TSALLIS METRICS ON DENSITY MATRICES VIA THE S-DIVERGENCE

SUVRIT SRA*

ABSTRACT. We study symmetrized divergences induced by centroids of α-Tsallis relative entropies between density matrices σ and ρ; up to scaling, the induced divergence equals $|2 - \|\rho^\alpha + \sigma^\alpha\|_{1/\alpha}|$. Earlier work (S. Sra. *"Metrics induced by Jensen–Shannon and related divergences on positive definite matrices."* Linear Algebra Appl., 2021) studied such divergences and metrics induced by them. That paper allowed centroids to range over the positive definite cone, while we require that the centroid itself be a density matrix. This (more natural) choice makes the metricity in this paper considerably stronger than that of the aforementioned paper. Specifically, we prove that $|2 - \|\rho^\alpha + \sigma^\alpha\|_{1/\alpha}|^{1/2}$ is a metric on density matrices for $1/2 \le \alpha \le 2$, $\alpha \neq 1$. For $\alpha < 1$, the proof uses a new four-point theorem for the S-divergence that might be of independent interest. For $\alpha > 1$, we first prove the endpoint $\alpha = 2$, and then transfer it to the other orders by a tensoring trick.

hexagon:2610.00602v1

1. INTRODUCTION

We continue the study of metric properties of symmetrized divergences on quantum states started in (Sra, 2021b). Square roots of symmetrized divergences sometimes turn out to be metrics, for instance Jensen–Shannon divergences for commuting states (Endres and Schindelin, 2003; Fuglede and Topsøe, 2004). For noncommuting states Lamberti et al. (2008) conjectured the metric property, which was proved by both Sra (2021b) and Virosztek (2021) independently.

We study herein a symmetrized version of Tsallis relative entropy obtained by computing centroids, and ask whether the resulting divergence is a metric. We denote via $\mathbb{P}^d$ ($\mathbb{P}^d_+$) the set of $d \times d$ positive semidefinite (definite) Hermitian matrices; and let $\mathcal{D}_d := \{\rho \in \mathbb{P}^d : \operatorname{tr} \rho = 1\}$ be the density matrices, or states. For $\rho, \tau \in \mathcal{D}_d$, the Tsallis relative entropy (Tsallis, 1988; Petz, 1986) is

$$T_\alpha(\rho, \tau) := \frac{1 - \operatorname{tr} \rho^\alpha \tau^{1-\alpha}}{1 - \alpha}, \qquad \alpha > 0, \quad \alpha \neq 1. \tag{1.1}$$

The symmetrized divergence obtained by computing the Tsallis centroid of ρ and σ is then:

$$d_\alpha^2(\rho, \sigma) := |1 - \alpha| \min_{\tau \in \mathcal{D}_d} \{T_\alpha(\rho, \tau) + T_\alpha(\sigma, \tau)\} \quad = \quad \big|2 - \|\rho^\alpha + \sigma^\alpha\|_{1/\alpha}\big|, \tag{1.2}$$

where $\|X\|_p = (\operatorname{tr} |X|^p)^{1/p}$ is the Schatten-p norm. The equality in (1.2) is the quantum Sibson identity (König and Wehner, 2009; Sharma and Warsi, 2013; Dalai, 2013); Proposition 3.1 verifies it for $0 < \alpha < 1$ by Hölder's inequality; the reverse Hölder inequality gives the case $\alpha > 1$ (Section 4).

It is evident that d_α is symmetric and nonnegative. The challenge is to show that it satisfies the triangle inequality. When the center τ in (1.2) is allowed to range over all $\mathbb{P}^d$, the minimum value is $2\alpha(1 - h)$, where $h = \operatorname{tr}((\rho^\alpha + \sigma^\alpha)/2)^{1/\alpha}$, and its square root is a metric for $1/2 \le \alpha < 1$ (Sra, 2021b, Theorem 4.1). But in (1.2), we require the center to lie in $\mathcal{D}_d$, which changes the minimum to $d_\alpha^2 = 2(1 - h^\alpha)$. It turns out to be much more difficult to establish that d_α is a metric in this case. The proof of this metricity is our first main result.

2020 *Mathematics Subject Classification.* 15A42, 46L51, 47A63, 94A17.

Key words and phrases. Tsallis relative entropy, quantum Jensen–Shannon divergence, Sibson information radius, S-divergence, negative type, Schatten norm, quantum Rényi entropy, log-determinant inequality.

* TU Munich, Dept. of Mathematics, School of CIT; Garching, Germany. Email: s.sra@tum.de.

First public release: 13 October 2026, as v1 in the author's preprint repository, https://github.com/suvrit/preprints/tree/main/preprints/tsallis; every version is kept there with its timestamp.

Theorem 1.1. *For every d and $1/2 \le \alpha \le 2$, $\alpha \neq 1$, d_α is a metric on $\mathcal{D}_d$.*

Theorem 1.1 also implies the previous results of Sra (2021b); indeed, as $\alpha \uparrow 1$, $d_\alpha^2/(2(1-\alpha)) \to$ QJSD, which gives the Jensen–Shannon metric (Corollary A.3), and the proof gives the cone metric of Sra (2021b, Theorem 4.1). At $\alpha = 1/2$, the distance is also of negative type (Proposition 3.5).

For $\alpha < 1$, with $p = 1/\alpha$, Theorem 1.1 follows from positivity of the trace-power kernel $\operatorname{tr}(A+B)^p - \operatorname{tr} A^p - \operatorname{tr} B^p$ on every three positive semidefinite matrices (Theorem 3.2). For $1 < p < 2$, we write this kernel as a superposition of the kernels $-\log\det(I - XY)$ on strict contractions X and Y in $\mathbb{P}^d$, which turns our task into proving that $[-\log\det(I - X_iX_j)]_{i,j=1}^3$ is positive semidefinite. This task is nontrivial, and its proof uncovers a hitherto unknown property of the S-divergence (Sra, 2016), denoted δ_S^2 below, which leads to our second main result.

Theorem 1.2. *Let U_1, U_2, U_3 be Hermitian $d \times d$ matrices with $\|U_i\|_{\mathrm{op}} < 1$. Then,*

$$K := \left[-\log\det(I - U_iU_j)\right]_{i,j=1}^3 \succeq 0. \tag{1.3}$$

Equivalently, if $P_0, P_1, P_2, P_3 \in \mathbb{P}_+^d$ and $\xi \in \mathbb{R}^4$ satisfies $\xi_0 + \xi_1 + \xi_2 + \xi_3 = 0$, then

$$\sum_{a,c=0}^{3} \xi_a\xi_c\, \delta_S^2(P_a, P_c) \le 0. \tag{1.4}$$

The S-divergence δ_S^2 used in (1.4) is defined for any two matrices $P, Q \in \mathbb{P}_+^d$ as

$$\delta_S^2(P,Q) := \log\det\tfrac{P+Q}{2} - \tfrac{1}{2}\log\det P - \tfrac{1}{2}\log\det Q. \tag{1.5}$$

It is a Jensen–Bregman divergence (Cherian et al., 2013) and δ_S is a metric (Sra, 2016).

Recall that a symmetric function N with $N(x,x) = 0$ is *of negative type* if $\sum_{i,j} c_ic_jN(x_i,x_j) \le 0$ for every finite family of points and real coefficients with $\sum_i c_i = 0$. On three points, negative type of δ_S^2 is equivalent to the triangle inequality for δ_S, which is key to the results of Sra (2021b). However, the log-determinant inequality (1.3) needs four points. Indeed, using the Cayley transform $P_i = (I + U_i)(I - U_i)^{-1}$ we map Hermitian strict contractions U_i onto $\mathbb{P}_+^d$, and

$$-\log\det(I - U_iU_j) = \delta_S^2(P_i, I) + \delta_S^2(P_j, I) - \delta_S^2(P_i, P_j). \tag{1.6}$$

Hence positivity of $[-\log\det(I - U_iU_j)]_{i,j=1}^3$ is the same as negative type of δ_S^2 on the four matrices I, P_1, P_2, P_3. By Theorem 1.2 and Schoenberg's criterion (Schoenberg, 1935), every four-point subset of $(\mathbb{P}_+^d, \delta_S)$ embeds isometrically in $\mathbb{R}^3$. The number four is sharp when $d \ge 2$: Sra (2016) gave a numerical five-point obstruction in dimension two; Proposition 2.5 gives an analytic one.

