Quantum Stud.: Math. Found.
https://doi.org/10.1007/s40509-020-00222-w

CHAPMAN UNIVERSITY | INSTITUTE FOR QUANTUM STUDIES

REGULAR PAPER

From thermodynamic sufficiency to information causality

Peter Harremoës

Received: 1 October 2019 / Accepted: 6 February 2020

Abstract The principle called information causality has been used to deduce Tsirelson's bound. In this paper, we derive information causality from monotonicity of divergence and relate it to more basic principles related to measurements on thermodynamic systems. These principles are more fundamental in the sense that they can be formulated for both unipartite systems and multipartite systems while information causality is only defined for multipartite systems. Thermodynamic sufficiency is a strong condition that puts severe restrictions to shape of the state space to an extent that we conjecture that under very weak regularity conditions it can be used to deduce the complex Hilbert space formalism of quantum theory. Since the notion of sufficiency is relevant to all convex optimization problems, there are many examples where it does not apply.

Keywords Bregman divergence · Multipartite system · Information causality · Thermodynamic sufficiency

Mathematics Subject Classification 81P16 · 94A17

1 Introduction

Entanglement is a resource that may allow agents to solve certain game problems in a more efficient way than what is possible without entanglement. Such tasks could be solved even more efficiently if the agents had access to a fictive resource called PR boxes. Such boxes cannot be used for signaling, but they can create correlations that are stronger than the correlations that can be created using entanglement. To be more precise, all quantum mechanical correlations satisfy Tsirelson's bound while PR-boxes can violate Tsirelson's bound.

The goal is to explain Tsirelson's bound and other bounds on correlations from more basic physical principles. One such principle is called information causality, and it may be formulated as "one bit of communication cannot create more than one bit of correlation". In [16], this principle was introduced and it was proved that it can be used to derive Tsirelson's bound. In [16], information causality was formulated and derived from the existence of the function conditional mutual information that is assumed to satisfy some basic properties. In [17], two ways of defining entropy were specified, and they were used to formulate the principle of information causality.

P. Harremoës (✉)
Copenhagen Business College, Nørre Voldgade 34, Copenhagen, Denmark
e-mail: harremoes@ieee.org

Published online: 24 February 2020

This paper uses properties of Bregman divergences rather than entropy or mutual information as basic principles. These divergences have several advantages compared with entropy and mutual information.

To each convex optimization problem, one can associate a Bregman divergence. If the optimization problem is energy extracting in thermodynamics, the Bregman divergence is proportional to quantum relative entropy that has some very desirable properties. These properties may be violated if one looks at different optimization problems. Therefore, one may ask what is so special about energy extraction in thermodynamics, but this important problem will not be covered in the present paper. One advantage of studying divergence (and entropy) rather than conditional mutual information is that divergence and its properties can be studied for unipartite systems while conditional mutual information only makes sense for multipartite systems. This is important because we do not have a canonical way of forming product spaces in generalized probabilistic theories. Bregman divergences with nice properties can be defined on Jordan algebras and the existence of a nice Bregman divergence rule out most other convex bodies as potential state spaces. Finally, both entropy and conditional mutual information may be considered as derived concepts based on divergence. This aspect will be the focus of the present paper.

The paper is organized as follows. In Sect. 2, we specify concepts like state space and measurement and we fix notation. Jordan algebras and their most important properties are described in Sect. 3. In Sect. 4, it is proved that several different ways of defining entropy coincide for Jordan algebras. Bregman divergences and their relation to optimization are described in Sect. 5. Several conditions related to the notion of sufficiency are defined. For Jordan algebras, these conditions are equivalent and the Bregman divergence is generated by the entropy function. In Sect. 6, we define conditional mutual information based on a Bregman divergence and we demonstrate that the conditional mutual information has the properties that are needed for information causality to be satisfied. We conclude with Sect. 7 by summarizing our results and state some open problems.

2 State spaces

Let $\mathcal{P}$ denote a set of *preparations* of a physical experiment. A *mixed preparation* is a formal mixture $\sum s_i \cdot p_i$ where p_i are preparations and $(s_i)_i$ is a probability vector. The mixture $\sum s_i \cdot p_i$ is identified with the preparation where p_i is chosen with probability s_i. A measurement m maps each preparation in $\mathcal{P}$ into a probability measure on the set of possible outcomes of the experiment. We assume that $m\left(\sum s_i \cdot p_i\right) = \sum s_i \cdot m\left(p_i\right)$. Let $\mathcal{M}$ denote the set of measurements that can be performed by an observer (or a group of observers). If $m\left(p_1\right) = m\left(p_2\right)$ for all measurements $m \in \mathcal{M}$, then we say that p_1 and p_2 represent the same *state*. The set of states is called the state space, and with this Bayesian definition of a state the state space will depend on the set of feasible measurements. In particular, the state spaces of two different observers may be different because they may have different sets of measurements. A group of observers may have a different state space than any of the individual observers because the set of joint measurements may be larger than the set of measurements that can be performed by any of the individual observers.

For simplicity, we will assume that the state spaces are *convex bodies* Ω, i.e. convex compact sets spanned by finitely many elements. The extreme point are called *pure states*. Any convex body can be embedded in the pointed cone Ω_+ consisting of formal products $t \cdot \sigma$ where σ is a state and t is a positive real number called the *trace* of $t \cdot \sigma$. The notation is $\operatorname{tr}\left(t \cdot \sigma\right) = t$. The elements in the cone are called *positive operators* or un-normalized states. The cone is called the *state cone*. Positive elements can be added by

$$t_1 \cdot \sigma_1 + t_2 \cdot \sigma_2 = \left(t_1 + t_2\right) \cdot \left(\frac{t_1}{t_1 + t_2}\sigma_1 + \frac{t_2}{t_1 + t_2}\sigma_2\right).$$

The state cone spans a partially ordered vector space V_Ω and the trace extends linearly to V_Ω. Thus, the states may be considered as positive elements of an ordered vector space with trace 1.

Let $m \in \mathcal{M}$ denote a measurement with values v in some set $\mathcal{V}$. If σ is a state then the measurement is given by a probability measure $m\left(\sigma\right)$ over $\mathcal{V}$. Thus, for each $v \in \mathcal{V}$, we have a probability $m\left(\sigma\right)\left(v\right) \in [0, 1]$. For each v, the measurement m maps Ω into $[0, 1]$ and such a mapping is called a *test* and it is an element in Ω_+^*, i.e. the dual cone of the positive elements. In the literature on generalized probabilistic theories, a test is often called an effect, but in this paper, it is called a test, which is well established in the statistical literature. The test that maps $x \in V_\Omega$ into $\lambda \mathrm{tr}(x)$ will be denoted λ. In particular, the test 1 maps Ω into 1. Since the total probability of a measurement is 1 we have $\sum_v m(\cdot)(v) = 1$. A measurement can be represented as a test valued measure. In the Hilbert space formalism the tests are given by positive operators and the measurements are given by positive operator valued measures (POVM). We say that two states ρ and σ are mutually singular if there exists a test ϕ such that $\phi(\rho) = 0$ and $\phi(\sigma) = 1$.

