**, 629 (2005)
45. Luo, S.: Heisenberg uncertainty relation for mixed states. *Phys. Rev. A* **72**, 042110 (2005)
46. Luo, S.: Quantum versus classical uncertainty. *Theor. Math. Phys.* **143**, 681 (2005)
47. Chen, Z.: Wigner–Yanase skew information as tests for quantum entanglement. *Phys. Rev. A* **71**, 052302 (2005)
48. Luo, S.: Quantum uncertainty of mixed states based on skew information. *Phys. Rev. A* **73**, 022324 (2006)
49. Gibilisco, P., Isola, T.: Uncertainty principle and quantum Fisher information. *Ann. Inst. Stat. Math.* **59**, 147 (2007)
50. Gibilisco, P., Imparato, D., Isola, T.: Uncertainty principle and quantum Fisher information. II. *J. Math. Phys.* **48**, 072109 (2007)

51. Chen, P., Luo, S.: Direct approach to quantum extensions of Fisher information. *Front. Math. China* **2**, 359 (2007)
52. Hansen, F.: The Wigner and Yanase entropy is not subadditive. *J. Stat. Phys.* **126**, 643 (2007)
53. Hansen, F.: Metric adjusted skew information. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9909 (2007)
54. Luo, S., Zhang, Q.: Skew information decreases under quantum measurements. *Theor. Math. Phys.* **151**, 529 (2007)
55. Luo, S.: Notes on superadditivity of Wigner–Yanase–Dyson information. *J. Stat. Phys.* **128**, 1177 (2007)
56. Luo, S., Zhang, Q.: Superadditivity of Wigner–Yanase–Dyson information revisited. *J. Stat. Phys.* **131**, 1169 (2008)
57. Cai, L., Luo, S.: On convexity of generalized Wigner–Yanase–Dyson information. *Lett. Math. Phys.* **83**, 253 (2008)
58. Cai, L., Li, N., Luo, S.: Weak superadditivity of skew information. *J. Phys. A* **41**, 135301 (2008)
59. Cai, L., Hansen, F.: Metric-adjusted skew information: convexity and restricted forms of superadditivity. *Lett. Math. Phys.* **93**, 1 (2010)
60. Gibilisco, P., Hiai, F., Petz, D.: Quantum covariance, quantum Fisher information and uncertainty relations. *IEEE Trans. Inf. Theory* **55**, 439 (2009)
61. Gibilisco, P., Imperato, D., Isola, T.: Inequalities for quantum Fisher information. *Proc. Am. Math. Soc.* **137**, 317 (2009)
62. Gibilisco, P., Isola, T.: On a refinement of Heisenberg uncertainty relation by means of quantum Fisher information. *J. Math. Anal. Appl.* **375**, 270 (2011)
63. Furuchi, S., Yanagi, K., Kuriyama, K.: Trace inequalities on a generalized Wigner–Yanase skew information. *J. Math. Anal. Appl.* **356**, 179 (2009)
64. Furuchi, S.: Schrödinger uncertainty relation with Wigner–Yanase skew information. *Phys. Rev. A* **82**, 034101 (2010)
65. Furuchi, S.: Inequalities for Tsallis relative entropy and generalized skew information. *Linear Multilinear Algebra* **59**, 1143 (2011)
66. Yanagi, K.: Uncertainty relation on Wigner–Yanase–Dyson skew information. *J. Math. Anal. Appl.* **365**, 12 (2010)
67. Yanagi, K.: Metric adjusted skew information and uncertainty relation. *J. Math. Anal. Appl.* **380**, 888 (2011)
68. Li, D., Li, X., Wang, F., Huang, H., Li, X., Kwek, L.C.: Uncertainty relation of mixed states by means of Wigner–Yanase–Dyson information. *Phys. Rev. A* **79**, 052106 (2009)
69. Li, X., Li, D., Huang, H., Li, X., Kwek, L.C.: Averaged Wigner–Yanase–Dyson information as a quantum uncertainty measure. *Eur. Phys. J. D* **64**, 147 (2011)
70. Luo, S., Fu, S., Oh, C.H.: Quantifying correlations via the Wigner–Yanase skew information. *Phys. Rev. A* **85**, 032117 (2012)
71. Hansen, F.: WYD-like skew information measures. *J. Stat. Phys.* **151**, 974 (2013)
72. Girolami, D., Tufarelli, T., Adesso, G.: Characterizing nonclassical correlations via local quantum uncertainty. *Phys. Rev. Lett.* **110**, 240402 (2013)
73. Yu, C.S., Wu, S.X., Wang, X., Yi, X.X., Song, H.S.: Quantum correlation measure in arbitrary bipartite systems. *Europhys. Lett.* **107**, 10007 (2014)
74. Du, S., Bai, Z.: The Wigner–Yanase information can increase under phase sensitive incoherent operations. *Ann. Phys.* **359**, 136 (2015)
75. Fan, Y.J., Cao, H.X.: Quantifying correlations via the Wigner–Yanase–Dyson skew information. *Int. J. Theor. Phys.* **55**, 3843 (2016)
76. Fisher, R.A.: Theory of statistical estimation. *Proc. Camb. Philos. Soc.* **22**, 700 (1925)
77. Cramér, H.: Mathematical Methods of Statistics. Princeton University Press, New Jersey (1946)
78. Stam, A.: Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control.* **2**, 101 (1959)
79. Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic, New York (1976)

80. Holevo, A.S.: Probabilistic and Statistical Aspects of Quantum Theory. North-Holland, Amsterdam (1982)
81. Frieden, B.R.: Physics from Fisher Information: A Unification. Cambridge University, Cambridge (1998)
82. Luo, S.: Fisher information, kinetic energy and uncertainty relation inequalities. *J. Phys. A* **35**, 5181 (2002)
83. Carlen, A.E.: Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Func. Anal.* **101**, 194 (1991)
84. Braunstein, S.L., Caves, C.M.: Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72**, 3439 (1994)
85. Luo, S.: Uncertainty relations in terms of Fisher information. *Commun. Theor. Phys.* **36**, 257 (2001)
86. Hall, M.J.W.: Exact uncertainty relations. *Phys. Rev. A* **64**, 052103 (2001)
87. Luo, S.: Maximum Shannon entropy, minimum Fisher information, and an elementary game. *Found. Phys.* **32**, 1757 (2002)
88. Luo, S.: Statistics of local value in quantum mechanics. *Int. J. Theor. Phys.* **41**, 1713 (2002)
89. Petz, D.: Quantum Information Theory and Quantum Statistics. Springer, Berlin (2008)
90. Cencov, N.N.: Statistical Decision Rules and Optimal Inference. Am. Math. Soc, Providence (1982)
91. Petz, D.: Monotone metrics on matrix spaces. *Linear Algebra Appl.* **244**, 81 (1996)
92. Luo, S., Sun, Y.: Coherence and complementarity in state-channel interaction. *Phys. Rev. A* **98**, 012113 (2018)
93. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2010)
94. Baumgratz, T., Cramer, M., Plenio, M.B.: Quantifying coherence. *Phys. Rev. Lett.* **113**, 140401 (2014)
95. Girolami, D.: Observable measure of quantum coherence in finite dimensional systems. *Phys. Rev. Lett.* **113**, 170401 (2014)
96. Marvian, I.: Symmetry, asymmetry and quantum information, Ph.D. thesis, University of Waterloo (2012)
97. Marvian, I., Spekkens, R.W.: Extending Noether's theorem by quantifying the asymmetry of quantum states. *Nat. Commun.* **5**, 3821 (2014)
98. Lostaglio, M., Korzekwa, K., Jennings, D., Rudolph, T.: Quantum coherence, time-translation symmetry, and thermodynamics. *Phys. Rev. X* **5**, 021001 (2015)
99. Shao, L.-H., Xi, Z., Fan, H., Li, Y.: Fidelity and trace-norm distances for quantifying coherence. *Phys. Rev. A* **91**, 042120 (2015)
100. Yuan, X., Zhou, H., Cao, Z., Ma, X.: Intrinsic randomness as a measure of quantum coherence. *Phys. Rev. A* **92**, 022124 (2015)
101. Marvian, I., Spekkens, R.W., Zanardi, P.: Quantum speed limits, coherence and asymmetry. *Phys. Rev. A* **93**, 052331 (2016)
102. Marvian, I., Spekkens, R.W.: How to quantify coherence: distinguishing speakable and unspeakable notions. *Phys. Rev. A* **94**, 052324 (2016)
103. Winter, A., Yang, D.: Operational resource theory of coherence. *Phys. Rev. Lett.* **116**, 120404 (2016)
104. Napoli, C., Bromley, T.R., Cianciaruso, M., Piani, M., Johnston, N., Adesso, G.: Robustness of coherence: an operational and observable measure of quantum coherence. *Phys. Rev. Lett.* **116**, 150502 (2016)
105. Chitambar, E., Hsieh, M.H.: Relating the resource theories of entanglement and quantum coherence. *Phys. Rev. Lett.* **117**, 020402 (2016)
106. Chitambar, E., Gour, G.: Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence. *Phys. Rev. Lett.* **117**, 030401 (2016)
107. Rana, S., Parashar, P., Lewenstein, M.: Trace-distance measure of coherence. *Phys. Rev. A* **93**, 012110 (2016)

108. Yu, C.S.: Quantum coherence via skew information and its polygamy. *Phys. Rev. A* **95**, 042337 (2017)
109. Zhao, H., Yu, C.: Remedying the strong monotonicity of the coherence measure in terms of the Tsallis relative α -entropy (2017). [arXiv:1704.04876v1](https://arxiv.org/abs/1704.04876v1)
110. Luo, S., Sun, Y.: Partial coherence with application to the monotonicity problem of coherence involving skew information. *Phys. Rev. A* **96**, 022136 (2017)
111. Luo, S., Sun, Y.: Quantum coherence versus quantum uncertainty. *Phys. Rev. A* **96**, 022130 (2017)
112. Sun, Y., Mao, Y., Luo, S.: From quantum coherence to quantum correlations. *Europhys. Lett.* **118**, 60007 (2017)
113. Streltsov, A., Adesso, G., Plenio, M.B.: Quantum coherence as a resource (2017). [arXiv:1609.02439v3](https://arxiv.org/abs/1609.02439v3)
114. Rényi, A.: Probability Theory. North-Holland, Amsterdam (1970)
115. Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479 (1988)
116. Luo, S.: Brukner–Zeilinger invariant information. *Theor. Math. Phys.* **151**, 693 (2007)
117. Gibilisco, P.: Fisher information and means: some questions in the classical and quantum settings. *Int. J. Softw. Inform.* **8**, 265 (2014)
118. Cai, L.: Quantum uncertainty based on metric adjusted skew information (2017). [arXiv:1708.00978](https://arxiv.org/abs/1708.00978)
119. Modi, K., Brodutch, A., Cable, H., Paterek, T., Vedral, V.: The classical-quantum boundary for correlations: discord and related measures. *Rev. Mod. Phys.* **84**, 1655 (2012)
120. Li, L., Wang, Q.W., Shen, S.Q., Li, M.: Measurement-induced nonlocality based on Wigner–Yanase skew information. *Europhys. Lett.* **114**, 10007 (2016)
121. Luo, S., Fu, S.: Measurement-induced nonlocality. *Phys. Rev. Lett.* **106**, 120401 (2011)
122. Chang, L., Luo, S.: Remedying the local ancilla problem with geometric discord. *Phys. Rev. A* **87**, 062303 (2013)
123. Girolami, D., Souza, A.M., Giovannetti, V., Tufarelli, T., Filgueiras, J.G., Sarthour, R.S., Soares-Pinto, D.O., Oliveira, I.S., Adesso, D.: Quantum discord determines the interferometric power of quantum states. *Phys. Rev. Lett.* **112**, 210401 (2014)

Information Geometry of Quantum Resources



Davide Girolami

Abstract I review recent works showing that information geometry is a useful framework to characterize quantum coherence and entanglement. 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. The information geometry viewpoint elucidates the advantage provided by quantum superpositions in phase estimation. Also, it enables to link measures of coherence and entanglement to observables, which can be evaluated in a laboratory by a limited number of measurements.

Keywords Quantum information · Quantum coherence · Quantum correlations · Quantum metrology

1 Introduction

The possibility to prepare and manipulate even single, isolated atoms and photons makes possible to exploit quantum effects to speed-up information processing. In particular, the ability to create coherent superpositions of quantum states, enlightened by the iconic “Schrödinger’s cat” [1], is the most fundamental difference between classical and quantum systems. Quantum information theory established coherence (the quantum label is omitted, from now on) as a key resource for obtaining an advantage in information processing [2–5]. Another critical property of quantum systems is entanglement, a kind of correlation which yields speed-up in information

D. Girolami (✉)

Los Alamos National Laboratory, Theoretical Division, PO BOX 1663, Los Alamos,
NM 87545, USA

e-mail: davegirolami@gmail.com

URL: <https://www.sites.google.com/site/davegirolami/>

processing, as well as improves the precision of measurement devices [6]. Yet, to quantify the coherence of quantum states, and the coherence consumed and created by quantum dynamics, access to the full state of a system is usually required. As its degrees of freedom are exponentially growing with the number of constituents, the task is computationally and experimentally challenging.

The works I here review employed ideas developed in classical and quantum information geometry [7, 8], which visualize physical processes as paths on an abstract space, to develop efficient strategies to evaluate the coherence and the entanglement of a quantum state. The main result is a certification scheme, enabling to determine, by means of a limited number of measurements, the amount of coherence in a quantum state which is useful to phase estimation, the problem of reconstructing the value of an unknown parameter controlling the dynamics of a system. By generalizing the analysis to systems of many particles, an entanglement witness is obtained. The proposal can be experimentally demonstrated by performing standard measurement procedures, being no a priori information available, thus outperforming methods involving state and channel tomography, i.e. full reconstruction of the state of system. In fact, the test has been recently implemented in an all-optical setup via a network of Bell state projections [9]. I also discuss an alternative experimental architecture which makes use of spin polarization measurements, being suitable, for example, to Nuclear Magnetic Resonance (NMR) systems [10].

2 Making the Usefulness of Coherence Manifest via Information Geometry

2.1 Coherence as Complementarity Between State and Observable

The state of a finite dimensional quantum system is described by a self-adjoint semi-positive density matrix ρ , $\rho = \rho^\dagger$, $\text{Tr}\{\rho\} = 1$, $\rho \geq 0$. Coherence can emerge whenever the system is prepared in a mixture of superpositions of two or more states. In other words, the density matrix representing the state is not diagonal with respect to a reference basis $\{h_i\}$, $\rho \neq \sum_i p_i |h_i\rangle\langle h_i|$, $\langle h_i | h_j \rangle = \delta_{ij}$, $\sum_i p_i = 1$. Surprisingly, coherence has been characterised as a resource for information processing only recently. A consistent body of research identified the coherence of a quantum state as the ability of a system to break a symmetry generated by a Hamiltonian $H = \sum_i h_i |h_i\rangle\langle h_i|$, i.e. a phase reference frame under a superselection rule, where H acts as the charge operator [3, 4]. Coherence was then relabelled as *asymmetry*. Concurrent works proposed an alternative definition of coherence, as the distance of a state from the set of diagonal states in the reference basis [5]. In the following, I embrace the former interpretation. An important question is how to quantify asymmetry. A non-negative, contractive under noisy maps function of the state-observable commutator $[\rho, H]$ is arguably a good measure of asymmetry. Indeed, whenever the state is an eigenstate

or a mixture of eigenstates of the observable, then it is diagonal in the Hamiltonian eigenbasis, which I assume here to be non-degenerate. However, it is desirable to link asymmetry to the performance in an information processing task. In other words, the asymmetry quantifier should be the figure of merit of a protocol, benchmarking the usefulness of the system under scrutiny to complete the task. It is in fact possible to link asymmetry to the precision in phase estimation, as I explain in the next section.

2.2 *Quantum Phase Estimation*

Metrology is the discipline at the boundary between Physics and Statistics studying how to access information about a system by efficient measurement strategies and data analysis [11]. Quantum metrology investigates how to improve the precision of measurements by employing quantum systems. Results obtained in quantum metrology have found a use in interferometry, atomic spectroscopy, and gravitometry [12, 13]. An important metrology primitive, as well as a frequent subroutine in computation protocols, is parameter estimation, which can be interpreted as a dynamical process. First, a probe system is prepared in an input state. Then, a controlled interaction imprints information about the parameter to estimate in the system state. Finally, a measurement is performed, to extract information about the parameter. A question to answer is what is the key property of the input state to maximise the precision of the estimation. It is known that quantum probes outperform of metrology tasks. In particular, asymmetry is the key resource to phase estimation, a kind of parameter estimation where the perturbation of the system is described by a unitary dynamics. For the sake of clarity, I here review the protocol of parameter estimation, starting from the classical scenario [11]. A sample of independent measurement outcomes assigns values x to a random variable X . The goal is to construct a probability function $p_\theta(x)$. The exact value of the coordinate θ in the probability function space is not accessible. An estimator $\hat{\theta}(x)$, and thus $p_{\hat{\theta}}(x)$, can be built yet. The estimator is assumed to be unbiased. This means that its average value corresponds to the actual value of the parameter, $\int(\theta - \hat{\theta}(x)) p_\theta(x) dx = 0$. The estimation precision is benchmarked by the variance of the estimator $\hat{\theta}$. There exists a fundamental limit to the estimator performance. One defines the optimal estimator $\hat{\theta}_{\text{best}}$ as the maximiser of the log-likelihood function $\max_{\hat{\theta}} \ln l(\hat{\theta}|x) = \ln l(\hat{\theta}_{\text{best}}|x)$, $l(\hat{\theta}|x) \equiv p_{\hat{\theta}}(x)$. The information about θ extracted by the measurements is quantified by the score function $\frac{\partial \ln l(\theta|x)}{\partial \theta}$, which is the rate of change of the likelihood function with the parameter value. The second moment of the score is called the Fisher Information:

$$F(\theta) = \int \left(\frac{\partial}{\partial \theta} \log p(x, \theta) \right)^2 p(x, \theta) dx. \quad (1)$$

The Cramér–Rao bound establishes a lower limit to the variance of $\hat{\theta}$ in terms of such quantity,

$$V(p_\theta, \hat{\theta}) \geq \frac{1}{nF(\theta)}, \quad (2)$$

where n is the number of repetitions of the experiment. Hence, the Fisher information is a figure of merit of the classical estimation protocol.

In the quantum scenario, the state of the system is represented by the density matrix ρ_θ . Suppose to encode information about the parameter via a unitary transformation $\rho_\theta = U_\theta \rho_0 U_\theta^\dagger$, $U_\theta = e^{-iH\theta}$. The process corresponds to a path on the stratified manifold of the density matrices [7]. Assumed full knowledge of the Hamiltonian, but being the initial state unknown, what is the best strategy to extract the value of θ ? One performs a generalized positive operator value measure (POVM) $\{\Pi_x\}$ on the rotated state ρ_θ [2], where the $\{\Pi_x\}$ are the measurement operators corresponding to the outcome x . One has $p_\theta(x) = \text{Tr}\{\rho_\theta \Pi_x\}$, and thus

$$F(\rho_\theta) := \int dx \frac{1}{\text{Tr}\{\rho_\theta \Pi_x\}} (\text{Tr}\{\partial_\theta \rho_\theta \Pi_x\})^2. \quad (3)$$

The optimal estimator, i.e. the most informative POVM, is a projection into the eigenbasis of the symmetric-logarithmic derivative L , which solves the equation $\frac{\partial}{\partial \theta} \rho_\theta = \frac{1}{2}(\rho_\theta L + L \rho_\theta)$. Indeed, one has $F(\rho_\theta) \leq \mathcal{F}(\rho, H) := \text{Tr}\{\rho_\theta L^2\}$, where $\mathcal{F}(\rho, H)$ is the symmetric-logarithmic derivative quantum Fisher information (SLDF) [11]. Note that I omitted the parameter label for the state of the system, as the SLDF is independent of its value. The quantum version of the Cramér–Rao bound is then given by

$$V(\rho, \hat{\theta}) \geq \frac{1}{n\mathcal{F}(\rho, H)}. \quad (4)$$

Given the spectral decomposition $\rho = \sum_k \lambda_k |k\rangle\langle k|$, the SLDF takes the expression

$$\mathcal{F}(\rho, H) = \sum_{k < l} \frac{(\lambda_k - \lambda_l)^2}{2(\lambda_k + \lambda_l)} |\langle k | H | l \rangle|^2, \quad (5)$$

where each term in the sum is taken to be zero whenever $\lambda_i = \lambda_j$.

