The general form of γ-family of quantum relative entropies[1]

Ryszard Paweł Kostecki

Institute of Theoretical Physics, University of Warsaw, Hoża 69, 00-681 Warszawa, Poland

August 30, 2011

Abstract

We use the Falcone–Takesaki non-commutative flow of weights and the resulting theory of non-commutative L_p spaces in order to define the family of relative entropy functionals that naturally generalise the quantum relative entropies of Jenčová–Ojima and the classical relative entropies of Zhu–Rohwer, and belong to an intersection of families of Petz relative entropies with Bregman relative entropies. For the purpose of this task, we generalise the notion of Bregman entropy to the infinite-dimensional non-commutative case using the Legendre–Fenchel duality. In addition, we use the Falcone–Takesaki duality to extend the duality between coarse–grainings and Markov maps to the infinite-dimensional non-commutative case. Following the recent result of Amari for the Zhu–Rohwer entropies, we conjecture that the proposed family of relative entropies is uniquely characterised by the Markov monotonicity *and* the Legendre–Fenchel duality. The role of these results in the foundations and applications of quantum information theory is discussed.

dedicated to Professor Roman S. Ingarden on the occasion of his 90th birthday

1 Introduction

The information theory can be separated into two parts: information kinematics (described in terms of information geometry theory) and information dynamics (described in terms of inductive inference theory). Information kinematics incorporates the description of convex and smooth geometric structures on the spaces of information states, while information dynamics incorporates the methods of statistical inference such as Bayes' rule, maximum likelihood, and constrained relative entropic updating. Both parts of information theory share two underlying mathematical notions: information model and information deviation (relative entropy). Following the approach initiated by Ingarden et al [73] and Eguchi [44, 45] (and continued in particular in [99, 69, 135, 83]), we think that the further structures of the information geometry theory, such as riemannian metrics or affine connections on information models, should follow from the additional requirements imposed on these two notions. However, this requires us to define information models and information deviations on the level of generality that covers all regimes of information geometry, ranging from the finite-dimensional commutative normalised case to the infinite-dimensional non-commutative non-normalised case.

We define an *information model* as a subset $\mathcal{M}$ of the space $\mathcal{N}_*^+$ of all positive, finite, normal, linear $\mathbb{C}$-valued functions on a given commutative or non-commutative

---

[1]Published in: Open Systems and Information Dynamics **18**:2 (2011), 191-221.

W^*-algebra $\mathcal{N}$.[2] Elements of $\mathcal{N}_*^+$ will be called *states* (or *information states*). If the additional normalisation condition is assmed, then the resulting space $\mathcal{N}_{*1}^+ \subset \mathcal{N}_*^+$ consists of all normal algebraic states on $\mathcal{N}$. For commutative $\mathcal{N}$, $\mathcal{N}_*^+$ reduces to a space $L_1(\mho)$ of Daniell–Stone integrals on a Banach vector lattice of characteristic functions on Stone spectrum of a countably finite Dedekind complete maharanisable algebra $\mho$ that is canonically associated with $\mathcal{N} \cong L_\infty(\mho)$. In this case $\mathcal{N}_*^+$ reduces to a space $L_1(\mho)_1^+$ of all normal (hence, monotonically continuous) expectation functionals on $\mathcal{N}$, which bijectively corresponds to the space of all positive probability measures on $\mho$.

In the commutative and normalised case $\mathcal{N}_*^+$ reduces to a space of all normal (hence, monotonically continuous) expectation functionals on $\mathcal{N}$, which is a space of Daniell–Stone integrals on a Banach vector lattice that bijectively corresponds to the space $\mathcal{P}$ of all positive probability measures on a countably finite Dedekind complete boolean algebra $\mho$ that is derived from $\mathcal{N}$.

An *information deviation* (also referred to as: negative relative entropy, information divergence, information gain, loss function, risk function, contrast functional) is defined as a map $D : \mathcal{N}_*^+ \times \mathcal{N}_*^+ \to [0, \infty]$ such that $D(\omega, \phi) = 0$ iff $\omega = \phi$. It plays a role of a non-symmetric distance functional on $\mathcal{N}_*^+$, and serves as a principal tool for quantification of relative information content of states.

The above definitions of information model and information deviation are suitable also for the purposes of information dynamics theory (see [91, 92]). The main goal of information theory is to provide particular instances of inductive inference (information dynamics), so the derivation of the particular form of deviation functional from the requirements imposed on the information dynamics is an important task. This method of axiomatisation was advanced in particular in [143, 144, 85, 159, 160, 145, 113, 114, 37, 156, 23]. However, we would like to keep the kinematic and dynamic properties of information theory separated for a while, allowing for a better understanding of information geometric kinematics on its own right. Hence, in this work we will consider only the information kinematics and, more specifically, the problem of unique selection of a particular form of information deviation functional by the mathematical conditions that are plausible from the perspective of information kinematic applications. There exist also many axiomatisations of information deviations provided this way (see e.g., [1, 36, 42, 121, 111, 123, 38, 4]).

The virtue of (kinematic or dynamic) axiomatic characterisation is that it allows to

---

[2] An additional subscript $\lrcorner_1$ will be used to denote the subset of *normalised* ($\omega(\mathbb{I}) = 1$) elements, while an additional subscript $\lrcorner_0$ will be used to denote the subset of *faithful* ($\omega(x^*x) = 0 \Rightarrow x = 0 \ \forall x \in \mathcal{N}$) elements. A function $\omega : \mathcal{B} \to \mathbb{K}$, where $\mathcal{B}$ is a Banach space and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, is called *normal* iff $\omega(\sup \mathcal{F}) = \sup_{x \in \mathcal{F}} \omega(x)$ for each directed filter $\mathcal{F} \subseteq \mathcal{B}$ with the upper bound $\sup \mathcal{F}$. The *weak-* topology* on Banach dual $\mathcal{B}^{\mathbf{B}}$ of $\mathcal{B}$ is defined as the weakest topology on $\mathcal{B}^{\mathbf{B}}$ such that the linear functions $\mathcal{B}^{\mathbf{B}} \ni \omega \mapsto \omega(x) \in \mathbb{K}$ are continuous for every $x \in \mathcal{B}$. A $\mathbb{K}$-linear function on $\mathcal{B}$ is normal iff it is continuous in weak-* topology. A C^*-algebra is a Banach space $\mathcal{A}$ over $\mathbb{K}$ such that it is also a ring over $\mathbb{K}$ with unit $\mathbb{I}$ and operation $*$ defined by $(ab)^* := b^*a^*$, $(\lambda_1 a + \lambda b)^* := \lambda_1^* a^* + \lambda_2^* b^*$ (where $* : \mathbb{K} \to \mathbb{K}$ denotes the complex conjugation), such that $\|xy\| \leq \|x\|\|y\|$, $\|\mathbb{I}\| = 1$, $\|x^*x\| = \|x\|^2$ $\forall x, y \in \mathcal{A}$. A W^*-*algebra* is a C^*-algebra $\mathcal{N}$ such that there exists a Banach space $\mathcal{N}_*$, called *predual*, satisfying $\mathcal{N}_*^{\mathbf{B}} = \mathcal{N}$. The set of normal linear $\mathbb{K}$-valued (hence, weak-* continuous) functions on $\mathcal{N}$ is precisely a predual of $\mathcal{N}$. A *weight* on a W^*-algebra $\mathcal{N}$ is a function $\omega : \mathcal{N}^+ \to [0, +\infty]$ such that $\omega(0) = 0$, $\omega(x + y) = \omega(x) + \omega(y)$, $\lambda > 0 \Rightarrow \omega(\lambda x) = \lambda\omega(x)$. A weight ω is called *semi-finite* iff $\forall x \in \mathcal{N}^+ \ \exists y \in \mathcal{N}^+ \ (x \geq y, \ y \neq 0, \ \omega(y) < +\infty)$. A weight is called a *trace* iff $\omega(u^*xu) = \omega(x)$ for all unitary $u \in \mathcal{N}$. The space of all semi-finite normal weights on $\mathcal{N}$ will be denoted $\mathcal{W}(\mathcal{N})$.

equip the characterised object with an unambiguous justification that is grounded in the particular interpretation of the elementary axioms. Yet, a single mathematical axiomatisation can be equipped with arbitrary many conceptual interpretations, hence the relevance of this point can be questioned. Nevertheless, for the purposes of applications of the information theory, at least some minimal purely operational interpretation is required anyway, hence the axiomatic characterisation of the information geometric (and information dynamic) structures is important. The variety of different applications and axiomatisations suggests strongly that there is no 'absolute' or 'universal' method of model construction, as well as there is no 'universal' choice of information deviation and no 'universal' choice of a method of inductive inference. In consequence, the aim of mathematical inquiry is to identify various sets of conditions such that each set is necessary and sufficient to select some particular category of models and some particular information deviation (as well as some particular inductive inference rule). While the statistical inference (dynamics) and information geometric modelling (kinematics) based on different sets of conditions is incommensurable, each set on its own is a valid universe for information theoretic (statistical, inferential, probabilistic, quantum) inquiry.

This paper aims at the problem of identification of a set of conditions underlying the particular information geometry theory that was originally developed along the lines of requirement of monotonicity of information geometric structures (deviations, metrics, connections) under the Markov morphisms of information models. The results of Chentsov [25, 26] showed that the constraint of monotonicity under Markov maps is strong enough to provide interesting characterisations of unique (Fisher–Rao [53, 130, 14, 79]) riemannian metrics and one-parameter family of (Chentsov–Amari [25, 26, 2]) affine connections in normalised finite-dimensional commutative case. Due to work of Morozova and Chentsov [105] and final result of Petz [122, 124], a family of riemannian metrics in normalised finite-dimensional non-commutative case was characterised by the same condition (an explicit integral representation of the Morozova–Chentsov–Petz family was given by Hansen [67]). However, further results showed that monotonicity under Markov maps is insufficient for fine analysis of geometries of quantum states in non-commutative case, even in finite-dimensional normalised regime. This follows from the results of Jenčová [80, 82, 83] characterising the family of markovian monotone connections, as well as related results of Hasegawa and Petz [127, 70, 69], Grasselli and Streater [60, 59, 58] and Gibilisco and Isola [55, 56, 57] on characterisation of a subfamily of the Morozova–Chentsov–Petz metrics known as the Wigner–Yanase–Dyson metrics [164, 68], together with its boundary known as the Bogolyubov–Kubo–Mori metric [15, 93, 103]. This leads to a question about the additional principle of structural properties which would enable an improved characterisation.

As already remarked, Ingarden et al [73] and Eguchi [44, 45] observed that the structures of riemannian metric and affine connections on information manifold can be derived from information deviation functional by suitable differentiation of the arguments and passing to the limit with one argument converging to other. In particular, the Fisher–Rao metric and the family of Chentsov–Amari connections can be derived in finite-dimensional commutative case by differentiation of the family D_p of Zhu–Rohwer deviations [171, 172, 173]. This approach was continued also in the finite-dimensional non-commutative case [99, 69, 83]. This leads us to wonder whether (and which of)

3

the properties of information deviation could be used to characterise the structures of information geometry. Amari [4] recently showed that the Zhu–Rohwer deviations are characterised (under some mild auxiliary conditions) as an intersection of the families of Csiszár–Morimoto [34, 35, 104] and Bregman [19] deviations. While the former family is characterised [36, 37] by monotonicity under Markov maps, the latter is characterised [86, 37] by additive decomposition under projection on the subset that is 'orthogonal' to the projection in the sense of Legendre duality.

This leads us to consider duality as a principle of characterisation that is equally fundamental as markovian monotonicity. However, as opposed to perspective present in standard treatments such as [107, 44, 45, 98, 97, 5, 169], we require also that the dualistic properties should be independent of differential properties, because the former are essentially related to variational component of the theory, which in infinite-dimensional setting becomes quite independent of the smooth component of the theory. Consideration of non-smooth infinite-dimensional duality as a principle that is as important as Markov monotonicity leads us to deny the foundational role of differential geometric Norden–Sen duality [109, 142] in favour of variational convex Legendre–Fenchel duality [51, 20]. Turning this idea into structure, we define the generalised Bregman deviation in terms of the convex Legendre–Fenchel duality on the dualised vector spaces, with no differentiability and no topological continuity properties required a priori.

Jenčová [81, 84] and Ojima [112] introduced a family D_p of quantum deviation functionals on $\mathcal{N}_*^+$, defined using the dual structure of non-commutative Araki–Masuda $L_p(\mathcal{N}, \phi)$ spaces [10, 101]. These deviations are well-defined in the non-commutative infinite-dimensional case, and belong to an intersection of the families of generalised Bregman deviations and Markov monotone Petz deviations [119, 120] (the latter are non-commutative generalisation of the Csiszár–Morimoto deviations). In particular cases, the Jenčová–Ojima deviations reduce to relative entropies of Hasegawa [68] and Umegaki–Araki [157, 7, 8]. However, the Araki–Masuda $L_p(\mathcal{N}, \phi)$ space *is not* a canonical non-commutative L_p space construction, due to its dependence on an arbitrary weight ϕ. In consequence, the Jenčová–Ojima deviation *is not* a canonical non-commutative generalisation of the Zhu–Rohwer deviation, because the construction of the former depends on the choice of reference weight ϕ, while the latter is independent of any choice of this kind.

We propose a solution to this problem, which is based on the Falcone–Takesaki theory [48] of non-commutative flow of weights and associated construction of non-commutative $L_p(\mathcal{N})$ spaces that does not depend on a choice of reference weight. Using its remarkable properties, we define the family D_p of information deviation (relative entropy) functionals which is the canonical generalisation of Jenčová–Ojima *and* Zhu–Rohwer families of deviations. We show that our family D_p of information deviations belongs not only to Petz's class of Markov monotone deviations, but also to the class of generalised Bregman deviations. As a non-commutative counterpart to the result of Amari [4], we conjecture that the family of D_p deviations is uniquely characterised as intersection of Markov monotone deviations with generalised Bregman deviations. This requirement states that quantum information deviations should be monotone under (Markov) coarse-graining of information *and* should allow additive decompositions under (Legendre–Fenchel) orthogonal projections on closed convex affine information subspaces.

This way the family D_p of information deviations on $\mathcal{M} \subseteq \mathcal{N}_*^+$ becomes equipped

with an information theoretic interpretation, and the same is true for further information geometric structures on $\mathcal{M}$ derived from these deviations, such as riemannian metrics and affine connections. This interpretation refers to the notion of information as a fundamental entity, and is not involved in the conceptual problems associated with the interpretations of probability and the interpretations of Hilbert space based quantum theory.