Let $m_1, m_2 \in \mathcal{M}$ with values in $\mathcal{V}_1$ and $\mathcal{V}_2$. If $M : V_1 \to V_2$ is some map such that

$$m_2(\cdot)(v_2) = \sum_{v_1 : M(v_1) = v_2} m_1(\cdot)(v_1)$$

then the measurement m_1 is at least as informative about the state as m_2, and m_1 is called a *fine-graining* of m_2. If

$$m_2(\cdot)(v_2) \propto m_1(\cdot)(v_1)$$

for all values v_1 for which $M(v_1) = v_2$, then the fine-graining is said to be *trivial*. A measurement is *fine grained* if all fine-grainings are trivial. Note that a measurement m is fine grained if all tests $m(\cdot)(v)$ lie on extreme rays of Ω_+^*. Therefore, any measurement has a fine-graining that is fine grained when the state space Ω is a convex body.

Let Ω_1 and Ω_2 denote two state spaces. An affine map $\Phi : \Omega_1 \to \Omega_2$ is called and *affinity*. Let $S : \Omega_1 \to \Omega_2$ and $R : \Omega_2 \to \Omega_1$ denote affinities. If $R \circ S = id_{\Omega_1}$ then S is called a *section* and R is called a *retraction*. A *frame* is a section $S : \Omega_1 \to \Omega_2$ where Ω_1 is a simplex.

Let Ω denote the state space of a group of observers. The set of measurements M_A of a single observer Alice is a subset of the set of all measurements M of the whole group of observers. Therefore, then there is a surjective affinity $\mathbb{E}_A : \Omega \to \Omega_A$. Assume that Alice and Bob are observers that can perform measurements independently. Further assume that the choice of measurement made by Alice does not influence the outcome of a measurement made by Bob and that a choice of measurement made by Bob does not influence the outcome of a measurement made by Alice. This is called the *no-signaling condition*. If Alice performs the measurement m_A and Bob performs the measurement m_B, then the joint measurement is denoted $m_A \otimes m_B$. Further assume that Alice and Bob can communicate. Then Alice and Bob can perform any measurement of the form $\sum s_i \cdot m_A \otimes m_B$. If Alice and Bob together can only perform measurements of the form $\sum s_i \cdot m_A \otimes m_B$ their joint state space is a subset of $V_{\Omega_A} \otimes V_{\Omega_B}$. Assume further that Alice and Bob can prepare states individually. If Alice prepares the state σ_A and Bob prepares the state σ_B then their joints state is $\sigma_A \otimes \sigma_B \in V_{\Omega_A} \otimes V_{\Omega_B}$. The convex hull of $\{\sigma_A \otimes \sigma_B \,|\, \sigma_A \in \Omega_A \text{ and } \sigma_B \in \Omega_B\}$ is denoted $\Omega_A \otimes_{min} \Omega_B$ and the elements are called separable states. We assume that $\Omega_A \otimes_{\min} \Omega_B \subseteq \Omega$.

3 Jordan algebras

Here we will recall some fact and concepts related to Jordan algebras. A more detailed exposition can be found in [2,14]. In the Hilbert space formalism of quantum physics, the states are represented as density matrices on a complex Hilbert space. Classical probability distributions can be identified with density matrices that are diagonal. In the set of self-adjoint matrices, one may define a product $\bullet$ by

$$A \bullet B = \frac{1}{2}(AB + BA).$$

This product makes the set of Hermitean matrices into an algebra over the real numbers and the product $\bullet$ satisfies

$$A \bullet (B \bullet (A \bullet A)) = (A \bullet B) \bullet (A \bullet A). \tag{1}$$

With this equation fulfilled, it is possible to define $A^n = A \bullet A \bullet \cdots \bullet A$ without specifying where the parenthesis have to be placed. Further we have that

$$\sum_i A_i^2 = 0 \tag{2}$$

if and only if $A_i = 0$ for all i. The dimension of the algebra is defined as the dimension of the Jordan algebra as a real vector space. A finite dimensional algebra over the real numbers with a product $\bullet$ satisfying the properties (1) and (2) is called an *Euclidean Jordan algebra*.

Elements in an Euclidean Jordan algebra of the form $A \bullet A$ are called positive elements and they form a pointed cone. Further, an Euclidean Jordan algebra has a *trace* tr that maps positive elements into positive numbers and such

$$\operatorname{tr}((A \bullet B) \bullet C) = \operatorname{tr}(A \bullet (B \bullet C)).$$

A *state in a Jordan algebra* is a positive element of trace 1. The *rank of a Jordan algebra* is the Caratheodory rank of the state space of algebra. An Euclidean Jordan algebra has an inner product defined by

$$\langle A, B \rangle = \operatorname{tr}(A \bullet B).$$

With this inner product, the positive cone becomes *self-dual*.

An element E of a Jordan algebra is *idempotent* if $E^2 = E$. Elements A and B are *orthogonal* if $A \bullet B = 0$. With these definitions any element A has a spectral decomposition

$$A = \sum \lambda_i E_i,$$

where E_i are orthogonal idempotent. If the spectral values λ_i are different, the decomposition is unique. Therefore, one can define

$$f(A) = \sum f(\lambda_i) E_i.$$

The associative Euclidean Jordan algebras correspond to classical probability theory, where the state space is a simplex. Any Euclidean Jordan algebra $\mathcal{J}$ can be written as a direct sum $\bigoplus \mathcal{J}_i$ of Jordan algebras where each of the Jordan algebras $\mathcal{J}_i$ is simple. The simple Euclidean Jordan algebras belong to one of the following five types.

- $M_n(\mathbb{R})$ Real-valued Hermitean $n \times n$ matrices.
- $M_n(\mathbb{C})$ Complex-valued Hermitean $n \times n$ matrices.
- $M_n(\mathbb{H})$ Quaternionic-valued Hermitean $n \times n$ matrices.
- $M_3(\mathbb{O})$ Octonionic-valued Hermitean 3×3 matrices.
- $Jspin(d)$ Spin factors where the state space has the shape of a d-dimensional solid ball.