Remark 1.3. *Theorem 1.2 also implies the positivity of the Hua–Bellman matrices. For strict contractions Z_1, Z_2, Z_3, Sra (2021a, Corollary 3.2) proves that the Hua–Bellman matrix (Hua, 1955) $H_\gamma = [|\det(I - Z_i^*Z_j)|^{-\gamma}]_{i,j=1}^3$ is positive semidefinite for every $\gamma \ge 0$. For Hermitian Z_i, Theorem 1.2 recovers this entire family (where K is as in (1.3)) by the Schur product theorem:*

$$H_\gamma = \sum_{m=0}^{\infty} \frac{\gamma^m}{m!} K^{\circ m} \succeq 0 \qquad (\gamma \ge 0),$$

where $K^{\circ m}$ denotes entrywise power. The converse direction gives less: differentiating $c^\top H_\gamma c$ at zero gives $c^\top Kc \ge 0$ only when $\sum_i c_i = 0$, whereas Theorem 1.2 gives $K \succeq 0$.

Remark 1.4. *Theorem 1.2 also applies to entropy-based Jensen divergences, which compare the entropy of $(\rho+\sigma)/2$ with the mean entropy of ρ and σ. The Jensen–Tsallis and Jensen–Rényi divergences of this form are of negative type on every four states for orders $0 < q \le 2$ and $0 < q \le 1$, respectively (Theorem B.1). This strengthens the metric results of Sra (2021b).*

Below, we first prove Theorem 1.2 as it forms the basis of the proof of Theorem 1.1.

2. Proof of Theorem 1.2: four-point negative type of the S-divergence

For $\xi \in \mathbb{R}^4$ with $\sum_a \xi_a = 0$ and $P_0, \dots, P_3 \in \mathbb{P}_+^d$ put

$$F_\xi(P_0, P_1, P_2, P_3) := -\sum_{a,c=0}^{3} \xi_a \xi_c \, \delta_S^2(P_a, P_c),$$

so that (1.4) says $F_\xi \geq 0$. Since F_ξ vanishes when the four matrices coincide, we show that zero is its minimum. We prove that a minimizer exists and that the only critical point is the trivial one.

2.1. Existence of minimizer.

We start with three elementary properties of δ_S that we will need.

Lemma 2.1. *Let $P, Q, R \in \mathbb{P}_+^d$.*
(a) $\delta_S^2(GPG^*, GQG^*) = \delta_S^2(P, Q)$ *for every invertible G.*
(b) $\delta_S^2(I, P) = \sum_{\ell=1}^{d} \log\cosh\big(\frac{1}{2}\log\lambda_\ell(P)\big)$, *where $\lambda_1(P), \dots, \lambda_d(P)$ are the eigenvalues of P.*
(c) $\delta_S^2(P, R) \leq \delta_S^2(P, Q) + \delta_S^2(Q, R) + d\log 2$.

By continuity, it suffices to treat ξ with nonzero coordinates and a unique largest absolute value, say $|\xi_0| > |\xi_j|$ for $j = 1, 2, 3$; by Lemma 2.1(a), we may take $P_0 = I$. The domain $(\mathbb{P}_+^d)^3$ of $F_\xi(I, \cdot)$ is not compact, so a minimizer need not exist. A coercive lower bound rules this out. The bound depends on ξ through $\nu_\xi := \min_{\varnothing \neq T \subsetneq \{0,1,2,3\}} \left|\sum_{a \in T} \xi_a\right|$, which is positive for the fixed ξ.

Proposition 2.2. *If $\nu_\xi > 0$, then $F_\xi(I, P_1, P_2, P_3) \geq 2\nu_\xi^2 \max_{1 \leq i \leq 3} \delta_S^2(I, P_i) - d\log 2 \sum_a \xi_a^2$. Consequently, $F_\xi(I, \cdot)$ attains its minimum on $(\mathbb{P}_+^d)^3$.*

Proof. The values $\delta_S^2(P_a, P_c)$ need not satisfy the triangle inequality, but by Lemma 2.1(b) and (c) they do after a shift: on $\{0, 1, 2, 3\}$, the function $D(a, c) = \delta_S^2(P_a, P_c) + d\log 2$ for $a \neq c$, with $D(a, a) = 0$, is a metric. Since $\sum_{a \neq c} \xi_a \xi_c = -\sum_a \xi_a^2$, the shift changes F_ξ only by a constant:

$$-\sum_{a,c} \xi_a \xi_c \, D(a, c) = F_\xi(I, P_1, P_2, P_3) + d\log 2 \sum_a \xi_a^2.$$

Metrics on four points are conic combinations of cut metrics (Deza and Laurent, 1997, p. 422), so,

$$D(a, c) = \sum_{\varnothing \neq T \subsetneq \{0,1,2,3\}} w_T \big(\mathbf{1}_T(a) - \mathbf{1}_T(c)\big)^2, \qquad w_T \geq 0,$$

where $\mathbf{1}_T$ is the indicator vector of T. Each cut contributes at least $2\nu_\xi^2 w_T$, because $\sum_a \xi_a = 0$ gives

$$-\sum_{a,c} \xi_a \xi_c \big(\mathbf{1}_T(a) - \mathbf{1}_T(c)\big)^2 = 2\Big(\sum_{a \in T} \xi_a\Big)^2 \geq 2\nu_\xi^2,$$

and $\sum_T w_T \geq \max_{a,c} D(a, c) \geq \max_i \delta_S^2(I, P_i)$ as each cut takes values in $\{0, 1\}$. Together, these prove the bound. Finally, if $F_\xi(I, P_1, P_2, P_3) \leq 0$, the bound caps every $\delta_S^2(I, P_i)$, so by Lemma 2.1(b) and $\log\cosh t \geq |t| - \log 2$, eigenvalues of each P_i lie in a compact interval that depends only on ξ and d. The triples with this property form a compact set containing (I, I, I), where F_ξ vanishes, and $F_\xi(I, \cdot) > 0$ outside it. Hence the minimum of $F_\xi(I, \cdot)$ over this set is its minimum over $(\mathbb{P}_+^d)^3$. □

2.2. Potential minimizers (critical points of F_ξ).

Since $(\mathbb{P}_+^d)^3$ is open, the minimizer of Proposition 2.2 is a critical point. Lemma 2.3 describes these critical points.

Lemma 2.3. *For the ξ fixed above, put $\lambda_j = -\xi_j/\xi_0$, so that $\lambda_1 + \lambda_2 + \lambda_3 = 1$, $\lambda_j \neq 0$, and $|\lambda_j| < 1$. If (P_1, P_2, P_3) is a critical point of $F_\xi(I, \cdot)$, then the coordinates $U_j = (P_j - I)(P_j + I)^{-1}$ satisfy*

$$\sum_{j=1}^{3} \lambda_j U_j = 0, \qquad \sum_{j=1}^{3} \lambda_j (I - U_i U_j)^{-1} = I \quad (i = 1, 2, 3). \tag{2.1}$$

Proof. Noting $\sum_a \xi_a = 0$, and expanding $\delta_S^2(P_a, P_c)$, we see that only joint terms survive, so that

$$F_\xi = -\sum_{a,c=0}^{3} \xi_a \xi_c \log\det(P_a + P_c).$$

Taking derivatives of F_ξ with respect to the P_i we obtain the criticality condition

$$\sum_{a=0}^{3} \xi_a (P_i + P_a)^{-1} = 0 \qquad (i = 1, 2, 3). \tag{2.2}$$

The equation also holds for $i = 0$. Indeed, by Lemma 2.1(a), F_ξ is unchanged when we replace P_a by $e^{tH/2} P_a e^{tH/2}$. The derivative $(d/dt)F_\xi|_{t=0}$ along this curve is the sum of its four partial derivatives. At a critical point the last three vanish, so the derivative in P_0 at I in the direction H also vanishes.

Now divide (2.2) by $-\xi_0$ and rewrite it as the set of equations

$$\sum_{j=1}^{3} \lambda_j (P_i + P_j)^{-1} = (P_i + I)^{-1} \quad (i = 1, 2, 3), \qquad \sum_{j=1}^{3} \lambda_j (I + P_j)^{-1} = \tfrac{1}{2} I. \tag{2.3}$$

The second equation in Cayley coordinates is $\sum_j \lambda_j (I - U_j) = I$, thus $\sum_j \lambda_j U_j = 0$. Expanding $(I - U_i)(P_i + P_j)(I - U_j)$ with $(I - U_i)P_i = I + U_i$ and $P_j(I - U_j) = I + U_j$ and rearranging gives $(P_i + P_j)^{-1} = \frac{1}{2}(I - U_j)(I - U_i U_j)^{-1}(I - U_i)$; so, the first equation becomes $\sum_j \lambda_j (I - U_i U_j)^{-1} = I$. □

2.3. **Solving the critical-point equations.** On a common eigenvector of the products $U_i U_j$, the system (2.1) becomes scalar, so we first show that these products commute.