The quantity is well-known to the colleagues working in information geometry. It represents the norm related to the Bures metric, one of the quantum generalizations of the classical Fisher–Rao metric. These are special functions, being proven to be the unique Riemannian metrics which are contractive under quantum operations [14, 15]. Consequently, the resource of the quantum protocol is the speed of evolution of the system undergoing the phase shift, i.e. how fast its state changes, as quantified by the SLDF. I remind that for generic quantum operations the most general expression of the quantum Fisher norms reads

$$\begin{aligned}
||\partial_\theta \rho_\theta||_f^2 &= \sum_{k,l} \frac{|\langle k(\theta) | \partial_\theta \rho_\theta | l(\theta) \rangle|^2}{\lambda_l(\theta) f(\lambda_k(\theta)/\lambda_l(\theta))} \\
&= \sum_k (d_\theta \lambda_k(\theta))^2 / 4 \lambda_k(\theta) \\
&\quad + \sum_{k < l} c_f(\lambda_k(\theta), \lambda_l(\theta)) / 2 |\langle k(\theta) | \partial_\theta \rho_\theta | l(\theta) \rangle|^2,
\end{aligned} \tag{6}$$

where $c_f(i, j) = (jf(i/j))^{-1}$, being f s the Chentsov–Morozova functions [16]. The first term of the right hand side is the classical Fisher–Rao metric $\sum_k (d_\theta \lambda_k(\theta))^2 / (4 \lambda_k(\theta))$, which is the only one surviving for classical stochastic processes. On the other hand, for unitary transformations only the second, purely quantum term remains, as only the eigenbasis of the state evolves. One has $||\partial_\theta \rho_\theta||_f^2 = f(0)/2 ||i[\rho_\theta, H]||_f^2$. Then, the norm obtained by fixing $f(x) = \mathcal{F}(x) = (1+x)/2$ corresponds to the SLDF, $||\partial_\theta \rho_\theta||_F^2 = \mathcal{F}(\rho, H)$. In the more general case, when the quantum channel is not a unitary transformation, classical and quantum contributions co-exist.

To summarize, the SLDF is a function of the commutator between state and Hamiltonian which has a natural interpretation as speed of the evolution of the system along a unitary dynamics. Also, it is a figure of merit of the phase estimation protocol. To verify that the SLDF is a consistent measure of asymmetry, therefore completing the characterization of asymmetry as an information processing resource, it has been proven that the SLDF satisfies a set of required properties [17]:

- The SLDF is upper bounded by the variance, $\mathcal{F}(\rho, H) \leq V(\rho, H)$, $V(\rho, H) = \text{Tr}\{\rho H^2\} - \text{Tr}\{\rho H\}^2$, where the equality is reached for pure states. More precisely, the SLDF is the variance convex roof, $\mathcal{F}(\sum_i p_i |\psi_i\rangle, H) = \inf_{\{p_i, |\psi_i\rangle\}} \sum_i p_i V(|\psi_i\rangle, H)$ [18, 19].
- The SLDF is convex: $\mathcal{F}(p\rho_1 + (1-p)\rho_2, H) \leq p\mathcal{F}(\rho_1, H) + (1-p)\mathcal{F}(\rho_2, H)$.
- For unitaries U , $\mathcal{F}(U\rho U^\dagger, H) = \mathcal{F}(\rho, U^\dagger H U)$.
- The SLDF is non-increasing under operations commuting with the phase shift, $\mathcal{F}(\rho, H) \geq \mathcal{F}(\Phi(\rho), H)$, $\forall \Phi : [\Phi, U_\theta] = 0$. Note that an even stronger constraint, contractivity on average under commuting operations, has been proven [17].

Such properties are in fact met by all the regular quantum Fisher metrics, which are topologically equivalent to the SLDF [20]. Hence, they are legit measures of asymmetry. While the SLDF metric is the most employed one due to its operational interpretations in metrology, I observe that all the parent metrics may find, or have already found, their own operational interpretations. For example, the skew information $-1/2\text{Tr}\{[\sqrt{\rho}, H]^2\}$ was introduced by Wigner and Yanase [21], and then generalized by Dyson [22], while discussing the implications of superselection rules in the measurement process.

One may note that the variance enjoys both a simple expression and a close tie to experimental practice. However, it encodes a classical contribution due to the mixedness of the state, such that it takes arbitrary high values even for states commuting with the observable, vanishing if and only if the state is an observable

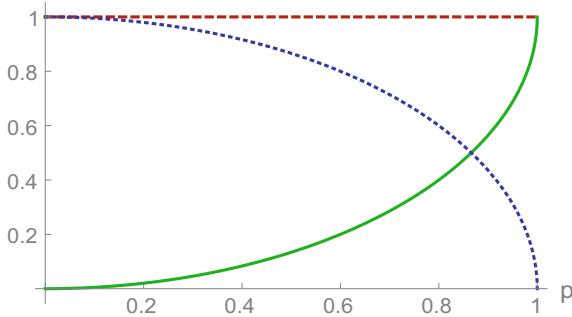


Fig. 1 Asymmetry and classical contribution to the variance are identified for the observable $\sigma_y = -i|0\rangle\langle 1| + i|1\rangle\langle 0|$ in the state $\rho = (1-p)I_2/2 + p|\phi\rangle\langle\phi|$, $|\phi\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$, $p \in [0, 1]$. The red dashed line is the variance, while the green continuous curve is the SLDF. The blue dotted curve represents the difference between the two quantities, a classical mixedness measure. By varying the value of the noise parameter p , while the variance is constant, its quantum and classical components change

eigenstate. The variance is therefore not suitable to quantify asymmetry, apart from the pure state case. Conversely, the SLDF appears as the truly quantum contribution to the variance. I clarify the interplay between variance and SLDF by a simple example, discussed in Fig. 1.

3 Making Coherence Experimentally Observable via Information Geometry

3.1 An Observable Lower Bound to Asymmetry

I here discuss how the information-geometric characterization of asymmetry as speed of evolution of a system yields an experimentally friendly strategy to evaluate the asymmetry of an unknown state. I start by recalling a simple yet powerful algebraic result. Any degree k polynomial function of a quantum state $f_k(\rho)$ equals the mean value of a self-adjoint operator O_f , measured on k copies of the state: $f_k(\rho) = \text{Tr}(O_f \rho^{\otimes k})$ [23, 24]. This is useful because in Quantum Mechanics observables, i.e. measurable quantities, are represented by self-adjoint operators. Then, searching for polynomial approximations is a convenient strategy to overcome expensive state tomography when one wants to determine non-directly observable quantities, e.g. quantifiers of non-linear properties as coherence.

Measuring the corresponding observable O_f is not guaranteed to be practicable. However, this is provably possible for the simplest case of quadratic polynomials, e.g. the overlap between two states $\text{Tr}\{\rho\sigma\}$. By selecting the swap operator $V(\phi_1 \otimes \phi_2) = \phi_2 \otimes \phi_1$, $\forall \phi_{1,2}$, as probe observable acting on the tensor product of

two copies of the system Hilbert space, one has $\text{Tr}\{\rho\sigma\} = \text{Tr}\{V(\rho \otimes \sigma)\}$, $\forall \rho, \sigma$. The swap can be measured by single qubit interferometry. Two copies of the system of interest, or two different degrees of freedom of a single replica, say spin and linear momentum, are prepared in the states ρ, σ . They are correlated by a controlled-swap gate to an ancillary qubit in the initial state α_{in} , which acts as the control qubit. The mean value of the swap is then encoded in the polarisation of the output state of the ancilla, $\text{Tr}\{\alpha_{\text{out}}\sigma_z\} = \text{Tr}(\alpha_{\text{in}}\sigma_z)\text{Tr}\{V(\rho \otimes \sigma)\}$, where $\sigma_{x,y,z}$ are the spin-1/2 Pauli operators. (See the scheme in [25].) A shortcoming of the scheme is that it is currently hard to engineer high fidelity controlled-swaps. The minimal three qubit architecture has been experimentally demonstrated only recently [26]. It is nevertheless possible to overcome the problem whenever the system of interest displays a partition in N subsystems $\{A_i\}$, $i = 1, \dots, N$, e.g. it is an N -qubit computational register [27]. By observing that the swap is factorizable, $V_{A_1 \dots A_N} = \otimes_i V_{A_i}$, one has $\text{Tr}\{\rho_{A_1 \dots A_N} \sigma_{A_1 \dots A_N}\} = \text{Tr}(\otimes_i V_{A_i} (\rho_{A_1 \dots A_N} \otimes \sigma_{A_1 \dots A_N}))$. The state overlap is then obtained by a collective detection of $O(2N)$ local observables on two copies of the N -partite register. The scheme performs exponentially better than the $O(4^N)$ measurements needed to state tomography. Note that state reconstruction also needs an equivalent number of system copies to perform the measurements. This is an experimentally exhausting task already for few qubits, even without considering the exposure to error sources affecting the detection, which arguably grows with the number of measurements. Moreover, one usually finds that protocols involving a limited number of measurements are faster and easier to control. Indeed, the very same existence of full-fledged research lines is devoted to avoid state tomography, e.g. the works in compressed sensing and state discrimination. In the qubit case, local Bell singlet projections are sufficient to evaluate the swap, $V = 1 - 2P_-$, $P_- = |\phi^-\rangle\langle\phi^-|$, $|\phi^-\rangle = 1/\sqrt{2}(|01\rangle - |10\rangle)$. They are implemented by beam splitter interactions between each subsystem A_i copy pair, and single-site polarization detections. A further alternative scheme relying on correlating the system with an array of ancillary qubits has been proposed [25].

I now show that picking the SLDF as a quantifier of asymmetry is useful for experimental practice. No measure of asymmetry can take the form of self-adjoint operators, as it is a non-linear property of a state. Yet, it is possible to construct a geometric lower bound to the SLDF, and in general to any regular quantum Fisher metric (up to a factor), which is a function of observables [9, 25, 28]. By employing the Hilbert–Schmidt norm $\|A\|_2 = \sqrt{\text{Tr}\{AA^\dagger\}}$, one has

$$\begin{aligned} \mathcal{S}_\theta(\rho, H) &\leq \mathcal{F}(\rho, H), \\ \mathcal{S}_\theta(\rho, H) &= \|U_\theta \rho U_\theta^\dagger - \rho\|_2^2 / (2\theta^2) = (\text{Tr}\{\rho^2\} - \text{Tr}\{\rho U_\theta \rho U_\theta^\dagger\}) / \theta^2. \end{aligned} \quad (7)$$

The proof of the result is given in Ref. [9]. Thus, a lower bound to asymmetry is given as a function of purity and overlap, as well as the parameter θ , whose value is experimentally controllable. As discussed before, quadratic polynomials are directly measurable, provided two system replicas $\rho_{1,2} \equiv \rho$. One has $\text{Tr}\{\rho^2\} = \text{Tr}\{V(\rho_1 \otimes \rho_2)\}$, $\text{Tr}\{\rho U_\theta \rho U_\theta^\dagger\} = \text{Tr}\{V(\rho_1 \otimes U_\theta \rho_2 U_\theta^\dagger)\}$. The result is valid for

arbitrary input states. In particular, the method is suitable for large scale detection of asymmetry, as it requires a limited number of measurements regardless the dimension of the system under scrutiny.

3.2 Asymmetry Witnesses Entanglement

The notion of asymmetry (coherence) can be applied to systems which display a structure, e.g. described a partition $S \rightarrow \{S_i\}$. The partition is usually determined by the particulars of the many-body system, as the spatial separation between the parts. One may ask what implies that the state of a multipartite system has asymmetry, and what a measure of asymmetry can reveal about the interdependence between the system parts. In spite of being a basis-dependent feature, asymmetry is affected by quantum correlations, which are basis independent properties of multipartite systems. Indeed, by measuring the observable $\mathcal{S}_\theta(\rho, H)$, entanglement between the particles can be witnessed. There are several entanglement indicators written in terms of the Fisher information, among the many strategies proposed to detect entanglement [29]. In particular, one has that, for N qubits, if $\bar{\mathcal{F}}(\rho) = 1/3(\mathcal{F}(\rho, J_x) + \mathcal{F}(\rho, J_y) + \mathcal{F}(\rho, J_z)) > N/6$, $J_{x(y,z)} = \sum_i 1/2\sigma_{x(y,z)}^i$, $\sigma_{x(y,z)}^1 = \sigma_{x(y,z)} \otimes I_{23}$, $\sigma_{x(y,z)}^2 = I_1 \otimes \sigma_{x(y,z)} \otimes I_3$, $\sigma_{x(y,z)}^3 = I_{12} \otimes \sigma_{x(y,z)} \otimes I_3$, then the state is entangled [30]. Also, a so called k -separable state of N qubits cannot satisfy $\mathcal{F}(\rho, J_{x(y,z)}) \geq (nk^2 + (N - nk)^2)/4$, where $n = \lfloor \frac{N}{k} \rfloor$. For example, given $N = 3$, one has $k = 1 \Rightarrow \mathcal{F} \geq 3/4$ and $k = 2 \Rightarrow \mathcal{F} \geq 5/4$. These bounds represent a witness of entanglement and genuine tripartite entanglement, respectively. Then, the asymmetry lower bound, for additive spin Hamiltonians, is also an entanglement witness:

$$\begin{aligned} \bar{\mathcal{S}}_\theta(\rho) &= 1/3(\mathcal{S}_\theta(\rho, J_x) + \mathcal{S}_\theta(\rho, J_y) + \mathcal{S}_\theta(\rho, J_z)) > N/6 \\ \mathcal{S}_\theta(\rho, J_{x(y,z)}) &\geq 1/4(nk^2 + (N - nk)^2). \end{aligned} \quad (8)$$

It is remarkable that superlinear scaling of asymmetry in multipartite systems witnesses entanglement in action, not just non-separability of the state. In other words, it detects when entanglement provides a tangible advantage, making the state evolve faster under a phase encoding evolution.

4 Experimental Implementation

4.1 Detecting Asymmetry via Bell State Projections

The result calls for experimental demonstration in standard quantum information testbeds. Let us apply the scheme to simulate the detection of coherence and

Table 1 Theoretical values of the SLDF, the lower bound, and the conditions witnessing entanglement, for the spin observables J_K , in ρ_{AB}^p , by fixing $\theta = \pi/6$. The lower bound is an entanglement witness almost as efficient as the quantum Fisher information, being not able to detect metrologically useful entanglement for $p \in [0.3, 0.427517]$. Note that the state is entangled for $p > 1/3$

J_k	J_x	J_y	J_z
$\mathcal{F}(\rho_{AB}^p, J_K)$	$2p^2/(p+1)$	0	$2p^2/(p+1)$
$\mathcal{S}_\theta(\rho_{AB}^p, J_K)$	$p^2 \sin \theta^2 / \theta^2$	0	$p^2 \sin \theta^2 / \theta^2$
$\mathcal{F}(\rho_{AB}^p, J_K) > 1/2$	$p > 0.640388$	/	$p > 0.640388$
$\mathcal{S}_{\pi/6}(\rho_{AB}^p, J_K) > 1/2$	$p > 0.74048$	/	$p > 0.74048$
$\tilde{\mathcal{F}}(\rho_{AB}^p) > 1/3$, $\tilde{\mathcal{S}}_{\pi/6}(\rho_{AB}^p) > 1/3$	$p > 0.3$, $p > 0.427517$		

entanglement in a two-qubit state. A similar experiment has been recently implemented in optical set-up [9]. For convenience, I choose a Bell-diagonal state $\rho^p = p(|\psi\rangle\langle\psi|) + (1-p)/4I_4$, $|\psi\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$, $p \in [0, 1]$, as a probe state for the test. This allows for investigating the behavior of the asymmetry lower bound and entanglement witness in the presence of noise. Two copies of a Bell diagonal state $\rho_{A_1B_1}^p$, $\rho_{A_2B_2}^p$ are prepared. One can verify the behaviour of asymmetry with respect to a set of spin observables $J_K = \sum_{i=A,B} j_K^i$, $j_A^K = j_A^K \otimes I_B$, $j_B^K = I_A \otimes j_B^K$, $j_K = 1/2\sigma_{K=x,y,z}$, by varying the value of the mixing parameter p . This is done by implementing the unitary gate $U_{K,\theta}^A \otimes U_{K,\theta}^B$, $U_{K,\theta} = e^{-iJ_K\theta}$, on a copy of the state. One then needs to evaluate the purity of the state of interest and an overlap with a shifted copy after a rotation has been applied. Note that to evaluate the purity, no gate has to be engineered. For optical setups, one rewrites the two quantities in terms of projections on the antisymmetric subspace. The swap acting on the register $A_1B_1A_2B_2$ is the product of two-qubit swaps on each subsystem, $V_{A_1B_1A_2B_2} = V_{A_1A_2} \otimes V_{B_1B_2}$. Also, for two qubit swaps, one has $V_{12} = I - 2P_{12}^-$, $P_{12}^- = |\psi\rangle\langle\psi|_{12}, |\psi\rangle = 1/\sqrt{2}(|01\rangle - |10\rangle)$. Thus, the observables O_i to be measured are

$$\begin{aligned}
 O_1 &= P_{A_1A_2}^- \\
 O_2 &= P_{B_1B_2}^- \\
 O_3 &= P_{A_1A_2}^- \otimes P_{B_1B_2}^- \\
 \text{Tr}\{\rho^2\} &= 1 + 4\text{Tr}\{O_3\rho_{A_1B_1} \otimes \rho_{A_2B_2}\} - 2\text{Tr}\{O_1\rho_{A_1B_1} \otimes \rho_{A_2B_2}\} \\
 &\quad - 2\text{Tr}\{O_2\rho_{A_1B_1} \otimes \rho_{A_2B_2}\} \\
 \text{Tr}\{\rho_{A_1B_1}U_\theta\rho_{A_2B_2}U_\theta^\dagger\} &= 1 + 4\text{Tr}\{O_3\rho_{A_1B_1} \otimes U_\theta\rho_{A_2B_2}U_\theta^\dagger\} - 2\text{Tr}\{O_1\rho_{A_1B_1} \otimes U_\theta\rho_{A_2B_2}U_\theta^\dagger\} \\
 &\quad - 2\text{Tr}\{O_2\rho_{A_1B_1} \otimes U_\theta\rho_{A_2B_2}U_\theta^\dagger\}.
 \end{aligned} \tag{9}$$

No further action is necessary to verify the presence of entanglement through the witnesses in Eq. (8). For $N = 2$, one has to verify $\mathcal{S}_\theta(\rho, J_K) \geq 1/2$ and $\tilde{\mathcal{S}}_\theta(\rho) > 1/3$. The results are summarised in Table 1.

4.2 Alternative Scheme

Let us suppose that the expectation values of spin magnetizations are measurable, e.g. as it happens in an NMR system [10]. Given a n -qubit register, without the possibility to perform projections, a full state reconstruction would require $2^{2n} - 1$ measurements. However, one can always retrieve the value of the overlap of any pair of states $\text{Tr}\{\rho\sigma\}$ by the same amount of measurements required by the tomography of a *single* state. In our case, by retaining the possibility to implement two state copies (our bound is a polynomial of degree two of the density matrix coefficients), one can evaluate the purity and the overlaps with $2^{2n} - 1$ measurements, avoiding to perform tomography on both the states. There is also a further advantage: there is no need to apply any additional (controlled or not) gate to the network. Let us suppose one wants to extract information about the asymmetry and the entanglement of a three-qubit state ρ_{ABC} . I assume for the sake of simplicity that our state has an X-like density matrix, i.e. it is completely determined by 15 parameters. Thus, 15 measurements are sufficient for state reconstruction. The parameters are the expectation values of magnetization measurements: $\rho_{ABC}^X = 1/8(I_8 + \sum_i \text{Tr}\{\rho_{ABC}^X m_i\})$, where:

$$\begin{aligned} m_1 &= 4\sigma_A^z \otimes I_B \otimes I_C \\ m_2 &= 4I_A \otimes \sigma_B^z \otimes I_C \\ m_3 &= 4I_A \otimes I_B \otimes \sigma_C^z \\ m_4 &= 16\sigma_A^z \otimes \sigma_B^z \otimes I_C \\ m_5 &= 16I_A \otimes \sigma_B^z \otimes \sigma_C^z \\ m_6 &= 16\sigma_A^z \otimes I_B \otimes \sigma_C^z \\ m_7 &= 64\sigma_A^z \otimes \sigma_B^z \otimes \sigma_C^z \\ m_8 &= 64\sigma_A^x \otimes \sigma_B^x \otimes \sigma_C^x \\ m_9 &= 64\sigma_A^x \otimes \sigma_B^x \otimes \sigma_C^y \\ m_{10} &= 64\sigma_A^x \otimes \sigma_B^y \otimes \sigma_C^x \\ m_{11} &= 64\sigma_A^y \otimes \sigma_B^x \otimes \sigma_C^x \\ m_{12} &= 64\sigma_A^y \otimes \sigma_B^y \otimes \sigma_C^x \\ m_{13} &= 64\sigma_A^y \otimes \sigma_B^x \otimes \sigma_C^y \\ m_{14} &= 64\sigma_A^x \otimes \sigma_B^y \otimes \sigma_C^y \\ m_{15} &= 64\sigma_A^y \otimes \sigma_B^y \otimes \sigma_C^y. \end{aligned} \tag{10}$$

As the swap is factorizable, any overlap $\text{Tr}\{\rho_{ABC}^X \sigma_{ABC}\}$ is fully determined by the very same measurements $\{M_i = m_i \otimes m_i\}$, regardless of the density matrix of σ . Also, the measurements to perform are independent of the specific observable J_K . By noting that $e^{i\sigma_l\theta} \sigma_j e^{-i\sigma_l\theta} = \cos(2\theta) \sigma_j - \sin(2\theta) \sigma_k \epsilon_{ijk}$, one finds that for any overlap

$$\text{Tr}\{\rho_{ABC}^P U_\theta \rho_{ABC}^X U_\theta^\dagger\} = 1/8 \text{Tr}\{\rho_{ABC}^X \otimes U_\theta \rho_{ABC}^X U_\theta^\dagger M_i\}. \quad (11)$$

The argument can be generalized to states of any shape and dimension.