From the perspective of applications, this allows to discuss the algebraic models of quantum statistical mechanics and quantum field theory as particular cases of quantum information theory. The use of the Falcone–Takesaki theory and the Legendre–Fenchel duality advocated in this paper is a key novel technique that allows to derive various new results in the infinite-dimensional non-commutative case, where neither the suitable differentiability properties nor the description in terms of density operators is available. From the perspective of foundations, this allows to provide a major change of perspective: instead of separate consideration of statistical theory and quantum theory, with their separate problems of model construction, one obtains a single framework covering all theories at once. The model construction techniques based on spectral theory can be replaced by the model construction techniques based on quantum information geometry (which includes the spectral theory as a special case), with the family D_p of information deviations playing the central role (and replacing the role played by the Hilbert space norm).

2 Information models

2.1 Non-commutative flow of weights

In what follows, we will often identify the W^*-algebra $\mathcal{N}$ with its standard representation von Neumann algebra $\pi(\mathcal{N})$ on the standard representation Hilbert space $\mathcal{H}$ [63]. Every W^*-algebra $\mathcal{N}$ has a unique standard representation, up to unitary isomorphism. This representation is faithful ($\ker(\pi) = \{0\}$) and is unitarily isomorphic with the Gel'fand–Naĭmark–Segal representation [138] for a pair $(\mathcal{N}, \omega)$, whenever $\omega \in \mathcal{N}_{*0}^+$ or $\omega \in \mathcal{W}_0(\mathcal{N})$. While the set $\mathcal{N}_{*0}^+$ is non-empty iff $\mathcal{N}$ is countably finite (i.e., it is isomorphic to a von Neumann algebra possessing a cyclic and separating vector), $\mathcal{W}_0(\mathcal{N}) \neq \emptyset$ for every W^*-algebra $\mathcal{N}$.

Two crucial elements of the Falcone–Takesaki theory [48] are the *core* von Neumann algebra $\widetilde{\mathcal{N}}$ associated functorially to any von Neumann algebra $\mathcal{N}$, and the resulting canonical construction of non-commutative $L_p(\mathcal{N})$ spaces, which is independent of any reference weight or state.

For $x, x' \in \mathcal{N}$, $\omega, \omega' \in \mathcal{W}_0(\mathcal{N})$, and $\left(\frac{\mathrm{D}\omega}{\mathrm{D}\omega'}\right)_t$ denoting Connes' cocycle [28] (which is a generalisation of the Radon–Nikodým derivative to the non-commutative case), one defines the equivalence relation $\sim_t$ on $\mathcal{N} \times \mathcal{W}_0(\mathcal{N})$ by

$$(x, \omega) \sim_t (x', \omega') \iff x' = x \left(\frac{\mathrm{D}\omega}{\mathrm{D}\omega'}\right)_t. \tag{1}$$

The equivalence class $(\mathcal{N} \times \mathcal{W}_0(\mathcal{N}))/\sim_t$ is denoted by $\mathcal{N}(t)$, and its elements are denoted by $x\omega^{it}$. The definitions $(x\omega^{it} + y\omega^{it}) := (x+y)\omega^{it}$, $\lambda(x\omega^{it}) := (\lambda x)\omega^{it}$ for $\lambda \in \mathbb{C}$, and

$\|x\omega^{it}\| := \|x\|$ equip $\mathcal{N}(t)$ with the structure of the Banach space, which is isometrically isomorphic to the Banach space structure of $\mathcal{N}$. By definition, $\mathcal{N}(0)$ is isomorphic to $\mathcal{N}$ also as a von Neumann algebra. The product topology on $\mathcal{N} \times \mathbb{R}$ (with $\mathcal{N}$ endowed with the weak-$*$ topology or Arens–Mackey topology, but not with norm topology) allows to use the bijections $\mathcal{N}(t) \ni x\varphi^{it} \mapsto (x, t) \in \mathcal{N} \times \mathbb{R}$ to form Fell's Banach $*$-algebra bundle $F(\mathcal{N}) := \coprod_{t \in \mathbb{R}} \mathcal{N}(t)$ over $\mathcal{N} \times \mathbb{R}$ [49, 50]. The *core* von Neumann algebra $\widetilde{\mathcal{N}}$ is constructed from $\mathcal{N}$ as a result of action of the Banach $*$-algebra of suitable sections of $F(\mathcal{N})$ on a suitably defined auxiliary Hilbert space (see [48] for details). The construction of $\widetilde{\mathcal{N}}$ does not depend on any weight $\mathcal{W}_0(\mathcal{N})$. However, for every choice of a particular ω, there exists a unitary map providing the isomorphism $\widetilde{\mathcal{N}} \cong \mathcal{N} \rtimes_{\sigma^\omega} \mathbb{R}$, where $\sigma^\omega(\cdot) = \Delta_\omega^{it}(\cdot)\Delta_\omega^{-it}$ denotes the Tomita–Takesaki modular automorphism of $\mathcal{N}$ associated with ω [154, 149], $\mathcal{N} \rtimes_{\sigma^\omega} \mathbb{R}$ is a crossed product von Neumann algebra acting on $\mathcal{H} \otimes L_2(\mathbb{R}, dt)$ [150, 158], and $\mathcal{H}$ is a standard representation Hilbert space of $\mathcal{N}$ [63].

The one-parameter automorphism group of $F(\mathcal{N})$, defined by

$$\tilde{\sigma}_s(x\phi^{it}) := e^{-its}x\phi^{it} \quad \forall x\phi^{it} \in \mathcal{N}(t), \tag{2}$$

extends uniquely to an automorphism group $\tilde{\sigma}_s : \widetilde{\mathcal{N}} \to \widetilde{\mathcal{N}}$. It allows to define a *grade* $\mathrm{grad}(T)$ of a closed densely defined operator T affiliated with $\widetilde{\mathcal{N}}$ as such $\gamma \in \mathbb{C}$ that

$$\tilde{\sigma}_s(T) = e^{-\gamma s}T \quad \forall s \in \mathbb{R}. \tag{3}$$

If $\mathrm{grad}(T) = 0$, then T is bounded, but if $\mathrm{re}\,\mathrm{grad}(T) \neq 0$, then T is unbounded. The action of $\tilde{\sigma}_s$ on $\widetilde{\mathcal{N}}$ is integrable over $s \in \mathbb{R}$, and $I_{\tilde{\sigma}}(x) := \int_{\mathbb{R}} ds\, \tilde{\sigma}_s(x)$, $x \in \widetilde{\mathcal{N}}^+$, is an operator valued weight from $\widetilde{\mathcal{N}}$ to $\mathcal{N}$ (for details on operator valued weights, see [65, 66, 47]). This allows to equip $\widetilde{\mathcal{N}}$ with a faithful normal semi-finite trace $\tilde{\tau}$, defined by

$$\tilde{\tau}_\varphi(x) := \lim_{\epsilon \to +0} \varphi \circ I_{\tilde{\sigma}}(\varphi^{-1/2}(1 + \epsilon\varphi^{-1})^{-1/2} x \varphi^{-1/2}(1 + \epsilon\varphi^{-1})^{-1/2}). \tag{4}$$

This definition is independent of the choice of weight ($\tilde{\tau}_\varphi = \tilde{\tau}_\psi \; \forall \varphi, \psi \in \mathcal{W}_0(\mathcal{N})$), which allows to write $\tilde{\tau}$ instead of $\tilde{\tau}_\varphi$. Moreover, $\tilde{\tau}$ has the scaling property

$$\tilde{\tau} \circ \tilde{\sigma}_s = e^{-s}\tilde{\tau} \quad \forall s \in \mathbb{R}. \tag{5}$$

This allows to consider $\tilde{\tau}$ as a *canonical trace* of $\widetilde{\mathcal{N}}$. Falcone and Takesaki call the system $(\widetilde{\mathcal{N}}, \mathbb{R}, \tilde{\sigma}, \tilde{\tau})$ the *non-commutative flow of weights*.

2.2 Non-commutative $L_p(\mathcal{N})$ spaces

Let $\mathfrak{M}^p(\widetilde{\mathcal{N}})$ denote the space of all $\tilde{\tau}$-measurable operators of grade $1/p$ affiliated to $\widetilde{\mathcal{N}}$ (for details on measurable operators, see [141, 108, 152]), and let the set $\mathfrak{m}_{I_{\tilde{\sigma}}}^+ \subseteq \widetilde{\mathcal{N}}^+$ be defined as

$$\mathfrak{m}_{I_{\tilde{\sigma}}}^+ := \{x^*y \in \widetilde{\mathcal{N}} \mid \|I_{\tilde{\sigma}}(x^*x)\| < \infty, \|I_{\tilde{\sigma}}(y^*y)\| < \infty\}. \tag{6}$$

Given weight $\varphi \in \mathcal{W}(\mathcal{N})$ and a canonical trace $\tilde{\tau}$, the construction

$$h_\varphi^{it} := \left(\frac{\mathrm{D}(\varphi \circ I_{\tilde{\sigma}})}{\mathrm{D}\tilde{\tau}}\right)_t \quad \forall t \in \mathbb{R} \tag{7}$$

defines the operator valued maps $\varphi \mapsto h_\varphi$ on $\mathcal{W}(\mathcal{N})$, called the Haagerup correspondence [64, 166]. From the Pedersen–Takesaki non-commutative generalisation [115] of the Radon–Nikodým theorem it follows that h_φ is a unique positive self-adjoint operator affiliated with $\widetilde{\mathcal{N}}$ which satisfies

$$\varphi \circ I_{\tilde{\sigma}}(\cdot) = \tilde{\tau}(h_\varphi \,\cdot). \tag{8}$$

Moreover, for any closed and densely defined positive operator T affiliated with $\widetilde{\mathcal{N}}$ with $p := \mathrm{re}\,\mathrm{grad}(T) > 0$ there exists a unique weight φ such that h_φ is of grade 1 and $h_\varphi = |T|^{1/p}$. Given a polar decomposition of $\varphi = u|\varphi|$ as a linear form, the weight φ is finite ($|\varphi|(\mathbb{I}) < \infty$) iff h_φ is $\tilde{\tau}$-measurable [108, 64]. Hence, h_φ can be considered as (a reference-independent) 'operator density' of φ. For $\varphi \in \mathcal{N}_*^+$, an assignment $T(\varphi) := uT(|\varphi|)$ defines a unique extension of the Haagerup correspondence map to a natural bijection (linear isomorphism)

$$\mathcal{N}_* \ni \omega \mapsto T(\omega) \in \mathfrak{M}^1(\widetilde{\mathcal{N}}). \tag{9}$$

Falcone and Takesaki show that the integral on T, given by

$$\int T := \tilde{\tau}(a^{1/2} T a^{1/2}), \ \ T \in \mathfrak{M}^1(\widetilde{\mathcal{N}}), \ \ a \in \mathfrak{m}_{I_{\tilde{\sigma}}}^+, \ \ I_{\tilde{\sigma}}(a) = 1, \tag{10}$$

is well defined. While $\tilde{\tau}$ takes only the $+\infty$ value on non-zero elements of $\mathfrak{M}^1(\widetilde{\mathcal{N}})$, the integral $\int$ takes finite values. This allows to extend (9) to an isometric isomorphism of Banach spaces, with the norm on $\mathfrak{M}^1(\widetilde{\mathcal{N}})$ defined by $\|T\|_1 := \int |T|$, and with $\|T(\varphi)\|_1 = \varphi(\mathbb{I})$. The duality pairing between Banach spaces $\mathcal{N}$ and $\mathfrak{M}^1(\widetilde{\mathcal{N}})$ that identifies $\mathcal{N}_*$ with $\mathfrak{M}^1(\widetilde{\mathcal{N}})$ is given by the bilinear form

$$\mathcal{N} \times \mathfrak{M}^1(\widetilde{\mathcal{N}}) \ni (x, T) \mapsto [\![x, T]\!]_{\widetilde{\mathcal{N}}} := \int xT \in \mathbb{C}. \tag{11}$$

The non-commutative $L_p(\mathcal{N})$ spaces for $p \in [1, \infty[$ are defined as the spaces $\mathfrak{M}^p(\widetilde{\mathcal{N}})$ equipped with, and Cauchy complete in, the norm

$$\|T\|_p := \left(\int |T|^p \right)^{1/p}, \ \ T \in \mathfrak{M}^p(\widetilde{\mathcal{N}}). \tag{12}$$

By (9) and (11), $L_1(\mathcal{N}) \cong \mathcal{N}_*$, and it is natural to define $L_\infty(\mathcal{N}) \cong \mathcal{N} \cong \mathcal{N}(0)$, using the definition (3) of grade with $\tilde{\sigma}_s(T) = T$ for $\mathrm{grad}(T) = 0$. The space $L_2(\mathcal{N})$ is isometrically isomorphic to the Hilbert space $\mathcal{H}$ of standard representation of $\mathcal{N}$ [63], and the inner product on $L_2(\mathcal{N})$ defined by

$$L_2(\mathcal{N}) \times L_2(\mathcal{N}) \ni (T_1, T_2) \mapsto \langle T_1, T_2 \rangle_{L_2(\mathcal{N})} := \int T_2^* T_1 \in \mathbb{C} \tag{13}$$

allows to identify them. For any choice of $\omega \in \mathcal{W}_0(\mathcal{N})$ (or $\omega \in \mathcal{N}_{*0}^+$), $L_2(\mathcal{N})$ is unitarily isomorphic to the GNS Hilbert space $\mathcal{H}_\omega$. By definition, all $L_p(\mathcal{N})$ for $p \in [1, \infty]$ are Banach spaces. For any choice of a reference weight $\psi \in \mathcal{W}_0(\mathcal{N})$, they are isometrically isomorphic to the non-commutative $L_p(\mathcal{N}, \psi)$ spaces of Kosaki and Terp [89, 153], Araki

and Masuda [10, 101], Connes and Hilsum [30, 72], and Haagerup [64, 152]. If the von Neumann algebra $\mathcal{N}$ is semi-finite and some faithful normal semi-finite trace τ_ψ on $\mathcal{N}$ is chosen, then the Falcone–Takesaki $L_p(\mathcal{N})$ spaces are isometrically isomorphic and order-isomorphic to non-commutative $L_p(\mathcal{N}, \tau_\psi)$ spaces developed in [43, 141, 41, 110, 96, 147, 108, 167].

The duality (11) extends to non-commutative $L_p(\mathcal{N})$ space duality, given by the bilinear map

$$L_p(\mathcal{N}) \times L_q(\mathcal{N}) \ni (S, T) \mapsto [\![S, T]\!]_{\widetilde{\mathcal{N}}} := \int ST \in \mathbb{C}, \tag{14}$$

with $1/p + 1/q = 1$, where $p \in \{z \in \mathbb{C} \mid \mathrm{re}\,(z) > 0\}$, and $L_{1/it}(\mathcal{N}) = \mathcal{N}(t)$ for all $t \in \mathbb{R}$. This way the Falcone–Takesaki theory incorporates also Izumi's [76, 77, 78] complex extension of the weight-dependent Kosaki–Terp theory.

