THE p-BURES–WASSERSTEIN DISTANCE VIA 3×3 BLOCK MATRICES
(with application to p-Bures–Wasserstein barycenters)

SUVRIT SRA*

ABSTRACT. We study the p-Bures–Wasserstein distance $D_p(A,B) = (\operatorname{tr} A^p + \operatorname{tr} B^p - 2\|A^{1/2}B^{1/2}\|_p^p)^{\frac{1}{2}}$ on positive semidefinite matrices; it equals the Bures–Wasserstein distance for $p=1$, and the Hilbert–Schmidt distance for $p=2$. We prove that D_p is a metric for every $1 \le p \le 2$, by proving a stronger 2014 conjecture of Lin and van den Driessche on 3×3 block matrices. Our main tool is a partition inequality for Schatten norms, proved by complex interpolation. We also characterize Fréchet means for D_p by a fixed-point equation and prove geometric convergence of the associated iteration.

hexagon:2610.00726v1

1. INTRODUCTION

Let $\mathbb{P}_n$ denote the cone of $n \times n$ complex positive semidefinite matrices, and write $\|\cdot\|_p$ for the Schatten p-norm. Fix $1 \le p \le 2$ and write $t := \frac{p}{2} \in \left[\frac{1}{2}, 1\right]$. For $A, B \in \mathbb{P}_n$, define the *affinity*

$$\Phi_p(A,B) := \operatorname{tr}\big(A^{1/2}BA^{1/2}\big)^t = \|A^{1/2}B^{1/2}\|_p^p. \tag{1.1}$$

We study the *p-Bures–Wasserstein distance*

$$D_p(A,B) := \left(\operatorname{tr} A^p + \operatorname{tr} B^p - 2\Phi_p(A,B)\right)^{1/2}. \tag{1.2}$$

At $p=1$, this is the Bures–Wasserstein distance (Bures, 1969; Olkin and Pukelsheim, 1982; Bhatia et al., 2019), while at $p=2$ it is the Hilbert–Schmidt distance. The main result interpolates the triangle inequality between these two endpoints.

The pth root of Φ_p, $\|A^{1/2}B^{1/2}\|_p$, is the p-norm fidelity of Wei et al. (2024), who record its elementary properties. At $p=1$ it is the root fidelity of quantum information. Other generalizations include the alpha Procrustes distances of Minh (2022), which are based on matrix powers and approach the Log-Euclidean distance, and the fidelities of Afham and Ferrie (2025), built from Bures–Wasserstein Riemannian geometry. The family D_p varies the Schatten exponent. Its triangle inequality is implicit in a conjecture of Lin and van den Driessche (2014) on positive semidefinite 3×3 block matrices, which we call the *LvD conjecture* (Section 4).

The same family reappeared later in work on matrix divergences, without reference to the LvD conjecture: for $1 \le p < 2$, $D_p(A,B)^2$ is twice the α-z Bures–Wasserstein divergence of Dinh et al. (2021) at $\alpha = 1/2$ and $z = p/2$, evaluated at (A^p, B^p), and Komálovics and Molnár (2024, Problem 2) asked whether the resulting distance is a metric for $1 < p < 2$.

Theorem 1.1 (p-Bures–Wasserstein metric)**.** *For every $1 \le p \le 2$ and all $A, B, C \in \mathbb{P}_n$,*

$$D_p(A,C) \le D_p(A,B) + D_p(B,C). \tag{1.3}$$

To control a weighted sum of affinities, we use the following inequality for Schatten norms.

Key words and phrases. Bures–Wasserstein distance, Schatten norm, positive semidefinite block matrices, Lin–van den Driessche conjecture, partition inequality, complex interpolation, barycenter, Fréchet mean.

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

Also on arXiv: https://arxiv.org/abs/2610.15889.

Theorem 1.2 (Partition inequality)**.** *Let $1 \le p \le 2$, let $H_1, \dots, H_m \succeq 0$, and suppose $H := \sum_{j=1}^m H_j \succ 0$. Then, for every matrix X,*

$$\sum_{j=1}^{m} \|H_j^{1/p} H^{-1/p} X\|_p^p \ge \|X\|_p^p. \tag{1.4}$$

At $p = 2$, inequality (1.4) is an identity. At $p = 1$ it follows from the triangle inequality. For intermediate p, it follows by interpolating a row contraction and then using a left-inverse relation.

