Linear Algebra and its Applications 616 (2021) 125–138

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Metrics induced by Jensen-Shannon and related divergences on positive definite matrices ☆

Suvrit Sra

Laboratory for Information and Decision Systems, MIT, United States of America

ARTICLE INFO

Article history:
Received 20 December 2019
Accepted 18 December 2020
Available online 13 January 2021
Submitted by P. Semrl

MSC:
15A45
52A99
47B65
65F60

Keywords:
Jensen-Shannon divergence
Jensen-Rényi divergence
Quantum information theory
Triangle inequality
Positive definite matrices

ABSTRACT

We study metric properties of symmetric divergences on Hermitian positive definite matrices. In particular, we prove that the square root of these divergences is a distance metric. As a corollary we obtain a proof of the metric property for Quantum Jensen-Shannon-(Tsallis) divergences (parameterized by $\alpha \in [0, 2]$). When specialized to $\alpha = 1$, we obtain as a corollary a proof of the metric property of the Quantum Jensen-Shannon divergence that was conjectured by Lamberti et al. (2008) [13], and recently also proved by Virosztek (2019) [28]. A more intricate argument also establishes metric properties of Jensen-Rényi divergences (for $\alpha \in (0, 1)$); this argument develops a technique that may be of independent interest.

1. Introduction

We study metric properties of symmetrized divergence measures on hermitian positive definite (hpd) matrices. Such divergence measures are widely useful, ranging from

☆ Submitted to the editors Dec 14, 2019.

E-mail address: suvrit@mit.edu.

https://doi.org/10.1016/j.laa.2020.12.023

quantum information theory [3,8,13], to optimization [22,25], to machine learning and computer vision [7,29,31], among others [3–5,10,11,14,16,19,21,30].

Our focus on studying metric properties of these divergences was inspired by the aim to build a theory that answers Conjecture 1.1 as a special case. A secondary aim is to obtain a family of metrics closely related to the S-Divergence [21] (which has found a variety of applications), and thus obtain a new family of potentially useful metrics. Remarkably, the metric property of the S-Divergence, which was the central result of [21], plays a crucial role in the present paper too.

The divergence underlying our primary aim is obtained by symmetrizing a Bregman divergence (see Section 2), or equivalently, by using midpoint convexity. For instance, consider the *von Neumann entropy*

$$S(X) := -\operatorname{tr}(X\log X), \qquad X \in \mathbb{P}_d, \tag{1.1}$$

which leads to the so-called the *Quantum Jensen-Shannon divergence* [13]:

$$\mathrm{QJSD}(X,Y) := S\left(\tfrac{X+Y}{2}\right) - \tfrac{1}{2}\big(S(X)+S(Y)\big). \tag{1.2}$$

Divergence (1.2) has found a variety of applications, including several cited above. While it is clearly symmetric and nonnegative, it is not a true distance; nevertheless, empirically its square root $\mathrm{QJSD}^{1/2}$ has been long observed to satisfy the triangle inequality [3,13].

A formal study $\mathrm{QJSD}^{1/2}$ as a metric was started by Lamberti et al. [13], who used it for measuring distances between quantum states. They also showed that QJSD is the square of a metric for pure states. Shortly thereafter, Briët and Harremoës [3] claimed that (3.1) is the square of a Hilbertian metric for qubits and pure states of any dimension; their proof, apparently contains an error, and a proof was furnished by Carlen, Lieb and Seiringer—please see [28, §3] for more details. For general quantum states, the work [28] (which appeared 2 weeks before a version of this paper appeared online [24]) furnished a proof of Conjecture 1.1). Our work is completely independent of [28], and it recovers not only Conjecture 1.1 as a special case, but also proves the metric properties of Jensen-Rényi divergence, a task that proves to be more challenging.

Specifically, Lamberti et al. had made the following conjecture:

Conjecture 1.1 *(Lamberti et al. [13]).* $\mathrm{QJSD}^{1/2}$ *is a metric on* $\mathbb{P}_d$ *(see also [3]).*

1.1. Summary of contributions

The main contributions of this paper are as follows:

- We prove in Theorem 3.1 the metric property for a rich class of Jensen divergences (please see Section 2 for background). This class includes QJSD_α ($\alpha \in [0,2]$) as a special case, and thus *a fortiori* also includes QJSD, yielding another proof of Conjecture 1.1 (the first publicly circulated proof of this conjecture is due to [28]). Moreover, our proof extends to more general settings based on certain convex functions (Theorem 3.6). Both Theorems 3.1 and 3.6 rely on integrals related to Pick functions (Section 5).