Lemma 2.4. *Let $\lambda_1, \lambda_2, \lambda_3$ be as in Lemma 2.3, and let U_1, U_2, U_3 be Hermitian strict contractions satisfying (2.1). Then $U_1 = U_2 = U_3 = 0$.*

Proof. **The squares commute with all U_j.** Fix i, let j, k be the other two indices, and write $X = U_i^2$, $B_j = I - U_i U_j$ and $B_k = I - U_i U_k$. Multiplying the first equation in (2.1) on the left by U_i gives the first identity below, and the second is the equation of (2.1) for this i:

$$\lambda_i (I - X) + \lambda_j B_j + \lambda_k B_k = I, \qquad \lambda_i (I - X)^{-1} + \lambda_j B_j^{-1} + \lambda_k B_k^{-1} = I. \tag{2.4}$$

Here $H = (1 - \lambda_i) I + \lambda_i X$ is positive definite, since $|\lambda_i| < 1$. With $B_j = HZ$, the first identity gives $\lambda_k B_k = H(I - \lambda_j Z)$; substituting into the second and multiplying on the right by H yields

$$\lambda_j Z^{-1} + \lambda_k^2 (I - \lambda_j Z)^{-1} = \big(I - \lambda_i (I - X)^{-1}\big) H = \varphi(X), \qquad \varphi(x) := (1 - \lambda_i)^2 - \lambda_i \frac{x^2}{1 - x}. \tag{2.5}$$

The left side commutes with Z. Since $\lambda_i \neq 0$ and $x^2/(1 - x)$ increases strictly on $[0, 1)$, φ is injective there, so X is a polynomial in $\varphi(X)$ and commutes with Z, hence with $B_j = HZ$. For $C = [U_i^2, U_j]$ this says $U_i C = [U_i^2, U_i U_j] = 0$, and taking adjoints gives $C U_i = 0$. So $C = PCP$ for the projection P onto $\ker U_i$, and $PCP = 0$ because $U_i^2 P = P U_i^2 = 0$. By symmetry in j and k,

$$[U_i^2, U_j] = 0 \qquad \text{for all } i, j. \tag{2.6}$$

Shared eigenbasis. By the first equation in (2.1), $U_3 = -(\lambda_1 U_1 + \lambda_2 U_2)/\lambda_3$, so every product $U_i U_j$ lies in the span of U_1^2, U_2^2, $U_1 U_2$ and $U_2 U_1$. The first two commute with all U_j by (2.6), and $U_1 U_2$ commutes with its adjoint $U_2 U_1$, since both products equal $U_1^2 U_2^2$. So these four matrices are normal and commute, and have a common orthonormal eigenbasis. Let v be a vector in it. The eigenvalue of $U_i U_j$ on v is $(U_i v)^* (U_j v)$, and since $(U_i U_j)(U_j U_i) = U_i^2 U_j^2$, its modulus is $\|U_i v\| \, \|U_j v\|$. By the equality case of Cauchy–Schwarz, the vectors $U_1 v, U_2 v, U_3 v$ are parallel: $U_i v = z_i e$ for a unit vector e and scalars z_i with $|z_i| = \|U_i v\| < 1$. Thus $U_i U_j$ acts on v as $\overline{z_i}\, z_j$.

The scalar moments vanish. On v, the second set of equations in (2.1) becomes

$$\sum_{j=1}^{3} \frac{\lambda_j}{1 - \overline{z_i}\, z_j} = 1 \qquad (i = 1, 2, 3).$$

Expand the denominators as geometric series, multiply by λ_i and sum over i. Since $\sum_i \lambda_i = 1$ and $|z_i| < 1$, we obtain

$$0 = \sum_{n \geq 1} \Big| \sum_{j=1}^{3} \lambda_j z_j^n \Big|^2.$$

Let $\zeta_1, \ldots, \zeta_m$ be the distinct nonzero values among the z_j, with weights $w_\ell = \sum_{j : z_j = \zeta_\ell} \lambda_j$. Then $\sum_\ell w_\ell \zeta_\ell^n = 0$ for $n = 1, \ldots, m$, and this Vandermonde system forces every $w_\ell = 0$. But no nonempty

subset of the λ_j sums to zero: such a sum is $\lambda_j \neq 0$, $1-\lambda_j > 0$, or 1. Thus $m = 0$, so all $z_j = 0$ and $U_j v = 0$. As v runs over an orthonormal basis, $U_1 = U_2 = U_3 = 0$. □

Proof of (1.4). We can now prove the four-point negative-type condition (1.4). By Proposition 2.2 and Lemmas 2.3 and 2.4, the minimum of $F_\xi(I, \cdot)$ is attained at a critical point, and every critical point has $U_1 = U_2 = U_3 = 0$, that is, $P_1 = P_2 = P_3 = I$. The minimum is therefore $F_\xi(I, I, I, I) = 0$. This proves (1.4) for the ξ fixed above, and continuity in ξ gives it for every zero-sum ξ; by the equivalence established via identity (1.6), the truth of (1.3) also follows.

2.4. **Number of points.** The following configuration gives the sharpness of both forms of Theorem 1.2 for $d = 2$. The embeddings $X \mapsto X \oplus 0$ and $P \mapsto P \oplus I$ extend it to every $d \geq 2$.

Proposition 2.5. *Let* $u_k = (\cos\frac{k\pi}{3}, \sin\frac{k\pi}{3})^\top$ *for* $k = 0, 1, 2$, *and let* $c = (1, 1, 1, -\frac{3}{2})^\top$.

(a) *Let* $x = \sqrt{3}/2$, $X_k = x\, u_k u_k^\top$ *for* $k = 0, 1, 2$, *and* $X_3 = xI$. *Then* $0 \preceq X_k \prec I$, *and the matrix* $K = [-\log\det(I - X_i X_j)]_{i,j=0}^3$ *satisfies* $c^\top K c = \log(2^{21}/13^6) < 0$.

(b) *With* X_k *as in* (a), *let* $P_0 = I$ *and* $P_{k+1} = (I + X_k)(I - X_k)^{-1}$ *for* $k = 0, \ldots, 3$. *Then* δ_S^2 *is not of negative type on* $P_0, \ldots, P_4$.

Proof. (a) For $i, j \leq 2$, $X_i X_j$ has rank one and $\det(I - X_i X_j) = 1 - x^2(u_i^\top u_j)^2$. Since $(u_i^\top u_j)^2 = \frac{1}{4}$ for $i \neq j$, this determinant is $\frac{1}{4}$ for $i = j$ and $\frac{13}{16}$ for $i \neq j$. Also $\det(I - X_k X_3) = 1 - x^2 = \frac{1}{4}$ and $\det(I - X_3^2) = \frac{1}{16}$. With $L = \log 4$ and $\ell = \log\frac{16}{13}$,

$$K = \begin{pmatrix} L & \ell & \ell & L \\ \ell & L & \ell & L \\ \ell & \ell & L & L \\ L & L & L & 2L \end{pmatrix}, \qquad c^\top K c = 6\ell - \tfrac{3}{2}L = \log\frac{2^{21}}{13^6} < 0.$$

(b) The vector $\xi = (-\frac{3}{2}, 1, 1, 1, -\frac{3}{2})$ ($\xi^T 1 = 0$) gives $\sum_{a,e=0}^4 \xi_a \xi_e\, \delta_S^2(P_a, P_e) = -c^\top K c > 0$. □

3. Proof of Theorem 1.1 for $1/2 \leq \alpha < 1$

We prove Theorem 1.1 for $\alpha < 1$ on the positive unit sphere of a Schatten norm. The passage to the sphere rests on the equality in (1.2), which we prove first.

3.1. **Reduction to a Schatten sphere.**

Proposition 3.1. *Let* $0 < \alpha < 1$, *let* $\rho, \sigma \in \mathcal{D}_d$, *and put* $M = \rho^\alpha + \sigma^\alpha$. *Then*

$$(1-\alpha) \min_{\tau \in \mathcal{D}_d} \{T_\alpha(\rho, \tau) + T_\alpha(\sigma, \tau)\} = 2 - \|M\|_{1/\alpha}, \tag{3.1}$$

and the minimum is attained only at

$$\tau_\star = \frac{M^{1/\alpha}}{\operatorname{tr} M^{1/\alpha}}. \tag{3.2}$$

Proof. The scaled objective is $2 - \operatorname{tr}(M\tau^{1-\alpha})$, and Schatten Hölder with exponents $1/\alpha$ and $1/(1-\alpha)$ gives $\operatorname{tr}(M\tau^{1-\alpha}) \leq \|M\|_{1/\alpha}\|\tau^{1-\alpha}\|_{1/(1-\alpha)} = \|M\|_{1/\alpha}$. The state (3.2) attains equality, and strict convexity of the Schatten $1/(1-\alpha)$-norm makes it the only maximizer. □

For $1 < p < \infty$, let $\mathcal{S}_p = \{A \in \mathbb{P}^d : \operatorname{tr} A^p = 1\}$ and, for $A, B \in \mathcal{S}_p$, let $D_p(A, B) = (2 - \|A+B\|_p)^{1/2}$. For $p = 1/\alpha$, $\rho \mapsto \rho^\alpha$ is a bijection from $\mathcal{D}_d$ onto $\mathcal{S}_p$, and Proposition 3.1 gives $d_\alpha(\rho, \sigma) = D_p(\rho^\alpha, \sigma^\alpha)$. It therefore suffices to prove the triangle inequality for $1 < p \leq 2$; Proposition 3.6 treats $p > 2$.