The spaces $\mathfrak{M}^p(\widetilde{\mathcal{N}})$, $p \in \mathbb{C} \setminus \{0\}$, embed naturally into the $*$-algebra $\mathfrak{M}(\widetilde{\mathcal{N}})$ of all $\tilde{\tau}$-measurable operators affiliated to $\widetilde{\mathcal{N}}$. Due to the properties of grade function, the elements of $\mathfrak{M}(\widetilde{\mathcal{N}})$ possess remarkable algebraic properties. The grade function satisfies:

$$\begin{aligned}
\mathrm{grad}(T^*) &= (\mathrm{grad}(T))^*, \quad \mathrm{grad}(\overline{ST}) = \mathrm{grad}(S) + \mathrm{grad}(T), \\
\mathrm{grad}(|T|) &= \mathrm{re}\,(\mathrm{grad}(T)) = \tfrac{1}{2}(\mathrm{grad}(T) + \mathrm{grad}(T)^*), \\
\mathrm{re}\,(p) < 0 \Rightarrow L_{\frac{1}{p}}(\mathcal{N}) &= \{0\}, \quad \mathrm{re}\,(\mathrm{grad}(T)) \geq 0 \Rightarrow |T|^{\frac{1}{\mathrm{re}\,(\mathrm{grad}(T))}} \in \mathcal{N}_*^+,
\end{aligned} \tag{15}$$

where $\overline{ST}$ is the closure of ST, and the last property is understood via the bijection between $\omega \in \mathcal{N}_*^+$ and the corresponding $h_\omega \in L_1(\mathcal{N})$. If $\phi, \psi \in \mathcal{N}_*^+$ and $x, y \in \mathcal{N}$, then the elements $x\phi^\gamma$ and $y\psi^\lambda$ can be added and multiplied freely inside $\mathfrak{M}(\widetilde{\mathcal{N}})$ for all $\gamma, \lambda \in \mathbb{C}$ such that $\mathrm{re}\,(\gamma) \geq 0$ and $\mathrm{re}\,(\lambda) \geq 0$. Moreover, for any $\{T_i\}_{i=1}^n \subseteq \mathfrak{M}(\widetilde{\mathcal{N}})$ such that $\sum_i \mathrm{grad}(T_i) = 1$,

$$\prod_i T_i \in L_1(\mathcal{N}), \quad \int T_1 \cdots T_n = \int T_n T_1 \cdots T_{n-1}, \quad \|T_1 \cdots T_n\| \leq \|T_1\| \cdots \|T_n\|, \tag{16}$$

holds. In particular, one can use the formal algebraic relations of $\mathcal{N}(t)$ to rewrite the Tomita–Takesaki modular automorphism as $\sigma_{-t}^\phi(x) = \Delta_\phi^{-it} x \Delta_\phi^{it} = \phi^{-it} x \phi^{it}$, and to rewrite Connes' cocycle as $\left(\frac{\mathrm{D}\omega}{\mathrm{D}\phi}\right)_t = \Delta_{\omega,\phi}^{it} \Delta_\phi^{-it} = \omega^{it} \phi^{-it}$, where $\Delta_{\omega,\phi}$ denotes the Connes–Digernes relative modular operator [6, 29, 40]. These remarkable algebraic properties were first observed by Woronowicz [165] and were later analysed in details by Yamagami [166].

2.3 Comparison with traditional approaches

With these results at hand, let us provide a brief discussion of the definition of information model which was given in Introduction.

From the mathematical point of view, the setup of W^*-algebras and their preduals is uniquely selected by three requirements:

1) the complex non-commutative generalisation of the abstract (Daniell–Stone) integration theory on abstract (Riesz) vector lattices should be provided in terms of the abstract $*$-algebras,

2) the duality pairing between the $*$-algebra and the space of integrals on it should be a Banach space duality between Banach spaces, with a Banach $*$-algebra being a Banach dual of the Banach space of integrals,

3) every $*$-homomorphism of $*$-algebra should be continuous in terms of Banach space norm $\|\cdot\|$ (this is assumed through the condition $\|x^*x\| = \|x\|^2$).

Thus, as opposed to the usual algebraic approach to foundations of quantum theory, we disregard here the foundational character of C^*-algebras (c.f. [139]) or Jordan algebras (c.f. [46]) and associated interpretation based on spectral properties and commutative measure theory in favour of W^*-algebras and associated interpretation based on dualistic properties and non-commutative integration theory. On the pragmatic level, this choice can be motivated by the requirement of providing the unified treatment of commutative and non-commutative cases (see below), and by an observation that only W^*-algebras provide a Banach space duality between states and algebras that allows to introduce the duality between Markov maps and coarse-grainings in the infinite-dimensional non-commutative case (see next section).

From the physical point of view, the use of non-commutative $*$-algebras instead of Hilbert spaces in the definition of information model is necessary in order to deal with the issue of unitarily inequivalent Hilbert space representations of algebras of operators, which is one of the key features of the quantum theory in infinite-dimensional (continuous) regime. This includes the continuous equilibrium statistical mechanics [46, 18] and relativistic quantum field theory [61, 163, 9] as two most prominent examples. Further restriction to W^*-algebras is caused by the fact that C^*-algebras are suitable only for description of fermionic models, while the description of bosonic models necessarily requires to use the W^*-algebras [18, 129].

From the statistical point of view, the use of positive parts of preduals of W^*-algebras instead of measure spaces $(\mathcal{X}, \mho(\mathcal{X}), \mu)$ has also an important reason. The traditional approach [52, 88] assumes a particular choice of a background measurable space $(\mathcal{X}, \mho(\mathcal{X}))$, where $\mho(\mathcal{X})$ is some particular countably additive algebra of subsets of some 'sample space' $\mathcal{X}$ (e.g., algebra of Borel subsets of a topological space), and defines a statistical or probabilistic model as a subset of the space

$$\mathcal{M}(\mathcal{X}, \mho(\mathcal{X}), \mu) \subseteq L_1(\mathcal{X}, \mho(\mathcal{X}), \mu)^+$$

of all measures that are absolutely continuous with respect to a fixed measure μ. However, there exist many different choices of $(\mathcal{X}, \mho(\mathcal{X}), \mu)$ such that the associated $L_1(\mathcal{X}, \mho(\mathcal{X}), \mu)$ spaces are isomorphic to the same abstract $L_1(\mho)$ space. The abstract $L_1(\mho)$ space is functorially associated to a countably additive Dedekind complete boolean (caDcb) algebra $\mho$, and the latter can be defined in terms of $(\mathcal{X}, \mho(\mathcal{X}), \mu)$ by [87]

$$\mho := \mho(X)/\{A \in \mho(\mathcal{X}) \mid \mu(A) = 0\}.$$

Hence, the notion of statistical model is in fact independent of the choice of a particular sample space $\mathcal{X}$ and a particular measure space $(\mathcal{X}, \mho(\mathcal{X}), \mu)$. It depends only on the choice of a particular algebra $\mho$. Furthermore, the well-defined statistical models should allow to use the Steinhaus–Nikodým isomorphism of Banach spaces

$$L_1(\mathcal{X}, \mho(\mathcal{X}), \mu)^{\mathbf{B}} \cong L_\infty(\mathcal{X}, \mho(\mathcal{X}), \mu).$$

9

This requirement is crucial for considering the elements of $\mathcal{M}(\mathcal{X}, \mho(\mathcal{X}), \mu)$ as the Radon–Nikodým derivatives with respect to measure μ, because the latter are elements of $L_\infty(\mathcal{X}, \mho(\mathcal{X}), \mu)$. However, this isomorphism does not hold for arbitrary measure spaces [137], but only for those $(\mathcal{X}, \mho(\mathcal{X}), \mu)$ which are localisable, what is equivalent with localisability of a pair $(\mho, \nu)$, where ν is an arbitrary semi-finite strictly positive measure on caDcb algebra $\mho$ [140]. Thus, the definition of the statistical model which lacks any mathematical ambiguity should use the localisable pairs $(\mho, \nu)$ instead of measure spaces $(\mathcal{X}, \mho(\mathcal{X}), \mu)$. In consequence, we can restrict the considerations to the *maharanisable* caDcb algebras, defined as such caDcb algebras that allow semi-finite strictly positive measures. We will call them *camDcb-algebras*.

Every W^*-algebra is equipped with at least one normal faithful semi-finite weight ψ [62]. Moreover, every commutative W^*-algebra $\mathcal{N}$ is isomorphic to $L_\infty(\mho)$ space for some camDcb-algebra $\mho$. This follows from [136] and the fact that every localisable measure space $(\mathcal{X}, \mho(\mathcal{X}), \mu)$ generates a unique corresponding camDcb-algebra [54, 128] (the strictly positive semi-finite measure μ on $\mho$ can be derived from ψ, see e.g. [148]). But, in face of the functorial association of $L_p(\mho)$ spaces with $\mho$ [54], this selects (up to isomorphism) a unique corresponding $\mho$. In such case, the space $L_1(\mathcal{N})$ of integrals on $\mathcal{N}$ reduces to the space $L_1(\mho)$. As a result, for commutative W^*-algebras $\mathcal{N}$, the quantum information models $\mathcal{M}(\mathcal{N}) \subseteq \mathcal{N}_*^+$ turn into subspaces $\mathcal{M}(\mho) \subseteq L_1(\mho)^+$, which are precisely the well-defined statistical models.

The Falcone–Takesaki theory provides a striking generalisation of the representation independent construction of spaces of integrals to the general non-commutative case, including the generalisation of functorial association of measure-independent $L_p(\mho)$ spaces with $\mho$ to functorial association of weight-independent $L_p(\mathcal{N})$ spaces with $\mathcal{N}$, which is based on the non-commutative flow of weights. In face of such deep structural relationships between commutative and non-commutative integration theory it is quite hard to find convincing arguments in favour of consideration of statistical theory and quantum theory as two separate theories (we do not count here the philosophical prejudices, but only the mathematical structure and the range of its applicability in concrete problems). From this perspective, relying on spectral theory, which represents quantum theoretic models in terms of commutative statistical models via the $L_2(\mathcal{X}, \mho(\mathcal{X}), \mu)$ space, cannot be longer justified. Quantum theoretic models can be quantitatively constructed and analysed as information models on its own right. Note that these conceptual changes are consequences of a single mathematical change: replacement of measure theory by integration theory.

Finally, let us note that we disregard the normalisation of states in favour of finiteness, because we consider the notion of information, quantified by finite positive integral, as more fundamental than the notion of probability, quantified by normalised measure. The quantitative evaluations $\omega \in \mathcal{N}_*^+$ can be used to store and transform information irrespectively of the normalisation. This perspective follows the ideas of Ingarden and Urbanik [74, 75], but is caused also by impossibility of maintaining the relative frequency interpretation for normalised integrals on non-commutative W^*-algebras beyond the type II_1 factors, what was observed already by von Neumann [162, 132]. Yet, without relative frequencies associated on the level of interpretation, the normalisation condition is just irrelevant. Thus, if one decides to consider the non-commutative integration as equally

fundamental as the commutative one, there is just no mathematical or conceptual reason why one should restrict the information theoretic considerations to normalised integrals instead of finite ones.

3   Information deviations

Two crucial properties that can be used to characterise the structures of information geometry (information deviations, metrics, connections) are the monotonicity under coarse graining and the generalised orthogonal decomposition under projection onto convex subspaces (that is, generalised cosine theorem). These properties encode, respectively, the requirements that "the coarse-graining of information should always be indicated by the decrease of relative measure of information", and "the restriction of relative measure of information to subinformation (submodel) should be expressible in terms of its additive decomposition".

In this section we will consider the families of deviations that are defined separately by each of these conditions. The information deviations that satisfy both of these conditions will be considered in Section 4.

3.1   Markov monotone deviations

In the commutative normalised case the coarse graining $T^{\mathbf{c}}$ is defined in terms of the *Markov map* T by $T^{\mathbf{c}} : \mathcal{P} \ni \omega \mapsto \omega \circ T \in \mathcal{P}$ such that, given $\omega : L_\infty(\mho, \mu; \mathbb{C}) \ni f \mapsto \int \mu f \in \mathbb{C}$, $T : L_\infty(\mho, \mu; \mathbb{C}) \to L_\infty(\mho', \mu'; \mathbb{C})$ is a normal (monotonically continuous) positive linear map satisfying $T(1) = 1$. In the non-commutative case the Markov maps are defined by normal unital completely positive linear maps $T : \mathcal{N}' \to \mathcal{N}$ between W^*-algebras $\mathcal{N}, \mathcal{N}'$ [146]. In the finite-dimensional normalised non-commutative case it is known (see e.g. [106, 122]) that there is a duality between Markov maps on W^*-algebras and coarse-grainings defined as normal trace-preserving completely positive maps on the corresponding spaces of density operators. However, this duality depends on a choice of underlying Hilbert space (or some reference state or weight), and the same is true for various infinite-dimensional generalisations of it.

Using the Falcone–Takesaki theory [48] and the results of Petz [117] we can show that this duality holds in the general non-commutative case without any additional assumptions. Consider a pair $(\mathcal{N}, \phi)$, where $\mathcal{N}$ is a von Neumann algebra and $\phi \in \mathcal{N}_*$. Define *contraction* as a morphism of pairs $(\mathcal{N}, \phi) \to (\mathcal{N}', \phi')$ given by the linear map $T : \mathcal{N} \to \mathcal{N}'$ such that $0 \leq x \leq \mathbb{I} \Rightarrow 0 \leq T(x) \leq \mathbb{I}$, $y \in \mathcal{N}^+ \Rightarrow \phi'(T(y)) \leq \phi(y)$, and T is weakly continuous. The pairs $(\mathcal{N}, \phi)$ and contractions between them define a category **Cont**. For any $T \in \mathrm{Mor}(\mathbf{Cont})$ we define a *dual map* of T as such $T^{\mathbf{c}} : \mathcal{N}'_* \to \mathcal{N}_*$ that

$$[\![T(x), \phi']\!]_{\widetilde{\mathcal{N}'}} = [\![x, T^{\mathbf{c}}(\phi')]\!]_{\widetilde{\mathcal{N}}} \quad \forall x \in \mathcal{N} \ \ \forall \phi' \in \mathcal{N}'_*, \tag{17}$$

where $[\![\cdot, \cdot]\!]_{\widetilde{\mathcal{N}}}$ is a Falcone–Takesaki duality (11). The map $T^{\mathbf{c}}$ is a normal positive contraction, hence $(\cdot)^{\mathbf{c}} : \mathbf{Cont} \to \mathbf{Cont}$ is a contravariant functor such that $(\cdot)^{\mathbf{cc}} = \mathrm{id}_{\mathbf{Cont}}$. Moreover, following the proof in [117], it is easy to prove that T is unital and completely positive iff $T^{\mathbf{c}}$ is trace-preserving and completely positive. Hence, we can replace the

discussion in terms of the Markov maps (communication channels) by the discussion in terms of the dual Markov maps (coarse-grainings).