5 Conclusion

I here presented an overview of recent works employing information geometry concepts to characterize the most fundamental quantum resources, i.e. coherence and entanglement. It is proven that the SLDF is an asymmetry measure, quantifying the coherence of a state with respect to the eigenbasis of a Hamiltonian H . The results holds for any regular quantum Fisher information metric. Furthermore, the SLDF, as well as the parent metrics, is lower bounded by a function of observable mean values. Such geometric quantity can be then related to experimentally testable effects, i.e. statistical relations in measurement outcomes. When the Hamiltonian is an additive spin operator generating a many-body system dynamics, the lower bound is an entanglement witness. Experimental schemes to detect asymmetry and entanglement, which are implementable with current technology, have been described.

It would be interesting to investigate whether the sensitivity of a system to non-unitary but still quantum evolutions is also linked to the presence of quantum resources. The scenario is certainly closer to realistic experimental practice, where the system is disturbed by uncontrollable error sources, accounting for imperfections in state preparation and dynamics implementation. I also anticipate that employing the information geometry toolbox may critically advance our understanding of genuinely quantum information processing. For example, by establishing fundamental geometric limits and optimal strategies to the control of quantum systems.

Notes and Comments. I thank the organizers and the participants of the IGAIA IV conference for the hospitality and the stimulating discussions. This work was supported by the Los Alamos National Laboratory, project 20170675PRD2. Part of this work was carried out at the University of Oxford, supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under the Grant No. EP/L01405X/1, and by the Wolfson College.

References

1. Schrödinger, E.: The present status of quantum mechanics. *Naturwissenschaften* **23**, 823807 (1935)
2. Nielsen, M.A., Chuang, I.L.: *Quantum Computation and Quantum Information*. Cambridge University Press, New York (2000)
3. Bartlett, S.D., Rudolph, T., Spekkens, R.W.: Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.* **79**, 555 (2007). <https://doi.org/10.1103/RevModPhys.79.555>

4. Marvian, I.: Symmetry, asymmetry and quantum information. Ph.D. thesis, University of Waterloo (2012)
5. Streltsov, A., Adesso, G., Plenio, M.B.: Quantum coherence as a resource. [arxiv:1609.02439](https://arxiv.org/abs/1609.02439)
6. Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K.: Quantum entanglement. *Rev. Mod. Phys.* **81**, 865–942 (2009). <https://doi.org/10.1103/RevModPhys.81.865>
7. Amari, S.: Differential-Geometrical Methods of Statistics. Springer, Berlin (1985)
8. Bengtsson, I., Zyczkowski, K.: Geometry of Quantum States. Cambridge University Press, Cambridge (2007)
9. Zhang, C., et al.: Detecting metrologically useful asymmetry and entanglement by few local measurements. [arxiv:1611.02004](https://arxiv.org/abs/1611.02004)
10. Abragam, A.: The Principles of Nuclear Magnetism. Oxford University Press, Oxford (1978)
11. Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic, New York (1976)
12. Giovannetti, V., Lloyd, S., Maccone, L.: Advances in quantum metrology. *Nat. Photon.* **5**, 222 (2011). <https://doi.org/10.1038/nphoton.2011.35>
13. Tóth, G., Apellaniz, I.: Quantum metrology from a quantum information science perspective. *J. Phys. A: Math. Theor.* **47**, 424006 (2014). <https://doi.org/10.1088/1751-8113/47/42/424006>
14. Petz, D.: Monotone metrics on matrix spaces. *Linear Algebra Appl.* **244**, 81 (1996). [https://doi.org/10.1016/0024-3795\(94\)00211-8](https://doi.org/10.1016/0024-3795(94)00211-8)
15. Petz, D., Ghinea, C.: Introduction to quantum Fisher information. *QP–PQ: Quantum Probab. White Noise Anal.* **27**, 261 (2011)
16. Morozova, E.A., Chentsov, N.N.: Markov invariant geometry on state manifolds. *Itogi Nauki i Tehniki* **36**, 69 (1990). (in Russian)
17. Yadin, B., Vedral, V.: A general framework for quantum macroscopicity in terms of coherence. *Phys. Rev. A* **93**, 022122 (2016). <https://doi.org/10.1103/PhysRevA.93.022122>
18. Tóth, G., Petz, D.: Extremal properties of the variance and the quantum Fisher information. *Phys. Rev. A* **87**, 032324 (2013). <https://doi.org/10.1103/PhysRevA.87.032324>
19. Yu, S.: Quantum Fisher information as the convex roof of variance. [arXiv:1302.5311](https://arxiv.org/abs/1302.5311)
20. Gibilisco, P., Imparato, D., Isola, T.: Inequalities for quantum Fisher information. *Proc. Am. Math. Soc.* **137**, 317 (2008)
21. Wigner, E.P., Yanase, M.M.: Information content of distributions. *PNAS* **49**, 910–918 (1963)
22. Wherl, A.: General properties of entropy. *Rev. Mod. Phys.* **50**, 221 (1978). <https://doi.org/10.1103/RevModPhys.50.221>
23. Paz, J.P., Roncaglia, A.: A quantum gate array can be programmed to evaluate the expectation value of any operator. *Phys. Rev. A* **68**, 052316 (2003). <https://doi.org/10.1103/PhysRevA.68.052316>
24. Brun, T.: Measuring polynomial functions of states. *Quantum Inf. Comp.* **4**, 401 (2004)
25. Girolami, D.: Observable measure of quantum coherence in finite dimensional systems. *Phys. Rev. Lett.* **113**, 170401 (2014). <https://doi.org/10.1103/PhysRevLett.113.170401>
26. Patel, R.B., Ho, J., Ferreyrol, F., Ralph, T. C., Pryde, G.J.: A quantum Fredkin gate. *Sci. Adv.* **25**(2), e1501531. <https://doi.org/10.1126/sciadv.1501531>
27. Moura Alves, C., Jaksch, D.: Multipartite entanglement detection in bosons. *Phys. Rev. Lett.* **93**, 110501 (2004). <https://doi.org/10.1103/PhysRevLett.93.110501>
28. Girolami, D., Yadin, B.: Witnessing multipartite entanglement by detecting asymmetry. *Entropy* **19**(3), 124 (2017). <https://doi.org/10.3390/e19030124>
29. Gühne, O., Tóth, G.: Entanglement detection. *Phys. Rep.* **474**, 1 (2009). <https://doi.org/10.1016/j.physrep.2009.02.004>
30. Pezzé, L., Smerzi, A.: Ultrasensitive two-mode interferometry with single-mode number squeezing. *Phys. Rev. Lett.* **110**, 163604 (2013). <https://doi.org/10.1103/PhysRevLett.110.163604>

Characterising Two-Sided Quantum Correlations Beyond Entanglement via Metric-Adjusted f -Correlations



Marco Cianciaruso, Irénée Frérot, Tommaso Tufarelli
and Gerardo Adesso

Abstract We introduce an infinite family of quantifiers of quantum correlations beyond entanglement which vanish on both classical-quantum and quantum-classical states and are in one-to-one correspondence with the metric-adjusted skew informations. The ‘quantum f -correlations’ are defined as the maximum metric-adjusted f -correlations between pairs of local observables with the same fixed equispaced spectrum. We show that these quantifiers are entanglement monotones when restricted to pure states of qubit-qudit systems. We also evaluate the quantum f -correlations in closed form for two-qubit systems and discuss their behaviour under local commutativity preserving channels. We finally provide a physical interpretation for the quantifier corresponding to the average of the Wigner–Yanase–Dyson skew informations.

Keywords Information geometry · Quantum correlations

1 Introduction

Nonclassical correlations in quantum systems manifest themselves in several forms such as non-locality [1, 2], steering [3, 4], entanglement [5], and discord-type quantum correlations beyond entanglement [6–9]. The purposes of identifying these various manifestations of quantumness are manifold. From a theoretical viewpoint, it is crucial to explore the classical-quantum boundary and the quantum origins of our everyday classical world [10]. From a pragmatic perspective, all such forms of

M. Cianciaruso · T. Tufarelli · G. Adesso (✉)

School of Mathematical Sciences, Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, The University of Nottingham, University Park, Nottingham NG7 2RD, UK

e-mail: gerardo.adesso@nottingham.ac.uk

I. Frérot

Laboratoire de Physique, Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Lyon F-69342, France

quantumness represent resources for some operational tasks and allow us to achieve them with an efficiency that is unreachable by any classical means [11].

In particular, quantum correlations beyond entanglement can be linked to the figure of merit in several operational tasks such as local broadcasting [12, 13], entanglement distribution [14, 15], quantum state merging [16–18], quantum state redistribution [19], quantum state discrimination [20–25], black box quantum parameter estimation [26], quantum data hiding [27], entanglement activation [28–31], device-dependent quantum cryptography [32, 33], quantum work extraction [34–37], quantum refrigeration [38, 39] and quantum predictive processes [40].

The quantification of quantum correlations is thus necessary to gauge the quantum enhancement when performing the aforementioned operational tasks. An intuitive way to measure the quantum correlations present in a state is to quantify the extent to which it violates a property characterising classically correlated states. For example, quantum correlated states cannot be expressed as a statistical mixture of locally, classically distinguishable states [20, 21, 41–44]; are altered by any local measurement [34, 45–47], any non-degenerate local unitary [48, 49] and any local entanglement-breaking channel [50]; always lead to creation of entanglement with an apparatus during a local measurement [28–31, 51]; manifest quantum asymmetry with respect to all local non-degenerate observables [26, 52] and coherence with respect to all local bases [8, 53, 54].

The ensuing measures of quantum correlations mostly belong to the following two categories: (i) asymmetric quantifiers, also known as one-sided measures, which vanish only on classical-quantum (resp., quantum-classical) states and thus capture the quantum correlations with respect to subsystem A (resp., B) only; (ii) symmetric quantifiers which vanish only on classical-classical states and thus capture the quantum correlations with respect to *either* subsystem A or B . The latter category of measures have also been improperly referred to as two-sided quantifiers, even though they do not actually capture the quantum correlations with respect to *both* subsystems A and B .

In this paper we instead introduce an infinite family of quantifiers of quantum correlations beyond entanglement which vanish on both classical-quantum and quantum-classical states and thus properly capture the quantum correlations with respect to both subsystems. More precisely, the ‘quantum f -correlations’ are here defined as the maximum metric-adjusted f -correlations between pairs of local observables with the same fixed equispaced spectrum and are in one-to-one correspondence with the family of metric-adjusted skew informations [55–61]. While similar ideas were explored earlier in [62, 63] to quantify entanglement, here we show that our quantifiers only reduce to entanglement monotones when restricted to pure states. The latter property is one of the desiderata for general measures of quantum correlations beyond entanglement [8]. Other desiderata, such as monotonicity under sets of operations which cannot create quantum correlations, are also critically assessed. We find in particular that the quantum f -correlations, while endowed with strong physical motivations, are not monotone under all such operations in general, although we show in the concluding part of the paper that their definition may be amended to cure this potential drawback.

The paper is organised as follows. In Sect. 2 we briefly review the characterisation and quantification of quantum correlations beyond entanglement by adopting a resource-theoretic framework. In Sect. 3 we define the quantum f -correlations and show that they vanish on both classical-quantum and quantum-classical states and are invariant under local unitaries for any bipartite quantum system. We further prove that they are entanglement monotones when restricted to pure states of qubit-qudit systems. We also analytically evaluate these quantifiers for two-qubit systems and analyse their behaviour under local commutativity preserving channels, showing that they are not monotone in general. In Sect. 4 we provide a physical interpretation for the special quantifier corresponding to the average of the Wigner–Yanase–Dyson skew informations and explore applications to statistical mechanics and many-body systems. We draw our conclusions and outline possible extensions of this work in Sect. 5, including a more general definition for a class of quantifiers of two-sided quantum correlations based on the metric-adjusted f -correlations. The latter quantities are proven in Appendix A to be monotone under local commutativity preserving channels for two-qubit systems, hence fulfilling *all* the resource-theoretic requirements for quantum correlations beyond entanglement.

2 Quantifying Quantum Correlations Beyond Entanglement

In this Section we concisely review the theory of the quantification of quantum correlations beyond entanglement by resorting to a resource-theoretic perspective [64, 65], even though the resource theory of this sort of correlations is still far from being established [66, 67].

From a minimalistic viewpoint, a resource theory relies on the following two ingredients: the sets of free states and free operations, both of which are considered to be freely implementable and are thus such that no resourceful state can be prepared through free operations. A fundamental question that any resource theory must address is how to quantify the resource present in any state. One could naively think that there should be a unique quantifier of a given resource, determining a universal ordering of the resourceful states. However, this should not be the case for the following two reasons. First, the same resource can be exploited for different operational tasks, such that a given resourceful state can be more successful than another one in order to achieve a given operational task, and viceversa when considering another task. Second, it is desirable to assign an operational meaning to any quantifier of a resource, in the sense that it needs to quantify how much the resource possessed by a given state will be useful for achieving a given operational task. An immediate consequence is that, in general, the various quantifiers disagree on the ordering of the resourceful states. Nevertheless, in order to have an operational significance, any *bona fide* quantifier of a resource must be compatible with the sets of free states

and free operations in the following sense: it must be zero for any free state and monotonically non-increasing under free operations.

Let us start by identifying the set of free states corresponding to quantum correlations beyond entanglement. As we have already mentioned in Sect. 1, there are at least four settings that we can consider. Within the asymmetric/one-sided setting, when measuring quantum correlations with respect to subsystem A only, the free states are the so-called *classical-quantum* (CQ) states, i.e. particular instances of biseparable states than can be written as follows

$$\chi_{cq}^{AB} = \sum_i p_i^A |i\rangle\langle i|^A \otimes \tau_i^B, \quad (1)$$

where $\{p_i^A\}$ is a probability distribution, $\{|i\rangle^A\}$ denotes an orthonormal basis for subsystem A , and $\{\tau_i^B\}$ are arbitrary states for subsystem B . CQ states represent the embedding of a classical probability distribution $\{p_i^A\}$ relating to only subsystem A into the quantum state space of a bipartite quantum system AB .

Analogously, when measuring quantum correlations with respect to subsystem B only, the free states are the so-called *quantum-classical* (QC) states, which are of the form

$$\chi_{qc}^{AB} = \sum_j p_j^B \tau_j^A \otimes |j\rangle\langle j|^B, \quad (2)$$

where $\{p_j^B\}$ is a probability distribution, $\{|j\rangle^B\}$ denotes an orthonormal basis for subsystem B , and $\{\tau_j^A\}$ are arbitrary states for subsystem A .

Within the symmetric setting, and when measuring quantum correlations with respect to either subsystem A or B , the free states are the so-called *classical-classical* (CC) states that can be written in the following form

$$\chi_{cc}^{AB} = \sum_{i,j} p_{ij}^{AB} |i\rangle\langle i|^A \otimes |j\rangle\langle j|^B, \quad (3)$$

where p_{ij}^{AB} is a joint probability distribution, while $\{|i\rangle^A\}$ and $\{|j\rangle^B\}$ denote orthonormal bases for subsystem A and B , respectively. CC states correspond to the embedding of classical bipartite probability distributions $\{p_{ij}^{AB}\}$ into a bipartite quantum state space.

Finally, within the symmetric and properly two-sided setting, wherein one measures quantum correlations with respect to both subsystems A and B , the free states are given by the union of the sets of CQ and QC states.

While the free states of the resource theory of general quantum correlations are well identified, the corresponding free operations are still under debate. Having said that, in [68] it has been shown that all, and only, the *local* operations that leave the set of CQ states invariant are the local commutativity preserving operations (LCPOs) on subsystem A , $\Phi_A^{LCPO} \equiv \Lambda_A \otimes \mathbb{I}_B$, where Λ_A acts on subsystem A is such a way that $[\Lambda_A(\rho^A), \Lambda_A(\sigma^A)] = 0$ when $[\rho^A, \sigma^A] = 0$ for arbitrary marginal states ρ^A and

σ^A . Analogously, in [68] it has been also shown that the LCPOs on subsystem B , Φ_B^{LCPO} , are all and only the local operations leaving the set of QC states invariant, while the LCPOs on both subsystems A and B , Φ_{AB}^{LCPO} , are all and only the local operations preserving the set of CC states. Consequently, due to the fact that free operations cannot create a resourceful state out of a free state, the free operations of the resource theory of quantum correlations beyond entanglement must be within the set of LCPOs, if one imposes a priori the locality of such free operations.

In the case of a qubit, the commutativity preserving operations are constituted by unital and semi-classical channels [69]. Unital channels are defined as those maps that leave the maximally mixed state invariant, whereas semi-classical channels transform the set of all states into a subset of states which are all diagonal in the same basis. More generally, for higher dimensional quantum systems, the commutativity preserving operations are either isotropic or completely decohering channels [70].

By considering a resource theory of general quantum correlations corresponding to the largest possible set of local free operations, and taking into account that entanglement is the only kind of quantum correlations that pure states can have, we define any non-negative function Q on the set of states ρ to be a *bona fide* quantifier of two-sided quantum correlations beyond entanglement if it satisfies the following desiderata:

- (Q1) $Q(\rho) = 0$ if ρ is either CQ or QC;
- (Q2) Q is invariant under local unitaries, i.e. $Q\left((U_A \otimes U_B)\rho(U_A^\dagger \otimes U_B^\dagger)\right) = Q(\rho)$ for any state ρ and any local unitary operation U_A (U_B) acting on subsystem A (B);
- (Q3) $Q(\Phi_{AB}^{LCPO}(\rho)) \leq Q(\rho)$ for any LCPO Φ_{AB}^{LCPO} on both subsystems A and B ;
- (Q4) Q reduces to an entanglement monotone when restricted to pure states.

We remark that, while (Q1), (Q2) and (Q4) are well established requirements, (Q3) may be too strong to impose, as monotonicity under a smaller set of free operations might be sufficient if justified on physical grounds. We will discuss this point further in the following.

For completeness, let us mention that when considering an asymmetric/one-sided measure with respect to subsystem A (resp., B), two of the above desiderata have to be slightly modified. Specifically, property (Q1) becomes: $Q(\rho) = 0$ if ρ is a CQ (resp., QC) state, while an even stricter monotonicity requirement may replace (Q3), namely being monotonically non-increasing under LCPOs on subsystem A (resp., B) and arbitrary local operations on subsystem B (resp., A). When considering instead symmetric measures with respect to either subsystem A or B , property (Q3) stays the same, while property (Q1) becomes: $Q(\rho) = 0$ if ρ is a CC state. On the other hand, properties (Q2) and (Q4) apply equally to all of the aforementioned four settings.

3 Quantum f -Correlations

In this Section we define the family of ‘quantum f -correlations’ and show that they all satisfy requirements (Q1) and (Q2) for any bipartite quantum system as well as property (Q4) for any qubit-qudit system. We also evaluate these quantifiers in closed form for two-qubit systems and discuss their behaviour under LCPOs, which reveals violations to (Q3), even though these violations can be cured by a suitable reformulation as shown in Sect. 5.1.

3.1 Metric-Adjusted Skew Informations

Let us start by introducing the family of metric-adjusted skew informations (MASIs). The Petz classification theorem provides us with a characterisation of the MASIs [55–61], by establishing a one-to-one correspondence between them and the Morozova–Čencov (MC) functions

$$c^f(x, y) = \frac{1}{yf(x/y)} \quad (4)$$

parametrized by any function $f(t) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that is

- (i) operator monotone (or *standard*), i.e. for any positive semi-definite operators A and B such that $A \leq B$, then $f(A) \leq f(B)$;
- (ii) symmetric (or *self-inverse*), i.e. $f(t) = tf(1/t)$;
- (iii) normalised, i.e. $f(1) = 1$.

The set of all normalised symmetric operator monotone functions f on the interval $(0, +\infty)$ is usually denoted by \mathcal{F}_{op} . It follows that any MC function is symmetric in its arguments, i.e. $c^f(x, y) = c^f(y, x)$, and homogeneous of degree -1 , i.e. $c^f(\alpha x, \alpha y) = \alpha^{-1}c^f(x, y)$.

In this formalism, the MASI of a quantum state $\rho > 0$ with respect to an observable O , corresponding to the MC function c^f , can be defined as follows [58]:

$$I^f(\rho, O) = \frac{f(0)}{2} \sum_{ij} c^f(p_i, p_j)(p_i - p_j)^2 \langle i|O|j\rangle \langle j|O|i\rangle, \quad (5)$$

where $\rho = \sum_i p_i |i\rangle\langle i|$ is the spectral decomposition of ρ and we have assumed f to be *regular*, i.e. $\lim_{t \rightarrow 0^+} f(t) \equiv f(0) > 0$. Notable examples of MASIs are the Bures–Uhlmann information [71], corresponding to the maximal function $f^{BU}(t) = (1+t)/2$, and the Wigner–Yanase–Dyson skew informations [72], corresponding to the functions

$$f_\alpha^{WYD}(t) = \frac{\alpha(1-\alpha)(1-t)^2}{(1-t^\alpha)(1-t^{1-\alpha})}, \quad (6)$$

for any $0 < \alpha < 1$.