3.2. **An integral representation.** We transfer the log-determinant form (1.3) of Theorem 1.2 to the trace-power kernel

$$L_p(A,B) := \operatorname{tr}(A+B)^p - \operatorname{tr} A^p - \operatorname{tr} B^p, \qquad A,B \in \mathbb{P}^d, \tag{3.3}$$

and then use its three-point positivity on the sphere.

Theorem 3.2. *Let $A_1, A_2, A_3 \in \mathbb{P}^d$. For $1 < p \le 2$,*

$$[L_p(A_i,A_j)]_{i,j=1}^3 \succeq 0. \tag{3.4}$$

Proof. At $p=2$, $L_2(A_i,A_j) = 2\operatorname{tr}(A_iA_j)$ is a Gram matrix. For $1<p<2$, put $\kappa_p = p\sin(\pi(p-1))/\pi > 0$ and use the scalar identity (Sra, 2021b)

$$x^p = \kappa_p \int_0^\infty (x/t - \log(1+x/t)) t^{p-1}\,\mathrm{d}t \qquad (x \ge 0). \tag{3.5}$$

For $t>0$, set $X_A(t) = A(A+tI)^{-1}$ and

$$K_t(A,B) := \log \frac{\det(A+tI)\det(B+tI)}{t^d \det(A+B+tI)} = -\log\det\bigl(I - X_A(t)X_B(t)\bigr). \tag{3.6}$$

Applying (3.5) to $A+B$, A and B and taking traces cancels the linear terms and gives

$$L_p(A,B) = \kappa_p \int_0^\infty t^{p-1} K_t(A,B)\,\mathrm{d}t. \tag{3.7}$$

Since $0 \preceq X_{A_i}(t) \prec I$, Theorem 1.2 gives $[K_t(A_i,A_j)] \succeq 0$, and integrating proves (3.4). □

3.3. **The triangle inequality.** Three-point positivity of L_p places any three points of $\mathcal{S}_p$ on a sphere, where D_p becomes a function of an angle. Let $1<p\le 2$, let $A_1,A_2,A_3 \in \mathcal{S}_p$, and put

$$G_{ij} := 2^{1-p}\operatorname{tr}(A_i+A_j)^p - 1. \tag{3.8}$$

Since $\operatorname{tr} A_i^p = 1$,

$$G = 2^{1-p}[L_p(A_i,A_j)]_{i,j=1}^3 + (2^{2-p}-1)\mathbf{1}\mathbf{1}^\top,$$

which is positive semidefinite by Theorem 3.2 and $2^{2-p}\ge 1$. Since $G_{ii}=1$, the matrix G is the Gram matrix of three unit vectors $u_1,u_2,u_3 \in \mathbb{R}^3$; write $G_{ij} = \cos\theta_{ij}$ with $0 \le \theta_{ij} \le \pi$. These angles satisfy the spherical triangle inequality

$$\theta_{13} \le \min\{\theta_{12}+\theta_{23}, \pi\}. \tag{3.9}$$

By (3.8), $\|A_i + A_j\|_p = 2\cos^{2/p}(\theta_{ij}/2)$, so

$$D_p(A_i,A_j) = f_p(\theta_{ij}), \qquad f_p(\theta) := \bigl(2(1-\cos^{2/p}(\theta/2))\bigr)^{1/2} \quad (0\le\theta\le\pi). \tag{3.10}$$

The triangle inequality for D_p therefore reduces to a scalar property of f_p.

Lemma 3.3. *For $1<p\le 2$, the function f_p in (3.10) is increasing and concave, $f_p(0)=0$, and*

$$f_p(\min\{x+y,\pi\}) \le f_p(x) + f_p(y) \qquad (0\le x,y\le\pi). \tag{3.11}$$

Proof. Since $\cos(\theta/2)$ decreases on $[0,\pi]$, the function f_p is increasing, and $f_p(0)=0$. For $0<\theta<\pi$, put $z = \cos^{2/p}(\theta/2)$. Differentiation gives

$$f_p''(\theta) \le 0 \quad\Longleftrightarrow\quad (2-z)(1-z^p) \ge p(1-z). \tag{3.12}$$

Since $t \mapsto t^{p-1}$ is concave on $[z,1]$,

$$\frac{1-z^p}{p(1-z)} = \frac{1}{1-z}\int_z^1 t^{p-1}\,\mathrm{d}t \ge \frac{1+z^{p-1}}{2} \ge \frac{1+z}{2} \ge \frac{1}{2-z}.$$

Thus f_p is concave.

Being concave with $f_p(0)=0$, f_p is subadditive; if $x+y\ge\pi$, monotonicity gives $f_p(\pi) \le f_p(x)+f_p(\pi-x) \le f_p(x)+f_p(y)$. This proves (3.11). □

Proof of Theorem 1.1 for $1/2 \le \alpha < 1$. Minkowski's inequality and strict convexity of the Schatten p-norm (Bhatia, 1997) give $D_p \ge 0$, with equality only when $A = B$. For the triangle inequality, let $A_1, A_2, A_3 \in \mathcal{S}_p$, with angles θ_{ij} as above. By (3.9), the monotonicity of f_p, and Lemma 3.3,

$$D_p(A_1, A_3) = f_p(\theta_{13}) \le f_p\big(\min\{\theta_{12} + \theta_{23}, \pi\}\big) \le f_p(\theta_{12}) + f_p(\theta_{23}) = D_p(A_1, A_2) + D_p(A_2, A_3). \quad \square$$

The angles θ_{ij} give more than the triangle inequality: three states lie on a sphere with chordal distance d_α.

Theorem 3.4. *Let* $1 < p \le 2$ *and* $A_1, A_2, A_3 \in \mathcal{S}_p$. *Then*

$$\big[\|(A_i + A_j)/2\|_p\big]_{i,j=1}^3 \succeq 0.$$

In particular, for $1/2 \le \alpha < 1$, *any three states with the metric* d_α *embed isometrically in the unit sphere of* $\mathbb{R}^3$.

Proof. With the angles of (3.8), $\|(A_i + A_j)/2\|_p = \cos\psi_p(\theta_{ij})$, where $\psi_p(\theta) := \arccos\cos^{2/p}(\theta/2)$ is increasing on $[0, \pi]$, with $\psi_p(0) = 0$ and $\psi_p \le \pi/2$. With $\beta = 2/p$ and $y = \cos^2(\theta/2)$, $\psi_p'(\theta)^2 = \frac{\beta^2}{4}\, y^{\beta-1}(1-y)/(1-y^\beta)$, and a brief calculation shows that this increases in y because $(1-y^\beta)/(1-y) \le \beta$. So ψ_p is concave, and (3.9) with the argument of Lemma 3.3 gives $\psi_p(\theta_{13}) \le \psi_p(\theta_{12}) + \psi_p(\theta_{23})$, and likewise for the other two angles. Three angles in $[0, \pi/2]$ with this property are the angles between three unit vectors v_1, v_2, v_3, whose Gram matrix is $[\cos\psi_p(\theta_{ij})]$. For states, $A_i = \rho_i^\alpha$ gives $\|(\rho_i^\alpha + \rho_j^\alpha)/2\|_{1/\alpha} = 1 - d_\alpha(\rho_i, \rho_j)^2/2$, so these unit vectors satisfy $\|v_i - v_j\| = d_\alpha(\rho_i, \rho_j)$. $\square$

3.4. **The case** $p = 2$**.** At $p = 2$, which corresponds to the endpoint $\alpha = 1/2$ of Theorem 1.1, the distance $D_2(A, B) = \big(2 - (2 + 2\operatorname{tr} AB)^{1/2}\big)^{1/2}$ depends only on the Hilbert–Schmidt inner product $\operatorname{tr} AB$. We can therefore work on the unit sphere of any real Hilbert space, where the distance is also of negative type.