- In Section 4 we prove the Jensen-Shannon divergence generated by the α-Tsallis relative entropy is also the square of a metric.
- Finally, in Theorem 6.3 we prove the harder result that the quantum Jensen-Rényi divergence QJRD_α is the square of a metric for $\alpha \in (0, 1)$. Our technique relies on integral representations of completely monotonic functions and an argument based on 3×3 matrices that may be of independent interest.

2. Background

We begin by recalling some basic facts about divergences. Perhaps the most well-known divergence is the *Bregman divergence* [6],[1] which is generated by differentiable and convex function $f : \mathbb{R}^n \to \mathbb{R}$ as follows,

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y),\, x - y\rangle. \tag{2.1}$$

By construction, $D_f(x, y)$ is nonnegative, convex in x, and equals 0 if $x = y$. It is typically asymmetric and does not satisfy the triangle inequality, which explains the name "divergence" as opposed to "distance."

Example 2.1. Some common Bregman divergences are listed below.

- **Squared ℓ_2-distance**: Let $f(x) = \frac{1}{2}x^Tx$, then $D_f(x, y) = \frac{1}{2}\|x - y\|_2^2$.
- **KL divergence** on $\mathbb{R}_{++}$. $f(x) = x \log x$, so $D_f(x, y) = x \log(x/y) - x + y$.
- **Burg divergence** on $\mathbb{R}_{++}$. $f(x) = -\log x$, so $D_f(x, y) = \log(y/x) + x/y - 1$.

The Bregman divergence (2.1) extends naturally to hermitian matrices. Let X, Y be hermitian, and let the scalar function f be defined on hermitian matrices the usual way (via spectral decomposition), then the *Bregman matrix divergence* is defined as

$$D_f(X, Y) := \operatorname{tr} f(X) - \operatorname{tr} f(Y) - \langle f'(Y),\, X - Y\rangle. \tag{2.2}$$

It is an instructive exercise to verify that $D_f(X, Y) \geq 0$.

Example 2.2. The matrix versions of Example 2.1 are:

[1] Bregman divergences over scalars and vectors have been well-studied; see e.g., [1,6]. They are called divergences because they are not distances.

- **Squared Frobenius**: here $\operatorname{tr} f(X) = \operatorname{tr}(X^2)$, so that $D_f(X,Y) = \frac{1}{2}\|X - Y\|_{\mathrm{F}}^2$.
- **von Neumann divergence (Umegaki relative entropy)**: here $\operatorname{tr} f(X) = \operatorname{tr}(X \log X)$, so $D_f(X,Y)$ yields the von Neumann divergence of quantum information theory [17]:

$$D_{\mathrm{vN}}(X,Y) = \operatorname{tr}(X \log X - X \log Y - X + Y).$$

- **Stein's loss**: here $\operatorname{tr} f(X) = -\log\det(X)$, so $D_f(X,Y)$ becomes

$$D_{\ell d}(X,Y) = \operatorname{tr}(Y^{-1}(X - Y)) - \log\det(XY^{-1}),$$

which is also known as the *LogDet Divergence* [12], or *Stein's loss* [26].

2.1. Jensen and Jensen-Shannon divergences

Although Bregman divergences are widely useful, their asymmetry can be undesirable. A popular symmetric alternative is the *Jensen divergence* (sometimes called Jensen-Bregman divergence [7]):

$$S_f(X,Y) := \tfrac{1}{2}\big(D_f(X, \tfrac{X+Y}{2}) + D_f(Y, \tfrac{X+Y}{2})\big). \tag{2.3}$$

This divergence has two possibly more transparent representations:

$$S_f(X,Y) = \tfrac{1}{2}\big[\operatorname{tr} f(X) + \operatorname{tr} f(Y)\big] - \operatorname{tr} f\big(\tfrac{X+Y}{2}\big), \tag{2.4}$$

$$S_f(X,Y) = \min_Z \ \tfrac{1}{2}[D_f(X,Z) + D_f(Y,Z)]. \tag{2.5}$$

Remark 2.3. In some contexts (2.4) is also called the *Jensen-Shannon divergence*. But for clarity within the context of the quantum setting, we reserve that name for symmetrization (2.5) applied to D_f being a suitable quantum relative entropy.

Example 2.4. The symmetric versions of Example 2.2 are:

- If $\operatorname{tr} f(X) = \frac{1}{2}\operatorname{tr} X^2$, we obtain $S_f(X,Y) = D_f(X,Y) = \frac{1}{2}\|X - Y\|_{\mathrm{F}}^2$.
- If $\operatorname{tr} f(X) = \operatorname{tr}(X \log X)$, both (2.4) and (2.5) yield the QJSD (1.2).
- For $\operatorname{tr} f(X) = -\log\det(X) \equiv -\ell d(X)$, we obtain the *S-Divergence* [21]:

$$S_f(X,Y) \equiv S_{\ell d}(X,Y) := \delta_S^2(X,Y) := \ell d\left(\tfrac{X+Y}{2}\right) - \tfrac{1}{2}\ell d(X) - \tfrac{1}{2}\ell d(Y). \tag{2.6}$$

With this background, we are now ready to present the main results on this paper.