Each MASI $I^f(\rho, O)$ can be interpreted as a genuinely quantum contribution to the uncertainty of the observable O in the state ρ [57, 59–61, 63, 73, 74]. Two important properties of $I^f(\rho, O)$, justifying this intuition, are that: (a) $I^f(\rho, O) = 0$ iff $[\rho, O] = 0$; and (b) $I^f(\rho, O) \leq \text{Var}_\rho(O) = \text{Tr}(\rho O^2) - [\text{Tr}(\rho O)]^2$ where the equality holds for pure states. Hence, a nonzero MASI indicates that the state ρ contains coherences among different eigenstates of O [property (a)]. For a pure state, any source of uncertainty has a quantum origin, and all MASIs coincide with the ordinary (Robertson–Schrödinger) variance of O in the state.

The MASI $I^f(\rho, O)$ may also be interpreted as asymmetry of the state ρ with respect to the observable O [75–77]. In a bipartite system ρ_{AB} , the minimum of $I^f(\rho_{AB}, O_A \otimes \mathbb{I}_B)$ over local non-degenerate observables O_A (with fixed spectrum) can be seen as a measure of asymmetric/one-sided quantum correlations of the state ρ_{AB} with respect to subsystem A , as investigated for special instances in [26, 52]. Applications of different MASIs to determining quantum speed limits for closed and open quantum system dynamics have been explored in [78, 79] and references therein.

3.2 Maximising Metric-Adjusted f -Correlations Over Pairs of Local Observables

We now recall the notion of *metric-adjusted f -correlations* between observables O_A and O_B in the quantum state $\rho = \sum_i p_i |i\rangle\langle i|$, defined by [58, 59]

$$\Upsilon^f(\rho, O_A, O_B) = \frac{f(0)}{2} \sum_{ij} c^f(p_i, p_j)(p_i - p_j)^2 \langle i|O_A|j\rangle\langle j|O_B|i\rangle, \quad (7)$$

Equivalently, one can write [80, 81]

$$\Upsilon^f(\rho, O_A, O_B) = \text{Cov}_\rho(O_A, O_B) - \text{Cov}_\rho^{\tilde{f}}(O_A, O_B), \quad (8)$$

where

$$\text{Cov}_\rho^f(O_A, O_B) = \text{Tr} \left\{ m^f \left[\rho(O_A - \text{Tr}(\rho O_A)), (O_A - \text{Tr}(\rho O_A))\rho \right] (O_B - \text{Tr}(\rho O_B)) \right\} \quad (9)$$

stands for the Petz f -covariance [57] associated with the Kubo–Ando operator mean $m^f[A, B] = A^{\frac{1}{2}} f(A^{-\frac{1}{2}} B A^{\frac{1}{2}}) A^{\frac{1}{2}}$ [82], reducing to the ordinary (Robertson–Schrödinger) covariance

$$\text{Cov}_\rho(O_A, O_B) = \frac{1}{2} \text{Tr} [\rho(O_A O_B + O_B O_A)] - \text{Tr}(\rho O_A) \text{Tr}(\rho O_B) \quad (10)$$

for $f(t) \equiv f^{BU}(t) = (1+t)/2$ (in which case m^f denotes the arithmetic mean), and [80, 83]

$$\tilde{f}(t) = \frac{1}{2} \left[(t+1) - (t-1)^2 \frac{f(0)}{f(t)} \right], \quad (11)$$

for any regular $f \in \mathcal{F}_{op}$. It follows from Eq. (8) or, alternatively, from Eq. (7) due to the symmetry of the MC functions $c^f(p_i, p_j)$, that

$$I^f(\rho, O_A + O_B) = I^f(\rho, O_A) + I^f(\rho, O_B) + 2\Upsilon^f(\rho, O_A, O_B). \quad (12)$$

In other words, the metric-adjusted f -correlations can be seen as measures of non-additivity of the corresponding MASIs.

We are now ready to define the *quantum f -correlations* of a state ρ as

$$Q^f(\rho) = \max_{O_A, O_B} \Upsilon^f(\rho, O_A \otimes \mathbb{I}_B, \mathbb{I}_A \otimes O_B), \quad (13)$$

where the maximisation is over all local observables O_A and O_B whose eigenvalues are equispaced with spacing $d/(d-1)$ and are given by $\{-d/2, -d/2 + d/(d-1), \dots, d/2 - d/(d-1), d/2\}$, with $d = \min\{d_A, d_B\}$. If the dimensions of the two subsystems are different, say $d_B > d_A$, the remaining eigenvalues of O_B are set to zero (and vice-versa if $d_A > d_B$).

3.3 The Quantum f -Correlations Satisfy Q1 and Q2

We now show that the quantity Q^f defined in Eq. (13) vanishes on both CQ and QC states and thus satisfies requirement (Q1) for any function $f \in \mathcal{F}_{op}$. This is due to the fact that the metric-adjusted f -correlations actually vanish on both CQ and QC states for any pair of local observables O_A and O_B . Indeed, consider a CQ state as in Eq. (1). This can also be written as follows:

$$\chi_{cq}^{AB} = \sum_{i,j} p_{i,j} |i^A\rangle\langle i^A| \otimes |\psi_{i,j}^B\rangle\langle\psi_{i,j}^B|, \quad (14)$$

where we have used the spectral decomposition of the states τ_i^B , i.e. $\tau_i^B = \sum_j q_{j|i} |\psi_{i,j}^B\rangle\langle\psi_{i,j}^B|$, and introduced the probabilities $p_{i,j} = p_i^A q_{j|i}$.

By using Eq. (14) we can see that (up to the factor $f(0)/2$)

$$\begin{aligned} & \Upsilon^f(\chi_{cq}^{AB}, O_A \otimes \mathbb{I}_B, \mathbb{I}_A \otimes O_B) \\ & \propto \sum_{i,j,k,l} g^f(p_{i,j}, p_{k,l}) \langle i^A | \langle\psi_{i,j}^B | O_A \otimes \mathbb{I}_B | k^A \rangle | \psi_{k,l}^B \rangle \langle k^A | \langle\psi_{k,l}^B | \mathbb{I}_A \otimes O_B | i^A \rangle | \psi_{i,j}^B \rangle \end{aligned}$$

$$\begin{aligned}
&= \sum_{i,j,k,l} g^f(p_{i,j}, p_{k,l}) \langle i^A | O_A | k^A \rangle \langle \psi_{i,j}^B | \psi_{k,l}^B \rangle \langle k^A | i^A \rangle \langle \psi_{k,l}^B | O_B | \psi_{i,j}^B \rangle \\
&= \sum_{i,j,l} g^f(p_{i,j}, p_{i,l}) \langle i^A | O_A | i^A \rangle \langle \psi_{i,j}^B | \psi_{i,l}^B \rangle \langle \psi_{i,l}^B | O_B | \psi_{i,j}^B \rangle \\
&= \sum_{i,j} g^f(p_{i,j}, p_{i,j}) \langle i^A | O_A | i^A \rangle \langle \psi_{i,j}^B | O_B | \psi_{i,j}^B \rangle \\
&= 0,
\end{aligned} \tag{15}$$

being $\langle k^A | i^A \rangle = \delta_{i,k}$, $\langle \psi_{i,j}^B | \psi_{i,l}^B \rangle = \delta_{j,l}$ for any i and $g^f(p_{i,j}, p_{i,j}) \equiv c^f(p_{i,j}, p_{i,j})$ ($p_{i,j} - p_{i,j}$) $^2 = 0$. An analogous reasoning applies when considering QC states, thus concluding our proof.

It is also clear that Q^f is by construction invariant under local unitaries, as the latter cannot vary the spectrum of the local observables involved in the optimisation in Eq.(13), so that Q^f satisfies requirement (Q2) for any bipartite system.

3.4 Quantum f -Correlations as Entanglement Monotones for Pure Qubit-Qudit States

Specialising our discussion to qubit-qudit systems, we now show that Q^f is an entanglement monotone [84, 85] when restricted to pure states, and thus satisfies requirement (Q4) for this special class of bipartite systems. For every MC function c^f , the quantity Q^f reduces to the maximum ordinary (Robertson–Schrödinger) covariance of local observables when calculated for pure states, i.e.

$$Q^f(|\psi\rangle) = E(|\psi\rangle) \equiv \max_{O_A, O_B} (\langle \psi | O_A \otimes O_B | \psi \rangle - \langle \psi | O_A \otimes \mathbb{I}_B | \psi \rangle \langle \psi | \mathbb{I}_A \otimes O_B | \psi \rangle), \tag{16}$$

where the maximisation is over all local observables O_A and O_B with equispaced eigenvalues. We thus want to prove that this is a pure state entanglement monotone.

It is known that if $E(|\psi\rangle)$ can be written as a Schur-concave function of the Schmidt coefficients $\{\lambda_i\}$ of $|\psi\rangle$, then $E(|\psi\rangle)$ is a pure state entanglement monotone [86, 87]. Let us recall that the Schmidt decomposition [88] of a bipartite pure state $|\psi\rangle$ is given by

$$|\psi\rangle = \sum_{i=1}^d \lambda_i |e_i^A\rangle \otimes |f_i^B\rangle, \tag{17}$$

where $\{|e_i^A\rangle\}$ and $\{|f_i^B\rangle\}$ are orthonormal states of subsystems A and B , and the Schmidt coefficients λ_i satisfy $\lambda_i \geq 0$ and $\sum_{i=1}^d \lambda_i^2 = 1$.

By substituting the Schmidt decomposition of $|\psi\rangle$ into Eq.(16) we get

$$E(|\psi\rangle) = \sum_{i,j=1}^d a_{ij} b_{ij} \lambda_i \lambda_j - \sum_{i,j=1}^d a_{ii} b_{jj} \lambda_i^2 \lambda_j^2, \quad (18)$$

where $a_{ij} = \langle e_i^A | O_A^* | e_j^A \rangle$, $b_{ij} = \langle f_i^B | O_B^* | f_j^B \rangle$, while O_A^* and O_B^* are the local observables achieving the maximum in Eq.(16). Moreover, by using the fact that $\sum_{i=1}^d \lambda_i^2 = 1$, we have that

$$E(|\psi\rangle) = \sum_{j>i=1}^d (a_{ii} - a_{jj})(b_{ii} - b_{jj}) \lambda_i^2 \lambda_j^2 + \sum_{j>i=1}^d (a_{ij} b_{ij} + a_{ji} b_{ji}) \lambda_i \lambda_j. \quad (19)$$

Now we just need to prove that the function expressed in Eq.(19) is Schur-concave. The Schur–Ostrowski criterion [89] says that a symmetric function $f(\lambda_1, \lambda_2, \dots, \lambda_d)$ is Schur-concave if, and only if,

$$(\lambda_i - \lambda_j) \left(\frac{\partial f}{\partial \lambda_i} - \frac{\partial f}{\partial \lambda_j} \right) \leq 0 \quad (20)$$

for any $j > i \in \{1, 2, \dots, d\}$. However, before applying this criterion to the function in Eq.(19), we need to find the optimal local observables O_A^* and O_B^* and thus the explicit form of the coefficients a_{ij} and b_{ij} .

We may now exploit some convenient simplifications occurring in the case $d = 2$, i.e., for any qubit-qudit bipartite system. In this case we have that

$$E(|\psi\rangle) = (a_{11} - a_{22})(b_{11} - b_{22}) \lambda_1^2 \lambda_2^2 + (a_{12} b_{12} + a_{21} b_{21}) \lambda_1 \lambda_2, \quad (21)$$

which is a symmetric function of λ_1 and λ_2 , regardless of the form of the local optimal observables O_A^* and O_B^* . Having O_A^* and O_B^* the same spectrum $\{-1, 1\}$ by construction, we can easily see that the optimal local observables must be of the form

$$O_A^* = \begin{pmatrix} a & e^{i\varphi} \sqrt{1-a^2} \\ e^{-i\varphi} \sqrt{1-a^2} & -a \end{pmatrix}, \quad (22)$$

$$O_B^* = \begin{pmatrix} b & e^{i\phi} \sqrt{1-b^2} & 0 & \cdots & 0 \\ e^{-i\phi} \sqrt{1-b^2} & -b & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad (23)$$

for some $-1 \leq a \leq 1$ and $-1 \leq b \leq 1$, so that Eq.(21) becomes

$$E(|\psi\rangle) = 4ab\lambda_1^2\lambda_2^2 + 2\sqrt{(1-a^2)(1-b^2)}\cos(\varphi+\phi)\lambda_1\lambda_2, \quad (24)$$

whose maximum is given by $\varphi = \phi = a = b = 0$, i.e.,

$$E(|\psi\rangle) = 2\lambda_1\lambda_2. \quad (25)$$

By applying the Schur–Ostrowski criterion to the function E in Eq. (25), we find that

$$(\lambda_1 - \lambda_2) \left(\frac{\partial E}{\partial \lambda_1} - \frac{\partial E}{\partial \lambda_2} \right) = -2(\lambda_1 - \lambda_2)^2 \leq 0, \quad (26)$$

so that $E(|\psi\rangle)$ is a Schur-concave function of the Schmidt coefficients of $|\psi\rangle$ and thus is a pure state entanglement monotone for any qubit-qudit system. In particular, for a qubit-qubit system the above function reduces to the well known concurrence [90].

3.5 Analytical Expression of the Two-Qubit Quantum f–Correlations

We now analytically evaluate $Q^f(\rho)$ for any MC function c^f when restricting to two-qubit states. We start by noting that in the two-qubit case any local observable whose spectrum is given by $\{-1, 1\}$ can be written as $O = \mathbf{n} \cdot \boldsymbol{\sigma}$, with $\mathbf{n} = \{n_1, n_2, n_3\}$ being a real unit vector, $\mathbf{n} \cdot \mathbf{n} = 1$, and $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \sigma_3\}$ is the vector of Pauli matrices. Therefore, Eq. (13) becomes

$$Q^f(\rho^{AB}) = \max_{\mathbf{n}_A, \mathbf{n}_B} \mathbf{n}_A^T M^f \mathbf{n}_B, \quad (27)$$

where the maximum is over all real unit vectors \mathbf{n}_A and \mathbf{n}_B , while M^f is the 3×3 matrix with elements

$$M_{ij}^f = \Upsilon^f(\rho, \sigma_i^A \otimes \mathbb{I}_B, \mathbb{I}_A \otimes \sigma_j^B), \quad (28)$$

so that we can formally write the result of the maximisation as

$$Q^f(\rho^{AB}) = s_{\max}(M^f), \quad (29)$$

where $s_{\max}(M^f)$ is the maximum singular value of the matrix M^f .

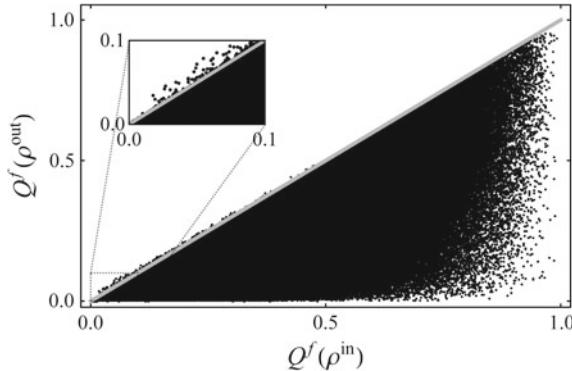


Fig. 1 Comparison between the quantum f -correlation $Q^f(\rho^{\text{in}})$ of 10^6 random two-qubit states (horizontal axis), and the quantum f -correlation $Q^f(\rho^{\text{out}})$ of the corresponding states after random local unital channels (vertical axis), for $f(t) = (1 + \sqrt{t})^2/4$ associated with the Wigner–Yanase skew information. The presence of points above the solid gray line $Q^f(\rho^{\text{out}}) = Q^f(\rho^{\text{in}})$, better highlighted in the zoomed-in inset, shows that the quantum f -correlations are in general not monotonically non-increasing under local unital channels for two qubits, as discussed further in the main text

3.6 Behaviour of the Quantum f -Correlations Under Local Commutativity Preserving Channels

Here we investigate whether the quantum f -correlations defined in Eq.(13) are monotonically nonincreasing under LCPOs, as would be demanded by the resource-theoretic desideratum (Q3). The answer is trivially affirmative in the case of local semi-classical channels, which map any state into one with vanishing Q^f . To investigate the non-trivial cases, we carry out a numerical exploration for two-qubit states subject to local unital channels. Figure 1 compares the input $Q^f(\rho^{\text{in}})$ with the output $Q^f(\rho^{\text{out}})$ for 10^6 randomly generated two-qubit states ρ^{in} , where $\rho^{\text{out}} = p(U_A \otimes \mathbb{I}_B)\rho^{\text{in}}(U_A^\dagger \otimes \mathbb{I}_B) + (1 - p)(V_A \otimes \mathbb{I}_B)\rho^{\text{in}}(V_A^\dagger \otimes \mathbb{I}_B)$, with random local unitaries U_A , V_A , and random probability $p \in [0, 1]$. The analysis reported in Fig. 1 has been done in particular using the Wigner–Yanase skew information [72], specified by $f^{WY} \equiv f_{1/2}^{WYD}$ [see Eq.(6)], although qualitatively similar results are obtained for other choices of f . As clear from the plot, while monotonicity under local unital channels appears to hold in most cases, narrow violations can still be identified in about 0.1% of the cases in our study. This shows that the quantifiers defined by Eq.(13) can increase under some LCPOs, thus generally failing to fulfil (Q3).

On one hand, this may suggest that the quantum f -correlations are not entirely satisfactory measures of general quantum correlations from a resource theory perspective, while still providing an approximately reliable quantitative estimate. On the other hand, this may indicate that a more narrow and possibly physically relevant subset of LCPOs may play a preferred role in identifying the free operations for the

resource theory of general quantum correlations, and the quantum f -correlations could still be monotone under such a restricted set.

In fact, the latter scenario resembles what happens in the resource theories of entanglement and coherence, wherein the chosen free operations do not cover the whole maximal set of operations leaving the set of free states invariant. For example, in the entanglement case, the free operations are the local operations and classical communication, which are only a restricted subset of the separability preserving operations [91]. In the coherence case, there are in fact many different definitions of free operations that are proper subsets of the maximal set of incoherence preserving operations [92], such as the incoherent operations [93], the strictly incoherent operations [94], the translationally invariant operations [79], and several others [95], with no consensus yet reached on the most representative set.

In the case of general quantum correlations, as already mentioned in Sect. 2, the quest for the physical justification to identify the right set of free operations is still open [8]. Based on our numerical analysis for two-qubit systems, there was no subset of local unital channels which clearly emerged as the one under which monotonicity could hold in general. Hence, a way to save (Q3) for the quantum f -correlations could be to impose only monotonicity under local semi-classical channels, which might be nonetheless too weak a constraint.

In Sect. 4, we discuss one possible physical setting that bolsters, from a different perspective, the interpretation of the quantum f -correlations as indicators of quantum correlations, leaving aside the critical resource-theoretic characterisation of the ensuing set of free operations. We return to the latter issue in the concluding Sect. 5, where an amended definition to cure the drawbacks of Eq. (13) is proposed and validated, leading in particular to generalised quantifiers for which monotonicity under local unital channels does hold for all two-qubit states.

4 Physical Interpretation and Applications

In this Section we provide a physical interpretation for the quantum f -correlation corresponding to the average of the Wigner–Yanase–Dyson skew informations, which is itself a member of the family of MASIs, as shown in the following. As we mentioned above, each ISI $I^f(\rho, O)$ defined in Eq. (5) can be used to quantify the coherent spread of the state ρ across the eigenstates of an observable O —or the quantum portion of the total uncertainty $\text{Var}_\rho(O)$. The metric-adjusted f -correlations defined in Eq. (12), which stem from the non-additivity of these MASIs, thus have the transparent meaning of *quantum contributions to the covariance of different observables*. Among the MASIs, one of them takes a special meaning for thermal equilibrium states

$$\rho = \frac{1}{Z} e^{-H/T}, \quad (30)$$

where T is the temperature (in natural units) and H the Hamiltonian of the AB system, while the partition function $Z = \text{Tr}(e^{-H/T})$ ensures that the density matrix is normalised, $\text{Tr}(\rho) = 1$. Indeed, let us consider the quantity (referred to as ‘quantum variance’ in [96])

$$I^{\bar{f}}(\rho, O) \equiv \int_0^1 d\alpha I^\alpha(\rho, O), \quad (31)$$

where $I^\alpha \equiv I^{f_\alpha^{WYD}}$ is the Wigner–Dyson–Yanase skew information [58] with f_α^{WYD} defined in Eq. (6). Then, using the methods of [58], it is straightforward to show that $I^{\bar{f}}(\rho, O)$ is a MASI in the form of Eq. (5), defined by the corresponding operator monotone function $\bar{f} \in \mathcal{F}_{op}$, which takes the expression

$$\bar{f}(t) = \frac{(1-t)^2}{12 \left(\frac{t+1}{2} - \frac{t-1}{\log(t)} \right)}. \quad (32)$$

It turns out that, for this instance, the associated metric-adjusted f –correlation can be defined, independently of Eq. (8), as [97]

$$\Upsilon^{\bar{f}}(\rho, O_A, O_B) \equiv \text{Cov}(O_A, O_B) - T \left. \frac{\partial \langle O_A \rangle}{\partial h_B} \right|_{h_B=0}. \quad (33)$$

Here, $\text{Cov}(O_A, O_B)$ is the ordinary covariance defined in Eq. (10), and $\frac{\partial \langle O_A \rangle}{\partial h_B}$ is the static susceptibility of $\langle O_A \rangle$ with respect to the application of a field h_B which couples to O_B in the Hamiltonian, $H(h_B) = H - h_B O_B$. The equality

$$\text{Cov}(O_A, O_B) = T \left. \frac{\partial \langle O_A \rangle}{\partial h_B} \right|_{h_B=0}, \quad (34)$$

is a thermodynamic identity (a “fluctuation-dissipation theorem” [98]) for classical systems at thermal equilibrium. Therefore, the genuinely quantum contribution to the covariance as defined in Eq. (33) quantifies those correlations between a pair of local observables O_A and O_B which cannot be accounted for by classical statistical mechanics. As we have proved in Sect. 3.3, the discrepancy between classical and quantum statistical mechanics can be traced back, within the framework of quantum information theory, to the state ρ not being CQ or QC.