The condition of monotonicity under coarse graining imposed on the information deviation D reads $D(\omega, \phi) \geq D(\omega \circ T, \phi \circ T)$, or equivalently,

$$D(\omega, \phi) \geq D(T^{\mathbf{c}}(\omega), T^{\mathbf{c}}(\phi)) \tag{18}$$

for all $\phi, \omega \in \mathcal{M}$ and for all coarse-grainings $T^{\mathbf{c}}$. As remarked in introduction, in commutative finite-dimensional case (and under some mild auxiliary conditions) (18) characterises [36] the class of Csiszár–Morimoto [34, 35, 104] deviations. Their direct generalisation to general non-commutative case is provided by the *Petz deviations* [119, 120]

$$D_{\mathfrak{f}}(\omega, \phi) := \langle \xi_\omega, \mathfrak{f}(\Delta_{\phi,\omega}) \xi_\omega \rangle_{\mathcal{H}}, \tag{19}$$

where $\mathfrak{f} : [0, \infty[\rightarrow \mathbb{R}$ is a continuous operator monotone function with $\mathfrak{f}(0) \geq 0$ ($\mathfrak{f}$ is called *operator monotone* iff $0 \leq x \leq y \Rightarrow 0 \leq \mathfrak{f}(x) \leq \mathfrak{f}(y)$ for any bounded operators $x, y \in \mathfrak{B}(\mathcal{H})$). The Hilbert space $\mathcal{H}$ is a standard representation of the W^*-algebra $\mathcal{N}$, and ξ_ω is the unique representative of ω in the natural positive cone of the standard representation Hilbert space $\mathcal{H}$. All Petz deviations (19) satisfy (18), but it is not known [126] whether these are the only deviations on $\mathcal{N}_*^+$ which have this property.

3.2 Generalised Bregman deviations

While the monotonicity under coarse graining requires to introduce the framework of Markov maps and the Falcone–Takesaki duality, the generalised orthogonal decomposition requires to introduce the framework of non-smooth variational analysis and the Legendre–Fenchel duality (for an exposition of the latter, see [51, 134, 17, 16]).

The existence of generalised orthogonal decompositions is involved in the representation properties of $\mathcal{M}$. Let L and $L^{\mathbf{d}}$ be two linear (vector) spaces over $\mathbb{R}$ (or $\mathbb{C}$), equipped with a bilinear duality pairing $[\![\cdot, \cdot]\!]_L : L \times L^{\mathbf{d}} \rightarrow \mathbb{R}$ (or $\mathbb{C}$). Let $\Psi : L \rightarrow [-\infty, +\infty]$ and let $\mathrm{dom}\Psi := \{x \in L \mid \Psi(x) < +\infty\}$. The *Fenchel subdifferential* [51, 102, 21] of Ψ at $x \in \mathrm{dom}\Psi$ is defined as a set

$$\partial\Psi(x) := \{y \in L^{\mathbf{d}} \mid \Psi(z) - \Psi(x) \geq [\![z - x, y]\!]_L \,\forall z \in L\}. \tag{20}$$

The *Legendre–Fenchel dual* of Ψ is defined as $\Psi^{\mathbf{L}} : L^{\mathbf{d}} \rightarrow [-\infty, +\infty]$ such that

$$\Psi^{\mathbf{L}}(y) := \sup_{x \in L} \{[\![x, y]\!]_L - \Psi(x)\} \,\forall y \in L^{\mathbf{d}}. \tag{21}$$

One defines $\Psi^{\mathbf{LL}} : L \rightarrow [-\infty, +\infty]$ by $\Psi^{\mathbf{LL}} := (\Psi^{\mathbf{L}})^{\mathbf{L}}$. The functions $\Psi^{\mathbf{L}}$ and $\Psi^{\mathbf{LL}}$ are convex for any Ψ, and $\Psi^{\mathbf{LL}} \leq \Psi$. If $\mathrm{dom}\Psi \neq \emptyset$, then $\Psi^{\mathbf{L}}(x) > -\infty \,\forall x \in L^{\mathbf{d}}$. If $(L, L^{\mathbf{d}})$ are separated locally convex vector spaces, equipped with weak-* and weak topologies, respectively, then $\Psi^{\mathbf{L}}$ is weak-* lower semi-continuous, $\Psi^{\mathbf{LL}}$ is weak lower semi-continuous, and ($\Psi^{\mathbf{LL}} = \Psi$ iff Ψ is weak lower semi-continuous and convex) [20]. In what follows, we will always assume $\mathrm{dom}\Psi \neq \emptyset$. If $\Psi : L \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex and $y \in L^{\mathbf{d}}$, then the *Young–Fenchel inequality*

$$\Psi(x) + \Psi^{\mathbf{L}}(y) - [\![x, y]\!]_L \geq 0 \tag{22}$$

holds, with equality iff $y \in \partial\Psi(x)$ and $x \in \mathrm{dom}\Psi$. If L and $L^{\mathbf{d}}$ are over $\mathbb{C}$, then $[\![x, y]\!]_L$ in (21) and (22) is replaced by re $[\![x, y]\!]_L$.

This allows to define the *Bregman functional* for a convex $\Psi : L \to \mathbb{R} \cup \{+\infty\}$ as a map

$$L \times L^{\mathbf{d}} \ni (x, y) \mapsto D_\Psi(x, y) := \Psi(x) + \Psi^{\mathbf{L}}(y) - [\![x, y]\!]_L \in [0, +\infty], \qquad (23)$$

if L is over $\mathbb{R}$, and with $[\![x, y]\!]_L$ replaced by re $[\![x, y]\!]_L$ if L is over $\mathbb{C}$. By definition, D_Ψ is convex in each variable separately, $D_\Psi(x, y) \geq 0 \ \forall (x, y) \in L \times L^{\mathbf{d}}$, and $D_\Psi(x, y) = 0 \iff y \in \partial\Psi(x)$. We will call Bregman functional D_Ψ *adequate for $V \subseteq L$* iff $\partial\Psi(x) \neq \emptyset \ \forall x \in \mathrm{dom}\Psi \cap V$. There exist various criteria for non-emptiness of Fenchel subdifferential which can be used to ensure adequacy of D_Ψ. In particular, if $(L, L^{\mathbf{d}})$ are Banach spaces, and Ψ is convex and lower semi-continuous, then the Fenchel–Rockafellar theorem [134, 17] states that $\partial\Psi(x) \neq \emptyset \ \forall x \in \mathrm{core}(\mathrm{dom}\Psi)$, where $x \in \mathrm{core}(X)$ for $\emptyset \neq X \subseteq L$ iff $\forall h \in S_L \ \exists \varepsilon > 0 \ \forall t \in [0, \varepsilon] \ \ x + th \in X$, with S_L denoting a unit sphere in L.

Define *dual coordinate system* (or *dual representation*) as a map $(r, s) : \mathcal{M} \times \mathcal{M} \ni (\omega, \phi) \mapsto (r(\omega), s(\phi)) \in L \times L^{\mathbf{d}}$. We define the *generalised Bregman deviation* as a function

$$D_\Psi : \mathcal{M} \times \mathcal{M} \ni (\omega, \phi) \mapsto D_\Psi(\omega, \phi) := D_\Psi(r(\omega), s(\phi)) \in [0, +\infty], \qquad (24)$$

where (r, s) is a dual coordinate system such that $r(\omega) \in \partial\Psi(r(\omega)) \ \forall \omega \in \mathcal{M}$, while D_Ψ is a Bregman functional adequate for $\mathrm{cod}(r)$. By definition, $D_\Psi(\omega, \phi)$ is convex in each variable separately, $D_\Psi(\omega, \phi) \geq 0 \ \forall \omega, \phi \in \mathcal{M}$, and $\omega = \phi \Rightarrow D_\Psi(\omega, \phi) = 0 \ \forall \omega \in \mathcal{M}$. This weakening of the usual property of deviation ($\omega = \phi \iff D(\omega, \phi) = 0$) to one-sided implication is the price paid for defining D_Ψ in terms of a weakly constrained (hence, quite general) variational problem. In order to impose an implication in opposite direction, one would have to require additional conditions that are not natural at this level of generality (they will be discussed below). Note also that due to adequacy of D_Ψ, the condition $s(\omega) \in \partial\Psi(r(\omega))$ is well defined for all $\omega \in \mathcal{M}$. It can be understood either as a condition on allowed dual coordinate systems if Ψ is given, or as a condition on Ψ if (r, s) is given. To summarise, an information deviation $D : \mathcal{M} \times \mathcal{M} \to [0, \infty]$ is a generalised Bregman deviation, denoted D_Ψ, iff there exists a pair of dualised vector spaces $(L, L^{\mathbf{d}}, [\![\cdot, \cdot]\!]_L)$, a pair of functions $r : \mathcal{M} \to L$ and $s : \mathcal{M} \to L^{\mathbf{d}}$ and a convex function $\Psi : L \to \mathbb{R} \cup \{+\infty\}$ satisfying the conditions

$$\partial\Psi(x) \neq \emptyset \ \forall x \in \mathrm{dom}\Psi \cap \mathrm{cod}r, \quad s(\omega) \in \partial\Psi(r(\omega)) \ \forall \omega \in \mathcal{M}, \qquad (25)$$

and such that

$$D(\omega, \phi) = \sup_{x \in L}\{[\![x, s(\phi)]\!]_L - \Psi(x)\} - [\![r(\omega), s(\phi)]\!]_L - \Psi(r(\omega)), \qquad (26)$$

with $[\![\cdot, \cdot]\!]_L$ replaced by re $[\![\cdot, \cdot]\!]_L$ if L is over $\mathbb{C}$. Note that one could also define the functional

$$\hat{D}_\Psi(\omega, \phi) = \sup_{\omega \in \mathcal{M}}\{[\![r(\omega), s(\phi)]\!]_L - \Psi(r(\omega))\} - [\![r(\omega), s(\phi)]\!]_L - \Psi(r(\omega)), \qquad (27)$$

but as long as $\mathrm{cod}(r) \subsetneq \mathrm{dom}\Psi \cap L$, (27) is essentially different from (26), due to non-linear character of the sup operation.

3.3 Dualisers of representation

The generalised Bregman deviation (26) exposes the underlying dualistic and variational properties of the deviation functional. However, often the other representation of Bregman deviation is used, which exposes its geometric properties at the price of the non-trivial restrictions on the domain of duality and convexity. Usually these restrictions are imposed in order to adapt to *presupposed* topological framework and some presupposed form of representation of Bregman deviation, and the goal is to show that such Bregman deviation encodes the Legendre case of the Legendre–Fenchel duality with the dual variable $y \in L^{\mathbf{d}}$ given by some suitably defined notion of derivative (e.g. Fréchet, Gâteaux, one-sided Gâteaux), see e.g. [24, 12, 22, 16] for standard examples in commutative case and [125] for an example in finite-dimensional non-commutative case. Our approach is different, because we do not assume any fixed framework for continuity or smoothness, so we can analyse the general relationship between explicitly dual (generalised) Bregman deviation and its *standard* (hence, restricted) version, which has both arguments represented on the same space. We will construct now a general framework for conversion between these two forms of the Bregman deviation, which is independent of any particular assumptions about continuity or differentiability. The key role in this setting will be played by the (generally, non-linear) *dualiser* function. It is an infinite-dimensional generalisation of the transformation between the dual coordinate systems of Bregman functional, which in the finite-dimensional case is usually encoded in terms of derivative.

For a convex $\Psi : L \to \mathbb{R} \cup \{+\infty\}$ define a function $f_\Psi : L \to L^{\mathbf{d}}$ such that there exists a set $\emptyset \neq V \subseteq L$ satisfying:

(1_f) $f_\Psi : V \to f_\Psi(V)$ is an isomorphism,

(2_f) $\forall y \in V$ $\Psi^{\mathbf{L}}(f_\Psi(y)) - \Psi(y) = [\![y, f_\Psi(y)]\!]_L$,

(3_f) $\forall x \in \mathrm{dom}\Psi \cap V$ $\exists! z \in f_\Psi(V)$ $z \in \partial\Psi(x)$.

Such function f_Ψ will be called a *dualiser*, while the set V will be denoted $\mathrm{adm}f_\Psi$ and called and *admissible domain* of f_Ψ. We will consider f_Ψ only on the set $\mathrm{adm}f_\Psi$, so we define $\mathrm{cod}f_\Psi := f_\Psi(\mathrm{adm}f_\Psi)$. The function Ψ will be called *dualisable* with respect to $(L, L^{\mathbf{d}}, [\![\cdot, \cdot]\!]_L)$ iff there exists at least one dualiser f_Ψ. Note that dualisability of Ψ is a relative property: a change of domain L or a change of a duality structure $[\![\cdot, \cdot]\!]_L$ on L changes the available dualisers. Note also that the condition (3_f) can be equivalently written as

$$\mathrm{cod}f_\Psi \cap \partial\Psi|_{\mathrm{adm}f_\Psi \cap \mathrm{dom}\Psi}(x) = \{*\}, \tag{28}$$

where $\{*\}$ is a singleton. For convenience of notation, till the end of this section we will assume that L is over $\mathbb{C}$ (if it is over $\mathbb{R}$, then one has just to drop 're' everywhere).

If D_Ψ is a Bregman functional and Ψ is dualisable with a dualiser f_Ψ, then the *unbounded standard Bregman functional* is defined as a map

$$\bar{D}_\Psi : L \times \mathrm{adm}f_\Psi \ni (x, y) \mapsto \bar{D}_\Psi(x, y) := \Psi(x) - \Psi(y) - \mathrm{re}\,[\![x - y, f_\Psi(x)]\!]_L \in \mathbb{R} \cup \{+\infty\}, \tag{29}$$

while the *bounded standard Bregman functional* is defined as a map

$$\bar{D}_\Psi : \mathrm{dom}\Psi \times \mathrm{adm}f_\Psi \ni (x, y) \mapsto \bar{D}_\Psi(x, y) := \Psi(x) - \Psi(y) - \mathrm{re}\,[\![x - y, f_\Psi(x)]\!]_L \in \mathbb{R}. \tag{30}$$

By definition, $\bar{D}_\Psi$ satisfies

    i) $\bar{D}_\Psi(x,y) = D_\Psi(x, f_\Psi(y))$,

    ii) $\bar{D}_\Psi(x,y) \geq 0$,

    iii) $\bar{D}_\Psi(x,y) = 0 \iff x = y$,

for all $(x,y) \in L \times \mathrm{adm} f_\Psi$ (or for all $(x,y) \in \mathrm{dom}\Psi \times \mathrm{adm} f_\Psi$, if $\bar{D}_\Psi$ is bounded). The two-sided implication appears here at the price of loss of convexity of $\bar{D}_\Psi$ in the second variable (what is a common problem in standard treatments, see e.g. [13]). This is because (the inverse of) a dualiser f_Ψ may not preserve the convexity properties.