3. Metric properties of quantum Jensen divergences

In this section we study symmetric divergence whose square roots are metrics. The class of divergences covered is chosen to capture the α-Tsallis generalization to QJSD (1.2):

$$\mathrm{QJSD}_\alpha(X,Y) := S_\alpha\left(\tfrac{X+Y}{2}\right) - \tfrac{1}{2}\big(S_\alpha(X) + S_\alpha(Y)\big), \tag{3.1}$$

where S_α is the α-*Tsallis entropy*

$$S_\alpha(X) := \frac{\operatorname{tr}(X^\alpha) - \operatorname{tr} X}{1-\alpha}, \quad \alpha \in [0,2] \setminus \{1\}. \tag{3.2}$$

Note that, $\lim_{\alpha\to 1^+} S_\alpha(X) = S(X)$; and that S_α is concave on hpd matrices.

We present a technique which implies that $\mathrm{QJSD}_\alpha^{1/2}$ is a metric as a special case; this result in turn immediately yields a proof of Conjecture 1.1 (which corresponds to $\alpha \to 1^+$). Specifically, consider a function f on $(0,\infty)$ that can be written as:

$$f(x) = a + bx + c\log x + \int_0^\infty \log\frac{(t+x)}{h(t)}d\mu(t), \tag{3.3}$$

where $a, b \in \mathbb{R}$, $c \geq 0$, $h(t) > 0$, and μ is a nonnegative measure. This function is concave, so using it we can define a *"quantum" Jensen divergence* (on hpd matrices):

$$\Delta_f(X,Y) := \operatorname{tr}\big[f\big(\tfrac{X+Y}{2}\big) - \tfrac{1}{2}f(X) - \tfrac{1}{2}f(Y)\big]. \tag{3.4}$$

Our first main result is Theorem 3.1.

Theorem 3.1. *Let f be a function that admits the representation (3.3), and let Δ_f be the Jensen-divergence (3.4). Then, $\Delta_f^{1/2}$ is a distance on $\mathbb{P}_d$.*

Crucial to our proof is the metric property of the S-Divergence (2.6).

Theorem 3.2 *(Sra (2016) [21]). δ_S given by (2.6) is a metric.*

Proof of Theorem 3.1. The only non-trivial part is to prove the triangle inequality for $\Delta_f^{1/2}$. Using (3.3) and noting that $\operatorname{tr}\log(X) = \ell d(X)$ we can express Δ_f as

$$\begin{aligned}\Delta_f(X,Y) &= c\big(\ell d\left(\tfrac{X+Y}{2}\right) - \tfrac{1}{2}\ell d(X) - \tfrac{1}{2}\ell d(Y)\big)\\ &+ \int_0^\infty \big[\ell d(tI + \tfrac{X+Y}{2}) - \tfrac{1}{2}\ell d((tI+X)(tI+Y))\big]d\mu(t),\end{aligned}$$

which may be written in terms of the S-Divergence as

$$\Delta_f(X,Y) = c\delta_S^2(X,Y) + \int_0^\infty \delta_S^2(tI+X, tI+Y)d\mu(t). \tag{3.5}$$

Since $c \geq 0$, it follows from (3.5) that Δ_f is a non-negatively weighted sum of squared distances, therefore $\Delta_f^{1/2}$ satisfies the triangle inequality, completing the proof. □

Corollary 3.3. *Let* $\alpha \in (0,1)$. *Then,* $\mathrm{QJSD}_\alpha^{1/2}$ *is a metric.*

Proof. For $0 < \alpha < 1$ and $x > 0$, we use the integral representation introduced in [23]:

$$x^\alpha = \frac{\alpha \sin(\alpha\pi)}{\pi} \int_0^\infty \log\left(\tfrac{t+x}{t}\right) t^{\alpha-1} dt, \tag{3.6}$$

which is an instance of (3.3). Now, Theorem 3.1 immediately yields the corollary. □

While Corollary 3.3 holds for $\alpha \in (0,1)$, a slightly different integral representation allows us to also obtain the following result:

Corollary 3.4. *Let* $\alpha \in (1,2)$. *Then,* $\mathrm{QJSD}_\alpha^{1/2}$ *is a metric.*

Proof. The key idea is to use the following integral representation (for $1 < \alpha < 2$):

$$x^\alpha = \frac{|\alpha \sin(\alpha\pi)|}{\pi} \int_0^\infty (tx - \log(1+tx)) t^{-\alpha-1} dt, \tag{3.7}$$

which was also noted in [23]; notice that this representation is not captured by (3.3). Using (3.7) and arguing as for Theorem 3.1, the proof readily follows. □

Observing that $\lim_{\alpha\to 1^+} \mathrm{QJSD}_\alpha = \mathrm{QJSD}$, we obtain a proof for Conjecture 1.1.