Proposition 3.5. *On the unit sphere* $\mathbb{S}$ *of a real Hilbert space* $\mathcal{H}$, *put* $D(x, y) = (2 - \|x + y\|)^{1/2}$. *Then* D *is a metric of negative type on* $\mathbb{S}$, *so* $(\mathbb{S}, D^{1/2})$ *embeds isometrically in a Hilbert space.*

Proof. For unit vectors at angle θ, $D(x, y) = 2\sin(\theta/4) = f_2(\theta)$, so (3.9) and Lemma 3.3 prove that D is a metric.

For negative type, write $D(x, y) = \varphi(\langle x,\, y\rangle)$, where $\varphi(\cos\theta) = 2\sin(\theta/4)$. Then $(1 - t^2)\varphi'' - t\varphi' + \varphi/16 = 0$, so the Taylor coefficients b_k of φ at 0, whose series converges on $|t| < 1$, satisfy

$$b_{k+2} = \frac{k^2 - 1/16}{(k+1)(k+2)}\, b_k \qquad (k \ge 0).$$

Since $b_0 = 2\sin(\pi/8) > 0$ and $b_1 = -\frac{1}{2}\cos(\pi/8) < 0$, the recurrence gives $b_k < 0$ for every $k \ge 1$. Let $x_1, \dots, x_n \in \mathbb{S}$, let $c \in \mathbb{R}^n$ with $\sum_i c_i = 0$, and let $0 < r < 1$. Since $\sum_i c_i = 0$ and $\langle x_i,\, x_j\rangle^k = \langle x_i^{\otimes k},\, x_j^{\otimes k}\rangle$ in the k-fold tensor power of $\mathcal{H}$,

$$\sum\nolimits_{i,j=1}^n c_i c_j\, \varphi(r\langle x_i,\, x_j\rangle) = \sum\nolimits_{k\ge1} b_k r^k \Big\|\sum\nolimits_{i=1}^n c_i x_i^{\otimes k}\Big\|^2 \le 0.$$

Letting $r \uparrow 1$ gives negative type, and Schoenberg's criterion gives the embedding of $(\mathbb{S}, D^{1/2})$. $\square$

For states, $d_{1/2}(\rho, \sigma) = D_2(\rho^{1/2}, \sigma^{1/2})$ is a function of the quantum affinity $\operatorname{tr}\rho^{1/2}\sigma^{1/2}$ (Luo and Zhang, 2004), so Proposition 3.5 shows that $d_{1/2}$ is of negative type.

3.5. **The range of p.** For $p > 2$, the triangle inequality fails even on diagonal matrices.

Proposition 3.6. *Let $p > 2$ and $d = 2$, and let $A = \operatorname{diag}(0,1)$, $C = \operatorname{diag}(1,0)$ and $B_\varepsilon = \operatorname{diag}(\varepsilon, (1-\varepsilon^p)^{1/p})$ for $0 < \varepsilon < 1$. Then $A, B_\varepsilon, C \in \mathcal{S}_p$, and*

$$D_p(C, B_\varepsilon) + D_p(B_\varepsilon, A) < D_p(C, A)$$

for all sufficiently small ε.

Proof. All three matrices are diagonal, so the norms are ℓ_p norms of the diagonals. As $\varepsilon \to 0$, expanding $(1-\varepsilon^p)^{1/p} = 1 - \varepsilon^p/p + O(\varepsilon^{2p})$ gives $D_p(A, B_\varepsilon)^2 = \frac{1-2^{1-p}}{p}\varepsilon^p + O(\varepsilon^{2p})$, so $D_p(A, B_\varepsilon) = O(\varepsilon^{p/2})$. On the other hand, $\|C + B_\varepsilon\|_p^p = 2 + p\varepsilon + O(\varepsilon^2)$ because $p > 2$, so

$$D_p(C, B_\varepsilon) = D_p(C, A) - \frac{2^{1/p}}{4 D_p(C, A)}\,\varepsilon + O(\varepsilon^2).$$

Since $p/2 > 1$, the sum $D_p(C, B_\varepsilon) + D_p(B_\varepsilon, A)$ is smaller than $D_p(C, A)$ for small ε. Adjoining zero blocks gives the same counterexample for every $d \geq 2$. □

4. Proof of Theorem 1.1 for $1 < \alpha \leq 2$

For $\alpha > 1$, the proof of Proposition 3.1 with the reverse Hölder inequality gives the equality in (1.2), with the same minimizing state (3.2). We prove the triangle inequality first at $\alpha = 2$. At a minimizing middle state, the stationarity equation forces two auxiliary matrices to commute; the trace-one constraints then reduce the problem to a scalar inequality. A positive linear construction on tensor powers transfers this endpoint to the remaining orders.

For $p = 1/\alpha$ and $A_i = \rho_i^\alpha$, strict concavity of $X \mapsto \operatorname{tr} X^p$ gives $\operatorname{tr}(A_i + A_j)^p \geq 2^p$, with equality only when $A_i = A_j$, so d_α separates states. It remains to prove the triangle inequality.

4.1. **The scalar inequality.** The trace-one constraints will give weighted scalar means of at least one. To use these inequalities directly, we subtract the tangent term $(x+y)/\sqrt{2}$ before averaging.

Lemma 4.1. *Let $w_k, x_k, y_k \geq 0$ satisfy $\sum_k w_k = 1$, $\sum_k w_k x_k \geq 1$, and $\sum_k w_k y_k \geq 1$. If $r = \sum_k w_k \sqrt{1+x_k^2}$ and $s = \sum_k w_k \sqrt{1+y_k^2}$, then*

$$\sum\nolimits_k w_k \sqrt{x_k^2 + y_k^2} \leq L(r,s) := \sqrt{2 + \left(\sqrt{r^2-2} + \sqrt{s^2-2}\right)^2}. \tag{4.1}$$

Proof. Jensen's inequality gives $r, s \geq \sqrt{2}$. For $a, b \geq 0$, put

$$\delta(a,b) = \sqrt{a^2+b^2} - \frac{a+b}{\sqrt{2}}, \qquad D(a,b) = \sqrt{\delta(a,b)\big(\delta(a,b) + 2\sqrt{2}\big)}.$$

We first prove the pointwise inequality

$$D(x,y) \leq D(x,1) + D(1,y), \qquad x, y \geq 0. \tag{4.2}$$

For $a \leq b$, both δ and D are nonincreasing in a and nondecreasing in b. Thus (4.2) holds when x, y lie on the same side of 1. By symmetry and continuity, it remains to treat $0 < x < 1 < y$.

We compare $-D_x(x,y)$ with its value at $y = 1$. For fixed $0 < x < 1$ and $y \geq 1$, write $H = \sqrt{2(x^2+y^2)}$, $S = x + y$, and $E = H - S > 0$. Then $D^2 = E(E+4)/2$, and differentiation gives

$$D_{xy}(x,y) = \frac{E^2}{H^3 D(x,y)^3}\big[6(S - xy) + (2 - xy)E\big].$$

Since $S - xy = x + (1-x)y > 0$, the bracket is positive if $xy \leq 2$. If $xy > 2$, it has the sign of

$$6 - \frac{xy - 2}{x + (1-x)y}E.$$

Both factors in the subtracted product are positive and increasing in y, since

$$\frac{\partial}{\partial y}\frac{xy-2}{x+(1-x)y} = \frac{x^2-2x+2}{(x+(1-x)y)^2} > 0, \qquad E_y = \frac{2y}{H} - 1 > 0.$$

Thus D_{xy} can change sign only from positive to negative, so $-D_x(x,y)$ has no interior maximum as a function of $y \geq 1$. Its limit at infinity is $1/\sqrt{2}$.