The Jordan algebra $M_3(\mathbb{O})$ is called the *exceptional Jordan algebra* and Jordan algebras that do not contain such an exceptional component are called *special Jordan algebras*. All special Jordan algebras appear as sections of $M_n(\mathbb{C})$ for some value of n. In this sense, all special Jordan algebras have representations as physical systems. If a section of the set of complex-valued Hermitean matrices is required to be completely positive then the section can be represented as a set of complex-valued Hermitean matrices.

It is an important question why exactly the complex-valued Hermitean matrices are so good in modeling quantum physics compared with the other simple Jordan algebras. Actually, Adler has attempted to model quantum theory using quaternions [1], and there have been a number of attempts to let the exceptional Jordan algebra play an active role in modeling physics [6,13]. One important property that singles out the complex-valued Hermitean matrices is that there is a canonical tensor product construction within the category of complex-valued Hermitean matrices with completely positive maps as morphisms [3].

Example 1 Assume that the whole state space Ω can be represented as real non-negative definite 4×4 matrices with trace 1. The dimension of this state space is 9. Let A and B denote a 2×2 real Hermitean matrices. Then $A \otimes B$ can embedded in Ω as

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}.$$

The vector space of Hermitean 2×2 matrices has dimension 3. Therefore, the tensor product has dimension 9. Hence, the set of tensors with trace 1 has dimension 8, so it has a lower dimension than set of states on the whole space. Therefore, there are joint states on the whole space that cannot be distinguished by local measurements. Hence, the tomography condition is not fulfilled.

There are a number of ways to characterize Jordan algebras. Above we have defined the Jordan algebras algebraically. A classic result is that a real vector space with a self-dual homogeneous cone can be represented as a Jordan algebra [11]. A new result is that a state space that is spectral and where any pair of frames can be mapped into each other can be represented by a Jordan algebra [4].

For Jordan algebras, it is possible to define a well-behaved entropy function and an associated divergence function. In [10] it was proved that if a state space has rank 2 and it has a monotone Bregman divergence then it can be represented as a Jordan algebra (spin factor). Similar representation theorems for state spaces of higher rank are not yet available; so, in this paper, we focus on other consequences of the existence of entropy function or Bregman divergences.

4 Entropy in Jordan algebras

In generalized probabilistic theories, there are two ways of defining entropy [17]. The *decomposition entropy* of a state σ is given by

$$\breve{H}(\sigma) = \inf_{\sum p_i \cdot \sigma_i = \sigma} H\left((p_i)_i\right).$$

Here the infimum is taken over all mixtures $\sum p_i \cdot \sigma_i = \sigma$ where σ_i are pure states and $H\left((p_i)_i\right)$ denotes the Shannon entropy of the probability vector $(p_i)_i$. Versions of this definition can also be found in [8], but they date back to [18]. Note that the definition of spectral entropy in [12] is closely related but slightly different.

Following [17], one can define the *fine grained entropy* of a state in a generalized probabilistic theory by

$$\hat{H}(\sigma) = \inf_m H(m(\sigma)),$$

where the infimum has been taken over all fine grained measurements m on Ω. This fine-grained entropy is a strictly concave function.

Lemma 1 *If the state space Ω is spectral a decomposition that minimizes the decomposition entropy is spectral.*

Proof This was essentially proved in [8] although the terminology regarding spectrality was slightly different. □

Theorem 1 *If the state space Ω is spectral, then for any state σ the following inequality holds:*

$$\hat{H}(\sigma) \leq \breve{H}(\sigma).$$

Proof Let $\sigma = \sum p_i \sigma_i$ be a decomposition of σ where the states σ_i are pure. To this decomposition there corresponds a measurement m such that

$$m(\sigma)(i) = p_i.$$

Since this measurement is fine grained, we have

$$\hat{H}(\sigma) \leq H(m(\sigma)) = H\left((p_i)_i\right).$$

Therefore,

$$\hat{H}(\sigma) \leq \inf_{\sum p_i \sigma_i = \sigma} H\left((p_i)_i\right) = \breve{H}(\sigma).$$

□

Theorem 2 *If the state space Ω is spectral and the cone Ω_+ is self-dual then*

$$\hat{H}(\sigma) = \breve{H}(\sigma) = -\langle \sigma, \ln(\sigma) \rangle . \tag{3}$$

Proof Let M denote a fine-grained measurement. The measurement is given by a positive test-valued measure, i.e. there exists $\rho_j \geq 0$ such that $\sum \rho_j = 1$ and such that

$$M(\rho)(j) = \langle \rho_j, \rho \rangle .$$

Since the measurement is fine grained, ρ_j must be states. Thus,

$$M(\sigma)(j) = \langle \rho_j, \sigma \rangle = \left\langle \rho_j, \sum_i p_i \sigma_i \right\rangle = \sum_i p_i \langle \rho_j, \sigma_i \rangle .$$

If $\tilde{\sigma}$ is the state $\sum_i \frac{1}{r} \cdot \sigma_i$, then

$$M(\tilde{\sigma}) = \left\langle \rho_j, \sum_i \frac{1}{r} \cdot \sigma_i \right\rangle_j = \frac{1}{r} \langle \rho_j, 1 \rangle_j = \frac{1}{r}.$$

The Markov kernel $(p_i)_i \to \sum_i p_i \langle \rho_j, \sigma_i \rangle_j$ maps the uniform distribution $\left(\frac{1}{r}\right)_i$ into the uniform distribution $\left(\frac{1}{r}\right)_i$, i.e. the Markov kernel is bi-stochastic. Since bi-stochastic Markov kernels increase entropy, we have

$$H(M(\sigma)) = H\left(\langle \rho_j, \sigma \rangle_j\right) \geq H\left((p_i)_i\right) = -\langle \sigma, \ln(\sigma) \rangle .$$

Therefore,

$$-\langle \sigma, \ln(\sigma) \rangle \leq \hat{H}(\sigma) . \tag{4}$$

Now the result is obtained by combining Lemma 1 and Theorem 1 with inequality (4). □

Definition 1 The entropy H of a state σ in a Jordan algebra is given as the common value of any of the expressions given in Eq. (3).

Corollary 1 *In a finite Euclidean Jordan algebra, the entropy $-\langle \sigma, \ln(\sigma) \rangle$ is a concave function.*

Proof Concavity of H follows because H equals the fine-grained entropy and the fine-grained entropy is concave [17]. □

Concavity of the entropy function H on Jordan algebras was proved in [8] with a more involved proof.