Corollary 3.5. $\mathrm{QJSD}^{1/2}$ *is a metric on* $\mathbb{P}_d$.

3.1. Generalizing Corollary 3.4

The reader has perhaps already realized that the argument used to prove Corollary 3.4 also holds for convex functions on $(0, \infty)$ that admit the representation

$$f(x) = a + bx - c\log x + \int_0^\infty (tx - \log(1+tx))d\mu(t), \tag{3.8}$$

where $a, b \in \mathbb{R}$, $c \geq 0$, and μ is a nonnegative measure. Then we have the following:

Theorem 3.6. *Let $f(x)$ be given by (3.8), and define the Jensen divergence*

$$\Delta_f(X,Y) := \tfrac{1}{2}\operatorname{tr} f(X) + \tfrac{1}{2}\operatorname{tr} f(Y) - \operatorname{tr} f\left(\tfrac{X+Y}{2}\right). \tag{3.9}$$

Then, $\Delta^{1/2}$ is a distance function on $\mathbb{P}_d$.

4. The Jensen-Shannon α-Tsallis relative entropy

This section addresses a question posed by a referee of an earlier version. They remarked that the channel capacity interpretation of the *Jensen-Shannon Tsallis relative entropy* (4.2) makes its corresponding metric property more valuable than that of QJSD_α. However, as shown below, this property easily follows from that of QJSD_α.

Consider the α-*Tsallis relative entropy*

$$S_\alpha(X,Y) := \frac{\alpha\operatorname{tr} X + (1-\alpha)\operatorname{tr} Y - \operatorname{tr} X^\alpha Y^{1-\alpha}}{1-\alpha}, \quad \alpha \in (0,\infty)\setminus\{1\}.$$

The "centroid" of its symmetrization reduces to a power-mean; more precisely,

$$\operatorname*{argmin}_{Z\in\mathbb{P}_d} \quad S_\alpha(X,Z) + S_\alpha(Y,Z) = \left(\tfrac{X^\alpha+Y^\alpha}{2}\right)^{1/\alpha}. \tag{4.1}$$

Using (4.1), we arrive at the main result of this section.

Theorem 4.1. *Let the Jensen-Shannon α-Tsallis divergence be defined as*

$$\begin{aligned}\Delta_\alpha(X,Y) &:= S_\alpha(X,M) + S_\alpha(Y,M)\\ &= \tfrac{1}{1-\alpha}\left(\alpha\operatorname{tr}(X+Y) + 2(1-\alpha)\operatorname{tr} Z - \operatorname{tr}(Z^{1-\alpha}(X^\alpha+Y^\alpha))\right),\end{aligned} \tag{4.2}$$

where M denotes the rhs of (4.1). Then, for $\alpha \geq \frac{1}{2}$, $\Delta_\alpha^{1/2}$ is a metric.

Proof. We prove this metricity by reducing it to Corollaries 3.3 and 3.4. To that end, write $A = X^\alpha$ and B^α, and $t = 1/\alpha$. Then, we see that

$$\Delta_\alpha(X,Y) = \tfrac{1}{1-\alpha}\left(\alpha\operatorname{tr}(A^t+B^t) - 2\alpha\operatorname{tr}\left(\tfrac{A+B}{2}\right)^t\right). \tag{4.3}$$

For $\alpha \in [\frac{1}{2}, 1)$, we have $t \in (1, 2]$, whereby we can apply Corollary 3.4 to deduce metricity of $\Delta_\alpha^{1/2}$. For the case $\alpha > 1$, we have $t < 1$, and also $1 - \alpha < 0$. In this case, we can apply Corollary 3.3 to obtain metricity of $\Delta_\alpha^{1/2}$. □

Remark 4.2. Similarly, one could consider Jensen-Shannon versions of Rényi relative entropies such as the Petz-Rényi relative entropy [18] and the sandwiched Rényi relative entropy [15]; we defer such a discussion to the future.

5. Divergences and Pick functions

We briefly remark below on the deeper connection that motivates our choice (3.3). This connection also provides a valuable converse, namely, conditions under which such a representation holds for a given function.