Defining the nonclassical contribution to the covariance in a thermal state via Eq. (33) is experimentally and computationally appealing, because one does not rely upon the tomographic reconstruction of the state, *a priori* needed in view of the general definition of Eq. (7), and which is prohibitive for large systems. Being defined in terms of measurable quantities (namely usual correlation and response functions), Eq. (33) provides a convenient tool to access two-sided quantum correlations in quantum systems at thermal equilibrium. Moreover, Eq. (33) is accessible also in the case of large-scale numerical calculations: in [97], the spatial structure of the quantity in

Eq.(33) has been investigated for many-body systems of thousands of qubits using quantum Monte Carlo methods. It should also be accessible to state-of-the-art cold-atom experiments as proposed in [96].

5 Discussion and Conclusions

We have defined an infinite family of quantitative indicators of two-sided quantum correlations beyond entanglement, which vanish on both classical-quantum and quantum-classical states and thus properly capture quantumness with respect to both subsystems. These quantifiers, named ‘quantum f -correlations’, are in one-to-one correspondence with the metric-adjusted skew informations [55–61]. We have shown that the quantum f -correlations are entanglement monotones for pure states of qubit-qudit systems, having also provided closed-form expressions for these quantifiers for two-qubit systems. We further analysed their behaviour under local commutativity preserving operations. Focusing on systems at thermal equilibrium, a situation especially relevant to many-body systems, we have physically interpreted the quantifier corresponding to the average of the Wigner–Yanase–Dyson skew informations by resorting to a quantum statistical mechanics perspective [96, 97].

The still unsolved characterisation of the subset of local commutativity preserving operations under which the quantum f -correlations are monotonically nonincreasing deserves special attention in light of the quest for the identification of physically relevant free operations within a resource theory of general quantum correlations [8]. Further investigation towards a deeper understanding of the quantum f -correlations for higher dimensional and multipartite quantum systems is also worthwhile. In particular, it could be interesting to explore the possible operational role played by these quantifiers in multiparameter quantum estimation [99–101].

5.1 Extending the Optimisation in the Definition of the Quantum f -Correlations

Finally, we note that the notion of two-sided quantum correlations we have introduced depends nontrivially on what portion of a multipartite system is assumed to be accessible. Indeed, if O_A and O_B in Eq.(7) are local observables acting on two different subsystems A and B of a larger system ABC , the quantum covariance between O_A and O_B will in general take a different value if calculated on the full tripartite state ρ^{ABC} , as compared to the original state $\rho^{AB} = \text{Tr}_C[\rho^{ABC}]$. Furthermore, we have verified numerically that in general the two quantities do not satisfy a particular ordering. This issue appears to be at the root of the violation of the monotonicity (Q3) for the quantum f -correlations under local commutativity preserving operations. In the interest of removing such an ambiguity, we conjecture that a general and *bona*

fide quantifier of two-sided quantum correlations, solely dependent on the state ρ^{AB} , may be defined as follows:

$$\tilde{Q}^f(\rho^{AB}) \equiv \inf_{\substack{\rho^{ABC} \text{ s.t.} \\ \text{Tr}_C[\rho^{ABC}] = \rho^{AB}}} \left[\max_{O_A, O_B} \Upsilon^f(\rho^{ABC}, O_A \otimes \mathbb{I}_B \otimes \mathbb{I}_C, \mathbb{I}_A \otimes O_B \otimes \mathbb{I}_C) \right]. \quad (35)$$

The optimisation problem above, performed over all the possible extensions of the state ρ^{AB} into a larger Hilbert space, appears to be a rather daunting task. Yet, trading computability for reliability [102], it is interesting to assess whether the quantity in Eq. (35) may serve as a meaningful tool to provide further insight on the operational interpretation and mathematical characterisation of two-sided quantum correlations beyond entanglement, in particular respecting all desiderata arising from a resource-theoretic approach while maintaining a clear physical motivation.

Here we provide a first affirmative answer. In particular, we prove in Appendix A that $\tilde{Q}^f(\rho^{AB})$ is in fact monotonically nonincreasing under all local unital channels for any two-qubit state ρ^{AB} , hence fulfilling requirement (Q3) in this prominent instance. A more general investigation into the monotonicity properties of $\tilde{Q}^f(\rho^{AB})$ under local commutativity preserving channels (or relevant subsets thereof) for states ρ^{AB} of arbitrary dimension will be the subject of future work.

Acknowledgements We thank Thomas Bromley and Tommaso Roscilde for stimulating discussions, as well as Paolo Gibilisco and an anonymous referee for very fruitful comments on a previous version of this manuscript. We acknowledge financial support from the European Research Council (Grant No. 637352 GQCOP), the Foundational Questions Institute (Grant No. FQXi-RFP-1601), and the Agence Nationale de la Recherche (“ArtiQ” project). T.T. acknowledges financial support from the University of Nottingham via a Nottingham Research Fellowship.

A Monotonicity of Eq. (35) Under Local Unital Channels

We will here prove that, if Λ_A is a unital channel on qubit A , then the following inequality holds:

$$\tilde{Q}^f(\Lambda_A(\rho^{AB})) \leq \tilde{Q}^f(\rho^{AB}). \quad (36)$$

In order to prevent the notation from becoming too cumbersome, in this Appendix we shall leave identity operators implicit wherever convenient: for example in the equation above we defined $\Lambda_A(\rho^{AB}) \equiv \Lambda_A \otimes \mathbb{I}_B(\rho^{AB})$.

To begin our proof, let us assume that ρ^{ABC} is the optimal dilation of ρ^{AB} for the sake of Eq. (35), that is, $\tilde{Q}^f(\rho^{AB}) = Q_{AB}^f(\rho^{ABC})$, where the subscript AB indicates what subsystems are involved in the calculation of the relevant quantum f -correlations. Consider now any dilation τ^{ABCD} of $\Lambda_A(\rho^{AB})$ into a larger space, including a further ancillary system D . We note that τ^{ABCD} is automatically also a dilation of $\Lambda_A(\rho^{AB})$. Hence, the following inequality holds by definition:

$$\tilde{Q}^f(\Lambda_A(\rho^{AB})) \leq Q_{AB}^f(\tau^{ABCD}), \quad (37)$$

Eq.(36) can then be proven by showing that $Q_{AB}^f(\tau^{ABCD}) \leq Q_{AB}^f(\rho^{ABC})$ for a particular choice of τ^{ABCD} .

To proceed, we use the fact that any unital qubit operation can be equivalently written as a random unitary channel [88], i.e.

$$\Lambda_A(\bullet) = \sum_k q_k U_A^{(k)} \bullet (U_A^{(k)})^\dagger, \quad (38)$$

for an appropriate collection of unitaries $\{U_A^{(k)}\}$ (acting on subsystem A) and probabilities $\{q_k\}$. A suitable dilation of $\Lambda_A(\rho^{ABC})$ may then be chosen as

$$\tau^{ABCD} = U_{AD}(\rho^{ABC} \otimes |\alpha\rangle\langle\alpha|_D)U_{AD}^\dagger, \quad (39)$$

$$\begin{aligned} U_{AD} &= \sum_k U_A^{(k)} \otimes |k\rangle\langle k|_D, \\ |\alpha\rangle_D &= \sum_k \sqrt{q_k} |k\rangle_D, \end{aligned} \quad (40)$$

where $\{|k\rangle_D\}$ is an orthonormal basis on system D . We shall now make use of Eqs.(7) and (28) to calculate the matrix M_τ^f corresponding to τ^{ABCD} , relating it to the matrix $M_\rho^f \equiv M^f$ of ρ^{ABC} . We will then show that the maximum singular value of M_τ^f is smaller than that of M^f .

To do so we infer from Eq.(39) that the nonzero eigenvalues of τ^{ABCD} are the same as those of ρ^{ABC} , say $\{p_i\}$, while the associated eigenvectors are

$$|\Phi_i\rangle_{ABCD} = U_{AD}|\phi_i\rangle_{ABC} \otimes |\alpha\rangle_D, \quad (41)$$

$|\phi_i\rangle_{ABC}$ being the eigenvectors of ρ^{ABC} . Using the shorthand $\sigma = \{\sigma_1, \sigma_2, \sigma_3\}$ as in the main text, we can then write

$$\begin{aligned} M_\tau^f &= \frac{f(0)}{2} \sum_{ij} c^f(p_i, p_j)(p_i - p_j)^2 \langle \Phi_i | \sigma_A | \Phi_j \rangle \langle \Phi_j | \sigma_B^T | \Phi_i \rangle \\ &= \frac{f(0)}{2} \sum_{ij} c^f(p_i, p_j)(p_i - p_j)^2 \langle \phi_i | \sum_k q_k (U_A^{(k)})^\dagger \sigma_A U_A^{(k)} | \phi_j \rangle \langle \phi_j | \sigma_B^T | \phi_i \rangle, \end{aligned} \quad (42)$$

where we have used the fact that $U_{AD}|\alpha\rangle_D = \sum_k \sqrt{q_k} U_A^{(k)}|k\rangle_D$. From the well known correspondence between the special unitary group $SU(2)$ and special orthogonal group $SO(3)$, it follows that for each k there exists an orthogonal matrix R_k such that $(U_A^{(k)})^\dagger \sigma_A U_A^{(k)} = R_k \sigma_A$. Applying this idea to the last line in Eq.(42) we thus obtain

$$\begin{aligned} M_\tau^f &= \frac{f(0)}{2} \sum_k q_k \sum_{ij} c^f(p_i, p_j) (p_i - p_j)^2 \langle \phi_i | R_k \boldsymbol{\sigma}_A | \phi_j \rangle \langle \phi_j | \boldsymbol{\sigma}_B^T | \phi_i \rangle \\ &= S M^f, \end{aligned} \quad (43)$$

where $S = \sum_k q_k R_k$ is a real matrix such that $SS^T \leq \mathbb{I}$, since it is a convex combination of orthogonal matrices. Since M^f and M_τ^f are real matrices, their singular values are found as the square roots of the eigenvalues of $Q = M^f(M^f)^T$ and $Q_\tau = M_\tau^f(M_\tau^f)^T = S Q S^T$, respectively. Let \mathbf{v} be the normalised eigenvector of Q_τ corresponding to its largest eigenvalue. Then

$$\lambda_{\max}(Q_\tau) = \mathbf{v}^T Q_\tau \mathbf{v} = \mathbf{v}^T S Q S^T \mathbf{v} \leq \lambda_{\max}(Q) \underbrace{\|S^T \mathbf{v}\|^2}_{\leq 1} \leq \lambda_{\max}(Q), \quad (44)$$

where we have used that \mathbf{v} is normalised and $SS^T \leq \mathbb{I}$. This in turn implies that $s_{\max}(M_\tau^f) \leq s_{\max}(M^f)$, concluding our proof.

The proof can be repeated to show monotonicity under unital channels on qubit B as well. This proves that the quantity $\tilde{Q}^f(\rho^{AB})$ defined in Eq. (35) is a *bona fide* quantifier of two-sided quantum correlations which obeys requirement (Q3) for any state ρ^{AB} of a two-qubit system.

References

1. Bell, J.S.: Rev. Mod. Phys. **38**(3), 447 (1966)
2. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V., Wehner, S.: Rev. Mod. Phys. **86**, 419 (2014)
3. Wiseman, H.M., Jones, S.J., Doherty, A.C.: Phys. Rev. Lett. **98**(14), 140402 (2007)
4. Cavalcanti, D., Skrzypczyk, P.: Rep. Prog. Phys. **80**(2), 024001 (2016)
5. Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K.: Rev. Mod. Phys. **81**(2), 865 (2009)
6. Modi, K., Brodutch, A., Cable, H., Paterek, T., Vedral, V.: Rev. Mod. Phys. **84**(4), 1655 (2012)
7. Quantum Correlations Beyond Entanglement and their Role in Quantum Information Theory. SpringerBriefs in Physics. Springer, Berlin (2015)
8. Adesso, G., Bromley, T.R., Cianciaruso, M.: J. Phys. A Math. Theor. **49**(47), 473001 (2016)
9. Fanchini, F.F., Soares Pinto, D.O., Adesso, G. (eds.): Lectures on General Quantum Correlations and their Applications. Springer, Berlin (2017)
10. Zurek, W.H.: Rev. Mod. Phys. **75**, 715 (2003)
11. Dowling, J.P., Milburn, G.J.: Philos. Trans. R. Soc. A **361**(1809), 1655 (2003)
12. Piani, M., Horodecki, P., Horodecki, R.: Phys. Rev. Lett. **100**(9), 090502 (2008)
13. Brandão, F.G., Piani, M., Horodecki, P.: Nat. Commun. **6** (2015)
14. Chuan, T., Maillard, J., Modi, K., Paterek, T., Paternostro, M., Piani, M.: Phys. Rev. Lett. **109**(7), 070501 (2012)
15. Streltsov, A., Kampermann, H., Bruß, D.: Phys. Rev. Lett. **108**(25), 250501 (2012)
16. Cavalcanti, D., Aolita, L., Boixo, S., Modi, K., Piani, M., Winter, A.: Phys. Rev. A **83**(3), 032324 (2011)
17. Madhok, V., Datta, A.: Phys. Rev. A **83**(3), 032323 (2011)
18. Streltsov, A., Lee, S., Adesso, G.: Phys. Rev. Lett. **115**, 030505 (2015)

19. Wilde, M.M.: Proceedings of the royal society of london a: mathematical. Phys. Eng. Sci. **471**(2177), 20140941 (2015)
20. Spehner, D., Orszag, M.: New J. Phys. **15**(10), 103001 (2013)
21. Spehner, D., Orszag, M.: J. Phys. A Math. Theor. **47**(3), 035302 (2013)
22. Spehner, D.: J. Math. Phys. **55**(7), 075211 (2014)
23. Weedbrook, C., Pirandola, S., Thompson, J., Vedral, V., Gu, M.: New J. Phys. **18**(4), 043027 (2016)
24. Farace, A., De Pasquale, A., Rigovacca, L., Giovannetti, V.: New J. Phys. **16**(7), 073010 (2014)
25. Roga, W., Buono, D., Illuminati, F.: New J. Phys. **17**(1), 013031 (2015)
26. Girolami, D., Souza, A.M., Giovannetti, V., Tufarelli, T., Filgueiras, J.G., Sarthour, R.S., Soares-Pinto, D.O., Oliveira, I.S., Adesso, G.: Phys. Rev. Lett. **112**(21), 210401 (2014)
27. Piani, M., Narasimhachar, V., Calsamiglia, J.: New J. Phys. **16**(11), 113001 (2014)
28. Piani, M., Ghariibian, S., Adesso, G., Calsamiglia, J., Horodecki, P., Winter, A.: Phys. Rev. Lett. **106**(22), 220403 (2011)
29. Ghariibian, S., Piani, M., Adesso, G., Calsamiglia, J., Horodecki, P.: Int. J. Quantum Inf. **9**(07n08), 1701 (2011)
30. Piani, M., Adesso, G.: Phys. Rev. A **85**(4), 040301 (2012)
31. Adesso, G., DAMBROSIO, V., Nagali, E., Piani, M., Sciarrino, F.: Phys. Rev. Lett. **112**(14), 140501 (2014)
32. Pirandola, S.: Sci. Rep. **4** (2014)
33. Pirandola, S., Laurenza, R., Ottaviani, C., Banchi, L.: Nat. Commun. **8** (2017)
34. Horodecki, M., Horodecki, P., Horodecki, R., Oppenheim, J., Sen, A., Sen, U., Synak-Radtke, B., et al.: Phys. Rev. A **71**(6), 062307 (2005)
35. Dillenschneider, R., Lutz, E.: Europhys. Lett. **88**(5), 50003 (2009)
36. Park, J.J., Kim, K.H., Sagawa, T., Kim, S.W.: Phys. Rev. Lett. **111**(23), 230402 (2013)
37. Leggio, B., Bellomo, B., Antezza, M.: Phys. Rev. A **91**(1), 012117 (2015)
38. Correa, L.A., Palao, J.P., Adesso, G., Alonso, D.: Phys. Rev. E **87**, 042131 (2013)
39. Liuzzo-Scorpo, P., Correa, L.A., Schmidt, R., Adesso, G.: Entropy **18**(2), 48 (2016)
40. Grimsmo, A.L.: Phys. Rev. A **87**(6), 060302 (2013)
41. Dakić, B., Vedral, V., Brukner, Č.: Phys. Rev. Lett. **105**(19), 190502 (2010)
42. Modi, K., Paterek, T., Son, W., Vedral, V., Williamson, M.: Phys. Rev. Lett. **104**(8), 080501 (2010)
43. Paula, F., de Oliveira, T.R., Sarandy, M.: Phys. Rev. A **87**(6), 064101 (2013)
44. Roga, W., Spehner, D., Illuminati, F.: J. Phys. A Math. Theor. **49**(23), 235301 (2016)
45. Ollivier, H., Zurek, W.H.: Phys. Rev. Lett. **88**(1), 017901 (2001)
46. Henderson, L., Vedral, V.: J. Phys. A Math. Gen. **34**(35), 6899 (2001)
47. Luo, S.: Phys. Rev. A **77**(2), 022301 (2008)
48. Giampaolo, S., Streltsov, A., Roga, W., Bruß, D., Illuminati, F.: Phys. Rev. A **87**(1), 012313 (2013)
49. Roga, W., Giampaolo, S., Illuminati, F.: J. Phys. A Math. Theor. **47**(36), 365301 (2014)
50. Seshadreesan, K.P., Wilde, M.M.: Phys. Rev. A **92**(4), 042321 (2015)
51. Streltsov, A., Kampermann, H., Bruß, D.: Phys. Rev. Lett. **106**(16), 160401 (2011)
52. Girolami, D., Tufarelli, T., Adesso, G.: Phys. Rev. Lett. **110**(24), 240402 (2013)
53. Bromley, T.R., Cianciaruso, M., Adesso, G.: Phys. Rev. Lett. **114**(21), 210401 (2015)
54. Ma, J., Yadin, B., Girolami, D., Vedral, V., Gu, M.: Phys. Rev. Lett. **116**, 160407 (2016)
55. Morozova, E.A., Chentsov, N.N.: J. Sov. Math. **56**(5), 2648 (1991)
56. Petz, D.: Linear Algebra Appl. **244**, 81 (1996)
57. Petz, D.: J. Phys. A Math. Gen. **35**(4), 929 (2002)
58. Hansen, F.: Proc. Natl. Acad. Sci. **105**(29), 9909 (2008)
59. Gibilisco, P., Imparato, D., Isola, T.: J. Math. Phys. **48**, 072109 (2007)
60. Gibilisco, P., Imparato, D., Isola, T.: Proc. Am. Math. Soc. **137**, 317 (2009)
61. Gibilisco, P., Hiai, F., Petz, D.: IEEE Trans. Inf. Theory **55**(1), 439 (2009)
62. Davis, R., Delbourgo, R., Jarvis, P.: J. Phys. A Math. Gen. **33**(9), 1895 (2000)