At $y = 1$, rationalization gives

$$D(x,1) = (1-x)h(x), \qquad h(x)^2 = \frac{1}{2} + \frac{1-x}{\sqrt{2(1+x^2)}+x+1}.$$

The fraction is positive and decreasing, so $h > 1/\sqrt{2}$, $h' < 0$, and $-D_x(x,1) = h(x) - (1-x)h'(x) > 1/\sqrt{2}$. Hence $-D_x(x,y) \leq -D_x(x,1)$ for $0 < x < 1$ and $y \geq 1$, and integrating proves (4.2):

$$D(x,y) - D(1,y) = \int_x^1 -D_a(a,y)\,\mathrm{d}a \leq \int_x^1 -D_a(a,1)\,\mathrm{d}a = D(x,1).$$

To average (4.2), we use concavity of L on $[\sqrt{2},\infty)^2$. In the interior, put $u = \sqrt{r^2-2}$, $v = \sqrt{s^2-2}$, $q = u+v$, and $\xi = (r/u, s/v)^\top$. Differentiation gives

$$\nabla^2 L = \frac{2}{L^3}\left[\xi\xi^\top - qL^2\,\mathrm{diag}(u^{-3}, v^{-3})\right] \preceq 0,$$

where the last inequality follows by weighted Cauchy–Schwarz from $r^2u + s^2v = u^3 + v^3 + 2q \leq q(q^2+2) = qL^2$. Also $L_r = qr/(uL) \geq 1$, because $q^2r^2 - u^2L^2 = 2(q^2-u^2) \geq 0$, and similarly $L_s \geq 1$. Integrating these derivative bounds gives

$$L(r+a, s+b) \geq L(r,s) + a + b, \qquad r,s \geq \sqrt{2}, \quad a,b \geq 0. \tag{4.3}$$

Let (X,Y) take the values (x_k,y_k) with probabilities w_k, and put $t = \mathbb{E}\sqrt{X^2+Y^2}$. Define

$$\widehat{R} = \sqrt{2} + \delta(X,1), \qquad \widehat{S} = \sqrt{2} + \delta(Y,1), \qquad \widehat{T} = \sqrt{2} + \delta(X,Y).$$

These variables are at least $\sqrt{2}$, and (4.2) is exactly $\widehat{T} \leq L(\widehat{R},\widehat{S})$. The mean constraints give $a := (\mathbb{E}X - 1)/\sqrt{2} \geq 0$ and $b := (\mathbb{E}Y - 1)/\sqrt{2} \geq 0$, so $\mathbb{E}\widehat{R} = r - a$, $\mathbb{E}\widehat{S} = s - b$ and $\mathbb{E}\widehat{T} = t - a - b$. Concavity and (4.3) now give (4.1):

$$t - a - b \leq \mathbb{E}L(\widehat{R},\widehat{S}) \leq L(r-a, s-b) \leq L(r,s) - a - b. \qquad \square$$

4.2. **The endpoint $\alpha = 2$.** The scalar means in Lemma 4.1 arise from a trace comparison. Commutativity is needed only for the auxiliary matrices in this comparison, not for the states.

Lemma 4.2. *Let $Z \succ 0$, and let $R_1, \ldots, R_\ell$ be commuting Hermitian matrices. If*

$$F^2 := \sum\nolimits_j R_j^2 - cI \succeq 0, \qquad C := Z^{1/2}\left(\sum\nolimits_j R_j Z R_j - cZ\right)Z^{1/2} \succeq 0,$$

where $F \succeq 0$ and $c \in \mathbb{R}$, then $\operatorname{tr}\sqrt{C} \leq \operatorname{tr}(ZF)$.

Proof. For $W \succ 0$, Hilbert–Schmidt Cauchy–Schwarz gives $(\operatorname{tr}\sqrt{C})^2 \leq \operatorname{tr}(CW^{-1})\operatorname{tr}W$. Choose $W = Z^{1/2}(F+\varepsilon I)Z^{1/2}$. Since each R_j commutes with F, cyclicity of trace gives

$$\operatorname{tr}(CW^{-1}) = \operatorname{tr}\left[ZF^2(F+\varepsilon I)^{-1}\right] \leq \operatorname{tr}(ZF), \qquad \operatorname{tr}W = \operatorname{tr}(ZF) + \varepsilon \operatorname{tr}Z.$$

Let $\varepsilon \downarrow 0$. $\square$

Proposition 4.3. *The function d_2 satisfies the triangle inequality on $\mathcal{D}_d$ for every d.*

Proof. Write $T(X,Y) = \operatorname{tr}\sqrt{X^2+Y^2}$, so $d_2(X,Y) = \sqrt{T(X,Y)^2 - 2}$ for states. By continuity, it suffices to treat distinct positive definite states ρ, σ. Minimize $f(z) = d_2(\rho,z) + d_2(z,\sigma)$ over $\mathcal{D}_d$. A minimizer exists by compactness. If it is an endpoint, there is nothing to prove. Otherwise put

$$u = d_2(\rho,z) > 0, \qquad v = d_2(\sigma,z) > 0, \qquad K_1 = \sqrt{\rho^2+z^2}, \quad K_2 = \sqrt{\sigma^2+z^2},$$

and $t_j = \operatorname{tr} K_j$, so $t_1^2 = u^2 + 2$ and $t_2^2 = v^2 + 2$. The gradient is

$$\nabla f(z) = \tfrac{1}{2}(zH + Hz), \qquad H = \frac{t_1}{u}K_1^{-1} + \frac{t_2}{v}K_2^{-1} \succ 0.$$

The minimizer is positive definite. Indeed, if $e \in \ker z$ is a unit vector, the feasible path $z_\varepsilon = (1-\varepsilon)z + \varepsilon ee^*$ has derivative $-\operatorname{tr}(z^2 H) < 0$.

Stationarity under $\operatorname{tr} z = 1$ gives $zH + Hz = 2\lambda I$, so $H = \lambda z^{-1}$, that is,

$$\frac{t_1}{u}R_1^{-1} + \frac{t_2}{v}R_2^{-1} = \lambda I, \qquad R_j := z^{-1/2}K_j z^{-1/2}. \tag{4.4}$$

Thus R_1, R_2 commute. Operator monotonicity of the square root also gives $K_j \succeq z$, so $R_j \succeq I$. In a common eigenbasis e_k, write

$$R_j e_k = r_{jk} e_k, \qquad w_k = \langle e_k, z e_k\rangle, \qquad x_k = \sqrt{r_{1k}^2 - 1}, \quad y_k = \sqrt{r_{2k}^2 - 1}.$$

Then $\sum_k w_k = 1$ and

$$t_1 = \sum\nolimits_k w_k \sqrt{1 + x_k^2}, \qquad t_2 = \sum\nolimits_k w_k \sqrt{1 + y_k^2}.$$

Since $\rho^2 = z^{1/2}(R_1 z R_1 - z)z^{1/2}$, Lemma 4.2 gives

$$1 = \operatorname{tr}\rho \le \sum\nolimits_k w_k x_k, \qquad 1 = \operatorname{tr}\sigma \le \sum\nolimits_k w_k y_k.$$

Applying that lemma once more to $\rho^2 + \sigma^2 = z^{1/2}(R_1 z R_1 + R_2 z R_2 - 2z)z^{1/2}$ yields

$$T(\rho,\sigma) \le \sum\nolimits_k w_k \sqrt{x_k^2 + y_k^2} \le L(t_1,t_2) = \sqrt{2 + (u+v)^2},$$

where the second inequality is Lemma 4.1. Thus $d_2(\rho,\sigma) \le f(z)$, and minimality gives the triangle inequality for every middle state. □

4.3. **Transfer to the other orders.** The transfer uses positive linear maps so that sums are preserved even when the inputs do not commute.

Lemma 4.4. *Let $1/2 < p < 1$, and let $A_1, A_2, A_3 \in \mathbb{P}^d$ satisfy $\operatorname{tr} A_i^p = 1$. There are finite-dimensional density matrices $B_i^{(N)}$ such that, for $i \ne j$,*

$$\operatorname{tr}\sqrt{(B_i^{(N)})^2 + (B_j^{(N)})^2} \longrightarrow \left[\operatorname{tr}(A_i + A_j)^p\right]^{1/(2p)}. \tag{4.5}$$

Proof. Let S have density ps^{-p-1} on $[1,\infty)$, and put $\phi(t) = \mathbb{E}e^{-tS}$. A change of variables gives

$$1 - \phi(t) = pt^p \int_t^\infty (1 - e^{-v}) v^{-p-1}\,\mathrm{d}v \sim \ell_p t^p, \qquad 1 - \phi(t) \le \ell_p t^p,$$

where $\ell_p = p\int_0^\infty (1-e^{-v})v^{-p-1}\,\mathrm{d}v \in (0,\infty)$.