LvD conjecture. We prove the triangle inequality behind Theorem 1.1 for angles on the normalized sphere (Theorem 3.4); in that form it resolves Conjecture 5.1 of Lin and van den Driessche (2014).

Theorem 1.3 (LvD conjecture)**.** *Let $\mathbf{H} = [H_{ij}]_{i,j=1}^3 \succeq 0$ be a positive semidefinite 3×3 block matrix, with arbitrary finite block sizes compatible with the block structure. Then, for every $1 \le p \le 2$,*

$$\big[\|H_{ij}\|_p^p\big]_{i,j=1}^3 \succeq 0. \tag{1.5}$$

The endpoints $p = 1$ and $p = 2$ were known (Drury, 2014; Marcus and Watkins, 1971). For semidefinite block input, the conclusion can be singular (Example 4.3).

Vuong and Osaka (2026) have also proved Theorem 1.3 and the metric property in Theorem 1.1, for finite tracial von Neumann algebras, by a different method.[1] Their proof uses tracial Carlen–Lieb convexity to control the associated 3×3 quadratic form. Our partition inequality gives an angular proof and also the sufficient decrease estimate for the barycenter iteration (Lemma 5.4).

1.1. **Barycenters.** The metric D_p defines Fréchet means (Fréchet, 1948). Given $A_1, \dots, A_m \in \mathbb{P}_n$ and weights w in the probability simplex Δ_m, a *barycenter* is a minimizer of $\mathcal{F}_w(X) = \sum_i w_i D_p(X, A_i)^2$ over $\mathbb{P}_n$. The functional $\mathcal{F}_w$ is coercive and convex in the cone's linear structure, and strictly convex for $p > 1$. Barycenters exist throughout the range, with uniqueness for $p > 1$ (Theorem 5.2). Positive definite barycenters are characterized by a fixed-point equation.

Theorem 1.4 (Fixed-point characterization)**.** *Let $1 \le p \le 2$, $A_1, \dots, A_m \in \mathbb{P}_n$, $w \in \Delta_m$, and let $X \succ 0$. Then X is a barycenter if and only if*

$$X^p = \sum_{i=1}^{m} w_i (X^{1/2} A_i X^{1/2})^{p/2}. \tag{1.6}$$

At $p = 2$ the equation reduces to $X = \sum_i w_i A_i$. At $p = 1$, for real symmetric data, it characterizes the barycenter of the Gaussians $N(0, A_i)$ in 2-Wasserstein space (Agueh and Carlier, 2011; Álvarez-Esteban et al., 2016) (Corollaries B.1 and B.2). For positive definite data and $1 \le p < 2$, writing the equation in the variables X^p and A_i^p recovers the characterization of the minimizer of a weighted sum of the α-z Bures–Wasserstein divergences above (Dinh et al., 2021; Jeong et al., 2023).

The equation suggests the iteration $X \mapsto \mathcal{T}(X)$ of (5.3). For positive definite data and positive weights, one step of this iteration decreases $\mathcal{F}_w$ by at least the square of the distance it moves (Lemma 5.4), and the iteration converges linearly from every positive definite start (Theorem 5.8, and Proposition D.1 at $p = 1$). An m-point form of the support inequality below also gives a lower bound for the minimum value of $\mathcal{F}_w$ in terms of the pairwise distances of the data (Theorem C.4).

[1] Our proofs of Theorems 1.1 and 1.3 were obtained independently of Vuong and Osaka (2026), which appeared on arXiv on 8 October 2026. They are recorded, together with the barycenter results of Section 5, in commits `02d4b2b5f5ae` and `8baca98a9389` of the author's repository, pushed on 12 August 2026 at 10:08 and 21:02 UTC.

1.2. **Notation and elementary facts.** All matrices are finite-dimensional. We write $X \succ 0$ and $X \succeq 0$ for positive definiteness and semidefiniteness, take matrix powers as the principal powers on the support, and approximate by $X + \varepsilon I$ when inverses appear. For $A \in \mathbb{P}_n$, the radial coordinate is $r_p(A) := (\operatorname{tr} A^p)^{1/2}$. The Hölder inequality (Simon, 2005; Bhatia, 1997) and (1.1) give

$$0 \le \Phi_p(A,B) \le \|A^{1/2}\|_{2p}^p \|B^{1/2}\|_{2p}^p = r_p(A) r_p(B), \tag{1.7}$$

so $D_p(A,B)^2 \ge (r_p(A) - r_p(B))^2$. On the normalized positive sphere $\mathbb{S}_{p,+} := \{A \in \mathbb{P}_n : \operatorname{tr} A^p = 1\}$, homogeneity gives, for nonzero A and B,

$$\widehat{A} := r_p(A)^{-2/p} A \in \mathbb{S}_{p,+}, \qquad \Phi_p(A,B) = r_p(A) r_p(B) \Phi_p(\widehat{A}, \widehat{B}). \tag{1.8}$$