5 Bregman divergences and sufficiency conditions

We consider a optimization problem where we want to optimize some quantity defined on the state space. In thermodynamics, the goal is typically to extract energy from the system by some feasible interaction with the system. Our approach makes sense for any convex optimization problem and in principle the function may represent other objectives such as the amount of money one may obtained by trading or the code length that is obtained after using a certain data compression procedure. Various examples of such optimization problems are given in [7]. In this paper, the objective function will be energy.

Assume that the system is in state $\rho \in \Omega$ and that we apply some action a from a set of feasible actions $\mathcal{A}$. Then the mean energy that we extract will be denoted

$$\langle a, \rho \rangle$$

and it is an affine function of the state ρ. An action a will be identified with this function $\rho \to \langle a, \rho \rangle$ so that the actions are considered as elements in the dual space of the state space. We can define the free energy of state ρ as

$$F(\rho) = \sup_{a \in \mathcal{A}} \langle a, \rho \rangle .$$

In thermodynamics Helmholz free energy is given as $F = U - TS$ so that the free energy is an affine function minus a term that is proportional to the entropy function. Then F is a convex function of ρ. The regret of doing action a if the state is ρ is defined as

$$D_F(\rho, a) = F(\rho) - \langle a, \rho \rangle .$$

The interpretation of the regret function is as follows. Assume that the system is in state ρ but one uses a sub-optimal action a. Then the regret measures the difference between the energy that one could have extracted $F(\rho)$ and the energy that one extracts using action a. For simplicity, we will assume that F is differentiable so that to each state ρ there exists a unique action a_ρ such that $F(\rho) = \langle a, \rho \rangle$. For states $\rho, \sigma \in \Omega$, the Bregman divergence is defined as

$$D_F(\rho, \sigma) = D_F(\rho, a_\sigma) .$$

It measures the regret of acting as if the state were σ if it actually is ρ. The Bregman divergence is given by

$$\begin{aligned} &D_F(\rho, \sigma) \\ &\quad = F(\rho) - \left(F(\sigma) + \frac{\mathrm{d}}{\mathrm{d}t} F((1-t)\sigma + t\rho)_{|t=0} \right). \end{aligned}$$

The formula for the Bregman divergence is often written in terms of the gradient.

$$D_F(\rho, \sigma) = F(\rho) - (F(\sigma) + \langle \nabla F(\sigma) | \, \rho - \sigma \rangle) .$$

Proposition 1 [8, Lemma 17] *For Hermitean matrices A and B, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} (\operatorname{tr}(f(A + tB)))_{|t=0} = \left\langle f'(A), B \right\rangle .$$

Example 2 Assume that the state space can be represented as a state space of a Jordan algebra. Let $F(\sigma) = \langle \sigma, \ln(\sigma) \rangle$ denote the negative of the entropy. The Bregman divergence corresponding to F can be computed as

$$\begin{aligned} D_F(\rho, \sigma) &= F(\rho) - \left\{ F(\sigma) + \frac{\mathrm{d}}{\mathrm{d}t} F((1-t)\sigma + t\rho)_{|t=0} \right\} = \langle \rho, \ln(\rho) \rangle \\ &\quad - \{ \langle \sigma, \ln(\sigma) \rangle + \langle \ln(\sigma) + 1, \rho - \sigma \rangle \} = \langle \rho, \ln(\rho) - \ln(\sigma) \rangle - \operatorname{tr}(\rho - \sigma) . \end{aligned} \tag{5}$$

We call this quantity the *information divergence* and denote it as $D(\rho \| \sigma)$. Note that the last term vanishes if ρ and σ are states. If the Jordan algebra is associative, we get Kullback–Leibler divergence given by

$$D(P \| Q) = \sum p_i \ln \frac{p_i}{q_i} .$$

If the Jordan algebra is a C^*-algebra F is minus the von Neumann entropy the information divergence equals quantum information divergence (quantum relative entropy) given by

$$D(\rho \| \sigma) = \operatorname{tr}(\rho (\ln \rho - \ln \sigma)) .$$

There are a number of conditions that some regret functions and Bregman divergences may have.

Definition 2 The Bregman divergence D_F is *monotone* if $D_F(\Phi(\rho), \Phi(\sigma)) \le D_F(\rho, \sigma)$ for any affinity $\Phi : \Omega \to \Omega$.

We note that monotonicity is associated with the decrease of free energy for a closed thermodynamic system. It is possible to define the regret $D_F(\rho, \sigma)$ even if the function F is not differentiable, but if such a regret function is monotone then F is automatically differentiable [7]. In the rest of this paper, we shall focus entirely on the case when F is differentiable and the regret between states is given by a Bregman divergence.

Theorem 3 *Information divergence is monotone on special Jordan algebras.*

Proof Let Ω denote the state space of a special Jordan algebra. Then there exists a section $S : \Omega \to M_n(\mathbb{C})^1_+$ with a corresponding retraction $R : M_n(\mathbb{C})^1_+ \to \Omega$. Let $\Phi : \Omega \to \Omega$ denote some affinity. Then $S \circ \Phi \circ R$ is an affinity $M_n(\mathbb{C})^1_+ \to M_n(\mathbb{C})^1_+$. Then

$$\begin{aligned} D(\Phi(\rho) \| \Phi(\sigma)) &= D(S(\Phi(\rho)) \| S(\Phi(\sigma))) = D((S \circ \Phi \circ R)(S(\rho)) \| (S \circ \Phi \circ R)(S(\sigma))) \\ &\leq D(S(\rho) \| S(\sigma)) = D(\rho \| \sigma). \end{aligned}$$

Here we have used that information divergence is monotone on $M_n(\mathbb{C})^1_+$ [15]. □

It is not known if information divergence is monotone on the exceptional Jordan algebra. Let ρ_θ denote a family of states and let Φ denote an affinity $\Phi : \Omega \to \Omega$. Then Φ is said to be *sufficient* for ρ_θ if there exists a *recovery map* $\Psi : \Omega \to \Omega$, i.e. an affinity such that $\Psi(\Phi(\rho_\theta)) = \rho_\theta$.

Definition 3 A Bregman divergence D_F is said to satisfy *sufficiency* if $D_F(\Phi(\rho), \Phi(\sigma)) = D_F(\rho, \sigma)$ whenever Φ is sufficient for ρ, σ.

It is easy to prove that monotonicity implies sufficiency. Further it is easy to prove that sufficiency implies the property called statistical locality as defined below.