63. Luo, S.: *Theor. Math. Phys.* **143**(2), 681 (2005)
64. Coecke, B., Fritz, T., Spekkens, R.W.: *Inf. Comput.* **250**, 59 (2016)
65. Brandão, F.G., Gour, G.: *Phys. Rev. Lett.* **115**(7), 070503 (2015)
66. Horodecki, M., Oppenheim, J.: *Int. J. Mod. Phys. B* **27**(01n03), 1345019 (2013)
67. Cianciaruso, M., Bromley, T.R., Roga, W., Lo Franco, R.: *Sci. Rep.* **5** (2015)
68. Hu, X., Fan, H., Zhou, D., Liu, W.M.: *Phys. Rev. A* **85**(3), 032102 (2012)
69. Streltsov, A., Kampermann, H., Bruß, D.: *Phys. Rev. Lett.* **107**(17), 170502 (2011)
70. Guo, Y., Hou, J.: *J. Phys. A Math. Theor.* **46**(15), 155301 (2013)
71. Uhlmann, A.: *Rep. Math. Phys.* **9**(2), 273 (1976)
72. Wigner, E.P., Yanase, M.M.: *Proc. Natl. Acad. Sci.* **49**, 910 (1963)
73. Luo, S.: *Phys. Rev. A* **73**(2), 022324 (2006)
74. Li, X., Li, D., Huang, H., Li, X., Kwek, L.C.: *Eur. Phys. J. D* **64**(1), 147 (2011)
75. Marvian, I., Spekkens, R.W.: *Nat. Commun.* **5**, 3821 (2014)
76. Girolami, D.: *Phys. Rev. Lett.* **113**, 170401 (2014)
77. Zhang, C., Yadin, B., Hou, Z.B., Cao, H., Liu, B.H., Huang, Y.F., Maity, R., Vedral, V., Li, C.F., Guo, G.C., et al.: *Phys. Rev. A* **96**(4), 042327 (2017)
78. Pires, D.P., Cianciaruso, M., Céleri, L.C., Adesso, G., Soares-Pinto, D.O.: *Phys. Rev. X* **6**, 021031 (2016)
79. Marvian, I., Spekkens, R.W., Zanardi, P.: *Phys. Rev. A* **93**(5), 052331 (2016)
80. Gibilisco, P., Isola, T.: *J. Math. Anal. Appl.* **384**(2), 670 (2011)
81. Audenaert, K., Cai, L., Hansen, F.: *Lett. Math. Phys.* **85**(2–3), 135 (2008)
82. Kubo, F., Ando, T.: *Math. Ann.* **246**(3), 205 (1980)
83. Gibilisco, P., Hansen, F., Isola, T.: *Linear Algebra Appl.* **430**(8–9), 2225 (2009)
84. Vedral, V., Plenio, M.B., Rippin, M.A., Knight, P.L.: *Phys. Rev. Lett.* **78**(12), 2275 (1997)
85. Vidal, G.: *J. Mod. Opt.* **47**(2–3), 355 (2000)
86. Nielsen, M.A.: *Phys. Rev. Lett.* **83**(2), 436 (1999)
87. Nielsen, M.A., Vidal, G.: *Quantum Inf. Comput.* **1**(1), 76 (2001)
88. Nielsen, M.A., Chuang, I.L.: *Quantum Computer Quantum Information*. Cambridge University Press, Cambridge (2000)
89. Ando, T.: *Linear Algebra Appl.* **118**, 163 (1989)
90. Hill, S., Wootters, W.K.: *Phys. Rev. Lett.* **78**, 5022 (1997)
91. Chitambar, E., Leung, D., Mančinska, L., Ozols, M., Winter, A.: *Commun. Math. Phys.* **328**(1), 303 (2014)
92. Åberg, J.: [arXiv:quant-ph/0612146](https://arxiv.org/abs/quant-ph/0612146) (2006)
93. Baumgratz, T., Cramer, M., Plenio, M.B.: *Phys. Rev. Lett.* **113**, 140401 (2014)
94. Yadin, B., Ma, J., Girolami, D., Gu, M., Vedral, V.: *Phys. Rev. X* **6**(4), 041028 (2016)
95. Streltsov, A., Adesso, G., Plenio, M.B.: *Rev. Mod. Phys.* **89**, 041003 (2017)
96. Frérot, I., Roscilde, T.: *Phys. Rev. B* **94**(7), 075121 (2016)
97. Malpetti, D., Roscilde, T.: *Phys. Rev. Lett.* **117**(13), 130401 (2016)
98. Kubo, R.: *Rep. Prog. Phys.* **29**, 255 (1966)
99. Braunstein, S.L., Caves, C.M.: *Phys. Rev. Lett.* **72**(22), 3439 (1994)
100. Giovannetti, V., Lloyd, S., Maccone, L.: *Phys. Rev. Lett.* **96**(1), 010401 (2006)
101. Ragy, S., Jarzyna, M., Demkowicz-Dobrzański, R.: *Phys. Rev. A* **94**(5), 052108 (2016)
102. Tufarelli, T., MacLean, T., Girolami, D., Vasile, R., Adesso, G.: *J. Phys. A Math. Theor.* **46**(27), 275308 (2013)

The Effects of Random Qubit-Qubit Quantum Channels to Entropy Gain, Fidelity and Trace Distance



Attila Andai

Abstract In quantum information theory, a qubit is the non-commutative analogue of the classical bit. We consider the space of qubit-qubit quantum channels, which are linear, completely positive and trace preserving maps. We present a natural probability measure on the space of quantum channels, which helps us to consider uniformly distributed random quantum channels. The main aim of this paper to study the effect of random qubit-qubit channels to entropy gain, fidelity and trace distance. We fix an arbitrary initial state and we apply a uniformly distributed random general or unital quantum channel on it, resulting the final state. It turns out that while studying the effect of random quantum channels, only the Bloch radius of the initial state counts. We point out that a uniformly distributed random quantum channel could decrease the entropy of special qubits in average. In details, if a qubit is highly mixed (i.e. its Bloch radius is small), then a randomly uniform general quantum channel will decrease its entropy in average. The fidelity of the initial state with Bloch radius r and the final state is a strictly monotonously decreasing function of r , decreasing from approximately 0.981–0.696. The fidelity takes its highest value for the most mixed state and the lowest one for pure states. We study the trace distance, which equals to the Hilbert–Schmidt distance for qubits, up to a multiplicative factor. We present formulas for the trace distance between the initial and the final states in average for general and unital channels. The trace distance of pure initial states and final states are nearly the same for unital and general quantum channels, but general quantum channels always cause bigger distance from the initial state.

Keywords Quantum channel

MSC2010 81P16 · 81P45 · 94A17

A. Andai (✉)

Department of Analysis, Budapest University of Technology and Economics,
Egry József street 1., Building H, Budapest 1111, Hungary
e-mail: andaia@math.bme.hu

1 Introduction

In quantum mechanical setting, a system is modeled on a Hilbert-space, which will be finite dimensional in our case. The system described by the Hilbert space $\mathcal{H}_n = \mathbb{C}^n$ has state space

$$\mathcal{M}_n = \{D \in M_n \mid D = D^*, D \geq 0, \text{Tr } D = 1\},$$

where M_n denotes the set of $n \times n$ complex matrices. The simplest quantum mechanical system consists of qubits, that is the $n = 2$ case. In quantum information theory, a qubit is the non-commutative analogue of the classical bit. A qubit can be represented by a 2×2 complex self-adjoint positive semidefinite matrix with trace one [10, 13, 14].

A quantum channel is a special $\mathcal{M}_n \rightarrow \mathcal{M}_n$ linear map [13]. We consider the space of qubit-qubit quantum channels (\mathcal{Q}), which are linear, completely positive and trace preserving maps. A qubit channel $\mathcal{M}_2 \rightarrow \mathcal{M}_2$ is called to be unital if maps the most mixed state to itself. The unital channels (\mathcal{Q}^1) form a submanifold of \mathcal{Q} .

The sets \mathcal{Q} and \mathcal{Q}^1 contain simplest building blocks of quantum information processing, namely quantum channels are intended to describe elementary quantum operations. There is a one to one correspondence between the qubit-qubit physical transitions and the elements of \mathcal{Q} [10]. Every element of \mathcal{Q} can be described uniquely by 12 real parameters, therefore \mathcal{Q} can be identified with a subset of \mathbb{R}^{12} . It is worth mentioning that the classical analogue of \mathcal{Q} is

$$\left\{ \begin{pmatrix} a & 1-a \\ 1-f & f \end{pmatrix} \mid a, f \in [0, 1] \right\},$$

which form a rather simple subset of \mathbb{R}^2 , therefore its geometry is trivial, while the geometrical properties of \mathcal{Q} are highly nontrivial and are widely studied [11–13]. Using the same idea as presented in [1], the volume of \mathcal{Q} with respect to the natural euclidean measure and the transition probability of qubits under uniformly distributed random quantum channels were given in [9] recently. These random quantum operations has considerable scientific interest [2], and can be used to study the spectral properties of random quantum channels [3]. The image of the most mixed state under random quantum operation plays a crucial role in superdense coding [6].

The main aim of this paper to study the effect of random qubit-qubit channels to entropy gain, fidelity and trace distance. We present a natural probability measure on the space of quantum channels, which helps us to consider uniformly distributed random quantum channels. We fix an arbitrary initial state and we apply a uniformly distributed random general or unital quantum channel on it and resulting the final state. It turns out that while studying the effect of random quantum channels, only the Bloch radius of the initial state counts. We point out that a uniformly distributed random quantum channel could decrease the entropy of special qubits in average. In

details, if a qubit is highly mixed (i.e. its Bloch radius is small), then a randomly uniform general quantum channel will decrease its entropy in average. The fidelity [15] of the initial state with Bloch radius r_0 and the final state is a strictly monotonously decreasing function of r_0 , decreasing from ≈ 0.981 to ≈ 0.696 . The fidelity takes its highest value for the most mixed state and the lowest one for pure states. We study the trace distance, which equals to the Hilbert–Schmidt distance for qubits, up to a $\sqrt{2}$ factor. We present formulas for the trace distance between the initial and the final states in average for general and unital channels too. The trace distance of pure initial states and final states are nearly the same for unital and general quantum channels, but general quantum channels always cause bigger distance from the initial state.

In Sect. 2 the qubit-qubit general and unital quantum channels will be introduced in detail, the exact meaning of uniformly distributed random quantum channel will be presented and we will cite the transition probabilities for random channels. Section 3 answers the question how a uniformly random quantum channel effects the entropy, and what is the average fidelity and trace distance between the initial and final state in average.

2 The Space of Qubit-Qubit Quantum Channels

A general quantum channel sends qubits to qubits. We consider only linear quantum channels. Let us start with linear maps between 2×2 complex matrices, which will be denoted by M_2 . Every linear map $\mathcal{M}_2 \rightarrow \mathcal{M}_2$ can be written in the form of

$$\varphi : \mathcal{M}_2 \rightarrow \mathcal{M}_2 \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto aQ_{11} + bQ_{12} + cQ_{21} + dQ_{22}, \quad (1)$$

where $(Q_{i,j})_{i,j=1,2}$ are 2×2 matrices. For a state $D \in \mathcal{M}_2$ the condition $\text{Tr } Q_{11} = \text{Tr } Q_{22} = 1$ imply that $\text{Tr } \varphi(D) = 1$ and the conditions $Q_{11} = Q_{11}^*$, $Q_{22} = Q_{22}^*$ and $Q_{12} = Q_{21}^*$ that $\varphi(D) = \varphi(D)^*$. The map φ is called to be positive if for every positive matrix X , $\varphi(X)$ is positive, because in this case for every state $D \in \mathcal{M}_2$ we have $\varphi(D) \in \mathcal{M}_2$. At first sight, it seems natural to call those operators quantum channels which fulfill the above mentioned conditions and are positive. In quantum setting we need a bit more, namely the completely positivity of φ , which means that for every $n \in \mathbb{N}^+$ the map

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \mapsto \begin{pmatrix} \varphi(a_{11}) & \varphi(a_{12}) & \dots & \varphi(a_{1n}) \\ \varphi(a_{21}) & \varphi(a_{22}) & \dots & \varphi(a_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(a_{n1}) & \varphi(a_{n2}) & \dots & \varphi(a_{nn}) \end{pmatrix} \quad (2)$$

is positive, where $(a_{i,j})_{i,j=1,\dots,n} \in M_2$. The idea behind this requirement is the following. If we have a quantum system in a given state with two subsystems, one of them

is a qubit that we want to transform, and we apply a quantum channel on this quantum system which transforms the qubit and leaves the other subsystem untouched, then we should get a state of the original compound quantum system. One of the postulates of quantum mechanics is that if a composite quantum system consists of two subsystems described by Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , then the composite system is described by the tensor product of Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$. In our case $\mathcal{H}_2 = \mathbb{C}^2$ and $\mathcal{H}_1 = \mathbb{C}^n$, where n is a positive natural number. In order to get a state of the composite system, we have to guarantee the positivity of the map

$$I \otimes \varphi : \mathcal{H}_1 \otimes \mathcal{H}_2 \rightarrow \mathcal{H}_1 \otimes \mathcal{H}_2,$$

which was given by the map (2).

According to Choi and Jamiołkowski's representation theorem for completely positive linear maps [4, 8] the map (1) is completely positive iff the 4×4 Choi matrix

$$\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \quad (3)$$

is positive. Now we can identify the set of qubit-qubit channels with special 4×4 matrices as

$$\mathcal{Q} = \left\{ \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^* & Q_{22} \end{pmatrix} \in M_4 \mid Q_{11}, Q_{22} \in \mathcal{M}_2, \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^* & Q_{22} \end{pmatrix} \geq 0 \right\}. \quad (4)$$

A quantum channel Q is called to be unital if

$$Q(I_2) = I_2,$$

where I_2 is the 2×2 identity matrix. This identity preserving property requires that $Q_{11} + Q_{22} = I$ must hold in the Choi representation (1), hence the space of unital qubit channels (\mathcal{Q}^1) can be identified with a convex submanifold of \mathcal{Q} .

For quantum channels the following parametrization of $\mathcal{Q} \subset \mathbb{R}^{12}$ is considered

$$Q = \begin{pmatrix} a & b & c & d \\ \bar{b} & 1-a & e & -c \\ \bar{c} & \bar{e} & f & g \\ \bar{d} & -\bar{c} & \bar{g} & 1-f \end{pmatrix}, \quad (5)$$

where $a, f \in [0, 1]$ and $b, c, d, e, g \in \mathbb{C}$. Let us denote by λ_n the Lebesgue measure in \mathbb{R}^n . The measure on the space of quantum channels will be the euclidean measure, where the volume element is $2^7 d\lambda_{12}$. In coordinates with respect to the Lebesgue measure we have

$$d\lambda_{12} = d\lambda_1(a) d\lambda_1(f) d\lambda_2(b) d\lambda_2(c) d\lambda_2(d) d\lambda_2(e) d\lambda_2(g). \quad (6)$$

For unital quantum channels we consider the parametrization of $\mathcal{Q}^1 \subset \mathbb{R}^9$ like

$$\mathcal{Q} = \begin{pmatrix} a & b & c & d \\ \bar{b} & 1-a & e & -c \\ \bar{c} & \bar{e} & 1-a & -b \\ \bar{d} & -\bar{c} & -\bar{b} & a \end{pmatrix}, \quad (7)$$

where $a \in [0, 1]$ and $b, c, d, e \in \mathbb{C}$. The studied measure on this space is again the euclidean measure, where the volume element is $2^7 d\lambda_9$.

We have the following theorem for the volumes of the general and unital quantum channels [9].

Theorem 1 *The volumes of the spaces \mathcal{Q} , \mathcal{Q}^1 with respect to the Lebesgue measure are*

$$V(\mathcal{Q}) = \frac{8\pi^4}{945} \quad \text{and} \quad V(\mathcal{Q}^1) = \frac{2\pi^5}{4725}.$$

The concept of uniformly distributed random quantum channel in our setting means a uniformly random variable with values in \mathcal{Q} (or in \mathcal{Q}^1) with respect to the normalized euclidean measure (i.e. $\frac{15120}{\pi^4} d\lambda_{12}$ on \mathcal{Q} and $\frac{302400}{\pi^5} d\lambda_9$ on \mathcal{Q}^1).

For further calculations, we need the Pauli basis representation of qubits. The Pauli matrices are the following.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

We use the Stokes representation of qubits which gives a bijective correspondence between qubits and the unit ball in \mathbb{R}^3 via the map

$$\{x \in \mathbb{R}^3 \mid \|x\|_2 \leq 1\} \rightarrow \mathcal{M}_2 \quad x \mapsto \frac{1}{2} (I_2 + x \cdot \sigma),$$

where $x \cdot \sigma = \sum_{j=1}^3 x_i \sigma_i$. The vector x , which describes the state called Bloch vector, and the unit ball in this setting is called Bloch sphere and the quantity $\|x\|_2$ is called the Bloch radius of the state.

According to the next theorem, the chosen normalized euclidean measures are unitary invariant in the space of qubits or in another form, they are invariant under rotations if the Stokes parameterization is used [9].

Theorem 2 *An orthogonal orientation preserving transformation O in \mathbb{R}^3 induces maps $\alpha_O, \beta_O : \mathcal{Q} \rightarrow \mathcal{Q}$ via Stokes parametrization $\alpha_O(Q) = O \circ Q$ and $\beta_O(Q) = Q \circ O$. The Jacobian of these transformations are 1. The Jacobian of the restricted transformations $\alpha'_O = \alpha_O |_{\mathcal{Q}^1}$ and $\beta'_O = \beta_O |_{\mathcal{Q}^1}$ are 1.*

In practice, the above mentioned theorem means that if considered a state given by (Bloch) vector x with Bloch radius $r_0 = \|x\|_2$, and applied a uniformly distributed random general or unital quantum channel to the state, then the distribution of the resulting states depends only on the Bloch radius. These distributions are described by the following theorems [9].

Theorem 3 *Assume that a uniformly distributed quantum channel is applied to a given state with Bloch radius r_0 . The radii distribution of the resulted quantum states is the following.*

$$\kappa(r, r_0) = \begin{cases} \text{If } 0 < r \leq r_0 : \\ \frac{40r^2}{r_0(1+r_0)^6}(21r^4 - 6r^2r_0^2 - 36r^2r_0 + r_0^4 + 6r_0^3 + 12r_0^2 + 2r_0), \\ \text{if } r_0 < r \leq 1 : \\ \frac{40r(r-1)^6}{(1-r_0^2)^6}(21r_0^4 - 6r^2r_0^2 - 36rr_0^2 + r^4 + 6r^3 + 12r^2 + 2r). \end{cases} \quad (8)$$

Theorem 4 *Assume that a uniformly distributed unital quantum channel is applied to a given state with Bloch radius r_0 . The radii distribution of the resulted quantum states is the following.*

$$\kappa_1(r, r_0) = \frac{315}{16} \times \frac{r^2(r_0^2 - r^2)^3}{r_0^9} \chi_{[0, r_0]}(r) \quad r \in [0, 1], \quad (9)$$

where $\chi_{[0, r_0]}$ denotes the indicator function of the set $[0, r_0]$.

The transition probability between different Bloch radii under uniformly distributed random channels $\kappa(r, r_0)$ is shown in Figs. 1 and 2 shows the unital case.

3 The Effect of Random Quantum Channels from Information Theoretical Point of View

The above mentioned theorem gives us a kind of transition functions, as they present a probability of starting from a Bloch radius r_0 ending at r . The probability measures of the final state in spherical coordinates are the following.

$$\frac{\kappa_{(1)}(r, r_0)}{4\pi} \sin \vartheta dr d\vartheta d\varphi \quad (10)$$

Fig. 1 The function $\kappa(r, r_0)$. The radii distribution (r) of the resulted quantum states if a random quantum channel was applied to a given state with Bloch radius r_0

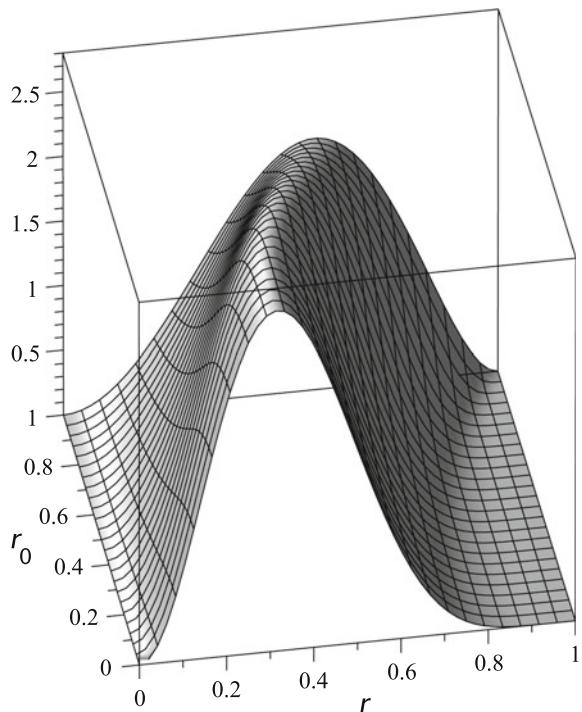
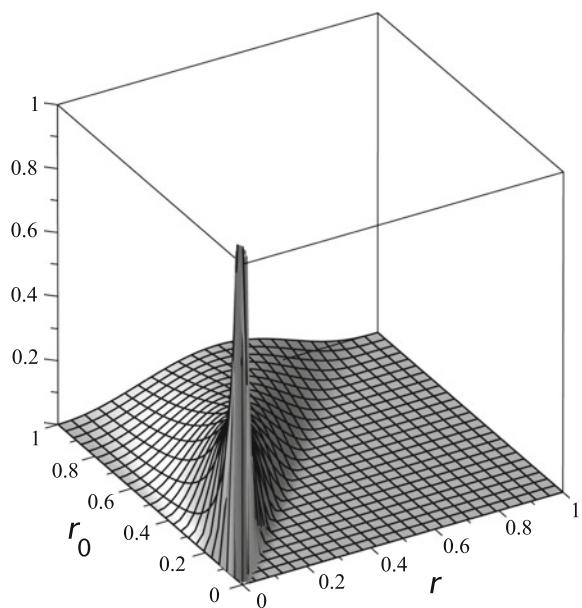


Fig. 2 The function $\kappa_1(r, r_0)$. The radii distribution (r) of the resulted quantum states if a random unital quantum channel was applied to a given state with Bloch radius r_0



3.1 Entropy Gain

von Neumann associated an entropy quantity to a density operator D in 1927 [16, 17] as

$$S(D) = -\text{Tr}(D \log D).$$

In qubit case the entropy for a state with Bloch radius r is

$$S(r) = -\frac{1+r}{2} \log \left(\frac{1+r}{2} \right) - \frac{1-r}{2} \log \left(\frac{1-r}{2} \right).$$

The entropy gain under uniformly random quantum channel is defined as the difference of the expected value of the entropy of the final state and the entropy of the original state

$$\begin{aligned} Sg_{(1)}(r_0) &= \int_0^{2\pi} \int_0^\pi \int_0^1 S(r) \times \frac{\kappa_{(1)}(r, r_0)}{4\pi} \sin \vartheta \, dr \, d\vartheta \, d\varphi - S(r_0) \\ &= \int_0^1 S(r) \kappa_{(1)}(r, r_0) \, dr - S(r_0). \end{aligned}$$

These integrals (for transition functions $\kappa(r, r_0)$ and $\kappa_1(r, r_0)$) were computed numerically. Figures 3 and 4 show the average entropy gain under a general and unital random quantum channel for a state with Bloch radius r_0 . As it is expected unital quantum channels (in average) always increase the the entropy, but this is not true for general quantum channels as Fig. 3 shows. Important to note that a general quantum channel likely decreases the entropy of highly mixed states ($r_0 \approx 0$) and increases for nearly pure states ($r_0 \approx 1$).