2. The Schatten partition inequality

The partition inequality follows once the column map below has a contractive left inverse. Write p' for the conjugate exponent, so $1/p + 1/p' = 1$.

Proof of Theorem 1.2. For $1 < p \le 2$ and $H_j \succ 0$ for every j, define the column and row maps

$$\mathcal{L}_p(X) := (H_j^{1/p} H^{-1/p} X)_{j=1}^m, \qquad \mathcal{R}_p((X_j)_{j=1}^m) := \sum_{j=1}^m H^{-1/p'} H_j^{1/p'} X_j.$$

They satisfy the exact left-inverse identity

$$\begin{aligned} \mathcal{R}_p \mathcal{L}_p(X) &= \sum_j H^{-1/p'} H_j^{1/p'} H_j^{1/p} H^{-1/p} X \\ &= H^{-1/p'} \left(\sum_j H_j \right) H^{-1/p} X = X. \end{aligned} \tag{2.1}$$

It therefore remains to show that

$$\|\mathcal{R}_p((X_j))\|_p \le \left(\sum_j \|X_j\|_p^p \right)^{1/p}. \tag{2.2}$$

Consider the analytic family $\mathcal{R}_z((X_j)) := \sum_j H^{-z} H_j^z X_j$ on the strip $0 \le \operatorname{Re} z \le 1/2$. On the boundary $z = is$, all imaginary powers are unitary, so

$$\|\mathcal{R}_{is}((X_j))\|_1 \le \sum_j \|X_j\|_1. \tag{2.3}$$

On the boundary $z = 1/2 + is$, put $V_j := H^{-1/2-is} H_j^{1/2+is}$. Then

$$\sum_j V_j V_j^* = H^{-1/2-is} \left(\sum_j H_j \right) H^{-1/2+is} = I. \tag{2.4}$$

Thus $[V_1 \ \cdots \ V_m]$ has operator norm one, and its Hilbert–Schmidt amplification gives

$$\|\mathcal{R}_{1/2+is}((X_j))\|_2^2 \le \sum_j \|X_j\|_2^2. \tag{2.5}$$

Interpolate (2.3) and (2.5) at $\theta = 2/p'$, so that $1/p = 1 - \theta/2$ and $\theta/2 = 1/p'$. The complex interpolation identities for finite Schatten classes and their ℓ_q sums (Bergh and Löfström, 1976; Simon, 2005) give $\mathcal{R}_{1/p'} : \ell_p^m(S_p) \longrightarrow S_p$ with norm at most one. Since $\mathcal{R}_{1/p'} = \mathcal{R}_p$, this proves (2.2). Combining it with (2.1),

$$\|X\|_p = \|\mathcal{R}_p \mathcal{L}_p(X)\|_p \le \|\mathcal{L}_p(X)\|_{\ell_p(S_p)}.$$

Taking the pth power proves (1.4).

At $p = 1$, the identity $X = \sum_j H_j H^{-1} X$ and the trace-norm triangle inequality give the result directly. Semidefinite H_j follow by replacing H_j by $H_j + \varepsilon I$ and passing to the limit. □

3. Proof of Theorem 1.1

We apply the partition inequality at a weighted maximizer to control angles on the sphere, then use a planar comparison to reach the full cone.

3.1. **The weighted support inequality.** To evaluate the partition inequality at a maximizer, we need its Lagrange condition and strict positivity. We first compute the differential.