In particular, in [20, Theorem 2.1] it was shown that if $xf'(x)$ has an analytic extension whose restriction to the upper half plane is a Pick function [9] and $xf'(x)$ is bounded, then f admits the representation (valid for $x > 0$):

$$f(x) = a + bx + c\log x + \frac{d}{x} + \int_0^\infty \Big[\log\frac{(t+x)}{(1+t)} - \frac{\log x}{1+t^2}\Big] d\mu(t), \tag{5.1}$$

with $a, c \in \mathbb{R}$, $b, d \geq 0$, and the nonnegative measure μ satisfies $\int_0^\infty t/(1+t^2)d\mu(t) < \infty$. Moreover, if in addition $f' \geq 0$, then (see [20, Thm. 4.2])

$$f(x) = a + bx + c\log x + \int_0^\infty \log\frac{t+x}{1+t} d\mu(t), \tag{5.2}$$

with $a \in \mathbb{R}$, $b, c \geq 0$ and $\int_0^\infty (1+t^2)^{-1}d\mu(t) < \infty$; this form is what motivates our slightly more general choice (3.3).

6. Quantum Jensen-Rényi divergence

Recall that the Quantum Rényi Entropy is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha}\log\frac{\mathrm{tr}(X^\alpha)}{\mathrm{tr}(X)}, \quad \alpha \geq 0, \alpha \neq 1. \tag{6.1}$$

Observe that (6.1) is concave for $\alpha \in (0, 1)$; thus, for such α we can define the *Quantum Jensen-Rényi Divergence* as:

$$\mathrm{QJRD}_\alpha := H_\alpha\left(\tfrac{X+Y}{2}\right) - \tfrac{1}{2}H_\alpha(X) - \tfrac{1}{2}H_\alpha(Y). \tag{6.2}$$

Proving that QJRD_α is the square of a metric (Theorem 6.3) turns out to be harder than analyzing QJSD_α. Indeed, QJRD_α is *not* directly amenable to the Pick function technique developed above, and it requires a more intricate argument that uses two more ingredients: *complete monotonicity* and the relation between *conditionally negative definite* matrices and metrics.

Definition 6.1 *(Complete monotonicity).* A function $F : (0,\infty) \to \mathbb{R}$ is called *completely monotonic (CM)* if $(-1)^k F^{(k)}(x) \geq 0$, for $k \geq 0$. Bernstein's theorem (see e.g., [27, Thm. 6.13]) shows that such an F can be written as

$$F(x) = \int_0^\infty e^{-tx} d\nu(t), \tag{6.3}$$

for a nonnegative measure ν on $[0,\infty)$.

Definition 6.2 *(CND). $X \in \mathbb{H}_d$ is conditionally negative definite (cnd)* if

$$v^* X v \leq 0, \quad \text{for all } v \in \mathbb{C}^d \text{ s.t. } v^* 1 = 0. \tag{6.4}$$

Theorem 6.3. $\mathrm{QJRD}_\alpha^{1/2}$ *is a metric on HPD matrices for $\alpha \in (0,1)$.*

Proof. Without loss of generality we may assume that $\mathrm{tr}(X) = 1$. Introduce now the shorthand $d_{xy} = \mathrm{tr}\left(\frac{X+Y}{2}\right)^\alpha$, $d_x = d_{xx}$; define d_{xz}, d_{yz}, and d_y, d_z similarly. Then, by Theorems 3.1 and A.1 it follows that the matrix

$$D = \begin{bmatrix} 0 & 2d_{xy} - d_x - d_y & 2d_{xz} - d_x - d_z \\ 2d_{xy} - d_x - d_y & 0 & 2d_{yz} - d_y - d_z \\ 2d_{xz} - d_x - d_z & 2d_{yz} - d_y - d_z & 0 \end{bmatrix},$$

is cnd. It is also known that (see e.g., [2, Thm. 4.4.2]) that if an elementwise nonnegative matrix $[m_{ij}]$ is cnd, and F is a CM function, then $[F(m_{ij})]$ is positive definite. Let $\boldsymbol{\theta} = [d_x, d_y, d_z]^T$; then, $M = \boldsymbol{\theta}\mathbf{1}^T + \mathbf{1}\boldsymbol{\theta}^T + D$ is also cnd, and so is $2t\mathbf{1}\mathbf{1}^T + M$ for all $t \geq 0$. Thus, using the CM function $F(s) = 2/s$ on M we see that

$$M' = \begin{bmatrix} \frac{1}{t+d_x} & \frac{1}{t+d_{xy}} & \frac{1}{t+d_{xz}} \\ \frac{1}{t+d_{xy}} & \frac{1}{t+d_y} & \frac{1}{t+d_{yz}} \\ \frac{1}{t+d_{xz}} & \frac{1}{t+d_{yz}} & \frac{1}{t+d_z} \end{bmatrix} \succeq 0. \tag{6.5}$$