For independent copies $S_1, \dots, S_N$, define the positive linear map

$$\Phi_N(X) = N^{-1/p} \sum_{r=1}^N S_r X^{(r)}$$

on $(\mathbb{C}^d)^{\otimes N}$, where $X^{(r)}$ acts in tensor factor r. Writing $\tau_m = \operatorname{tr}/m$, independence and commutativity of the tensor copies of each fixed $X \succeq 0$ give

$$\mathbb{E}\tau_{d^N} e^{-t\Phi_N(X)} = \left[\tau_d \phi(tN^{-1/p}X)\right]^N \longrightarrow e^{-\ell_p t^p \tau_d X^p}.$$

Moreover,

$$0 \le 1 - \mathbb{E}\tau_{d^N} e^{-t\Phi_N(X)} \le \min\{1, \ell_p t^p \tau_d X^p\}.$$

Multiply by $t^{-3/2}$ and use $\sqrt{x} = (2\sqrt{\pi})^{-1} \int_0^\infty (1 - e^{-tx}) t^{-3/2}\, \mathrm{d}t$. The bound is integrable because $p > 1/2$, so dominated convergence gives

$$\mathbb{E}\tau_{d^N} \sqrt{\Phi_N(X)} \longrightarrow \gamma_p (\tau_d X^p)^{1/(2p)}, \qquad \gamma_p := \frac{1}{2\sqrt{\pi}} \int_0^\infty (1 - e^{-\ell_p t^p}) t^{-3/2}\, \mathrm{d}t \in (0, \infty). \tag{4.6}$$

For each N, approximate these expectations simultaneously for the six inputs A_i and $A_i + A_j$ ($i < j$) by a finite weighted average, with error at most $1/N$; such averages exist because the integrands are continuous and dominated by $\|X\|_{\mathrm{op}}^{1/2} N^{-1/(2p)} \sum_r \sqrt{S_r}$. Let s_ℓ be the nodes and $q_\ell \ge 0$ their weights, and set

$$D_i^{(N)} = \bigoplus_\ell \frac{q_\ell}{d^N} \sqrt{\Phi_{N,s_\ell}(A_i)}.$$

Their traces are the chosen weighted averages. Linearity gives

$$\operatorname{tr} \sqrt{(D_i^{(N)})^2 + (D_j^{(N)})^2} = \sum_\ell q_\ell \tau_{d^N} \sqrt{\Phi_{N,s_\ell}(A_i + A_j)}.$$

Consequently, with $M = \gamma_p d^{-1/(2p)} > 0$,

$$m_i^{(N)} := \operatorname{tr} D_i^{(N)} \longrightarrow M, \qquad T_{ij}^{(N)} := \operatorname{tr} \sqrt{(D_i^{(N)})^2 + (D_j^{(N)})^2} \longrightarrow M \big[\operatorname{tr}(A_i + A_j)^p\big]^{1/(2p)}.$$

For sufficiently large N, normalize $B_i^{(N)} = D_i^{(N)}/m_i^{(N)}$. Operator monotonicity gives

$$\frac{T_{ij}^{(N)}}{\max\{m_i^{(N)}, m_j^{(N)}\}} \le \operatorname{tr} \sqrt{(B_i^{(N)})^2 + (B_j^{(N)})^2} \le \frac{T_{ij}^{(N)}}{\min\{m_i^{(N)}, m_j^{(N)}\}}.$$

Both bounds have the limit in (4.5). □

Proof of Theorem 1.1 for $1 < \alpha \le 2$. Proposition 4.3 proves the triangle inequality at $\alpha = 2$. For $1 < \alpha < 2$, set $p = 1/\alpha$ and $A_i = \rho_i^\alpha$. Lemma 4.4 gives

$$d_2(B_i^{(N)}, B_j^{(N)}) \longrightarrow \sqrt{\big[\operatorname{tr}(A_i + A_j)^p\big]^{1/p} - 2} = d_\alpha(\rho_i, \rho_j).$$

The endpoint triangle inequality passes to this limit. □

AI USAGE

The results in this paper were developed through extensive use of AI assistants: OpenAI's GPT Pro for the main line of argument, and Anthropic's Claude for the endpoint results in Section 3.4 and Appendix A.2, Remark A.2, Proposition 3.6, and the writing, positioning, and literature search. OpenAI assistants also developed the four-point consequences in Appendix B and assisted with proof simplification and reorganization. The arguments emerged through sustained work involving failed attempts, counterexamples, numerical checks, proof audits, and repairs.

ACKNOWLEDGEMENTS

The author (SS) thanks the Alexander von Humboldt (AvH) Foundation for their generous support via a Humboldt Professorship in AI.

References

Rajendra Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics 169. Springer, New York, 1997. doi: 10.1007/978-1-4612-0653-8.

Jop Briët and Peter Harremoës. Properties of classical and quantum Jensen–Shannon divergence. *Physical Review A*, 79(5):052311, 2009. doi: 10.1103/PhysRevA.79.052311.

Anoop Cherian, Suvrit Sra, Arindam Banerjee, and Nikolaos Papanikolopoulos. Jensen–Bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2161–2174, 2013. doi: 10.1109/TPAMI.2012.259.

Marco Dalai. Lower bounds on the probability of error for classical and classical-quantum channels. *IEEE Transactions on Information Theory*, 59(12):8027–8056, 2013. doi: 10.1109/TIT.2013.2283794.

Michel Marie Deza and Monique Laurent. *Geometry of Cuts and Metrics*, volume 15 of *Algorithms and Combinatorics*. Springer, Berlin, 1997. ISBN 3-540-61611-X. doi: 10.1007/978-3-642-04295-9.

D. M. Endres and J. E. Schindelin. A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49(7):1858–1860, 2003. doi: 10.1109/TIT.2003.813506.

Bent Fuglede and Flemming Topsøe. Jensen–Shannon divergence and Hilbert space embedding. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, page 31, 2004. doi: 10.1109/ISIT.2004.1365067.

Loo-Keng Hua. Inequalities involving determinants. *Acta Mathematica Sinica*, 5:463–470, 1955.

Robert König and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103:070504, 2009. doi: 10.1103/PhysRevLett.103.070504. Appendix of arXiv:0903.2838.

P. W. Lamberti, A. P. Majtey, A. Borras, M. Casas, and A. Plastino. Metric character of the quantum Jensen–Shannon divergence. *Physical Review A*, 77(5):052311, 2008. doi: 10.1103/PhysRevA.77.052311.

Shunlong Luo and Qiang Zhang. Informational distance on quantum-state space. *Physical Review A*, 69:032106, 2004. doi: 10.1103/PhysRevA.69.032106.

André F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research*, 10:935–975, 2009.

Ferdinand Österreicher and Igor Vajda. A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics*, 55(3):639–653, 2003. doi: 10.1007/BF02517812.

Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, 1986. doi: 10.1016/0034-4877(86)90067-4.

I. J. Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Annals of Mathematics*, 36 (3):724–732, 1935. doi: 10.2307/1968654.

Naresh Sharma and Naqueeb Ahmad Warsi. Fundamental bound on the reliability of quantum information transmission. *Physical Review Letters*, 110:080501, 2013. doi: 10.1103/PhysRevLett.110.080501.

Suvrit Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7):2787–2797, 2016. doi: 10.1090/proc/12953.

Suvrit Sra. Positive definite functions of noncommuting contractions, Hua–Bellman matrices, and a new distance metric, 2021a.

Suvrit Sra. Metrics induced by Jensen–Shannon and related divergences on positive definite matrices. *Linear Algebra and its Applications*, 616:125–138, 2021b. doi: 10.1016/j.laa.2020.12.023.

Constantino Tsallis. Possible generalization of Boltzmann–Gibbs statistics. *Journal of Statistical Physics*, 52(1–2):479–487, 1988. doi: 10.1007/BF01016429.

Dániel Virosztek. The metric property of the quantum Jensen–Shannon divergence. *Advances in Mathematics*, 380:107595, 2021. doi: 10.1016/j.aim.2021.107595.

Appendix A. Relation to the cone and Jensen–Shannon metrics

For $0 < \alpha < 1$, we relate d_α to the two earlier metrics of the introduction: the cone metric of Sra (2021b), which drops the trace constraint, and the Jensen–Shannon metric, the limit of the rescaled d_α as $\alpha \uparrow 1$.

A.1. **Comparison with the cone minimum.** To let the center range over the cone, extend T_α to $X, Z \in \mathbb{P}^d$ by

$$T_\alpha(X,Z) := \frac{\alpha \operatorname{tr} X + (1-\alpha)\operatorname{tr} Z - \operatorname{tr} X^\alpha Z^{1-\alpha}}{1-\alpha}. \tag{A.1}$$

Minimizing over the cone gives the relative-entropy Jensen–Tsallis divergence of Sra (2021b, Theorem 4.1); its commuting case was studied by Österreicher and Vajda (2003).