Lemma 3.1 (Differential of the affinity)**.** *Let $X \succ 0$, $D \succeq 0$ and $H = H^*$. Then*

$$D_X \Phi_p(X, D)[H] = t \operatorname{tr}\Big[X^{-1/2}(X^{1/2} D X^{1/2})^t X^{-1/2} H\Big]. \tag{3.1}$$

Proof. By symmetry, $\Phi_p(X, D) = \operatorname{tr}(D^{1/2} X D^{1/2})^t$. Let Π be the orthogonal projection onto $\operatorname{ran} D$. The matrix $D^{1/2} X D^{1/2}$ vanishes on $\operatorname{ran} \Pi^\perp$. Its compression to $\operatorname{ran} \Pi$ depends linearly on X and is positive definite whenever $X \succ 0$. Since $M \mapsto M^t$ is real analytic on positive definite matrices, $\Phi_p(\cdot, D)$ is real analytic near X, and differentiation under the trace gives

$$D_X \Phi_p(X, D)[H] = t \operatorname{tr}\Big[D^{1/2}(D^{1/2} X D^{1/2})^{t-1} D^{1/2} H\Big],$$

the power $(\cdot)^{t-1}$ being taken on the support. Let $M = X^{1/2} D^{1/2}$, with singular value decomposition $M = U \Sigma V^*$. Then $M(M^*M)^{t-1} M^* = U \Sigma^{2t} U^* = (MM^*)^t$, because $\sigma \cdot \sigma^{2t-2} \cdot \sigma = \sigma^{2t}$ for $\sigma > 0$ and both sides vanish for $\sigma = 0$. This identity shows that the coefficient of H equals the one in (3.1). □

Lemma 3.2 (Positivity of a weighted maximizer)**.** *Let $a_1, \dots, a_m > 0$ and $B_1, \dots, B_m \succ 0$. A maximizer of $F(X) := \sum_{i=1}^m a_i \Phi_p(X, B_i)$ over $X \succeq 0$ subject to $\operatorname{tr} X^p = 1$ is positive definite.*

Proof. Existence follows from compactness. Suppose that a maximizer X is singular, let E be the orthogonal projection onto $\ker X$, and put $k = \operatorname{rank} E > 0$. Let $m_0 > 0$ be the smallest positive eigenvalue of X. For $0 < \varepsilon < m_0$, define $X_\varepsilon := X + \varepsilon E$. Then $X_\varepsilon \geq X$ and $X_\varepsilon \geq \varepsilon I$. For each i, set $M_i := B_i^{1/2} X B_i^{1/2}$ and $M_{i,\varepsilon} := B_i^{1/2} X_\varepsilon B_i^{1/2}$. Since $B_i \succ 0$, $\operatorname{rank} M_i = \operatorname{rank} X = n - k$. Eigenvalue monotonicity gives $\lambda_j(M_{i,\varepsilon}) \geq \lambda_j(M_i)$ for all j, while $M_{i,\varepsilon} \geq \varepsilon \lambda_{\min}(B_i) I$. Hence

$$\operatorname{tr} M_{i,\varepsilon}^t \geq \operatorname{tr} M_i^t + k\big(\varepsilon \lambda_{\min}(B_i)\big)^t. \tag{3.2}$$

Since $\operatorname{tr} X_\varepsilon^p = 1 + k\varepsilon^p$, normalize by $\widetilde{X}_\varepsilon := (1 + k\varepsilon^p)^{-1/p} X_\varepsilon$. The objective is homogeneous of degree t in its first argument. By (3.2),

$$F(\widetilde{X}_\varepsilon) \geq (1 + k\varepsilon^p)^{-t/p} \Big(F(X) + c\varepsilon^t\Big), \qquad c := k \sum_i a_i \lambda_{\min}(B_i)^t > 0.$$

Since $t = p/2 < p$, the positive ε^t term dominates the $O(\varepsilon^p)$ normalization loss. Thus $F(\widetilde{X}_\varepsilon) > F(X)$ for all sufficiently small ε, a contradiction. □

We apply these lemmas to a weighted sum of two affinities.

Proposition 3.3 (Weighted support inequality)**.** *Let $A, B, C \in \mathbb{S}_{p,+}$. Then, for every $a, b \geq 0$,*

$$a\Phi_p(A, B) + b\Phi_p(A, C) \leq \sqrt{a^2 + b^2 + 2ab\Phi_p(B, C)}. \tag{3.3}$$

Proof. We first assume $a, b > 0$ and $B, C \succ 0$. Maximize $F(X) = a\Phi_p(X, B) + b\Phi_p(X, C)$ over $\mathbb{S}_{p,+}$, and let A_* be a maximizer. By Lemma 3.2, $A_* \succ 0$. Put $Y := A_*^{1/2} B A_*^{1/2}$ and $Z := A_*^{1/2} C A_*^{1/2}$. By Lemma 3.1, the Lagrange multiplier equation is

$$aY^t + bZ^t = \lambda A_*^p. \tag{3.4}$$

Taking traces and using $\operatorname{tr} A_*^p = 1$ identifies $\lambda = a\Phi_p(A_*, B) + b\Phi_p(A_*, C) = F(A_*)$.