Using $\boldsymbol{\eta} = \frac{1}{2}[t+d_x, t+d_y, t+d_z]^T$, we construct $\boldsymbol{\eta}\mathbf{1}^T + \mathbf{1}\boldsymbol{\eta}^T - M'$, which is clearly cnd. Explicitly, this matrix is given by (we suppress symmetric entries via $*$ for brevity):

$$\begin{bmatrix} 0 & \frac{1}{2}\left(\frac{1}{t+d_x} + \frac{1}{t+d_y}\right) - \frac{1}{t+d_{xy}} & \frac{1}{2}\left(\frac{1}{t+d_x} + \frac{1}{t+d_z}\right) - \frac{1}{t+d_{xz}} \\ * & 0 & \frac{1}{2}\left(\frac{1}{t+d_y} + \frac{1}{t+d_z}\right) - \frac{1}{t+d_{yz}} \\ * & * & 0 \end{bmatrix}.$$

Since $d_x = \mathrm{tr}(X^\alpha)$ is concave for $\alpha \in (0,1)$, it follows that $1/(t+d_x)$ is convex, whereby

$$\frac{1}{2}\Big(\frac{1}{t+d_x}+\frac{1}{t+d_y}\Big)-\frac{1}{t+d_{xy}} \geq 0.$$

Thus, we can invoke Theorem A.1 again to conclude that

$$\frac{1}{2}\Big(\frac{1}{t+d_x}+\frac{1}{t+d_y}\Big)-\frac{1}{t+d_{xy}} =: \delta_t^2(X,Y), \tag{6.6}$$

where δ_t is a distance metric. Next, recall the following integral representation

$$\log x = -\int_0^\infty \Big(\frac{1}{t+x}-\frac{t}{1+t^2}\Big)dt, \tag{6.7}$$

which can be obtained for instance by first writing $(\log x)^2$ using (5.1) and then differentiating [20]. Integrating (6.6) using representation (6.7) we can finally write

$$\mathrm{QJRD}_\alpha(X,Y) = \int_0^\infty \delta_t^2(X,Y)dt,$$

which proves that $\mathrm{QJRD}_\alpha^{1/2}$ is a metric. $\square$

6.1. Other extensions

The above proof actually also shows that if $d_{xy} = h\left(\frac{X+Y}{2}\right)$ where $h(X)$ is concave and $\Delta_h(X,Y) = d_{x,y} - \frac{1}{2}d_x - \frac{1}{2}d_y$ is the square of a metric, then

$$\Delta_F(X,Y) := \tfrac{1}{2}F(d_x) + \tfrac{1}{2}F(d_y) - F(d_{xy}),$$

is the square of a metric. Indeed, if h is concave, then e^{-th} is convex for $t \geq 0$. Thus, for a CM function F the map $F(h(X))$ is convex, whence $\Delta_F(X,Y) \geq 0$. The triangle inequality follows from a construction analogous to (6.5). We omit details for brevity.

7. Conclusions

In this paper, we identified sufficient conditions based on Pick-Nevanlinna integral representations for ensuring that the corresponding (quantum) Jensen divergence is the square of a metric. At this point, it is natural to consider the following (likely harder) task as an open problem:

Problem 7.1. Identify conditions that are both necessary and sufficient for a given (quantum) Jensen divergence to be the square of a metric.

Declaration of competing interest

None declared.

Acknowledgements

I would like to thank Dániel Virosztek for bringing [28] to my attention, and also for pointing out to me certain corrections in the attribution of the QJSD conjecture, as well as the reference to a proof by Carlen-Lieb-Seiringer (see [28, §3]) as the correct proof of metricity for qubits.

Appendix A. Distances and 3 × 3 cnd matrices*

In this section, we summarize the equivalence between 3 × 3 cnd matrices and corresponding distance metrics. This material is classical, but we include our proofs for keeping the description self-contained. Indeed, squared distances are intimately related with cnd matrices. The following theorem summarizes this connection.

Theorem A.1. *Let D be a 3×3 symmetric positive definite matrix that is elementwise nonnegative; we write explicitly*

$$D = \begin{bmatrix} d_x^2 & d_{xy}^2 & d_{xz}^2 \\ d_{xy}^2 & d_y^2 & d_{yz}^2 \\ d_{xz}^2 & d_{yz}^2 & d_z^2 \end{bmatrix}.$$

Let $\boldsymbol{\theta} = \frac{1}{2}[d_x^2, d_y^2, d_z^2]^T$ and define $M = \boldsymbol{\theta}\mathbf{1}^T + \mathbf{1}\boldsymbol{\theta}^T - D$. Then for the statements

(i) D is positive definite;
(ii) M is cnd and nonnegative; and
(iii) $d^2(x, y)$ is a squared metric,

the following claims hold: (i) $\implies$ (ii) $\iff$ (iii).