Definition 4 A Bregman divergence D_F satisfies *statistical locality* if $\rho \perp \sigma_i$ implies

$$D_F(\rho, (1-t) \cdot \rho + t \cdot \sigma_1) = D_F(\rho, (1-t) \cdot \rho + t \cdot \sigma_2).$$

Proposition 2 *In an Euclidean Jordan algebra, information divergence satisfies statistical locality.*

Proof Assume that ρ, σ_1, and σ_2 are states and that $\rho \perp \sigma_i$. Then

$$\begin{aligned} D(\rho \| (1-t) \cdot \rho + t \cdot \sigma_1) &= \langle \rho, \ln(\rho) - \ln((1-t) \cdot \rho + t \cdot \sigma_1) \rangle \\ &= \langle \rho, \ln(\rho) - \ln((1-t) \cdot \rho) \rangle = -\ln(1-t). \end{aligned}$$

□

Theorem 4 *If the state space Ω can be represented as the state space of a Jordan algebra of rank at least 3, then a statistically local Bregman divergence D_F is proportional to information divergence given by Eq. (5). There exists a constant $c > 0$ such that the function F equals $c \cdot \langle \rho, \ln \rho \rangle$ plus an affine function on Ω.*

Proof The theorem was proved for finite C^*-algebras in [7], but the proof is the same for more general Jordan algebras. □

The theorem implies that under certain conditions the following conditions are equivalent

- Monotonicity.
- Sufficiency.
- Statistical locality.
- The Bregman divergence is proportional to information divergence.
- The objective function F is proportional to entropy plus an affine function.

If the state space has rank 2, these conditions are not equivalent and this special case was studied in great detail in [10].

6 Information causality

Consider a bipartite system with Alice and Bob as observers. We assume that the no-signaling condition and local tomography are fulfilled so that a joint state can be described as an element in the tensor product of local vector spaces. Let U_A and U_b denote order units of Alice and Bob.

Let F denote some payoff function on a joint system with regret function D_F. We will assume that the regret function D_F satisfies monotonicity. Then F is differentiable and D_F is a Bregman divergence. Therefore, D_F is given by

$$D_F(\rho, \sigma) = F(\rho) - (F(\sigma) + \langle \nabla F(\sigma) | \rho - \sigma \rangle).$$

The following proposition is well known if the affine combination is a convex combination.

Proposition 3 *If $\sum_i t_i = 1$ and the affine combination $\bar{\rho} = \sum_i t_i \cdot \rho_i$ is a state, then the Bregman identity holds:*

$$\sum_i t_i \cdot D_F(\rho_i, \sigma) = \sum_i t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma)\,. \tag{6}$$

Proof We expand the right-hand side of (6) and get

$$\begin{aligned}&\sum_i t_i \cdot D_F(\rho_i, \bar{\rho}) + D_F(\bar{\rho}, \sigma)\\ &\quad = \sum_i t_i \cdot (F(\rho_i) - (F(\bar{\rho}) + \langle \nabla F(\bar{\rho}) |\, \rho_i - \bar{\rho}\rangle)) + F(\bar{\rho}) - (F(\sigma) + \langle \nabla F(\sigma) |\, \bar{\rho} - \sigma\rangle)\,.\end{aligned}$$

We can re-arrange the terms and use that

$$\bar{\rho} = \sum_i t_i \cdot \rho_i$$

to get

$$\begin{aligned}&\sum_i t_i \cdot F(\rho_i) - \left(\sum_i t_i \cdot F(\bar{\rho}) + \left\langle \nabla F(\bar{\rho}) \middle| \sum_i t_i \cdot \rho_i - \bar{\rho}\right\rangle\right) + F(\bar{\rho}) - (F(\sigma) + \langle \nabla F(\sigma) |\, \bar{\rho} - \sigma\rangle)\\ &\quad = \sum_i t_i \cdot F(\rho_i) - (F(\bar{\rho}) + \langle \nabla F(\bar{\rho}) |\, \bar{\rho} - \bar{\rho}\rangle) + F(\bar{\rho}) - (F(\sigma) + \langle \nabla F(\sigma) |\, \bar{\rho} - \sigma\rangle)\,.\end{aligned}$$

Therefore, the right-hand side of Eq. (6) reduces to

$$\begin{aligned}&\sum_i t_i \cdot F(\rho_i) - (F(\sigma) + \langle \nabla F(\sigma) |\, \bar{\rho} - \sigma\rangle)\\ &\quad = \sum_i t_i \cdot (F(\rho_i) - (F(\sigma) + \langle \nabla F(\sigma) |\, \rho_i - \sigma\rangle)) = \sum_i t_i \cdot D_F(\rho_i, \sigma)\,,\end{aligned}$$

which is the left-hand side of Eq. (6) and this completes the proof. □

Theorem 5 *Assume that $\Omega \subset V_A \otimes V_B$. If $\rho_1, \rho_2 \in \Omega_A$ and $\sigma_1, \sigma_2 \in \Omega_B$ and D_F satisfies sufficiency then*

$$D_F(\rho_1 \otimes \sigma_1, \rho_2 \otimes \sigma_1) = D_F(\rho_2 \otimes \sigma_2, \rho_2 \otimes \sigma_2)\,.$$

Proof To see this, define

$$\Phi(\pi) = \mathbb{E}_A(\pi) \otimes \sigma_1, \quad \Psi(\pi) = \mathbb{E}_A(\pi) \otimes \sigma_2.$$

Then

$$\Phi(\rho_i \otimes \sigma_2) = \rho_i \otimes \sigma_1, \Psi(\rho_i \otimes \sigma_1) = \rho_i \otimes \sigma_2.$$

The result is obtained by sufficiency of D_F. □

If $\rho_1, \rho_2 \in \Omega_A$, we may write $D_F(\rho_1, \rho_2)$ as an abbreviation for $D_F(\rho_2 \otimes \sigma, \rho_2 \otimes \sigma)$ where some arbitrary state $\sigma \in \Omega_B$ is used.

Definition 5 Let σ denote a state on a system with a bipartite subsystem composed of subsystems labeled A and B. Then the mutual information between the subsystem A and subsystem B is defined as

$$I_\sigma(A; B) = D_F(\sigma_{AB}, \sigma_A \otimes \sigma_B). \tag{7}$$

Theorem 6 *If the Bregman divergence D_F is monotone then mutual information satisfies the following two conditions.*

Consistency *If the system has a bipartite subsystem consisting of two classical subsystems A and B, then the mutual information restricted to the bipartite subsystem is proportional to classical mutual information.*

Data processing inequality *If $\Phi : V_B \to V_B$ is a positive trace conserving affinity, then*

$$I_\sigma(A; B) \geq I_{(id\otimes\Phi)(\sigma)}(A; B).$$

Proof **Consistency** If the subsystems defined by Alice and Bob are classical and non-trivial then the rank of their joint state space is at least $2 \times 2 = 4$. When the rank of the state space is least 3, the function F is a linear function of the Shannon entropy and, therefore, the mutual information defined by (7) is proportional to the classical mutual information.