The *(standard) Bregman deviation* is defined correspondingly as

$$\bar{D}_\Psi(\omega,\phi) := \bar{D}_\Psi(r(\omega), f_\Psi^{-1}\circ s(\phi)) = \Psi(r(\omega)) - \Psi(f_\Psi^{-1}\circ s(\phi)) - \mathrm{re}\left[\!\left[r(\omega) - f_\Psi^{-1}\circ s(\phi), s(\phi)\right]\!\right]_L, \tag{31}$$

where $(r,s) : \mathcal{M}\times\mathcal{M} \to L\times L^\mathbf{d}$ is a dual coordinate system such that $s(\omega) \in \partial\Psi(r(\omega)) \ \forall\omega \in \mathcal{M}$. By definition, $\bar{D}_\Psi(\omega,\phi) = D_\Psi(\omega,\phi)$, where D_Ψ is a Bregman functional adequate for $\mathrm{cod}(r)$. It follows that a single generalised Bregman deviation can have several different representations in terms of standard Bregman deviations, depending on the choice of dualiser $f_\Psi : L \supseteq \mathrm{adm} f_\Psi \to \mathrm{cod} f_\Psi \subseteq L^\mathbf{d}$. If $\bar{D}_{\Psi,f_1}$ and $\bar{D}_{\Psi,f_2}$ are two standard Bregman functionals defined from a single generalised Bregman functional D_Ψ by two dualisers f_1 and f_2, then they are equal to each other on $V \subseteq \mathrm{adm} f_1 \cap \mathrm{adm} f_2$ iff there exists a dualiser f_3 of Ψ such that $\mathrm{adm} f_3 = V$. The existence of different dualisers is equivalent to $\partial\Psi(r(\cdot))$ being a non-singleton, non-empty, set-valued function. From the geometric perspective, the choice of a particular dualiser (and an associated representation of generalised Bregman deviation in terms of standard Bregman deviation) corresponds to choice of a particular section of the presheaf $\mathcal{M} \ni \omega \mapsto \partial\Psi(r(\omega)) \subseteq L^\mathbf{d}$.

The definition of bounded standard Bregman functional $\bar{D}_\Psi$ implies that $\bar{D}_\Psi$ satisfies the *generalised cosine equation*

$$\bar{D}_\Psi(r_1,r_2) + \bar{D}_\Psi(r_2,r_3) - \bar{D}_\Psi(r_1,r_3) = \mathrm{re}\left[\!\left[r_1 - r_2, f_\Psi(r_3) - f_\Psi(r_2)\right]\!\right]_L \quad \forall r_1, r_2, r_3 \in \mathrm{adm} f_\Psi \cap \mathrm{dom}\Psi. \tag{32}$$

Under suitable assumptions that guarantee existence and uniqueness of the solution of the corresponding variational problem, this equation can be turned into a theorem on existence and uniqueness of generalised additive decomposition of information deviation under projection onto subspace (submodel).

Let $y \in \mathrm{adm} f_\Psi \cap \mathrm{dom}\Psi$, let $R \subseteq \mathrm{adm} f_\Psi \cap \mathrm{dom}\Psi$ be non-empty and convex, and let $r \in R$. The element $\bar{x} \in R$ will be called the *Bregman projection* of y on R, and denoted by $P_R^\Psi(y)$, iff

$$\bar{x} = \arg\inf_{r\in R}\{\bar{D}_\Psi(r,y)\}. \tag{33}$$

The problem with this definition is that in general case $P_R^\Psi(y)$ might not exist or might be non-unique. The existence and uniqueness of Bregman projection can be guaranteed under various assumptions. In particular, the existence can be guaranteed by means of Bauer's theorem [11] (if L is a locally convex space, R is weakly compact, and $\bar{D}_\Psi$ is weakly lower semi-continuous). On the other hand, if L is a reflexive Banach space, R is closed, $\bar{D}_\Psi$ is lower semi-continuous, strictly convex, and Gâteaux differentiable at y,

with int $\mathrm{dom}\bar{D}_\Psi \neq \emptyset$, $R \cap \mathrm{dom}\bar{D}_\Psi$ and $y \in \mathrm{int\,dom}\bar{D}_\Psi$, then $P_R^\Psi(y)$ is at most a singleton [16]. Unfortunately, we were unable to find the sufficient conditions for uniqueness and existence of the Bregman projections that would be phrased without appealing to particular topological framework. Expressing such conditions in general terms of dualisers would be a valuable result.

Let $(L, L^{\mathbf{d}})$ be a dual pair of separated locally convex spaces, let $R \subseteq \mathrm{adm}f_\Psi \cap \mathrm{dom}\Psi$ be a non-empty convex, affine, weakly closed set, and let $\bar{x} := P_R^\Psi(y)$ be a unique Bregman projection of $y \in \mathrm{adm}f_\Psi \cap \mathrm{dom}\Psi$. Then the *generalised pythagorean theorem*

$$\bar{D}_\Psi(r, \bar{x}) + \bar{D}_\Psi(\bar{x}, y) = \bar{D}_\Psi(r, y) \quad \forall r \in R \tag{34}$$

holds iff the orthogonality condition re $[\![r - \bar{x}, f_\Psi(y) - f_\Psi(\bar{x})]\!]_L = 0$ is satisfied. Such $\bar{x} = P_R^\Psi(y)$ is called a *(dually) orthogonal Bregman projection*. The property (34) generalises the additive decompositions of norm under linear projections on closed convex subsets in the Hilbert space to the class of non-linear Bregman projections on closed convex subsets in the linear space L. Note that the 'orthogonality' of projection is understood in the sense of the linear duality pairing $[\![\cdot, \cdot]\!]_L$, while the non-linearity of projection P_R^Ψ corresponds to the non-linear dualiser f_Ψ.

4 The γ-family of relative entropies

4.1 Families of γ-deviations

By imposing monotonicity under coarse-graining condition on the generalised Bregman deviation (or on a corresponding standard Bregman deviation) one obtains strong restriction on the allowed form of the deviation and corresponding dual coordinate systems (representations). For the finite-dimensional commutative case, Amari [4] has shown that the unique deviation that satisfies these constraints is the Zhu–Rohwer γ-deviation [171, 172, 173]

$$D_\gamma(\omega, \phi) = \int \frac{\gamma\mu_\omega + (1 - \gamma)\nu_\phi - \mu_\omega^\gamma \nu_\phi^{1-\gamma}}{\gamma(1 - \gamma)}, \tag{35}$$

where μ_ω and ν_ϕ are the finite positive measures corresponding to the Daniell–Stone integrals ω and ϕ, while $\mu_\omega^\gamma \nu_\phi^{-\gamma} = (\frac{\mu_\omega}{\nu_\phi})^\gamma$ is just the γ-th power of the Radon–Nikodým derivative $\frac{\mu_\omega}{\nu_\phi}$. The most general deviation satisfying both constraints in the non-commutative case that has been known so far is the Jenčová–Ojima γ-deviation [81, 84, 112]

$$D_\gamma(\omega, \phi) = \frac{\gamma\omega(\mathbb{I}) + (1 - \gamma)\phi(\mathbb{I}) - \mathrm{re}\,[\![u\Delta_{\omega,\psi}^\gamma, v\Delta_{\phi,\psi}^{1-\gamma}]\!]_\psi}{\gamma(1 - \gamma)}, \tag{36}$$

where $\psi \in \mathcal{W}_0(\mathcal{N})$ is an arbitrary reference functional, $[\![\cdot, \cdot]\!]_\psi$ is the Banach space duality pairing between the Araki–Masuda non-commutative $L_{1/\gamma}(\mathcal{N}, \psi)$ and $L_{1/(1-\gamma)}(\mathcal{N}, \psi)$ spaces, $\Delta_{\phi,\psi}$ is a relative modular operator, while u (resp., v) is an operator arising from the polar decomposition of ω (resp., ϕ), $\omega(xu) = [\![u\Delta_{\omega,\psi}, x^{\mathbf{d}}]\!]_\psi \,\forall x \in \mathcal{N}$, where $x^{\mathbf{d}}$ denotes a dual of x with respect to the Araki–Masuda duality $[\![\cdot, \cdot]\!]_\psi$.

The deviations given by (35) and (36) are well defined for $\gamma \in \mathbb{R} \setminus \{0, 1\}$, however the characterisation results include also the boundary case $\gamma \in \{0, 1\}$, for which the corresponding deviations are derived from D_γ by passing to the limit with γ under integral sign. If the normalisation condition is imposed, then the uniqueness results are stronger, selecting precisely the boundary functionals: the Kullback–Leibler [95, 94] deviation

$$D_0(\omega, \phi) = D_1(\phi, \omega) = \int \nu_\phi \log \frac{\nu_\phi}{\mu_\omega} \tag{37}$$

in the finite-dimensional commutative case [37], and the Umegaki–Araki [157, 7, 8] deviation

$$D_0(\omega, \phi) = D_1(\phi, \omega) = \mathrm{tr}(\rho_\phi(\log \rho_\phi - \log \rho_\omega)) = \mathrm{tr}\left(\rho_\phi^{1/2}(\log \Delta_{\phi,\omega})\rho_\phi^{1/2}\right) \tag{38}$$

in the finite-dimensional non-commutative case [125].

Despite these similarities, the Jenčová–Ojima deviation is not a canonical non-commutative generalisation of the Zhu–Rohwer deviation. The construction of the former is dependent on the choice of fixed reference weight ψ, while the latter does not depend on any additional measure. (Nevertheless, the values taken by the Jenčová–Ojima deviation are independent of the choice of reference ψ.) In what follows, we will use the Falcone–Takesaki theory in order to introduce a new family of information deviations that is a proper non-commutative generalisation of the Zhu–Rohwer deviations and is also a proper generalisation of the Jenčová–Ojima deviations.

4.2 General form of family of γ-deviations

Consider the γ-*embedding* (γ-*coordinate*) functions on $\mathcal{N}_*^+$ valued in $L_{1/\gamma}(\mathcal{N})$ spaces:

$$\ell_\gamma : \mathcal{N}_*^+ \ni \omega \mapsto \ell_\gamma(\omega) := \frac{\omega^\gamma}{\gamma} \in L_{1/\gamma}(\mathcal{N}), \tag{39}$$

with $\gamma \in]0, 1]$. By definition, the embeddings ℓ_γ and $\ell_{1-\gamma}$ encode the duality (14),

$$\mathcal{N}_*^+ \times \mathcal{N}_*^+ \ni (\omega, \phi) \mapsto \int \ell_\gamma(\omega)\ell_{1-\gamma}(\phi) \in \mathbb{C}. \tag{40}$$

The special case of these maps were introduced by Nagaoka and Amari [107] as coordinate systems in normalised commutative finite-dimensional setting, and since then they became a standard tool of information geometry theory. However, the Nagaoka–Amari formulation, as well as all its later applications, including the non-commutative generalisations, is based on using the γ-powers of densities (Radon–Nikodým derivatives or Pedersen–Takesaki operator densities) with respect to a fixed reference measure, state, or weight. An important attempt to circumvent this problem in the commutative case was made by Zhu [170, 173], who considered the spaces of measures constructed through an equivalence relation based on γ-powers of Radon–Nikodým derivatives, but without fixing any particular reference measure. However, his work remained unfinished and widely unknown, and it covered only the commutative case. The embeddings (39) solve these problems.

Due to the Falcone–Takesaki theory we are able to make the reference-independent approach strict and valid in all cases, including the infinite-dimensional non-commutative one. Let us define [91] the general form of *quantum γ-deviation* on $\mathcal{N}_*^+$, $D_\gamma : \mathcal{N}_*^+ \times \mathcal{N}_*^+ \ni (\omega, \phi) \mapsto D_\gamma(\omega, \phi) \in \mathbb{R}$, by

$$D_\gamma(\omega, \phi) := \int \left(\frac{\omega}{1-\gamma} + \frac{\phi}{\gamma} - \mathrm{re}\left(\ell_\gamma(\omega)\ell_{1-\gamma}(\phi) \right) \right) = \mathrm{re} \int \left(\frac{\gamma\omega + (1-\gamma)\phi - \omega^\gamma\phi^{1-\gamma}}{\gamma(1-\gamma)} \right),$$
(41)

where $\gamma \in]0, 1[$. This definition can be extended to include the boundary values $\gamma \in \{0, 1\}$, by

$$D_1(\phi, \omega) := D_0(\omega, \phi) := \int \lim_{\gamma \to 0} \left(\frac{\omega}{1-\gamma} + \frac{\phi}{\gamma} - \mathrm{re}\left(\ell_\gamma(\omega)\ell_{1-\gamma}(\phi) \right) \right).$$
(42)

The deviation D_γ reduces to the following special cases:

- the Jenčová–Ojima γ-deviation (36) when the choice of a reference state $\psi \in \mathcal{N}_{*0}^+$, or a reference weight $\psi \in \mathcal{W}_0(\mathcal{N})$, for the isometric isomorphism with the Araki–Masuda $L_p(\mathcal{N}, \psi)$ space is provided,

- the Umegaki–Araki deviation (38) in the representation-independent Petz's form [118]

$$D_0(\omega, \phi) = D_1(\phi, \omega) = i \lim_{t \to 0} \frac{\phi}{t} \left(\left(\frac{\mathrm{D}\omega}{\mathrm{D}\phi} \right)_t - \mathbb{I} \right),$$
(43)

for $\|\phi\| = \|\omega\| = 1$, $\mathrm{supp}(\omega) \geq \mathrm{supp}(\phi)$, and $\gamma \to 0$ limit, where $\mathrm{supp}(\omega)$ denotes the support projection of ω,

- the Hasegawa γ-deviation [68]

$$D_\gamma(\omega, \phi) = \frac{\psi(\rho_\omega - \rho_\omega^\gamma \rho_\phi^{1-\gamma})}{\gamma(1-\gamma)},$$
(44)

for semi-finite $\mathcal{N}$, and ρ_ϕ and ρ_ω defined as, respectively, the Pedersen–Takesaki densities of normalised states ϕ and ω with respect to a faithful normal semi-finite trace ψ that acts on a standard representation of $\mathcal{N} \cong L_\infty(\mathcal{N})$ on $\mathcal{H} \cong L_2(\mathcal{N})$. In this case the Pedersen–Takesaki densities reduce to the Dye–Segal densities [43, 141], hence $\phi(\cdot) = \psi(\rho_\phi \cdot)$ and $\omega(\cdot) = \psi(\rho_\omega \cdot)$. If the standard representation of $\mathcal{N}$ on $\mathcal{H}$ is isomorphic to $\mathfrak{B}(\mathcal{H})$ as a von Neumann algebra and trace ψ is normalised ($\psi(\mathbb{I}) = 1$), then $\psi(\cdot) = \mathrm{tr}(\cdot)$, where tr is a standard trace on $\mathfrak{B}(\mathcal{H})$,

- the Zhu–Rohwer γ-deviation (35) for commutative $\mathcal{N}$, and ν_ϕ absolutely continuous with respect to μ_ω,

- the Kullback–Leibler deviation (37) for commutative $\mathcal{N}$, ν_ϕ absolutely continuous with respect to μ_ω, $\int \mu_\omega = 1 = \int \nu_\phi$, and $\gamma \to 0$ limit,

- the Cressie–Read–Amari γ-deviation[3] [32, 3, 100, 131]

$$D_\gamma(\omega, \phi) = \frac{1}{\gamma(1-\gamma)} \int \upsilon(p_\omega - p_\omega^\gamma q_\phi^{1-\gamma}),$$
(45)

---

[3]The γ-deviation family (45) corresponds bijectively, but is not equal, to the γ-deviation families of Chernoff [27], Rényi [133], Pérez [116], Havrda–Chárvat [71], and Tsallis [155] (for a review with calculations, see e.g. [33]).

for semi-finite commutative $\mathcal{N}$, and normalised measures ν_ϕ and μ_ω ($\int \mu_\omega = 1 = \int \nu_\phi$) that are absolutely continuous with respect to a strictly positive semi-finite measure υ on the countably additive Dedekind complete boolean algebra $\mho$ derived from $\mathcal{N}$. The functions q_ϕ and p_ω are defined, respectively, as the Radon–Nikodým derivatives of ν_ϕ and μ_ω with respect to υ: $\phi(\cdot) = \int \nu_\phi(\cdot) = \int \upsilon(q_\phi \cdot)$, and $\omega(\cdot) = \int \mu_\omega(\cdot) = \int \upsilon(p_\omega \cdot)$.