Proof. *(i) $\implies$ (ii)*: Immediate, as $x^T M x = -x^T D x \le 0$ for any $x \in \mathbb{R}^3$ such that $x^T\mathbf{1} = 0$.

(ii) $\implies$ (iii): First, consider nonnegativty. It suffices to discuss α; the others follow similarly.

$$\alpha = \tfrac{1}{2}(d_x^2 + d_y^2) - d_{xy}^2 \ge 0 \quad \Leftrightarrow \quad d_{xy}^2 \le \tfrac{1}{2}(d_x^2 + d_y^2).$$

But D is psd, whereby $d_{xy}^4 \le d_x^2 d_y^2$.

What remains to show is that

$$M = \begin{bmatrix} 0 & \alpha & \beta \\ \alpha & 0 & \gamma \\ \beta & \gamma & 0 \end{bmatrix} \quad \text{is cnd,}$$

where α, β, γ are shorthand for the actual entries of M. For the vector $x = [-s-t, s, t]$, we have

$$-\tfrac{1}{2}x^T Mx = \alpha s^2 + st(\alpha + \beta - \gamma) + \beta t^2 \geq 0. \tag{A.1}$$

From this inequality we need to deduce that

$$\alpha^{1/2} \leq \beta^{1/2} + \gamma^{1/2}, \tag{A.2}$$

$$\beta^{1/2} \leq \alpha^{1/2} + \gamma^{1/2}, \tag{A.3}$$

$$\gamma^{1/2} \leq \alpha^{1/2} + \beta^{1/2}. \tag{A.4}$$

Assume without loss of generality that γ is the largest. Then, if we can prove that $\gamma^{1/2} \leq \alpha^{1/2} + \beta^{1/2}$, the other inequalities follow immediately since $\alpha, \beta \geq 0$. To that end, we can equivalently show that

$$\gamma \leq \alpha + \beta + 2\sqrt{\alpha\beta} \quad \Leftrightarrow \quad \alpha + \beta - \gamma \geq -2\sqrt{\alpha\beta}. \tag{A.5}$$

Let us see how to deduce (A.5) from (A.1), which says that

$$(\alpha + \beta - \gamma)st \geq -s^2\alpha - t^2\beta. \tag{A.6}$$

In particular, let $s^2 = \sqrt{\beta}/\sqrt{\alpha}$ and $t^2 = \sqrt{\alpha}/\sqrt{\beta}$; this yields $st = 1$ and $s^2\alpha + t^2\beta = 2\sqrt{\alpha\beta}$, so that (A.6) reduces to the desired inequality (A.4).

(iii) $\implies$ *(ii)*: Let $x = [-s-t, s, t]$ as before. We wish to show that $x^T Mx \leq 0$; we split this task into two subcases: *(a)* $st < 0$, and *(b)* $st > 0$.

Case (a). Let α, β, γ be squared distances as before, with γ that largest. Then, from inequality (A.4) it follows that

$$\gamma^{1/2} \geq |\alpha^{1/2} - \beta^{1/2}| \quad \implies \gamma \geq \alpha + \beta - 2\sqrt{\alpha\beta}. \tag{A.7}$$

The only way to violate cnd property of M is to choose s and t such that $x^T Mx \geq 0$, or equivalently to show that

$$s^2\alpha + t^2\beta + st(\alpha + \beta - \gamma) \leq 0. \tag{A.8}$$

Since $st < 0$, (A.8) turns into

$$\alpha + \beta - \gamma \geq \theta\alpha + \frac{1}{\theta}\beta \geq 2\sqrt{\alpha\beta}.$$

That is, to break the cnd property of M we need to have

$$\alpha + \beta - 2\sqrt{\alpha\beta} \geq \gamma,$$

which contradicts (A.7).

Case (b). If $st > 0$, then from (A.4) it follows that

$$(\alpha + \beta - \gamma)st \geq -2st\sqrt{\alpha\beta} = -2\sqrt{s^2\alpha t^2\beta} \geq -s^2\alpha - t^2\beta.$$

But this inequality can not contradict the cnd property, as it is just inequality (A.6) analyzed above. Thus, in both cases, we obtain that M must be cnd. □

References

[1] A. Banerjee, S. Merugu, I.S. Dhillon, J. Ghosh, Clustering with Bregman divergences, in: SIAM International Conf. on Data Mining, Lake Buena Vista, Florida, SIAM, April 2004.

[2] R.B. Bapat, T. Raghavan, Nonnegative Matrices and Applications, vol. 64, Cambridge University Press, 1997.

[3] J. Briët, P. Harremoës, Properties of classical and quantum Jensen-Shannon divergence, Phys. Rev. A 79 (5) (2009) 052311.