Data processing inequality Assume that

$$\Phi : V_B \to V_B$$

is a positive trace conserving affinity. Then $\tilde{\Phi} = id \otimes \Phi$ is given by $\tilde{\Phi}(\sigma_A \otimes \sigma_B) = \sigma \otimes \Phi(\sigma_B)$ and

$$\begin{aligned} I_\sigma(A; B) &= D_F(\sigma_{AB}, \sigma_A \otimes \sigma_B) \leq D_F\left(\tilde{\Phi}(\sigma_{AB}), \tilde{\Phi}(\sigma_A \otimes \sigma_B)\right) \\ &= D_F\left(\tilde{\Phi}(\sigma_{AB}), \sigma_A \otimes \Phi(\sigma_B)\right) = I_{\tilde{\Phi}(\sigma)}(A; B), \end{aligned}$$

which completes the proof. □

In probability theory, one may define entropy as self-information via

$$H(A) = I(A, A).$$

This is not possible in quantum theory because the different sub-spaces in a tensor product decomposition have to be distinct. In probability theory, this is not a problem and cloning is allowed, i.e. one is allowed to form identical copies in a state. In probability theory, one gets

$$\begin{aligned} H(AB) &= I(AB, AB) = I(A, AB) + I(B, AB \mid A) \geq I(A, AB) = I(A, A) + I(A, B \mid A) \\ &\geq I(A, A) = H(A). \end{aligned}$$

Therefore, in probability theory, the entropy of a subsystem is less than the entropy of the full system.

Definition 6 A Bregman divergence D_F on a bipartite system is *additive* if

$$D_F(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) = D_F(\rho_A, \sigma_A) + D_F(\rho_B, \sigma_B).$$

Theorem 7 *If the state spaces Ω_A and Ω_B can be represented as state spaces of Jordan algebras $\mathcal{J}_A$ and $\mathcal{J}_B$, and if D_F satisfies sufficiency then D_F is additive.*

Proof Let c_A and c_B denote distributions that maximize the fine grained entropy distributions in each of the algebras. Then D_F equals $D_{\tilde{F}}$, where

$$\tilde{F}(\sigma) = D_F(\sigma, c_A \otimes c_B).$$

Let ρ_A and ρ_B denote states in the state spaces Ω_A and Ω_B. Then ρ_A and ρ_B generate associative sub-algebras $\mathcal{A}_A \subseteq \mathcal{J}_A$ and $\mathcal{A}_B \subseteq \mathcal{J}_B$ with classical state spaces. Now the restriction of D_F to $\mathcal{A}_A \otimes \mathcal{A}_B$ satisfies sufficiency and according to Theorem 4 D_F is proportional to information divergence. Therefore,

$$D_F(\rho_A \otimes \rho_B, c_A \otimes c_B) = D_F(\rho_A, c_A) + D_F(\rho_B, c_B)$$

because information divergence is additive on classical state spaces. Define

$$\tilde{F}_A(\rho_A) = D_F(\rho_A, c_A), \quad \tilde{F}_B(\rho_B) = D_F(\rho_B, c_B).$$

With this notation

$$\tilde{F}(\rho_A \otimes \rho_B) = \tilde{F}_A(\rho_A) + \tilde{F}_B(\rho_B).$$

Thus,

$$\begin{aligned}
D_F(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) &= \tilde{F}(\rho_A \otimes \rho_B) - \begin{pmatrix} \tilde{F}(\sigma_A \otimes \sigma_B) + \\ \left\langle \nabla \tilde{F}(\sigma_A \otimes \sigma_B) \middle| \rho_A \otimes \rho_B - \sigma_A \otimes \sigma_B \right\rangle \end{pmatrix} \\
&= \tilde{F}_A(\rho_A) + \tilde{F}_B(\rho_B) - \begin{pmatrix} \tilde{F}_A(\sigma_A) + \tilde{F}_B(\sigma_B) + \\ \left\langle \nabla \tilde{F}_A(\sigma_A) + \nabla \tilde{F}_B(\sigma_B) \middle| \rho_A \otimes \rho_B - \sigma_A \otimes \sigma_B \right\rangle \end{pmatrix} \\
&= \tilde{F}_A(\rho_A) - \begin{pmatrix} \tilde{F}(\sigma_A) + \\ \left\langle \nabla \tilde{F}_A(\sigma_A) \middle| \rho_A \otimes \rho_B - \sigma_A \otimes \sigma_B \right\rangle \end{pmatrix} \\
&\quad + \tilde{F}_B(\rho_B) - \begin{pmatrix} \tilde{F}_B(\sigma_B) + \\ \left\langle \nabla \tilde{F}_B(\sigma_B) \middle| \rho_A \otimes \rho_B - \sigma_A \otimes \sigma_B \right\rangle \end{pmatrix} \\
&= D_F(\rho_A, \sigma_A) + D_F(\rho_B, \sigma_B).
\end{aligned}$$

□

Example 3 If tensor products of 2×2 Hermitean matrices are embedded in Hermitean 4×4 matrices as in Example 1, then mutual information is additive.

Lemma 2 *An additive monotone Bregman divergence satisfies the following identity:*

$$D_F(\sigma_{AB}, \rho_A \otimes \rho_B) = D_F(\sigma_{AB}, \sigma_A \otimes \rho_B) + D_F(\sigma_A, \rho_A). \tag{8}$$

Proof Any state σ_{AB} can be written as an affine combination of tensor products $\sigma_{AB} = \sum t_i \cdot \pi_{A,i} \otimes \pi_{B,i}$. Then

$$D_F(\sigma_{AB}, \rho_A \otimes \rho_B) = \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \rho_A \otimes \rho_B\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right).$$

Using additivity, it can be rewritten as

$$\begin{aligned}
&D_F(\sigma_{AB}, \rho_A \otimes \rho_B) \\
&\quad = \sum t_i \cdot \left(D_F\left(\pi_{A,i}, \rho_A\right) + D_F\left(\pi_{B,i}, \rho_B\right)\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right) \\
&\quad = \sum t_i \cdot D_F\left(\pi_{A,i}, \rho_A\right) + \sum t_i \cdot D_F\left(\pi_{B,i}, \rho_B\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right).
\end{aligned}$$