It is quite remarkable that the mathematical form of γ-deviation given by (41) resembles so strong similarity with its commutative special case (35).

4.3 Properties of γ-deviations

Let us turn to the discussion of the properties of the family (41). A direct calculation shows that (41) is a generalised Bregman deviation (26) with

$$\Psi_\gamma(x) = \frac{1}{1-\gamma} \int (\gamma x)^{1/\gamma} = \frac{1}{1-\gamma} \|\gamma x\|_{1/\gamma}^{1/\gamma}, \tag{46}$$

and $\Psi_\gamma(\ell_\gamma(\omega)) = \frac{1}{1-\gamma}\omega(\mathbb{I})$. This deviation has a dualiser $f_{\Psi_\gamma}(x) = \ell_{1-\gamma} \circ \ell_\gamma^{-1}(x)$, which is a homeomorphism $f_{\Psi_\gamma} : L_{1/\gamma}(\mathcal{N}) \to L_{1/(1-\gamma)}(\mathcal{N})$. Thus, D_γ can be written in terms of the generalised Bregman deviation (26), which takes the form

$$D_{\Psi_\gamma}(\omega, \phi) = \Psi_\gamma(\ell_\gamma(\omega)) + \Psi_{1-\gamma}(\ell_{1-\gamma}(\phi)) - \mathrm{re} \, [\![\ell_\gamma(\omega), \ell_{1-\gamma}(\phi)]\!]_{\widetilde{\mathcal{N}}}. \tag{47}$$

The straightforward calculation based on the definitions (41) and (39) shows that D_γ satisfies the generalised cosine equation

$$D_\gamma(\omega, \phi) + D_\gamma(\phi, \psi) = D_\gamma(\omega, \psi) + \mathrm{re} \int (\ell_\gamma(\omega) - \ell_\gamma(\phi))(\ell_{1-\gamma}(\psi) - \ell_{1-\gamma}(\phi)), \tag{48}$$

as well as

$$D_\gamma(\omega, \phi) = D_{1-\gamma}(\phi, \omega). \tag{49}$$

The equation (48) turns into the 'ordinary' generalised cosine equation (32) for the corresponding standard Bregman deviation (47), while the equation (49) turns into the 'representation-index duality' equation

$$\bar{D}_{\Psi_\gamma}(y, x) = \bar{D}_{\Psi_{1-\gamma}}(f_{\Psi_\gamma}(x), f_{\Psi_\gamma}(y)), \tag{50}$$

where $x, y \in L_{1/\gamma}(\mathcal{N})$. The finite-dimensional commutative version of the equation (50), with a dualiser given by derivative, was discussed in [168].

Furthermore, from the uniform convexity and the uniform smoothness of $L_{1/\gamma}(\mathcal{N})$ spaces it follows that the norm $\|\cdot\|_{1/\gamma}$ is Fréchet differentiable at all points $x \in L_{1/\gamma}(\mathcal{N})$ except $x = 0$. The Fréchet derivative of the norm allows to calculate the corresponding Fréchet derivative of the function $\Psi_\gamma(x)$ in the direction y, which reads

$$\mathfrak{D}_F(\Psi_\gamma(x))(y) = \mathrm{re} \, [\![y, f_{\Psi_\gamma}(x)]\!]_{\widetilde{\mathcal{N}}} = \mathrm{re} \, [\![y, \ell_{1-\gamma} \circ \ell_\gamma^{-1}(x)]\!]_{\widetilde{\mathcal{N}}}. \tag{51}$$

For $\gamma < 1$ the function $\Psi_\gamma(x)$ is Fréchet differentiable also at $x = 0$, with

$$\mathfrak{D}_F(\Psi_\gamma(0))(y) = \mathrm{re} \, [\![y, f_\Psi(0)]\!]_{\widetilde{\mathcal{N}}} = 0. \tag{52}$$

This gives a differential formula for the standard Bregman functional $\bar{D}_{\Psi_\gamma}$,

$$\bar{D}_{\Psi_\gamma}(y,x) = \Psi_\gamma(y) + \Psi_{1-\gamma}(\ell_{1-\gamma} \circ \ell_\gamma^{-1}(x)) - \mathfrak{D}_F(\Psi_\gamma(x))(y). \tag{53}$$

The uniform smoothness and uniform convexity of $L_p(\mathcal{N})$ allows to strictly follow Jenčová's proof of Proposition 6.1.(i)-(ii) in [84], what results in a proof that, for every $\phi \in \mathcal{N}_*^+$,

1) $D_\gamma(\omega, \phi) \geq 0$,

2) $D_\gamma(\omega, \phi) = 0 \iff \omega = \phi$.

As stated above, for any choice of a reference weight $\psi \in \mathcal{W}_0(\mathcal{N})$, D_γ turns into the Jenčová–Ojima deviation. The latter is in turn the particular case of Petz's deviation (19) with the operator convex function $\mathfrak{f}$ given by

$$\mathfrak{f}_\gamma(t) = \frac{1}{\gamma} + \frac{1}{1-\gamma}t - \frac{1}{\gamma(1-\gamma)}t^\gamma, \tag{54}$$

see [84] for a proof. The identification of D_γ as a member of Petz's family with the function (54) can be also obtained more directly, without any reference to the Araki–Masuda $L_p(\mathcal{N}, \psi)$ spaces. This is provided using the identification (13) of $L_2(\mathcal{N})$ with the standard representation Hilbert space $\mathcal{H}$, which gives

$$\int \omega^\gamma \phi^{1-\gamma} = \left\langle \xi_\phi, \left(\frac{\mathrm{D}\omega}{\mathrm{D}\phi} \right)_{-i\gamma} \xi_\phi \right\rangle_{\mathcal{H}} = \langle \xi_\phi, \Delta_{\omega,\phi}^\gamma \xi_\phi \rangle_{\mathcal{H}}, \tag{55}$$

where ξ_ϕ is a representative of ϕ in the standard cone $L_2(\mathcal{N})^+$ of the standard representation Hilbert space $\mathcal{H}$. The last equality follows from the fact that $\left(\frac{\mathrm{D}\omega}{\mathrm{D}\phi} \right)_z \xi_\phi$ agrees with $\Delta_{\omega,\phi}^{iz} \xi_\phi$ for $z \in \mathbb{R}$ [151]. From the fact that D_γ is a Petz deviation it follows that it satisfies the following properties [119, 84]:

1) $D_\gamma(\omega, \phi) \geq D_\gamma(T^{\mathbf{c}} \circ \omega, T^{\mathbf{c}} \circ \omega)$,

2) D_γ is jointly convex on $\mathcal{N}_*^+ \times \mathcal{N}_*^+$,

3) D_γ is lower semi-continuous on $\mathcal{N}_*^+ \times \mathcal{N}_{*0}^+$ endowed with the product of norm topologies.

Using the lower semi-continuity and convexity properties of D_γ together with the fact that $L_{1/\gamma}(\mathcal{N})$ spaces are Banach spaces, one can take the advantage of the theorems on existence and uniqueness of the solutions of the variational problems for the convex lower semi-continuous functionals in the Banach space setting. In particular, if $\psi \in \mathcal{N}_*^+$, $C \subseteq \mathcal{N}_*^+$, $\phi \in C$, and $\ell_\gamma(C)$ is a convex set, then the following conditions are equivalent:

i) $D_\gamma(\phi, \psi) = \inf_{\varphi \in C}\{D_\gamma(\varphi, \psi)\}$,

ii) $D_\gamma(\omega, \psi) \geq D_\gamma(\phi, \psi) + D_{1-\gamma}(\phi, \omega)$.

20

Moreover, if such $P_C^{\Psi^\gamma}(\psi) = \phi$ exists, then it is unique. This condition is satisfied if C is closed in the topology induced by the γ-embedding of the weak topology of $\mathcal{N}_*^+$ in $L_p(\mathcal{N})$. The explicit proof of these statements can be obtained by repeating the proofs of Propositions 7.1, 8.1 and 8.2 of [84] with the obvious replacement of $L_{1/\gamma}(\mathcal{N}, \psi)$ by $L_{1/\gamma}(\mathcal{N})$.

The γ-family of deviation functionals defined by (41) provides a proper infinite-dimensional non-commutative generalisation of the γ-families of Zhu–Rohwer deviations and Jenčová–Ojima deviations that is defined directly in terms of canonical non-commutative $L_{1/\gamma}(\mathcal{N})$ spaces and belongs to an intersection of the generalised Bregman deviations and Petz's deviations. The canonical character of the structures employed in the construction of D_γ, and the recent result of Amari [4] for Zhu–Rohwer deviation in finite-dimensional commutative case, leads us to state the following:

Conjecture ♣. The family $D_\gamma(\omega, \phi)$, $\gamma \in [0, 1]$, defined by (41) and (42) is the unique family of deviation functionals on $\mathcal{N}_*^+$ satisfying (18) and (26).

Further properties of D_γ family, the issue of selection of a particular $\gamma \in [0, 1]$, as well as the extension of other structures of quantum information geometry in infinite-dimensional non-commutative case based on Falcone–Takesaki theory will be the main topic of the next paper [90].

As a final remark, let us note that for $\gamma = 1/2$ the $L_{1/\gamma}(\mathcal{N})$ space turns into a Hilbert space $\mathcal{H}$, while the standard Bregman functional $\bar{D}_\gamma$ associated with D_γ turns into the norm distance

$$\bar{D}_{\Psi_{1/2}}(x, y) = \frac{1}{2}\|x - y\|_\mathcal{H}^2. \tag{56}$$

Hence, the Bregman projection of information geometric theory based on γ-deviations (41) turns in this particular case to the Hilbert space norm minimisation. This shows that the framework of Hilbert space geometry is just a special case of much more general framework of information geometry of spaces of non-commutative integrals. From this perspective, the traditional application of spectral theory to quantitative construction and analysis of quantum theoretic models is just a special case of relative entropic modelling. In consequence [91, 90], it is possible to consider information geometry of spaces of non-commutative integrals as a new framework for kinematics of quantum information theory (in particular) and quantum theory (in general), replacing the traditional and semi-spectral frameworks based on Hilbert spaces and spectral theory [161, 39] as well as the algebraic frameworks based on algebras of operators and spectral theory [139, 46, 61, 31].

References

[1] Aczél J., Daróczy Z., 1975, *On measures of information and their characterizations*, Academic Press, New York.

[2] Amari S.-i., 1980, *Theory of information spaces – a differential–geometrical foundation of statistics*, Post Res. Assoc. Appl. Geom. Mem. Report **106**.

[3] Amari S.-i., 1985, *Differential–geometrical methods in statistics*, Lecture Notes in Statistics **28**, Springer, Berlin.

[4] Amari S.-i., 2009, *α-divergence is unique, belonging to both f-divergence and Bregman divergence classes*, IEEE Trans. Inf. Theor. **55**, 4925.

[5] Amari S.-i., Nagaoka H., 1993, *Joho kika no hoho*, Iwanami Shoten, Tōkyō (engl. transl. rev. ed.: 2000, *Methods of information geometry*, American Mathematical Society, Providence).

[6] Araki H., 1973, *Relative hamiltonian for faithful normal states of von Neumann algebra*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **9**, 165. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=9&iss=1&rank=9.

[7] Araki H., 1976, *Relative entropy for states of von Neumann algebras I*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **11**, 809. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=11&iss=3&rank=9.

[8] Araki H., 1977, *Relative entropy for states of von Neumann algebras II*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **13**, 173. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=13&iss=1&rank=8.

[9] Araki H., 1993, *Ryoshiba no suri*, Iwanami Shoten, Tōkyō (engl. transl. 1999, *Mathematical theory of quantum fields*, Int. Ser. Monogr. Phys. **101**, Oxford University Press, Oxford).

[10] Araki H., Masuda T., 1982, *Positive cones and L_p-spaces for von Neumann algebras*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **18**, 339. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=18&iss=2&rank=11.

[11] Bauer H., 1958, *Minimalstellen von Funktionen und Extremalpunkte*, Arch. Math. **9**, 389.

[12] Bauschke H.H., Borwein J.M., 1997, *Legendre functions and the method of random Bregman projections*, J. Conv. Anal. **4**, 27. Available at: www.emis.de/journals/JCA/vol.4_no.1/j86.ps.gz.

[13] Bauschke H.H., Borwein J.M., 2001, *Joint and separate convexity of the Bregman distance*, Stud. Comp. Math. **8**, 23. Available at: people.ok.ubc.ca/bauschke/Research/c04.ps.

[14] Bhattacharyya A., 1943, *On a measure of divergence between two statistical populations defined by their probability distributions*, Bull. Calcutta Math. Soc. **35**, 99.