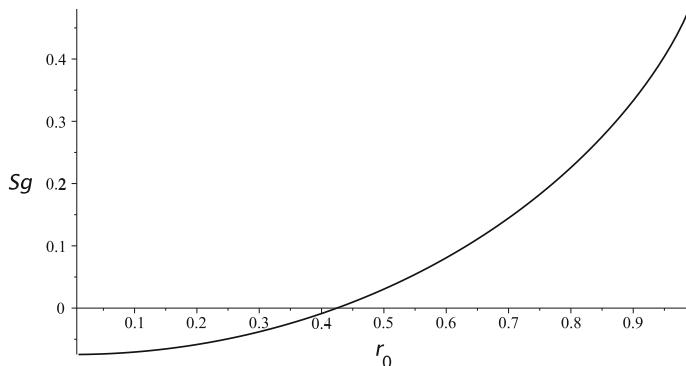


Fig. 3 The function $Sg(r_0)$. The average entropy gain if a random quantum channel was applied to a given state with Bloch radius r_0

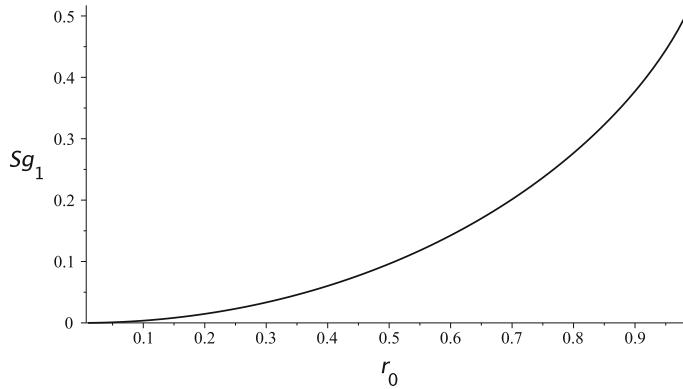


Fig. 4 The function $Sg_1(r_0)$. The average entropy gain if a random unital quantum channel was applied to a given state with Bloch radius r_0

3.2 Fidelity

How close are the resulted quantum states to the original one after a random quantum operation? There are many candidates to measure the distance between quantum states, one of them is the fidelity

$$F(D_1, D_2) = \text{Tr} \sqrt{\sqrt{D_1} D_2 \sqrt{D_1}} \quad D_1, D_2 \in \mathcal{M}_n.$$

The square root of the transition probability is called fidelity, and the above mentioned formula is its generalization to the quantum case. (For details see for example [15].) For qubits $D_1, D_2 \in \mathcal{M}_2$, fidelity can be computed as

$$F(D_1, D_2) = \sqrt{\text{Tr}(D_1 D_2) + 2\sqrt{\det(D_1 D_2)}}.$$

Using the unitary invariance of the measures chosen on quantum channels (Theorem 2), the initial state can be written in the form of

$$D_1 = \frac{1}{2} \begin{pmatrix} 1+r_0 & 0 \\ 0 & 1-r_0 \end{pmatrix}.$$

Let us denote the abbreviation

$$F(r, \vartheta, \varphi, r_0) = F \left(\frac{1}{2} \begin{pmatrix} 1+r_0 & 0 \\ 0 & 1-r_0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 1+r \cos \vartheta & r e^{i\varphi} \sin \vartheta \\ r e^{-i\varphi} \sin \vartheta & 1-r \cos \vartheta \end{pmatrix} \right)$$

and define the average fidelity distance caused by a random quantum channel as the expected value of $F(r, \vartheta, \varphi, r_0)$, that is,

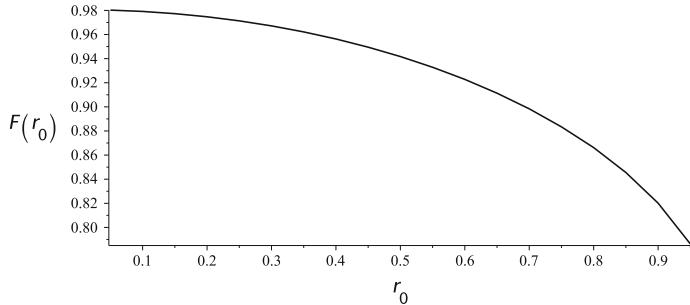


Fig. 5 The function $F(r_0)$. The average fidelity distance if a random quantum channel was applied to a given state

$$F_{(1)}(r_0) = \int_0^{2\pi} \int_0^\pi \int_0^1 F(r, \vartheta, \varphi, r_0) \frac{\kappa_{(1)}(r, r_0)}{4\pi} \sin \vartheta \, dr \, d\vartheta \, d\varphi.$$

The result of the integral for transition functions $\kappa(r, r_0)$ is shown in Fig. 5.

The fidelity of the initial state with Bloch radius r_0 and the final state under random a quantum channel is a strictly monotonously decreasing function of r_0 , decreasing from ≈ 0.981 to ≈ 0.696 . It takes its highest value for the most mixed state and the lowest one for pure states.

3.3 Trace and Hilbert–Schmidt Distances

The trace distance of states can be viewed as a natural generalization of Kolmogorov distance of classical probability distributions.

$$T(D_1, D_2) = \frac{1}{2} \text{Tr} \sqrt{(D_1 - D_2)^2} \quad D_1, D_2 \in \mathcal{M}_n$$

In qubit case, the trace distance is the usual euclidean distance in Bloch representation. We note the inequality between the trace distance and the fidelity.

$$1 - F(D_1, D_2) \leq T(D_1, D_2) \leq \sqrt{1 - F^2(D_1, D_2)} \quad D_1, D_2 \in \mathcal{M}_n$$

In quantum optics the Hilbert–Schmidt norm based distance seems to be the best for explicit calculations [5], for the definition of nonclassicality of states [7] and gives better insight to the relation of states than the trace distance in general. The Hilbert–Schmidt distance of states is defined as

$$d^{(\text{HS})}(D_1, D_2) = \sqrt{\text{Tr}(D_1 - D_2)^2} \quad D_1, D_2 \in \mathcal{M}_n.$$

In qubit case we have

$$d^{(\text{HS})}(D_1, D_2) = \sqrt{2}T(D_1, D_2) \quad D_1, D_2 \in \mathcal{M}_2,$$

so we can study the effect of random quantum channels to these distances together. The trace distance of the initial state and an arbitrary state is

$$\begin{aligned} T(r, \vartheta, \varphi, r_0) &= T\left(\frac{1}{2}\begin{pmatrix} 1+r_0 & 0 \\ 0 & 1-r_0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1+r \cos \vartheta & r e^{i\varphi} \sin \vartheta \\ r e^{-i\varphi} \sin \vartheta & 1-r \cos \vartheta \end{pmatrix}\right) \\ &= \frac{1}{2}\sqrt{r^2 - 2rr_0 \cos(\vartheta) + r_0^2}. \end{aligned}$$

The average trace distance between a qubit and its image under uniformly random quantum operation is

$$T_{(1)}(r_0) = \int_0^{2\pi} \int_0^\pi \int_0^1 T(r, \vartheta, \varphi, r_0) \frac{\kappa_{(1)}(r, r_0)}{4\pi} \sin \vartheta dr d\vartheta d\varphi.$$

For unital quantum channels we have

$$T_1(r_0) = \frac{6}{11}r_0 \approx 0.545r_0,$$

and for general quantum channels

$$T(r_0) = \frac{p(r_0)}{9009(1+r_0)^6},$$

where $p(x)$ is the polynomial

$$3456x^7 + 35751x^6 + 57834x^5 + 110089x^4 + 71820x^3 + 30345x^2 + 9450x + 1575.$$

For quantum channels the average trace distances $T_{(1)}(r_0)$ are shown in Fig. 6.

For general quantum channels we have the approximations

$$T(r_0) \approx \begin{cases} 0.048 + 0.508r_0 & \text{if } 0.6 \leq r \leq 1 \\ 0.175 + 0.746r_0^2 & \text{if } 0 \leq r \leq 0.2. \end{cases}$$

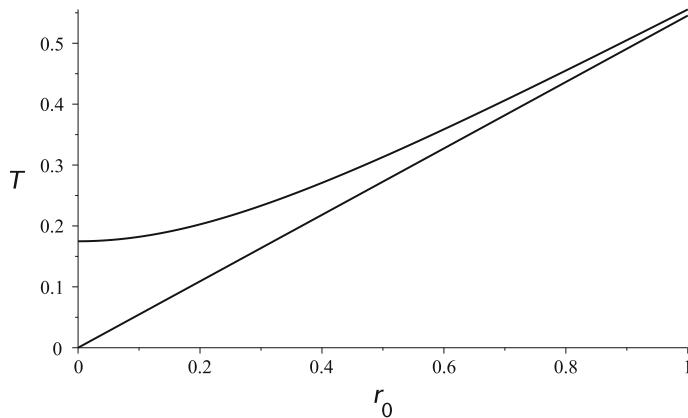


Fig. 6 The average trace distance if a random general ($T(r_0)$) or unital ($T_{(1)}(r_0)$) quantum channel was applied to a given state. The function $T(r_0)$ is the curved one and $T_{(1)}(r_0)$ is the linear one

From Fig. 6 it is clear that a uniformly random quantum channel will cause greater jump in Hilbert–Schmidt and trace distances than a unital channel. The average trace distance between a pure state ($r_0 = 1$) and its image caused by a uniformly random general channel is $\frac{5}{9}$ and $\frac{6}{11}$ for unital channel.

References

1. Andai, A.: Volume of the quantum mechanical state space. *J. Phys. A Math. Theor.* **39**, 13641–13657 (2006)
2. Bouda, J., Koniorczyk, M., Varga, A.: Random unitary qubit channels: entropy relations, private quantum channels and non-malleability. *Eur. Phys. J. D* **53**(3), 365–372 (2009)
3. Bruzda, W., Cappellini, V., Sommers, H.-J., Zyczkowski, K.: Random quantum operations. *Phys. Lett. A* **373**(3), 320–324 (2009)
4. Choi, M.-D.: Completely positive linear maps on complex matrices. *Linear Algebra Appl.* **10**, 285–290 (1975)
5. Dodonov, V.V., Man'ko, O.V., Man'ko, V.I., Wünsche, A.: Hilbert–Schmidt distance and non-classicality of states in quantum optics. *J. Mod. Opt.* **47**(4), 633–654 (2000)
6. Harrow, A., Hayden, P., Leung, D.: Superdense coding of quantum states. *Phys. Rev. Lett.* **92**(18) (2004)
7. Hillery, M.: Total noise and nonclassical states. *Phys. Rev. A* **39**, 2994–3002 (1989)
8. Jamiołkowski, A.: Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.* **3**(4), 275–278 (1972)
9. Lovas, A., Andai, A.: Volume of the space of qubit-qubit channels and state transformations under random quantum channels (2017)
10. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
11. Omkar, S., Srikanth, R., Banerjee, Subhashish: Dissipative and non-dissipative single-qubit channels: dynamics and geometry. *Quantum Inf. Process.* **12**(12), 3725–3744 (2013)

12. Pasieka, A., Kribs, D.W., Laflamme, R., Pereira, R.: On the geometric interpretation of single qubit quantum operations on the bloch sphere. *Acta Appl. Math.* **108**(697), 6 (2009)
13. Petz, D.: *Quantum Information Theory and Quantum Statistics*. Springer, Berlin (2008)
14. Ruskai, M.B., Szarek, S., Werner, E.: An analysis of completely positive trace- preserving maps on \mathcal{M}_2 . *Linear Algebra Appl.* **347**, 159–187 (2002)
15. Uhlmann, A.: The transition probability in the state space of a *-algebra. *Rep. Math. Phys.* **9**(2), 273–279 (1976)
16. von Neumann, J.: Thermodynamik quantenmechanischer gesamtheiten. *Gött. Nachr.* 273–291 (1927)
17. von Neumann, J.: *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, New Jersey (1955)

Robertson-Type Uncertainty Principles and Generalized Symmetric and Antisymmetric Covariances



Attila Lovas

Abstract Uncertainty principles are one of the basic relations of quantum mechanics. Robertson has discovered first that the Schrödinger uncertainty principle can be interpreted as a determinant inequality. Generalized quantum covariance has been previously presented by Gibilisco, Hiai and Petz. Gibilisco and Isola have proved that among these covariances the usual quantum covariance introduced by Schrödinger gives the sharpest inequalities for the determinants of covariance matrices. We have introduced the concept of symmetric and antisymmetric quantum f -covariances which give better uncertainty inequalities. In this paper we generalize the concept of symmetric and antisymmetric covariances considering a continuous path between them, called α , f -covariances. We derive uncertainty relations for α , f -covariances. Moreover, using a simple matrix analytical framework, we present here a short and tractable proof for the celebrated Robertson uncertainty principle. In our setting Robertson inequality is a special case of a determinant inequality, namely that the determinant of the (element-wise) real part of a positive self-adjoint matrix is greater or equal to the determinant of its imaginary part.

Keywords Uncertainty principle · Quantum fisher information · Quantum information geometry

MSC 62B10 · 94A17

1 Introduction

In quantum information theory, the most popular model of the quantum event algebra associated to an n -level system is the projection lattice of an n -dimensional Hilbert space ($\mathcal{L}(\mathbb{C}^n)$). According to Gleason's theorem [2], for $n > 2$ the states are of the form

A. Lovas (✉)

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

$$(\forall P \in \mathcal{L}(\mathbb{C}^n)) \quad P \mapsto \text{Tr}(DP),$$

where D is a positive semidefinite matrix with trace 1, hence the quantum mechanical state space arises as the intersection of the standard cone of positive semidefinite matrices and the hyperplane of trace one matrices. Let us denote by \mathcal{M}_n the set of $n \times n$ positive definite matrices and by \mathcal{M}_n^1 the interior of the n -level quantum mechanical state space, namely

$$\mathcal{M}_n^1 = \{D \in \mathcal{M}_n \mid \text{Tr } D = 1, D > 0\}.$$

Let $M_{n,\text{sa}}$ be the set of observables of the n -level quantum system, in other words, the set of $n \times n$ self adjoint matrices, and $M_{n,\text{sa}}^{(0)}$ stands for the set observables with zero trace. Observables are non commutative analogues of random variables known from Kolmogorovian probability. If A is an observable, then the expectation of A in $D \in \mathcal{M}_n^1$ is defined by $\mathbb{E}_D(A) = \text{Tr}(AD)$.

Spaces \mathcal{M}_n and \mathcal{M}_n^1 form convex sets in the space of self adjoint matrices, and they are obviously differentiable manifolds [10]. The tangent space of \mathcal{M}_n at a given state D can be identified with $M_{n,\text{sa}}$ and the tangent space of \mathcal{M}_n^1 with $M_{n,\text{sa}}^{(0)}$. Monotone metrics are the quantum analogues of the Fisher information matrix known from classical information geometry. Petz's classification theorem [14] establishes a connection between monotone metrics and the set of symmetric and normalized operator monotone functions \mathcal{F}_{op} . For the mean induced by the operator monotone function $f \in \mathcal{F}_{\text{op}}$, we also introduce the notation

$$m_f : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+ \quad (x, y) \mapsto yf\left(\frac{x}{y}\right).$$

The monotone metric associated to $f \in \mathcal{F}_{\text{op}}$ is given by

$$K_D^{(n)}(X, Y) = \text{Tr}\left(X m_f\left(L_{n,D}, R_{n,D}\right)^{-1}(Y)\right)$$

for all $n \in \mathbb{N}$ where $L_{n,D}(X) = DX$, $R_{n,D}(X) = XD$ for all $D, X \in M_n(\mathbb{C})$. The metric $K_D^{(n)}$ can be extended to the space \mathcal{M}_n . For every $D \in \mathcal{M}_n$ and matrices $A, B \in M_{n,\text{sa}}$, let us define

$$\langle A, B \rangle_{D,f} = \text{Tr}\left(A m_f\left(L_{n,D}, R_{n,D}\right)^{-1}(B)\right),$$

with this notion the pair $(\mathcal{M}_n, \langle \cdot, \cdot \rangle_{\cdot,f})$ will be a Riemannian manifold for every operator monotone function $f \in \mathcal{F}_{\text{op}}$.

Although the generalization of expectation and variance to the quantum case is straightforward, covariance has many different possible generalizations. Schrödinger has defined the (symmetric) covariance of the observables for a given state D as

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

In a recent paper [12] the concept of *symmetric f-covariance* was introduced as the scalar product of anti-commutators

$$\text{qCov}_{D,f}^s(A, B) = \frac{f(0)}{2} \langle \{D, A\}, \{D, B\} \rangle_{D,f}.$$

Note that, $\text{qCov}_{D,f}^s(A, B)$ coincides with $\text{Cov}_D(A, B)$ whenever $f(x) = \frac{1+x}{2}$. We have proved that for any f symmetric and normalized operator monotone function

$$\begin{aligned} & \det \left(\left[\frac{f(0)}{2} \langle \{D, A_h\}, \{D, A_j\} \rangle_{D,f} \right]_{h,j=1,\dots,N} \right) \\ & \geq \det \left(\left[\frac{f(0)}{2} \langle i[D, A_h], i[D, A_j] \rangle_{D,f} \right]_{h,j=1,\dots,N} \right) \end{aligned} \quad (1)$$

holds. Moreover it was shown that the function $f_0(x) = \frac{1}{2} \left(\frac{1+x}{2} + \frac{2x}{1+x} \right)$ gives the smallest universal upper bound for the right-hand side, that is, for every symmetric and normalized operator monotone function g ,

$$\begin{aligned} & \det \left(\left[\frac{f_0(0)}{2} \langle \{D, A_h\}, \{D, A_j\} \rangle_{D,f_0} \right]_{h,j=1,\dots,N} \right) \\ & \geq \det \left(\left[\frac{g(0)}{2} \langle i[D, A_h], i[D, A_j] \rangle_{D,g} \right]_{h,j=1,\dots,N} \right) \end{aligned} \quad (2)$$

holds.

The Eqs. (1) and (2) are Robertson-type uncertainty principles with clear geometric meaning, namely, they can be viewed as a kind of volume inequalities. The volume of the N dimensional parallelotope spanned by the vectors X_k ($k = 1, \dots, N$) with respect to the inner product $\langle \cdot, \cdot \rangle$ is

$$V_f(X_1, \dots, X_N) = \sqrt{\det \left(\left[\langle X_h, X_j \rangle_{D,f} \right]_{h,j=1,\dots,N} \right)}.$$

In this setting Eq. (1) and (2) can be written as

$$\begin{aligned} V_f(\{D, A_1\}, \dots, \{D, A_N\}) & \geq V_f(i[D, A_1], \dots, i[D, A_N]) \\ V_{f_0}(\{D, A_1\}, \dots, \{D, A_N\}) & \geq V_g(i[D, A_1], \dots, i[D, A_N]). \end{aligned}$$

First, in Sect. 1 we briefly outline the origin and development of Robertson-type uncertainty principles. In Sect. 2 we present a general framework for uncertainty relations which helps clearly and briefly present the well-known and new relations, we give continuous path of covariances between symmetric and antisymmetric covariances and prove some relations between them. Finally in Sect. 3, we present a simple and rather understandable proof for the original Robertson uncertainty principle.

2 Overview of Uncertainty Principles

The concept of uncertainty was introduced by Heisenberg in 1927 [9], who demonstrated the impossibility of simultaneous measurement of position (q) and momentum (p). He considered Gaussian distributions ($f(q)$), and defined uncertainty of f as its width D_f . If the width of the Fourier transform of f is denoted by $D_{\mathcal{F}(f)}$, then the first formalisation of the uncertainty principle can be written as

$$D_f D_{\mathcal{F}(f)} = \text{constant}.$$

In 1927, Kennard generalised Heisenberg's result [11], he proved the inequality

$$\text{Var}_D(A) \text{Var}_D(B) \geq \frac{1}{4}$$

for observables A, B which satisfy the relation $[A, B] = -i$, for every state D , where $\text{Var}_D(A) = \text{Tr}(DA^2) - (\text{Tr}(DA))^2$.