The Bregman identity (6) gives

$$D_F(\sigma_{AB}, \rho_A \otimes \rho_B) = \sum t_i \cdot D_F\left(\pi_{A,i}, \sigma_A\right) + D_F(\sigma_A, \rho_A) + \sum t_i \cdot D_F\left(\pi_{B,i}, \rho_B\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right).$$

This can be re-arranged as

$$D_F(\sigma_{AB}, \rho_A \otimes \rho_B) = \sum t_i \cdot \left(D_F\left(\pi_{A,i}, \sigma_A\right) + D_F\left(\pi_{B,i}, \rho_B\right)\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right) + D_F(\sigma_A, \rho_A).$$

Now additivity leads to

$$D_F(\sigma_{AB}, \rho_A \otimes \rho_B) = \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_A \otimes \rho_B\right) - \sum t_i \cdot D_F\left(\pi_{A,i} \otimes \pi_{B,i}, \sigma_{AB}\right) + D_F(\sigma_A, \rho_A) = D_F(\sigma_{AB}, \sigma_A \otimes \rho_B) + D_F(\sigma_A, \rho_A).$$

□

Definition 7 We define the *conditional mutual information* on a tripartite system as

$$I_\sigma(A; B \mid C) = D_F(\sigma_{ABC}, \sigma_A \otimes \sigma_B \otimes \sigma_C) - D_F(\sigma_{AC}, \sigma_A \otimes \sigma_C) - D_F(\sigma_{BC}, \sigma_B \otimes \sigma_C).$$

In our definition of conditional mutual information, the subsystems A, B, and C should be distinct so that the tensor products are defined. If the state space is a simplex, i.e. the system is classical, then one may allow the subsystems to overlap.

Definition 8 A function I_σ on a multipartite system is called a *separoid function* [5,9] if it satisfies the following three properties:

$$\begin{array}{ll} \textbf{Positivity} & I_\sigma(A; B \mid C) \geq 0. \\ \textbf{Symmetry} & I_\sigma(A; B \mid C) = I_\sigma(B; A \mid C). \\ \textbf{Chain rule} & I_\sigma(A; BC \mid D) = I_\sigma(A; B \mid D) \\ & \qquad + I_\sigma(A; C \mid BD). \end{array} \tag{9}$$

Theorem 8 *Assume that D_F is a monotone and additive Bregman divergence. Then conditional mutual information is a separoid function.*

Proof **Positivity** Conditional mutual information can be rewritten as

$$\begin{aligned} I_\sigma(A; B \mid C) &= D_F(\sigma_{ABC}, \sigma_A \otimes \sigma_B \otimes \sigma_C) - D_F(\sigma_{AC}, \sigma_A \otimes \sigma_C) \\ &\quad - D_F(\sigma_{BC}, \sigma_B \otimes \sigma_C) = D_F(\sigma_{ABC}, \sigma_B \otimes \sigma_{AC}) \\ &\quad - D_F(\sigma_{BC}, \sigma_B \otimes \sigma_C) = D_F(\sigma_{ABC}, \sigma_B \otimes \sigma_{AC}) \\ &\quad - D_F(\sigma_A \otimes \sigma_{BC}, \sigma_A \otimes \sigma_B \otimes \sigma_C). \end{aligned}$$

Let Φ denote the affinity $\Phi(\rho) = \sigma_A \otimes E_{BC}(\rho)$. Then

$$\Phi(\sigma_{ABC}) = \sigma_A \otimes \sigma_{BC}, \quad \Phi(\sigma_B \otimes \sigma_{AC}) = \sigma_A \otimes \sigma_B \otimes \sigma_C.$$

Therefore, monotonicity implies that $I_\sigma(A; B \mid C)$ cannot be negative.

Symmetry It follows directly from the definition that conditional mutual information is symmetric.

Chain rule To prove the chain rule, we expand the left-hand side of Eq. (9) as

$$I_\sigma(A; BC \mid D) = D_F(\sigma_{ABCD}, \sigma_A \otimes \sigma_{BC} \otimes \sigma_D) - D_F(\sigma_{AD}, \sigma_A \otimes \sigma_D) - D_F(\sigma_{BCD}, \sigma_{BC} \otimes \sigma_D).$$

Next we use Eq. (8) to get

$$I_\sigma(A;BC\mid D)=\begin{pmatrix}D_F(\sigma_{ABCD},\sigma_A\otimes\sigma_B\otimes\sigma_C\otimes\sigma_D)\\-D_F(\sigma_{BC},\sigma_B\otimes\sigma_C)\end{pmatrix}$$
$$-\;D_F(\sigma_{AD},\sigma_A\otimes\sigma_D)-\begin{pmatrix}D_F(\sigma_{BCD},\sigma_B\otimes\sigma_C\otimes\sigma_D)\\-D_F(\sigma_{BC},\sigma_B\otimes\sigma_C)\end{pmatrix}.$$

The left-hand side reduces to

$$I_\sigma(A;BC\mid D)=D_F(\sigma_{ABCD},\sigma_A\otimes\sigma_B\otimes\sigma_C\otimes\sigma_D)-D_F(\sigma_{AD},\sigma_A\otimes\sigma_D)-D_F(\sigma_{BCD},\sigma_B\otimes\sigma_C\otimes\sigma_D). \tag{10}$$

Similarly, we expand the right-hand side of Eq. (9) as

$$\begin{aligned}&I_\sigma(A;B\mid D)+I_\sigma(A;C\mid BD)\\&=D_F(\sigma_{ABD},\sigma_A\otimes\sigma_B\otimes\sigma_D)-D_F(\sigma_{AD},\sigma_A\otimes\sigma_D)-D_F(\sigma_{BD},\sigma_B\otimes\sigma_D)+D_F(\sigma_{ABCD},\sigma_A\otimes\sigma_C\otimes\sigma_{BD})\\&\quad-D_F(\sigma_{ABD},\sigma_A\otimes\sigma_{BD})-D_F(\sigma_{BCD},\sigma_C\otimes\sigma_{BD}).\end{aligned}$$