[15] Bogolyubov N.N., 1961, *Kvazisrednie v zadachakh statisticheskoĭ mekhaniki*, D-**871**, Ob'yedinyennyĭ Institut Yadernykh Issledovaniĭ, Laboratoriya Teoreticheskoĭ Fiziki, Dubna (germ. transl. 1962, *Quasimittelwerte in Problemen der Statistischen Mechanik I, II*, Physik. Abhandl. Sowjetunion **6**, 1, 229).

[16] Borwein J.M., Vanderwerff J.D., 2010, *Convex functions: constructions, characterizations and counterexamples*, Cambridge Universityu Press, Cambridge.

[17] Borwein J.M., Zhu Q.J., 2005, *Techniques of variational analysis*, Springer, Berlin.

[18] Bratelli O., Robinson D.W., 1979, 1981, *Operator algebras and quantum statistical mechanics*, Vol.1-2, Springer, Berlin.

[19] Bregman L.M., 1967, *Relaksatsionnyĭ metod nakhozdheniya obshcheĭ tochki vypuklykh mnozhestv i ego primenenie dla zadach vypuklogo programmirovaniya*, Zh. vychestl. matem. matem. fiz. **7**, 620. (engl. transl.: 1967, *The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys. **7**, 200.).

[20] Brønsted A., 1964, *Conjugate convex functions in topological vector spaces*, Mat. Fys. Medd. Dansk. Vid. Selsk. **34**, 1.

[21] Brønsted A., Rockafellar R.T., 1965, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc. **16**, 605. Available at: www.math.washington.edu/∼rtr/papers/rtr-Subdifferentiability.pdf.

[22] Butnariu D., Resmerita E., 2006, *Bregman distances, totally convex functions, and a method for solving operator equations in Banach spaces*, Abstr. Appl. Anal. **2006**, 84919. Available at: www.emis.de/journals/HOA/AAA/Volume2006/84919.pdf.

[23] Caticha A., 2004, *Relative entropy and inductive inference*, in: Erickson G., Zhai Y. (eds.) *Bayesian inference and maximum entropy methods in science and engineering*, AIP Conf. Proc. **707**, 75. arXiv:physics/0311093.

[24] Censor Y., Lent A., 1981, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl. **34**, 321.

[25] Chentsov N.N., 1969, *Obshchaya teoriya statisticheskogo vyvoda*, Mat. Zametki **5**, 635 (engl. transl. 1969, *General theory of statistical decisions*, Math. Notes **5**, 380).

[26] Chentsov N.N., 1972, *Statisticheskie reshayushchie pravila i optimal'nye vyvody*, Nauka, Moskva (engl. transl.: 1982, *Statistical decision rules and optimal inference*, American Mathematical Society, Providence).

[27] Chernoff H., 1952, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. **23**, 493. Available at: omega.albany.edu:8008/Classics/Chernoff52.pdf.

[28] Connes A., 1973, *Une classification des facteurs de type III*, Ann. Sci. École Norm. Sup. 4 ème sér. **6**, 133. numdam:ASENS_1973_4_6_2_133_0.

[29] Connes A., 1974, *Characterization des espaces vectoriels ordonnés sous-jacents aux algèbres de von Neumann*, Ann. Inst. Fourier Grenoble **24**, 121.

[30] Connes A., 1980, *On a spatial theory of von Neumann algebras*, J. Func. Anal. **35**, 153.

[31] Connes A., 1994, *Noncommutative geometry*, Academic Press, Boston.

[32] Cressie N., Read T.R.C., 1984, *Multinomial goodness-of-fit tests*, J. Roy. Statist. Soc. B **46**, 440.

[33] Crooks G.E., Sivak D.A., 2011, *Measures of trajectory ensemble disparity in nonequilibrium statistical dynamics*, J. Stat. Mech. (to appear). Available at: threeplusone.com/pubs/technote/CrooksTechNote009-TrajDiv.pdf.

[34] Csiszár I., 1963, *Eine informationstheorische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **8**, 85.

[35] Csiszár I., 1967, *Information-type measures of difference of probability distributions and indirect observation*, Stud.

Sci. Math. Hungar. **2**, 229.

[36] Csiszár I., 1978, *Information measures: a critical survey*, in: Kozesnik J. (ed.), *Transactions of the seventh Prague conference on information theory, statistical decision functions and random processes*, p.73.

[37] Csiszár I., 1991, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Stat. **19**, 2032.

[38] Csiszár I., 2008, *Axiomatic characterizations of information measures*, Entropy **10**, 261. Available at: www.mdpi.com/1099-4300/10/3/261/pdf.

[39] de Muynck W.M., 2002, *Foundations of quantum mechanics: an empiricist approach*, Kluwer, Dordrecht.

[40] Digernes T., 1975, *Duality for weights on covariant systems and its applications*, Ph.D. thesis, University of California, Los Angeles.

[41] Dixmier J., 1953, *Formes linéaires sur un anneau d'opérateurs*, Bull. Soc. Math. France **81**, 9.

[42] Donald M.J., 1986, *On the relative entropy*, Commun. Math. Phys. **105**, 13. euclid:1104115254.

[43] Dye H.A., 1952, *The Radon–Nikodym theorem for finite rings of operators*, Trans. Amer. Math. Soc. **72**, 243.

[44] Eguchi S., 1983, *Second order efficiency of minimum contrast estimators in a curved exponential family*, Ann. Statist. **11**, 793. euclid:1176346246.

[45] Eguchi S., 1985, *A differential geometric approach to statistical inference on the basis of contrast functionals*, Hiroshima Math. J. **15**, 341. euclid:1206130775.

[46] Emch G.G., 1972, *Algebraic methods in statistical mechanics and quantum field theory*, Wiley, New York.

[47] Falcone T., Takesaki M., 1999, *Operator valued weights without structure theory*, Trans. Amer. Math. Soc. **351**, 323. Available at: www.ams.org/journals/tran/1999-351-01/S0002-9947-99-02028-0/S0002-9947-99-02028-0.pdf.

[48] Falcone T., Takesaki M., 2001, *The non-commutative flow of weights on a von Neumann algebra*, J. Funct. Anal. **182**, 170. Available at: www.math.ucla.edu/~mt/papers/QFlow-Final.tex.pdf.

[49] Fell J.M.G., 1969, *An extension of Mackey's method to Banach *-algebraic bundles*, Memoirs of the American Mathematical Society **90**, American Mathematical Society, Providence.

[50] Fell J.M.G., 1977, *Induced representations and Banach ∗-algebraic bundles*, LNM **582**, Springer, Berlin.

[51] Fenchel W., 1949, *On conjugate convex functions*, Canadian J. Math. **1**, 73. Available at: math.ca/cjm/v1/cjm1949v01.0073-0077.pdf.

[52] Fisher R.A., 1922, *On the mathematical foundation of theoretical statistics*, Phil. Trans. Roy. Soc. London Ser. A **222**, 309.

[53] Fisher R.A., 1925, *Theory of statistical estimation*, Proc. Cambridge Phil. Soc. **22**, 700.

[54] Fremlin D.H., 2000, 2001, 2002, 2003, *Measure theory*, Vol.1-4, Torres Fremlin, Colchester.

[55] Gibilisco P., Isola T., 2001, *A characterisation of Wigner–Yanase skew information among statistically monotone metrics*, Infin. Dim. Anal. Quant. Prob. Relat. Topics **4**, 553. Available at: www.mat.uniroma2.it/~isola/research/preprints/GiIs03.pdf.

[56] Gibilisco P., Isola T., 2003, *Wigner–Yanase information on quantum state space: the geometric approach*, J. Math. Phys. **44**, 3752. Available at: /www.mat.uniroma2.it/~isola/research/preprints/GiIs05.pdf.

[57] Gibilisco P., Isola T., 2004, *On the characterisation of paired monotone metrics*, Ann. Inst. Stat. Math. **56**, 369. arXiv:math/0303059.

[58] Grasselli M.R., 2004, *Duality, monotonicity and the Wigner–Yanase–Dyson metrics*, Inf. Dim. Anal. Quant. Prob. Relat. Top. **7**, 215. arXiv:math-ph/0212022.

[59] Grasselli M.R., Streater R.F., 2001, *On the uniqueness of the Chentsov metric in quantum information geometry*, Inf. Dim. Anal. Quant. Prob. Rel. Top. **4**, 173. arXiv:math-ph/0006030.

[60] Grasselli M.R., Streater R.F., 2001, *The uniqueness of the Chentsov metric*, in: Sollich P., Hughston L.P., Streater R.F., Coolen A.C.C. (eds.), *Disordered and complex systems (King's College, London, UK, 10-14 July 2000)*, AIP Conf. Proc. **553**, AIP, Melville, p.165. Available at: www.math.mcmaster.ca/~grasselli/bkmshort.pdf.

[61] Haag R., 1992, *Local quantum physics*, Springer, Berlin. (1996, 2nd rev. enlarg. ed.).

[62] Haagerup U., 1975, *Normal weights on W^*-algebras*, J. Funct. Anal. **19**, 302.

[63] Haagerup U., 1975, *The standard form of von Neumann algebras*, Math. Scand. **37**, 271.

[64] Haagerup U., 1979, *L^p-spaces associated with an aribitrary von Neumann algebra*, in: *Algèbres d'opérateurs et leurs applications en physique mathématique (Proc. Colloques Internationaux, Marseille 20-24 juin 1977)*, Colloques Internationaux C.N.R.S. **274**, Éditions du C.N.R.S., Paris, p.175. Available at: math.berkeley.edu/~pavlov/scans/haagerup.pdf.

[65] Haagerup U., 1979, *Operator valued weights in von Neumann algebras I*, J. Funct. Anal. **32**, 175.

[66] Haagerup U., 1979, *Operator valued weights in von Neumann algebras II*, J. Funct. Anal. **33**, 339.

[67] Hansen F., 2006, *Characterization of symmetric monotone metrics on the state space of quantum systems*, Quant. Inf. Comput. **6**, 597. arXiv:math-ph/0601056.

[68] Hasegawa H., 1993, *α-divergence of the non-commutative information geometry*, Rep. Math. Phys. **33**, 87.

[69] Hasegawa H., 2003, *Dual geometry of the Wigner–Yanase–Dyson information content*, Inf. Dim. Anal. Quant. Prob. Relat. Top. **6**, 413.

[70] Hasegawa H., Petz D., 1997, *Non-commutative extension of information geometry II*, in: Hirota O. et al. (eds.), *Quantum communication, computing and measurements*, Plenum, New York, p.109. Available at: www.renyi.hu/~petz/pdf/75.pdf.

[71] Havrda J., Chárvat F., 1967, *Quantification method of classification processes. The concept of structural α-entropy*, Kybernetika **3**, 30. Available at: www.kybernetika.cz/content/1967/1/30/paper.pdf.

[72] Hilsum M., 1981, *Les espaces L^p d'une algèbre de von Neumann définies par la dérivée spatiale*, J. Funct. Anal. **40**, 151.

[73] Ingarden R.S., Janyszek H., Kossakowski A., Kawaguchi T., 1982, *Information geometry of quantum statistical systems*, Tensor (N.S.) **37**, 105.

[74] Ingarden R.S., Urbanik K., 1961, *Information as a fundamental notion of statistical physics*, Bull. Acad. Polon. Sci.,

Sér. Sci. Math. Astronom. Phys. **9**, 313.

[75] Ingarden R.S., Urbanik K., 1962, *Information without probability*, Colloq. Math. **9**, 131.

[76] Izumi H., 1996, *Non-commutative L^p-spaces*, Sūrikaisekikenkyūsho Kōkyūroku **956**, 24. Available at: www.kurims.kyoto-u.ac.jp/∼kyodo/kokyuroku/contents/pdf/0956-3.pdf.

[77] Izumi H., 1997, *Construction of non-commutative L^p-spaces with a complex parameter arising from modular actions*, Intern. J. Math. **8**, 1029.

[78] Izumi H., 1998, *Natural bilinear forms, natural sesquilinear forms and the associated duality on non-commutative L^p-spaces*, Intern. J. Math. **9**, 975.

[79] Jeffreys H., 1946, *An invariant form for the prior probability in estimation problems*, Proc. Roy. Soc. London A **186**, 453.

[80] Jenčová A., 2001, *Geometry of quantum states: dual connections and divergence functions*, Rep. Math. Phys. **47**, 121. Available at: www.mat.savba.sk/∼jencova/geom.pdf.

[81] Jenčová A., 2003, *Affine connections, duality and divergences for a von Neumann algebra*, arXiv:math-ph/0311004.

[82] Jenčová A., 2003, *Flat connections and Wigner–Yanase–Dyson metrics*, Rep. Math. Phys. **52**, 331. Available at: www.mat.savba.sk/∼jencova/wyd.pdf.

[83] Jenčová A., 2004, *Generalized relative entropies as contrast functionals on density matrices*, Int. J. Theor. Phys. **43**, 1635. Available at: www.mat.savba.sk/∼jencova/ijtp.pdf.

[84] Jenčová A., 2005, *Quantum information geometry and non-commutative L_p spaces*, Inf. Dim. Anal. Quant. Prob. Relat. Top. **8**, 215. Available at: www.mat.savba.sk/∼jencova/lpspaces.pdf.

[85] Johnson R.W., Shore J.E., 1983, *Comments and correction to "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy"*, IEEE Trans. Inf. Theory **29**, 942.

[86] Jones L.K., Byrne C.L., 1990, *General entropy criteria for inverse problems, with applications to ada compression, pattern classification and cluster analysis*, IEEE Trans. Inf. Theory **36**, 23.

[87] Kakutani S., 1941, *Concrete representation of abstract (L)-spaces and the mean ergodic theorem*, Ann. Math. **42**, 523.

[88] Kolmogorov A.N., 1933, *Grundbergriffe der Wahrscheinlichkeitrechnung*, Ergebnisse der Mathematik und Ihrer Grenzgebiete Bd. **2**, Springer, Berlin. (engl. transl. 1950, *Foundations of the theory of probability*, Chelsea, New York).

[89] Kosaki H., 1984, *Applications of the complex interpolation method to a von Neumann algebra: non-commutative L^p-spaces*, J. Funct. Anal. **56**, 29.

[90] Kostecki R.P., 2011, *On general form of quantum information geometry*, to be submitted.

[91] Kostecki R.P., 2011, *Quantum theory as inductive inference*, in: Mohammad-Djafari A., Bercher J., Bessière P. (eds.), *Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIP Conf. Proc. **1305**, Springer, Berlin, p.24. arXiv:1009.2423.

[92] Kostecki R.P., 2011, *Towards new foundations of quantum theory I. Formalism and principles*, to be submitted.

[93] Kubo R., 1957, *Statistical mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems*, J. Phys. Soc. Japan **12**, 570.

[94] Kullback S., 1959, *Information theory and statistics*, Wiley, New York.