In 1929, Robertson [15] extended Kennard's result for arbitrary two observables A, B

$$\text{Var}_D(A) \text{Var}_D(B) \geq \frac{1}{4} |\text{Tr}(D[A, B])|^2.$$

In 1930, Schrödinger [17] improved this relation including the correlation between observables A, B

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

The Schrödinger uncertainty principle can be formulated as

$$\det \begin{pmatrix} \text{Cov}_D(A, A) & \text{Cov}_D(A, B) \\ \text{Cov}_D(B, A) & \text{Cov}_D(B, B) \end{pmatrix} \geq \det \left(-\frac{i}{2} \begin{pmatrix} \text{Tr}(D[A, A]) & \text{Tr}(D[A, B]) \\ \text{Tr}(D[B, A]) & \text{Tr}(D[B, B]) \end{pmatrix} \right).$$

For the set of observables $(A_i)_{1,\dots,N}$ this inequality was generalized by Robertson in 1934 [16] as

$$\det \left([\text{Cov}_D(A_h, A_j)]_{h,j=1,\dots,N} \right) \geq \det \left(\left[-\frac{i}{2} \text{Tr}(D[A_h, A_j]) \right]_{h,j=1,\dots,N} \right). \quad (3)$$

The main drawback of this inequality is that the right-hand side is identical to zero whenever N is odd.

Gibilisco, Imparato and Isola in 2008 conjectured that

$$\det \left([\text{Cov}_D(A_h, A_j)]_{h,j=1,\dots,N} \right) \geq \det \left(\left[\frac{f(0)}{2} \langle i[D, A_h], i[D, A_j] \rangle_{D,f} \right]_{h,j=1,\dots,N} \right) \quad (4)$$

holds [6], where the scalar product $\langle \cdot, \cdot \rangle_{D,f}$ is induced by an operator monotone function f , according to Petz classification theorem [14]. We note that if the density matrix is not strictly positive, then the scalar product $\langle \cdot, \cdot \rangle_{D,f}$ is not defined. For arbitrary N the conjecture was proved by Andai [1] and Gibilisco, Imparato and Isola [4]. The inequality (4) is called *dynamical uncertainty principle* [3] because the right-hand side can be interpreted as the volume of a parallelepiped determined by the tangent vectors of the time-dependent observables $A_k(t) = e^{itD} A_k e^{-itD}$.

Gibilisco, Hiai and Petz studied the behaviour of a possible generalization of the covariance under coarse graining and they deduced that the covariance must have the following form for traceless observables A, B

$$\text{Cov}_D^f(A, B) = \text{Tr} (Af(L_{n,D}R_{n,D}^{-1})R_{n,D}(B)), \quad (5)$$

where $L_{n,D}$ and $R_{n,D}$ are superoperators acting on $n \times n$ matrices like $L_{n,D}(A) = DA$, $R_{n,D}(A) = AD$ and f is a symmetric and normalized operator monotone function [3]. Quantum covariances of the form (5) are called *quantum f-covariance*, which has been introduced for the first time by D. Petz [13]. It has been proved [3] that the generalized form of dynamical uncertainty principle holds true for an arbitrary quantum f -covariance

$$\det \left([\text{Cov}_D^g(A_h, A_j)]_{h,j=1,\dots,N} \right) \geq \det \left(\left[f(0)g(0) \langle i[D, A_h], i[D, A_j] \rangle_{D,f} \right]_{h,j=1,\dots,N} \right)$$

and for all g symmetric and normalized operator monotone function. If $g(x) = \frac{1+x}{2}$ is chosen, then we get the sharpest form of the inequality.

3 Generalized Covariances

After the historical overview we collect the definition of covariances and introduce a continuous path between symmetric and antisymmetric covariances using the idea

of q commutators

$$[A, B]_q = AB + qBA,$$

where $q \in \mathbb{C}$ in general. In the $q = 1$ case we write $\{A, B\}$ for $[A, B]_1$ and we omit q in the $q = -1$ case.

Definition 1 For observables $A, B \in M_{n,\text{sa}}$, state $D \in \mathcal{M}_n^1$ and function $f \in \mathcal{F}_{\text{op}}$, we define the *covariance* of A and B

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

the *quantum f -covariance*, which was introduced by Petz [13] and studied recently in this framework by Gibilisco and Isola [7]

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

the *antisymmetric f -covariance*

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

and the *symmetric f -covariance* as

$$\text{qCov}_{D,f}^s(A, B) = \frac{f(0)}{2} \langle \{D, A\}, \{D, B\} \rangle_{D,f}.$$

For $\alpha \in [0, \pi]$ let us define the α, f -covariance as

$$\text{qCov}_{D,f}^{(\alpha)}(A, B) = \frac{f(0)}{2} \langle e^{-i\alpha/2} [D, A]_{e^{i\alpha}}, e^{-i\alpha/2} [D, B]_{e^{i\alpha}} \rangle_{D,f}$$

and the *Robertson-covariance* as

$$\text{RCov}_D(A, B) = -\frac{i}{2} \text{Tr} (D[A, B]).$$

It is worthwhile to mention that the antisymmetric f -covariance indeed with the Metric Adjusted f -Correlation (I_D^f) introduced by Hansen in [8] and also studied by Gibilisco, Imparato and Isola [5].

We have $\text{qCov}_{D,f}^{(0)}(A, B) = \text{qCov}_{D,f}^s(A, B)$ and $\text{qCov}_{D,f}^{(\pi)}(A, B) = \text{qCov}_{D,f}^{as}(A, B)$. Note that the covariance $\text{qCov}_{D,f}^{(\alpha)}(A, B)$ is trivial if $f \in \mathcal{F}_{\text{op}}^n$ (that is $f(0) = 0$) and the α, f -covariance can be written in the form of

$$\text{qCov}_{D,f}^{(\alpha)}(A, B) = \frac{f(0)}{2} \langle e^{-i\alpha} [D, [D, A]_{e^{i\alpha}}]_{e^{i\alpha}}, B \rangle_{D,f}.$$

The unitary invariance is an important property of these covariances.

Theorem 1 For every state $D \in \mathcal{M}_n^1$, $A, B \in M_{n,\text{sa}}$, $f \in \mathcal{F}_{\text{op}}$, $\alpha \in [0, \pi]$ and $n \times n$ unitary operator U , we have

$$\begin{aligned}\text{Cov}_{UDU^*}(UAU^*, UBU^*) &= \text{Cov}_D(A, B) \\ \text{Cov}_{UDU^*}^f(UAU^*, UBU^*) &= \text{Cov}_D^f(A, B) \\ \text{qCov}_{UDU^*, f}^{(\alpha)}(UAU^*, UBU^*) &= \text{qCov}_{D, f}^{(\alpha)}(A, B) \\ \text{RCov}_{UDU^*}(UAU^*, UBU^*) &= \text{RCov}_D(A, B).\end{aligned}$$

Proof Simple computation using the invariance of trace under cyclic permutations. \square

This unitary invariance allows us to simplify the calculations assuming the state to be diagonal. The local form of the covariances at a given state is the following.

Theorem 2 Assume that the state $D \in \mathcal{M}_n^1$ is of the form $D = \text{Diag}(\lambda_1, \dots, \lambda_n)$. For every $A, B \in M_{n,\text{sa}}$, $f \in \mathcal{F}_{\text{op}}$ and $\alpha \in [0, \pi]$, we have

$$\begin{aligned}\text{Cov}_D(A, B) &= \sum_{k,l=1}^n \frac{\lambda_k + \lambda_l}{2} A_{lk} B_{kl} - \text{Tr}(DA) \text{Tr}(DB) \\ \text{Cov}_D^f(A, B) &= \sum_{k,l=1}^n m_f(\lambda_k, \lambda_l) A_{lk} B_{kl} \\ \text{qCov}_{D, f}^{(\alpha)}(A, B) &= \frac{f(0)}{2} \sum_{k,l=1}^n \frac{|\text{e}^{-i\alpha/2} \lambda_k + \text{e}^{i\alpha/2} \lambda_l|^2}{m_f(\lambda_k, \lambda_l)} A_{lk} B_{kl} \\ \text{RCov}_D(A, B) &= i \sum_{k,l=1}^n \frac{\lambda_l - \lambda_k}{2} A_{kl} B_{lk}\end{aligned}$$

Proof Simple matrix computation. \square

It is worth noting that the above mentioned covariances are real numbers.

For an observable $A \in M_{n,\text{sa}}$ and a state $D \in \mathcal{M}_n^1$ we define $A_0 = A - \text{Tr}(DA)I$, where I is the $n \times n$ identity matrix. Using this transformation we have $\text{Tr } DA_0 = 0$. Now we are in the position to define the covariance matrix of many observables at a given state.

Definition 2 For a fixed density matrix $D \in \mathcal{M}_n^1$, function $f \in \mathcal{F}_{\text{op}}$, $\alpha \in [0, \pi]$ and an N -tuple of nonzero matrices $(A^{(k)})_{k=1, \dots, N} \in M_{n,\text{sa}}$, we define the $N \times N$ matrices Cov_D , Cov_D^f , $\text{qCov}_{D, f}^{as}$ and $\text{qCov}_{D, f}^s$ with entries

$$\begin{aligned}[\text{Cov}_D]_{ij} &= \text{Cov}_D(A_0^{(i)}, A_0^{(j)}) \\ [\text{Cov}_D^f]_{ij} &= \text{Cov}_D^f(A_0^{(i)}, A_0^{(j)}) \\ [\text{qCov}_{D, f}^{(\alpha)}]_{ij} &= \text{qCov}_{D, f}^{(\alpha)}(A_0^{(i)}, A_0^{(j)}).\end{aligned}$$

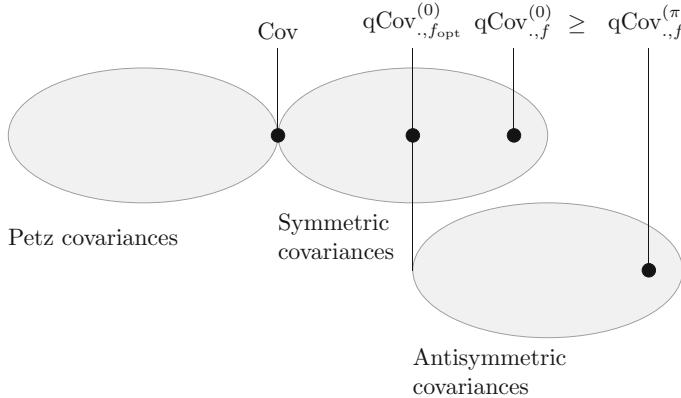


Fig. 1 Robertson-type uncertainty principles

The uncertainty relations mentioned in the overview can be formulated easily with this notations. The Robertson uncertainty principle (3) is

$$\det(\text{Cov}_D) \geq \det(\text{RCov}_D)$$

dynamical uncertainty principle (4) is

$$\det(\text{Cov}_D) \geq \det(q\text{Cov}_{D,f}^{(\pi)}).$$

It was shown in [12] that

$$\det(\text{Cov}_D) \geq \det(q\text{Cov}_{D,f}^{(0)}) \geq \det(q\text{Cov}_{D,f}^{(\pi)})$$

holds and for every function $g \in \mathcal{F}_{\text{op}}$

$$\det(q\text{Cov}_{D,f_{\text{opt}}}^{(0)}) \geq \det(q\text{Cov}_{D,g}^{(\pi)}),$$

where

$$f_{\text{opt}} = \frac{1}{2} \left(\frac{1+x}{2} + \frac{2x}{1+x} \right).$$

Uncertainty relations involving different type of covariances are illustrated in Fig. 1.

We show that the continuous path between symmetric and antisymmetric covariances has the same monotonicity property.

Theorem 3 Consider a fixed density matrix $D \in \mathcal{M}_n^1$, parameters $\alpha, \beta \in [0, \pi]$, $\alpha > \beta$ and an N -tuple of nonzero matrices $(A^{(k)})_{k=1,\dots,N} \in M_{n,\text{sa}}$. If f_1, f_2 are symmetric and normalized operator monotone functions for which

$$\frac{f_1(0)}{f_1(t)} \geq \frac{f_2(0)}{f_2(t)} \quad \forall t \in \mathbb{R}^+$$

holds, then we have

$$\begin{aligned} \det(\text{qCov}_{D,f_k}^{(\beta)}) &\geq \det(\text{qCov}_{D,f_k}^{(\alpha)}) \quad k = 1, 2 \\ \det(\text{qCov}_{D,f_1}^{(\alpha)}) &\geq \det(\text{qCov}_{D,f_2}^{(\alpha)}). \end{aligned}$$

Proof Consider the set of functions

$$\mathcal{C}_M = \left\{ g : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+ \mid \begin{array}{l} g \text{ is a symmetric smooth function, with analytical} \\ \text{extension defined on a neighbourhood of } \mathbb{R}^+ \times \mathbb{R}^+ \end{array} \right\}.$$

Fix a function $g \in \mathcal{C}_M$. Define for every $D \in M_n$ and for every $A, B \in M_{n,\text{sa}}$

$$(A, B)_{D,g} = \text{Tr} (Ag(L_{n,D}, R_{n,D})(B)).$$

and define the $N \times N$ matrix $\mathfrak{Cov}_{D,g}$ with entries

$$[\mathfrak{Cov}_{D,g}]_{ij} = (A_0^{(i)}, A_0^{(j)})_{D,g}.$$

At a diagonal state $D = \text{Diag}(\lambda_1, \dots, \lambda_n)$, for observables $A, B \in M_{n,\text{sa}}$ we have

$$(A, B)_{D,g} = \sum_{k,l=1}^n A_{lk} B_{kl} g(\lambda_k, \lambda_l)$$

(see [12]). With this formalism for an operator monotone symmetric and normalized function f and a parameter $\gamma \in [0, \pi]$ we have for observables A, B

$$(A, B)_{D,g_{f,\gamma}} = \text{qCov}_{D,f}^{(\gamma)}(A, B),$$

where

$$g_{f,\gamma}(x, y) = \frac{f(0) |x + e^{i\gamma} y|^2}{2y f\left(\frac{x}{y}\right)}.$$

For a fixed parameter $\gamma \in [0, \pi]$, if for every $t \in \mathbb{R}^+$

$$\frac{f_1(0)}{f_1(t)} \geq \frac{f_2(0)}{f_2(t)}$$

holds, then

$$g_{f_1,\gamma}(x, y) \geq g_{f_2,\gamma}(x, y) \quad \forall x, y \in \mathbb{R}^+.$$

For a fixed operator monotone function f , if $0 \leq \beta < \alpha \leq \pi$, then

$$g_{f,\alpha}(x, y) \leq g_{f,\beta}(x, y) \quad \forall x, y \in \mathbb{R}^+.$$

Completing the proof we use the fact ([12]) that if for functions $g_1, g_2 \in \mathcal{C}_M$

$$g_1(x, y) \geq g_2(x, y) \quad \forall x, y \in \mathbb{R}^+$$

holds, then we have

$$\det(\mathfrak{Cov}_{D,g_1}) \geq \det(\mathfrak{Cov}_{D,g_2}).$$

□

4 Robertson Uncertainty Principle

In this section, we give a very simple and understandable proof for the Robertson uncertainty principle. It turned out that the uncertainty principle in question can be originated from a more general determinant inequality between real and imaginary part of positive definite matrices.

Lemma 1 *Let $A \in M_n$ be a positive definite invertible matrix. The real and imaginary part of A satisfy the following determinant inequality.*

$$\det(\text{Re}(A)) \geq \det(\text{Im}(A)) \tag{6}$$

Proof For odd n , the right-hand side is identically 0 because $\text{Im}(A)$ is a real skew-symmetric matrix and thus we have nothing to prove.

Assume that n is even. The determinant of an even dimensional skew-symmetric matrix is obviously non-negative. The left-hand side is strictly positive because $\text{Re}(A)$ arises as the convex combination of A and \bar{A} that are positive definite invertible matrices, where \bar{A} stands for the element-wise conjugate of A .

The inequality (6) can be written in the form of

$$\det(A + \bar{A}) \geq \det\left(\frac{1}{i} \times (A - \bar{A})\right)$$

and multiplying by $A^{-1/2}$ from left and right we have

$$\det(I + A^{-1/2}\bar{A}A^{-1/2}) \geq \det\left(\frac{1}{i} \times (I - A^{-1/2}\bar{A}A^{-1/2})\right).$$

Now consider the matrix $B = A^{-1/2} \bar{A} A^{-1/2}$, which is clearly self-adjoint. For every vector $x \in \mathbb{C}^n$ define $y = A^{-1/2}x$ and denote by \bar{y} the coordinate-wise conjugate of y . The equalities

$$\langle x, Bx \rangle = \langle x, A^{-1/2} \bar{A} A^{-1/2} x \rangle = \langle y, \bar{A} y \rangle = \langle \bar{y}, A \bar{y} \rangle \geq 0$$

shows that B is positive definite hence its spectrum belongs to $]0, \infty[$. To prove inequality (6), we have to prove that for a positive definite operator B

$$1 \geq \det \left(\frac{1}{i} \times (I - B)(I + B)^{-1} \right) = |\det((I - B)(I + B)^{-1})| \quad (7)$$

holds. Consider the function

$$f : [0, \infty] \rightarrow \mathbb{R} \quad x \mapsto \frac{1-x}{1+x},$$

which is continuous and it maps $[0, \infty[$ onto $-1, 1]$.

By the spectral mapping theorem, we can write $\sigma(f(B)) = f(\sigma(B)) \subset [-1, 1]$ that implies immediately that

$$|\det(f(B))| = \left| \prod_{\lambda \in \sigma(f(B))} \lambda \right| \leq 1,$$

which gives back Eq.(7). \square

Now we are in the position to prove the Robertson uncertainty principle.

Theorem 4 (Robertson [16]) *In every state $D \in \mathcal{M}_n^1$ and for arbitrary set of observables $(A^{(k)})_{k=1, \dots, N}$,*

$$\det(\text{Cov}_D) \geq \det(\text{RCov}_D)$$

holds.

Proof We may assume that the A_k -s are linearly independent and centered i.e. $\text{Tr}(DA^{(k)}) = 0$ for $k = 1, \dots, N$. Consider a diagonal state $D = \text{Diag}(\lambda_1, \dots, \lambda_n)$, the V vector space of $n \times n$ self-adjoint centered matrices and the scalar product

$$(\cdot, \cdot) : V \times V \rightarrow \mathbb{C} \quad (A, B) \mapsto \text{Tr}(DAB).$$

Consider the Gram matrix $G = (A^{(i)}, A^{(j)})_{i,j=1, \dots, N}$. One can easily check that the following equalities hold.

$$\text{Re}(G) = [\text{Cov}_D(A^{(i)}, A^{(j)})]_{i,j=1,\dots,N}$$

$$\text{Im}(G) = [\text{RCov}_D(A^{(i)}, A^{(j)})]_{i,j=1,\dots,N}$$

By Lemma 1, $\det(\text{Re}(G)) \geq \det(\text{Im}(G))$, which completes the proof. \square

References

1. Andai, A.: Uncertainty principle with quantum Fisher information. *J. Math. Phys.* **49**(1), 7 (2008)
2. Dvurečenskij, A.: Gleason's Theorem and its Applications. Mathematics and its Applications (East European Series), vol. 60. Kluwer Academic Publishers Group, Dordrecht (1993)
3. Gibilisco, P., Hiai, F., Petz, D.: Quantum covariance, quantum Fisher information, and the uncertainty relations. *IEEE Trans. Inform. Theory* **55**(1), 439–443 (2009)
4. Gibilisco, P., Imparato, D., Isola, T.: A Robertson-type uncertainty principle and quantum Fisher information. *Linear Algebra Appl.* **428**(7), 1706–1724 (2008)
5. Gibilisco, P., Imparato, D., Isola, T.: Uncertainty principle and quantum Fisher information. ii. *J. Math. Phys.* **48**(7), 072109 (2007)
6. Gibilisco, P., Imparato, D., Isola, T.: A volume inequality for quantum Fisher information and the uncertainty principle. *J. Stat. Phys.* **130**(3), 545–559 (2008)
7. Gibilisco, P., Isola, T.: How to distinguish quantum covariances using uncertainty relations. *J. Math. Anal. Appl.* **384**(2), 670–676 (2011)
8. Hansen, F.: Metric adjusted skew information. *Proc. Natl. Acad. Sci.* **105**(29), 9909–9916 (2008)
9. Heisenberg, W.: Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.* **43**(3), 172–198 (1927)
10. Hiai, F., Petz, D., Toth, G.: Curvature in the geometry of canonical correlation. *Stud. Sci. Math. Hung.* **32**(1–2), 235–249 (1996)
11. Kennard, E.H.: Zur quantenmechanik einfacher bewegungstypen. *Z. für Phys.* **44**(4–5), 326–352 (1927)
12. Lovas, A., Andai, A.: Refinement of Robertson-type uncertainty principles with geometric interpretation. *Int. J. Quantum Inf.* **14**(02), 1650013 (2016)
13. Petz, D.: Covariance and Fisher information in quantum mechanics. *J. Phys. A* **35**(4), 929–939 (2002)
14. Petz, D., Sudár, Cs: Geometries of quantum states. *J. Math. Phys.* **37**(6), 2662–2673 (1996)
15. Robertson, H.P.: The uncertainty principle. *Phys. Rev.* **34**, 163–164 (1929)
16. Robertson, H.P.: An indeterminacy relation for several observables and its classical interpretation. *Phys. Rev.* **46**, 794–801 (1934)
17. Schrödinger, E.: About Heisenberg uncertainty relation (original annotation by A. Angelow and M.-C. Batoni). *Bulgar. J. Phys.* **26**(5–6), 193–203 (2000), (1999); *Transl. Proc. Prussian Acad. Sci. Phys. Math. Sect.* **19**, 296–303 (1930)