We use Eq. (8) to re-write the three last terms as

$$\begin{aligned}&I_\sigma(A;B\mid D)+I_\sigma(A;C\mid BD)\\&=D_F(\sigma_{ABD},\sigma_A\otimes\sigma_B\otimes\sigma_D)-D_F(\sigma_{AD},\sigma_A\otimes\sigma_D)-D_F(\sigma_{BD},\sigma_B\otimes\sigma_D)\\&\quad+\begin{pmatrix}D_F(\sigma_{ABCD},\sigma_A\otimes\sigma_B\otimes\sigma_C\otimes\sigma_D)\\-D_F(\sigma_{BD},\sigma_B\otimes\sigma_D)\end{pmatrix}-\begin{pmatrix}D_F(\sigma_{ABD},\sigma_A\otimes\sigma_B\otimes\sigma_D)\\-D_F(\sigma_{BD},\sigma_B\otimes\sigma_D)\end{pmatrix}\\&\quad-\begin{pmatrix}D_F(\sigma_{BCD},\sigma_B\otimes\sigma_C\otimes\sigma_D)\\-D_F(\sigma_{BD},\sigma_B\otimes\sigma_D)\end{pmatrix}.\end{aligned}$$

The right-hand side reduces to

$$\begin{aligned}&I_\sigma(A;B\mid D)+I_\sigma(A;C\mid BD)\\&=D_F(\sigma_{ABCD},\sigma_A\otimes\sigma_B\otimes\sigma_C\otimes\sigma_D)-D_F(\sigma_{AD},\sigma_A\otimes\sigma_D)-D_F(\sigma_{BCD},\sigma_B\otimes\sigma_C\otimes\sigma_D).\end{aligned} \tag{11}$$

Since the left-hand side (10) and the right-hand side (11) are equal, we have proved the chain rule (9). □

7 Conclusion

We have carefully described concepts like state space and introduced state spaces on Jordan algebras as the most important example. In general probabilistic theories, there are different ways of defining the entropy of a state, but these different definitions coincide on Jordan algebras. For any optimization problem there is an associated Bregman divergence, but with extra constraints like monotonicity, sufficiency, or statistical locality a Bregman divergence on a Jordan algebra is proportional to the Bregman divergence generated by the uniquely defined entropy function. A monotone Bregman divergence on a Jordan algebra is automatically additive. For composed systems, an additive and monotone Bregman divergence can be used to define conditional mutual information and this quantity will satisfy consistency, the data processing inequality and the chain rule. In [16], it was proved that if conditional mutual information can be defined in a way such that consistency, the data processing inequality and the chain rule are satisfied then the system will satisfy the condition called *information causality* [16]. In [16], it was also proved that a system that satisfies information causality cannot have super-quantum correlations, i.e. correlations violate Tsirelson's bound. The conclusion is that the existence of a monotone Bregman divergence implies that super-quantum correlations do not exist.

The results work out nicely on Jordan algebras, but maybe it will work in any generalized probabilistic theory. For instance, it would be interesting if the following conjecture holds.

Conjecture 1 All monotone Bregman divergences are additive.

A careful inspection of the proofs also reveal that the results involving Jordan algebras only involve that the cone is self-dual and that a Euclidean Jordan algebra is *strongly spectral* in the sense that $f(\sigma)$ is well defined for any function f. Apparently, monotonicity of a Bregman divergence implies spectrality, but the only solid result in this direction is the following theorem.

Theorem 9 [10] *If a state space has rank 2 and it has a strict and monotone Bregman divergence then the state space can be represented as a spin factor. In particular, the state space is strongly spectral.*

For most convex bodies, it is not possible to define a monotone Bregman divergence and it is not known if it is possible to define a monotone Bregman divergences on any convex body that cannot be represented by a Jordan algebra. It would be highly desirable to classify state spaces with monotone Bregman divergences in cases when the rank exceeds 2.

Compliance with ethical standards

Conflict of interest The corresponding author states that there is no conflict of interest.

References

1. Adler, S.: Quaternionic Quantum Mechanics and Quantum Fields. Oxford University Press, New York (1995)
2. Baes, M.: Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. Linear Algebra Appl. **422**, 664–700 (2007). https://doi.org/10.1016/j.laa.2006.11.025
3. Barnum, H., Graydon, M., Wilce, A.: Composites and categories of Euclidean Jordan algebras (2016). ArXiv preprint arXiv:1606.09331
4. Barnum, H., Hilgert, J.: Strongly symmetric spectral convex bodies are Jordan algebra state spaces (2019)
5. Dawid, A.P.: Separoids: a mathematical framework for conditional independence and irrelevance. Ann. Math. Artif. Intell. **32**, 335–372 (2001)
6. Günaydin, M., Gürsey, F.: Quark structure and octonions. J. Math. Phys. **14**(11), 1651–1667 (1973)
7. Harremoës, P.: Divergence and sufficiency for convex optimization. Entropy **19**(5), Article no. 206 (2017). https://doi.org/10.3390/e19050206
8. Harremoës, P.: Maximum entropy and sufficiency. AIP Conf. Proc. **1853**(1), 040001 (2017). https://doi.org/10.1063/1.4985352
9. Harremoës, P.: Entropy inequalities for lattices. Entropy **20**, 748 (2018). https://doi.org/10.3390/e20100784
10. Harremoës, P.: Entropy on spin factors. In: Ay, N., Gibilisco, P., Matúš, F. (eds.) Information Geometry and its Applications. Springer Proceedings in Mathematics & Statistics, vol. 252, pp. 247–278. Springer, Berlin (2018)
11. Jordan, P., von Neumann, J., Wigner, E.: On an algebraic generalization of the quantum mechanical formalism. Ann. Math. **35**(1), 29–64 (1934). https://doi.org/10.2307/1968117.JSTOR1968117
12. Krumm, M., Barnum, H., Barrett, J., Müller, M.P.: Thermodynamics and the structure of quantum theory. New J. Phys. **19**(4), 043025 (2017). https://doi.org/10.1088/1367-2630/aa68ef
13. Manogue, C.A., Dray, T.: Octonions, e6, and particle physics. In: Journal of Physics: Conference, p. 012005 (2010)
14. McCrimmon, K.: A Taste of Jordan Algebras. Springer, Berlin (2004)
15. Müller-Hermes, A., Reeb, D.: Monotonicity of the quantum relative entropy under positive maps. Ann. Henri Poincaré **18**(5), 1777–1788 (2017). https://doi.org/10.1007/s00023-017-0550-9
16. Pawlowski, M., Paterek, T., Kaszlikowski, D., Scarani, V., Winter, A., Zukowski, M.: Information causality as a physical principle. Nature **461**, 1101–1104 (2009). https://doi.org/10.1038/nature08400
17. Short, A.J., Wehner, S.: Entropy in general physical theories. New J. Phys. **12**, 033023 (2010)
18. Ullman, J.D.: Principles of Database and Knowledge-base Systems, vol. 1. Computer Science Press, Stanford (1989)