[95] Kullback S., Leibler R.A., 1951, *On information and sufficiency*, Ann. Math. Statist. **22**, 79. euclid:1177729694.

[96] Kunze R., 1958, *L_p Fourier transform on locally compact unimodular groups*, Trans. Amer. Math. Soc. **89**, 519.

[97] Lauritzen S.L., 1987, *Conjugate connections in statistical theory*, in: Dodson C.T.J. (ed.), *Geometrization of statistical theory*, Univ. Lancaster Dept. Math. Publ., Lancaster, p.33.

[98] Lauritzen S.L., 1987, *Statistical manifolds*, in: Gupta S.S. (ed.), *Differential geometry in statistical inference*, Institute of Mathematical Statistics, Hayward, p.163.

[99] Lesniewski A., Ruskai M.B., 1999, *Monotone riemannian metrics and relative entropy on non-commutative probability space*, J. Math. Phys. **40**, 5702. arXiv:math-ph/9808016.

[100] Liese F., Vajda I., 1987, *Convex statistical distances*, Teubner–Texte zur Mathematik **95**, Teubner, Leipzig.

[101] Masuda T., 1983, *L_p-spaces for von Neumann algebra with reference to a faithful normal semifinite weight*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **19**, 673. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=19&iss=2&rank=10.

[102] Moreau J.-J., 1963, *Fonctionelles sous-différentiables*, Compt. Rend. Acad. Sci. Paris **257**, 4117.

[103] Mori H., 1956, *A quantum-statistical theory of transport processes*, J. Phys. Soc. Japan **11**, 1029.

[104] Morimoto T., 1963, *Markov processes and the H-theorem*, J. Phys. Soc. Jap. **12**, 328.

[105] Morozova E.A., Chentsov N.N., 1989, *Markovskaya invaryantnaya geometriya na mnogoobraziyakh sostoyaniĭ*, Itogi Nauki i Tekhniki, Ser. Sovrem. Probl. Mat. Nov. Dostizh. **36**, 69 (engl. transl. 1991, *Markov invariant geometry on state manifolds*, J. Soviet Math. **56**, 2648). Available at: www.mathnet.ru/links/32fd0ad75d73f453dc3bfcbe90278097/intd121.pdf.

[106] Morozova E.A., Chentsov N.N., 1991, *Estestvennaya geometriya semeĭstv veroyatnostnykh zakonov*, Itogi Nauki i Tekhniki, Ser. Sovr. Probl. Mat. Fundam. Napravlen. **83**, 133. (*The natural geometry of families of probabilistic laws*) Available at: www.mathnet.ru/links/0b7482826af98f74f707088ff33082f6/intf210.pdf.

[107] Nagaoka H., Amari S.-i., 1982, *Differential geometry of smooth families of probability distributions*, Technical report METR **82-7**, University of Tōkyō, Tōkyō.

[108] Nelson E., 1974, *Notes on non-commutative integration*, J. Funct. Anal. **15**, 103.

[109] Norden A.P., 1945, *O parakh soprazhennykh parallel'nykh pereneseniĭ v n-mernykh prostranstvakh*, Dokl. Akad. Nauk SSSR **49**, 9.

[110] Ogasawara T., Yoshinaga K., 1955, *A non-commutative theory of integration for operators*, J. Sci. Hiroshima Univ. **18**, 311. Available at: math.berkeley.edu/∼pavlov/scans/ogasawara-yoshinaga-0.pdf.

[111] Ohya M., Petz D., 1993, *Quantum entropy and its use*, Springer, Berlin.

[112] Ojima I., 2004, *Temperature as order parameter of broken scale invariance*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **40**, 731. arXiv:math-ph/0311025.

[113] Paris J., Vecovská A., 1989, *Maximum entropy and inductive inference*, in: Skilling J. (ed.), *Maximum entropy and bayesian methods*, Kluwer, Dordrecht, p.397.

[114] Paris J., Vecovská A., 1989, *A note on the inevitability of maximum entropy*, Int. J. Inexact Reas. **4**, 183.

[115] Pedersen G.K., Takesaki M., 1973, *The Radon–Nikodym theorem for von Neumann algebras*, Acta Math. **130**, 53. Available at: www.kryakin.com/files/Acta_Mat_%282_55%29/acta150_107/130/130_02.pdf.

[116] Pérez A., 1967, *Information-theoretic risk estimates in statistical decision*, Kybernetika **3**, 1. Available at: www.kybernetika.cz/content/1967/1/1/paper.pdf.

[117] Petz D., 1984, *A dual in von Neumann algebras with weights*, Q. J. Math. Oxford **35**, 475.

[118] Petz D., 1985, *Properties of quantum entropy*, in: Accardi L., von Waldenfels W. (eds.), *Quantum probability and applications II*, LNM **1136**, Springer, Berlin, p.428.

[119] Petz D., 1985, *Quasi-entropies for states of a von Neumann algebra*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **21**, 787. Available at: www.math.bme.hu/∼petz/pdf/26quasi.pdf.

[120] Petz D., 1986, *Quasi-entropies for finite quantum systems*, Rep. Math. Phys. **23**, 57. Available at: www.renyi.hu/∼petz/pdf/52.pdf.

[121] Petz D., 1992, *Characterization of the relative entropy of states of matrix algebras*, Acta Math. Hung. **59**, 449. Available at: www.renyi.hu/∼petz/pdf/52.pdf.

[122] Petz D., 1994, *Geometry of canonical correlation on the state space of a quantum system*, J. Math. Phys. **35**, 780. Available at: www.renyi.hu/∼petz/pdf/61.pdf.

[123] Petz D., 1994, *On entropy functionals of states of operator algebras*, Acta Math. Hungar. **64**, 333. Available at: www.renyi.hu/∼petz/pdf/59.pdf.

[124] Petz D., 1996, *Monotone metrics on matrix spaces*, Lin. Alg. Appl. **224**, 81. Available at: www.renyi.hu/∼petz/pdf/68mon.pdf.

[125] Petz D., 2007, *Bregman divergence as relative operator entropy*, Acta Math. Hungar. **116**, 127. Available at: www.renyi.hu/∼petz/pdf/112bregman.pdf.

[126] Petz D., 2010, *From f-divergence to quantum quasi-entropies and their use*, Entropy **12**, 304. Available at: www.mdpi.com/1099-4300/12/3/304/pdf.

[127] Petz D., Hasegawa H., 1996, *On the riemannian metric of α-entropies of density matrices*, Lett. Math. Phys. **38**, 221. Available at: www.renyi.hu/∼petz/pdf/71.pdf.

[128] Pietsch A., 2007, *History of Banach spaces and linear operators*, Birkhäuser, Boston.

[129] Pillet C.-A., 2006, *Quantum dynamical systems*, in: Attal S. et al. (eds.), *Open quantum systems I: the hamiltonian approach*, LNM **1880**, Springer, Berlin, p.107. Available at: hal.archives-ouvertes.fr/docs/00/12/88/67/PDF/qds.pdf.

[130] Rao C.R., 1945, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc. **37**, 81.

[131] Read T.R.C., Cressie N., 1988, *Goodness-of-fit statistics for discrete multivariate data*, Springer, Berlin.

[132] Rédei M., 1999, *"Unsolved problems in mathematics": J. von Neumann's address to the International Congress of Mathematicians, Amsterdam, September 2–9, 1954*, Math. Int. **21**, 7. Available at: phil.elte.hu/∼redei/cikkek/intel.pdf.

[133] Rényi A., 1961, *On measures of entropy and information*, in: Neyman J. (ed.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol.1, University of California Press, Berkeley, p.547. euclid:1200512181.

[134] Rockafellar R.T., 1970, *Convex analysis*, Princeton University Press, Princeton. Available at: www.math.washington.edu/∼rtr/papers/rtr-ConvexAnalysis.djvu.

[135] Rodríguez C.C., 2003, *A geometric theory of ignorance*, preprint, available at: omega.albany.edu:8008/ignorance/ignorance03.pdf.

[136] Sakai S., 1971, *C*-algebras and W*-algebras*, Springer, Berlin.

[137] Saks S., 1933, *Théorie de l'integrale*, Monografie Matematyczne, ?, Warszawa. (2nd rev. ed., 1937, *Theory of the integral*, Stechert, New York).

[138] Segal I.E., 1947, *Irreducible representations of operator algebras*, Bull. Amer. Math. Soc. **61**, 69.

[139] Segal I.E., 1947, *Postulates for general quantum mechanics*, Ann. Math. **48**, 930.

[140] Segal I.E., 1951, *Equivalences of measure spaces*, Amer. J. Math. **73**, 275.

[141] Segal I.E., 1953, *A non-commutative extension of abstract integration*, Ann. Math. **57**, 401.

[142] Sen R.N., 1944, *On parallelism in riemannian space I*, Bull. Calcutta Math. Soc. **36**, 102.

[143] Shore J.E., Johnson R.W., 1980, *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*, IEEE Trans. Inf. Theory **26**, 26.

[144] Shore J.E., Johnson R.W., 1981, *Properties of cross-entropy minimization*, IEEE Trans. Inf. Theory **27**, 472.

[145] Skilling J., 1988, *The axioms of maximum entropy*, in: Erickson G.J. and Smith C.R. (eds.), *Maximum-entropy and bayesian methods in science and engineering*, Kluwer, Dordrecht, p. 173.

[146] Stinespring W.F., 1955, *Positive functions on C*-algebras*, Proc. Amer. Math. Soc. **6**, 211.

[147] Stinespring W.F., 1959, *Integration theorems for gages and duality for unimodular groups*, Trans. Amer. Math. Soc. **90**, 15.

[148] Ştrătilă S., Zsidó L., 1979, *Lectures on von Neumann algebras*, Edituria Academiei, Bucureşti.

[149] Takesaki M., 1970, *Tomita's theory of modular Hilbert algebras and its applications*, LNM **128**, Springer, Berlin.

[150] Takesaki M., 1973, *Duality for crossed products and the structure of von Neumann algebras of type III*, Acta Math. **131**, 249.

[151] Takesaki M., 2003, *Theory of operator algebras, Vol.1-3*, Springer, Berlin.

[152] Terp M., 1981, *L^p-spaces associated with von Neumann algebras*, Københavns Univ. Math. Inst. Rapp. No. **3a+3b**, Matematiske Institut, Københavns Universitet, København.

[153] Terp M., 1982, *Interpolation spaces between von Neumann algebra and its predual*, J. Oper. Theor. **8**, 327. Available

at: www.theta.ro/jot/archive/1982-008-002/1982-008-002-008.pdf.

[154] Tomita M., 1967, *Quasi-standard von Neumann algebras*, mimeographed notes. (unpublished).

[155] Tsallis C., 1988, *Possible generalization of Boltzmann–Gibbs statistics*, J. Stat. Phys. **52**, 479.

[156] Uffink J.B.M., 1996, *The constraint rule of the maximum entropy principle*, Stud. Hist. Phil. Mod. Phys. **27**, 47. Available at: www.phys.uu.nl/igg/jos/mep2def/mep2def.pdf.

[157] Umegaki H., 1962, *Conditional expectations in an operator algebra IV (entropy and information)*, Kōdai Math. Sem. Rep. **14**, 59.

[158] van Daele A., 1978, *Continuous crossed products and type III von Neumann algebras*, London Mathematical Society LNS **31**, Cambridge University Press, Cambridge.

[159] van Fraasen B.C., 1981, *A problem for relative information minimizers in probability kinematics*, Brit. J. Phil. Sci. **32**, 375.

[160] van Fraasen B.C., Hughes R.I.G., Harman G., 1986, *A problem for relative information minimizers, continued*, Brit. J. Phil. Sci. **37**, 453.

[161] von Neumann J., 1932, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin. (engl. transl. 1955, *Mathematical foundations of quantum mechanics*, Princeton University Press, Princeton).

[162] von Neumann J., 1937, *Quantum logic (strict- and probability logics)*, unpublished manuscript, John von Neumann Archive, Library of Congress, Washington (reviewed by Taub A.H., 1961, in: von Neumann J., 1961, *Collected Works*, Pergamon, London, p.195).

[163] Wald R.M., 1994, *Quantum field theory in curved spacetime and black hole thermodynamics*, The University of Chicago Press, Chicago.

[164] Wigner E., Yanase M., 1963, *Information content of distributions*, Proc. Nat. Acad. Sci. U.S.A. **49**, 910.

[165] Woronowicz S.L., 1979, *Operator systems and their application to the Tomita–Takesaki theory*, LNS **52**, Aarhus Universitet, Matematisk Institut, Aarhus. Reprinted in: J. Op. Theor. **2**, 169. Available at: www.theta.ro/jot/archive/1979-002-002/1979-002-002-003.pdf.

[166] Yamagami S., 1992, *Algebraic aspects in modular theory*, Publ. Res. Inst. Math. Sci. Kyōto Univ. **28**, 1075. Available at: www.ems-ph.org/journals/show_pdf.php?issn=0034-5318&vol=28&iss=6&rank=8.

[167] Yeadon F.J., 1975, *Non-commutative L_p-spaces*, Math. Proc. Cambridge Phil. Soc. **77**, 91.

[168] Zhang J., 2004, *Divergence function, duality, and convex analysis*, Neural Comp. **16**, 159. Available at: www.lsa.umich.edu/psych/junz/Neural%20Comp%202004.pdf.

[169] Zhang J., Matsuzoe H., 2009, *Dualistic riemannian manifold structure induced from convex functions*, in: *Advances in applied mathematics and global optimization in honor of Gilbert Strang*, Advances in Mechanics and Mathematics **17**, Springer, Berlin, p.1. Available at: www.lsa.umich.edu/psych/junz/Zhang%20and%20Matsuzoe%20hessian.pdf.

[170] Zhu H., 1998, *Generalized Lebesgue spaces and application to statistics*, Santa Fe Inst. Tech. Rep. **98-06-44**, Santa Fe. Available at: www.santafe.edu/media/workingpapers/98-06-044.ps.

[171] Zhu H., Rohwer R., 1995, *Information geometric measurements of generalisation*, Tech. Rep. Neural Comp. Res. Group **4350**, Aston. Available at: eprints.aston.ac.uk/507/1/NCRG_95_005.pdf.

[172] Zhu H., Rohwer R., 1997, *Measurements of generalisation based on information geometry*, in: Ellacott S.W. et al. (eds.), *Mathematics of neural networks: models, algorithms and applications*, Kluwer, Dordrecht, p.394. Available at: eprints.aston.ac.uk/514/1/NCRG_95_012.pdf.

[173] Zhu H., Rohwer R., 1998, *Information geometry, bayesian inference, ideal estimates and error decomposition*, Santa Fe Inst. Tech. Rep. **98-06-45**, Santa Fe. Available at: omega.albany.edu:8008/ignorance/zhu98.pdf.