Set $P := aY^t$, $Q := bZ^t$, and $T := P + Q = \lambda A_*^p$, and abbreviate $x_* := \Phi_p(A_*, B)$, $y_* := \Phi_p(A_*, C)$, and $z := \Phi_p(B, C)$. Since $t/p = 1/2$, we have $P^{1/p} = a^{1/p}Y^{1/2}$, $Q^{1/p} = b^{1/p}Z^{1/2}$, and $T^{-1/p} = \lambda^{-1/p}A_*^{-1}$. These identities give the exact norm evaluations

$$\|Q^{1/p}T^{-1/p}P^{1/p}\|_p^p = \frac{ab}{\lambda}\, z, \tag{3.5}$$

$$\|P^{1/p}T^{-1/p}P^{1/p}\|_p^p = \frac{a^2}{\lambda}, \tag{3.6}$$

$$\|Q^{1/p}T^{-1/p}Q^{1/p}\|_p^p = \frac{b^2}{\lambda}. \tag{3.7}$$

For the cross identity, put $X := Z^{1/2}A_*^{-1}Y^{1/2}$. Then X^*X has the same nonzero spectrum as $YA_*^{-1}ZA_*^{-1} = A_*^{1/2}BCA_*^{-1/2}$, which is similar to BC. Therefore $\|X\|_p^p = \operatorname{tr}(X^*X)^{p/2} = \Phi_p(B, C)$, and (3.5) follows from the scalar factors above. For the first self identity, the positive matrices $Y^{1/2}A_*^{-1}Y^{1/2}$ and $A_*^{-1/2}YA_*^{-1/2} = B$ have the same nonzero spectrum, so their pth power traces agree. The calculation for C is the same. The normalization $\operatorname{tr} B^p = \operatorname{tr} C^p = 1$ then yields (3.6)–(3.7).

Apply Theorem 1.2 to $T = Q + P$ with $X = P^{1/p}$. Using (3.5) and (3.6),

$$\frac{abz + a^2}{\lambda} \geq \|P^{1/p}\|_p^p = \operatorname{tr} P = ax_*.$$

Thus $a^2 + abz \geq \lambda a x_*$. Applying the same inequality with $X = Q^{1/p}$ gives $b^2 + abz \geq \lambda b y_*$. Adding and using $\lambda = ax_* + by_*$,

$$a^2 + b^2 + 2abz \geq \lambda(ax_* + by_*) = \lambda^2. \tag{3.8}$$

Since $\lambda = \max_{X \in \mathbb{S}_{p,+}} F(X)$, evaluating at the given A proves (3.3). If B or C is singular, replace it by a positive definite normalized approximation and pass to the limit. If $a = 0$ or $b = 0$, the claim follows directly from (1.7). □

3.2. **The angular triangle inequality.** To turn the support estimate into a triangle inequality, we express the affinity as the cosine of an angle: on $\mathbb{S}_{p,+}$, define $\vartheta_p(A, B) := \arccos \Phi_p(A, B)$. By (1.7), $\vartheta_p(A, B) \in [0, \pi/2]$.

Theorem 3.4 (Angular triangle inequality)**.** *For all $A, B, C \in \mathbb{S}_{p,+}$,*

$$\vartheta_p(B, C) \leq \vartheta_p(B, A) + \vartheta_p(A, C). \tag{3.9}$$

Proof. Set $\alpha := \vartheta_p(A, B)$, $\beta := \vartheta_p(A, C)$, and $\gamma := \vartheta_p(B, C)$. Thus $\Phi_p(A, B) = \cos\alpha$, $\Phi_p(A, C) = \cos\beta$, and $\Phi_p(B, C) = \cos\gamma$. If $\alpha + \beta \geq \pi/2$, then (3.9) is automatic because $\gamma \leq \pi/2$. Assume therefore that $\alpha + \beta < \pi/2$.