Proposition A.1. *Let $0 < \alpha < 1$ and $\rho, \sigma \in \mathcal{D}_d$, and put*

$$M_\alpha := \Big(\frac{\rho^\alpha + \sigma^\alpha}{2}\Big)^{1/\alpha}, \qquad h := \operatorname{tr} M_\alpha.$$

The unique minimizer over $\mathbb{P}^d$ is M_α, and

$$(1-\alpha) \min_{Z \in \mathbb{P}^d} \{T_\alpha(\rho, Z) + T_\alpha(\sigma, Z)\} = 2\alpha(1-h). \tag{A.2}$$

Proposition 3.1 gives the constrained center M_α/h and

$$d_\alpha(\rho,\sigma)^2 = 2(1-h^\alpha). \tag{A.3}$$

Proof. Write $Z = s\tau$, where $s = \operatorname{tr} Z$ and $\tau \in \mathcal{D}_d$. For fixed $s > 0$, Proposition 3.1 gives the unique optimal state $\tau = M_\alpha/h$. Since $\|\rho^\alpha + \sigma^\alpha\|_{1/\alpha} = 2h^\alpha$, the scaled objective reduces to $2\alpha + 2(1-\alpha)s - 2h^\alpha s^{1-\alpha}$, a strictly convex function of $s \geq 0$ minimized at $s = h$, with value $2\alpha(1-h)$. The minimizing matrix is $s\tau = M_\alpha$. □

Remark A.2. *For $1/2 \leq \alpha < 1$ and three states ρ_i, let $p = 1/\alpha$ and take $A_i = \rho_i^\alpha$ in the Gram matrix G of (3.8). Then $h_{ij} := \operatorname{tr}((\rho_i^\alpha + \rho_j^\alpha)/2)^{1/\alpha} = (1+G_{ij})/2$. Thus $H = [h_{ij}] = (\mathbf{1}\mathbf{1}^\top + G)/2$ is a Gram matrix of unit vectors v_i, and the scaled cone minimum satisfies $2\alpha(1-h_{ij}) = \alpha\|v_i - v_j\|^2$. Its square root satisfies the triangle inequality and separates states by (A.3) and Theorem 1.1, recovering the cone metric of Sra (2021b, Theorem 4.1) on states. In contrast, $d_\alpha(\rho_i,\rho_j)^2 = 2(1-h_{ij}^\alpha)$ by (A.3), and its triangle inequality needs Lemma 3.3.*

A.2. **The limit $\alpha \uparrow 1$.** For the Umegaki relative entropy $\operatorname{tr}\rho(\log\rho - \log\tau)$, the optimal center is $m = (\rho+\sigma)/2$ and the least total is $2\,\mathrm{QJSD}(\rho,\sigma)$, where

$$\mathrm{QJSD}(\rho,\sigma) := \tfrac{1}{2}\operatorname{tr}\rho(\log\rho - \log m) + \tfrac{1}{2}\operatorname{tr}\sigma(\log\sigma - \log m) \tag{A.4}$$

is the quantum Jensen–Shannon divergence.

Corollary A.3. *For $\rho, \sigma \in \mathcal{D}_d$,*

$$\lim_{\alpha\uparrow 1} \frac{d_\alpha(\rho,\sigma)^2}{2(1-\alpha)} = \mathrm{QJSD}(\rho,\sigma).$$

Consequently $\mathrm{QJSD}^{1/2}$ is a metric on $\mathcal{D}_d$.

Proof. Put $Y_\alpha = \rho^\alpha + \sigma^\alpha$, $g(\alpha) = \operatorname{tr} Y_\alpha^{1/\alpha}$ and $f(\alpha) = g(\alpha)^\alpha = \|Y_\alpha\|_{1/\alpha}$. Then $f(1) = 2$, so $d_\alpha(\rho,\sigma)^2 = f(1) - f(\alpha)$ by Proposition 3.1, and the limit is $f'(1)/2$. A brief calculation gives

$$g'(1) = \operatorname{tr}(\rho \log \rho + \sigma \log \sigma) - \operatorname{tr}\big((\rho+\sigma)\log(\rho+\sigma)\big) = 2\,\mathrm{QJSD}(\rho,\sigma) - 2\log 2,$$

so $(\log f)'(1) = \log g(1) + g'(1)/g(1) = \mathrm{QJSD}(\rho,\sigma)$ and $f'(1) = 2\,\mathrm{QJSD}(\rho,\sigma)$. Since $d_\alpha/(2(1-\alpha))^{1/2}$ is a metric for $1/2 \le \alpha < 1$ by Theorem 1.1, its pointwise limit $\mathrm{QJSD}^{1/2}$ satisfies the triangle inequality, and strict convexity of $X \mapsto \operatorname{tr} X \log X$ makes it separate states. □

Appendix B. Quantum Jensen–Tsallis and Jensen–Rényi divergences

We apply Theorem 1.2 to entropy-based Jensen divergences. For $\rho, \sigma \in \mathcal{D}_d$ and $m = (\rho+\sigma)/2$, let

$$\mathrm{QJSD}_q(\rho,\sigma) := \frac{\operatorname{tr} m^q - \frac{1}{2}\operatorname{tr}\rho^q - \frac{1}{2}\operatorname{tr}\sigma^q}{1-q} \qquad (0 < q \le 2,\ q \ne 1), \tag{B.1}$$

and put $\mathrm{QJSD}_1 = \mathrm{QJSD}$. The quantum Jensen–Rényi divergence is

$$\mathrm{QJRD}_q(\rho,\sigma) := H_q(m) - \tfrac{1}{2}H_q(\rho) - \tfrac{1}{2}H_q(\sigma), \qquad H_q(\rho) := \frac{\log \operatorname{tr}\rho^q}{1-q} \quad (0 < q < 1), \tag{B.2}$$

with $\mathrm{QJRD}_1 = \mathrm{QJSD}$. These divergences differ from both minima in Proposition A.1.

Theorem B.1. *On any four density matrices,* QJSD_q *is of negative type for* $0 < q \le 2$*, and* QJRD_q *is of negative type for* $0 < q \le 1$*. Their square roots are metrics, and every four states with either metric embed isometrically in* $\mathbb{R}^3$*.*

The square-root metric properties were proved in Sra (2021b, Corollaries 3.3–3.5 and Theorem 6.3). More generally, the proof below gives four-point negative type for every Jensen divergence that is a nonnegative integral of S-divergences of shifted matrices, which includes the families of Sra (2021b, Theorems 3.1 and 3.6). For commuting states, QJSD_q is of negative type on arbitrary finite sets (Briët and Harremoës, 2009; Martins et al., 2009).

Proof. For $0 < q < 2$,

$$\mathrm{QJSD}_q(\rho,\sigma) = c_q \int_0^\infty t^{q-1} \delta_S^2(\rho + tI, \sigma + tI)\,\mathrm{d}t, \qquad c_q = \frac{q\sin(\pi q)}{\pi(1-q)}, \quad c_1 = 1. \tag{B.3}$$

This follows from $c_q \int_0^\infty t^{q-1}(x+t)^{-2}\,\mathrm{d}t = q x^{q-2}$ after integrating twice in x and taking traces: the midpoint combination cancels the affine terms. The integral converges since $\delta_S^2(\rho+tI,\sigma+tI)$ is $O(1+|\log t|)$ at zero and $O(t^{-2})$ at infinity. Since $c_q > 0$, Theorem 1.2 proves four-point negative type for QJSD_q. At $q = 2$ the same conclusion follows from $\mathrm{QJSD}_2(\rho,\sigma) = \frac{1}{4}\|\rho-\sigma\|_2^2$.

For the Rényi case, we apply the logarithmic argument of Sra (2021b, Theorem 6.3) to four states. Fix $0 < q < 1$ and $\rho_0, \dots, \rho_3 \in \mathcal{D}_d$, and put $h_{ij} = \operatorname{tr}((\rho_i+\rho_j)/2)^q$, so that $h_{ij} = (1-q)\,\mathrm{QJSD}_q(\rho_i,\rho_j) + (h_{ii}+h_{jj})/2$. By the first part and Schoenberg's theorem, $[e^{-s(1-q)\,\mathrm{QJSD}_q(\rho_i,\rho_j)}] \succeq 0$ for every $s > 0$, and rescaling rows and columns by $e^{-sh_{ii}/2}$ gives $[e^{-sh_{ij}}] \succeq 0$. For any $\xi \in \mathbb{R}^4$ with $\sum_i \xi_i = 0$, the identity $\log r = \int_0^\infty (e^{-s} - e^{-sr})\,\mathrm{d}s/s$ therefore gives

$$\sum_{i,j} \xi_i \xi_j \log h_{ij} = -\int_0^\infty \sum_{i,j} \xi_i \xi_j e^{-sh_{ij}}\,\frac{\mathrm{d}s}{s} \le 0.$$

Since $(1-q)\,\mathrm{QJRD}_q(\rho_i,\rho_j) = \log h_{ij} - \frac{1}{2}\log h_{ii} - \frac{1}{2}\log h_{jj}$ and $\sum_i \xi_i = 0$, this proves four-point negative type for QJRD_q.

For distinct states, $\delta_S^2(\rho+tI,\sigma+tI) > 0$, so (B.3) gives $\mathrm{QJSD}_q(\rho,\sigma) > 0$. For $0 < q < 1$, this also gives $h_{ij} > (h_{ii}+h_{jj})/2 \ge \sqrt{h_{ii}h_{jj}}$, hence $\mathrm{QJRD}_q(\rho_i,\rho_j) > 0$ when $\rho_i \ne \rho_j$. Schoenberg's criterion now gives the claimed Euclidean embeddings, so the square roots are metrics. □