[4] J. Burbea, C. Rao, On the convexity of some divergence measures based on entropy functions, IEEE Trans. Inf. Theory 28 (3) (1982) 489–495.

[5] J. Burbea, C.R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: a unified approach, J. Multivar. Anal. 12 (4) (1982) 575–596.

[6] Y. Censor, S.A. Zenios, Parallel Optimization: Theory, Algorithms, and Applications. Numerical Mathematics and Scientific Computation, Oxford University Press, 1997.

[7] A. Cherian, S. Sra, A. Banerjee, N. Papanikolopoulos, Jensen-Bregman LogDet divergence with application to efficient similarity search for covariance matrices, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2012) 2161–2174.

[8] J. Dajka, J. Łuczka, P. Hänggi, Distance between quantum states in the presence of initial qubit-environment correlations: a comparative study, Phys. Rev. A (ISSN 1094-1622) 84 (3) (Sep 2011), https://doi.org/10.1103/physreva.84.032120.

[9] W.F. Donoghue, Monotone Matrix Functions and Analytic Continuation, Springer, 1974.

[10] B. Fuglede, F. Topsoe, Jensen-Shannon divergence and Hilbert space embedding, in: International Symposium on Information Theory, Proceedings, ISIT 2004, 2004, IEEE, 2004, p. 31.

[11] M. Harandi, M. Salzmann, R. Hartley, Dimensionality reduction on SPD manifolds: the emergence of geometry-aware methods, IEEE Trans. Pattern Anal. Mach. Intell. 40 (1) (2017) 48–62.

[12] B. Kulis, M. Sustik, I. Dhillon, Low-rank kernel learning with Bregman matrix divergences, J. Mach. Learn. Res. 10 (2009) 341–376.

[13] P. Lamberti, A. Majtey, A. Borras, M. Casas, A. Plastino, Metric character of the quantum Jensen-Shannon divergence, Phys. Rev. A 77 (5) (2008) 052311.

[14] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory 37 (1) (1991) 145–151.

[15] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, M. Tomamichel, On quantum Rényi entropies: a new generalization and some properties, J. Math. Phys. 54 (12) (2013) 122203, https://doi.org/10.1063/1.4838856.

[16] F. Nielsen, S. Boltz, The Burbea-Rao and Bhattacharyya centroids, IEEE Trans. Inf. Theory 57 (8) (2011) 5455–5466.

[17] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2000.

[18] D. Petz, Quasi-entropies for finite quantum systems, Rep. Math. Phys. 23 (1) (1986) 57–65.

[19] L. Rossi, A. Torsello, E.R. Hancock, R.C. Wilson, Characterizing graph symmetries through quantum Jensen-Shannon divergence, Phys. Rev. E 88 (3) (2013) 032806.
[20] M. Šilhavý, A functional inequality related to analytic continuation, Technical report, Preprint Institute of Mathematics AS CR IM-2015-37, 2015, www.math.cas.cz/fichier/preprints/IM_20150623102729_44.pdf.
[21] S. Sra, Positive definite matrices and the s-divergence, Proc. Am. Math. Soc. 144 (7) (2016) 2787–2797.
[22] S. Sra, On the matrix square root via geometric optimization, Electron. J. Linear Algebra 31 (1) (2016) 433–443.
[23] S. Sra, Logarithmic inequalities under a symmetric polynomial dominance order, Proc. Am. Math. Soc. 147 (2) (2019) 481–486.
[24] S. Sra, Metrics induced by quantum Jensen-Shannon-Renyí and related divergences, arXiv:1911.02643, 2019.
[25] S. Sra, R. Hosseini, Conic geometric optimization on the manifold of positive definite matrices, SIAM J. Optim. 25 (1) (2015) 713–739.
[26] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate distribution, in: Proc. Third Berkeley Symp. Math. Statist. Prob., vol. 1, 1956, pp. 197–206.
[27] C. van den Berg, J.P.R. Christensen, P. Ressel, Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions, Springer, 1984.
[28] D. Virosztek, The metric property of the quantum Jensen-Shannon divergence, arXiv:1910.10447, 2019.
[29] L. Wang, J. Zhang, L. Zhou, C. Tang, W. Li, Beyond covariance: feature representation with nonlinear kernel matrices, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4570–4578.
[30] F. Yger, M. Berar, F. Lotte, Riemannian approaches in brain-computer interfaces: a review, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (10) (2016) 1753–1762.
[31] J. Zhang, L. Wang, L. Zhou, W. Li, Learning discriminative Stein kernel for SPD matrices and its applications, IEEE Trans. Neural Netw. Learn. Syst. 27 (5) (2015) 1020–1033.