First suppose $\alpha, \beta > 0$. Apply Proposition 3.3 with $a = \sin\beta$ and $b = \sin\alpha$. The left-hand side becomes $\sin\beta\cos\alpha + \sin\alpha\cos\beta = \sin(\alpha + \beta)$. If $\gamma > \alpha + \beta$, then the difference between the square of the right-hand side of (3.3) and $\sin^2(\alpha + \beta)$ is

$$2\sin\alpha\sin\beta\big(\cos\gamma - \cos(\alpha + \beta)\big) < 0, \tag{3.10}$$

contradicting (3.3). Hence $\gamma \leq \alpha + \beta$.

If $\alpha = 0$, apply (3.3) with $a > 0$ and $b > 0$, divide the squared inequality by ab, and let $b/a \downarrow 0$. This gives $\cos\beta \leq \cos\gamma$, so $\gamma \leq \beta$. The case $\beta = 0$ is symmetric. □

Corollary 3.5 (Chordal triangle inequality on the normalized sphere)**.** *For $A, B, C \in \mathbb{S}_{p,+}$,*

$$D_p(A, C) \leq D_p(A, B) + D_p(B, C). \tag{3.11}$$

Proof. On $\mathbb{S}_{p,+}$,

$$D_p(A,B) = \sqrt{2 - 2\cos\vartheta_p(A,B)} = 2\sin\frac{\vartheta_p(A,B)}{2}. \tag{3.12}$$

Apply Theorem 3.4 and the elementary inequality $\sin\frac{u+v}{2} \le \sin\frac{u}{2} + \sin\frac{v}{2}$ for $u, v \ge 0$ with $u + v \le \pi$. □

3.3. **From the sphere to the cone.**

Proof of Theorem 1.1. First assume $A, B, C \neq 0$. Let $r_A := r_p(A)$, $r_B := r_p(B)$, and $r_C := r_p(C)$. Normalization as in (1.8) gives the law of cosines

$$D_p(A,B)^2 = r_A^2 + r_B^2 - 2r_A r_B \cos\vartheta_p(\widehat{A},\widehat{B}). \tag{3.13}$$

Put $a := \vartheta_p(\widehat{A},\widehat{B})$, $b := \vartheta_p(\widehat{B},\widehat{C})$, and $c := \vartheta_p(\widehat{A},\widehat{C})$. By Theorem 3.4, $c \le a + b$. Since all three angles lie in $[0, \pi/2]$, we have $a + b \le \pi$.

Choose planar vectors $u_A := r_A(1,0)$, $u_B := r_B(\cos a, \sin a)$, and $u_C := r_C(\cos(a+b), \sin(a+b))$. By (3.13), $\|u_A - u_B\| = D_p(A,B)$ and $\|u_B - u_C\| = D_p(B,C)$. Moreover, the Euclidean law-of-cosines expression is increasing in its angle on $[0,\pi]$, so $c \le a + b$ implies $D_p(A,C) \le \|u_A - u_C\|$. The ordinary Euclidean triangle inequality therefore gives

$$\begin{aligned} D_p(A,C) &\le \|u_A - u_C\| \\ &\le \|u_A - u_B\| + \|u_B - u_C\| \\ &= D_p(A,B) + D_p(B,C). \end{aligned}$$

The cases involving the zero matrix follow by continuity, or directly from $D_p(A,0) = r_p(A)$.

Symmetry and nonnegativity were noted in (1.1) and (1.7). Finally, if $D_p(A,B) = 0$, then (3.13) forces $r_A = r_B$ and equality in Schatten Hölder in (1.7). The standard equality characterization gives A^p proportional to B^p; the equal radii make the proportionality constant one, hence $A = B$. □

4. The LvD conjecture

To prove Theorem 1.3, we express the block-norm entries as affinities and use the angular triangle inequality. The conjecture of Lin and van den Driessche (2014, Conjecture 5.1) was known at both endpoints: $p = 1$ is a theorem of Drury (2014), with a short proof in Li and Zhang (2015), and $p = 2$ follows from Marcus and Watkins (1971, Theorem 1), for any number of blocks. For scalar blocks, it follows from the fractional Hadamard power theorem of FitzGerald and Horn (1977). The $p = 1$ statement had also been asserted for matrices of fidelities by Fannes et al. (2012), with an argument that Lin and van den Driessche (2014) found incomplete.

The pth powers in the conjecture are essential: Li and Zhang (2015) give a positive semidefinite block matrix for which $[\|H_{ij}\|_p]$ is indefinite for every $p > 1$. For the powered entries, the following three-point criterion converts the angular triangle inequality into positivity of a cosine matrix.

Lemma 4.1 (Three-point cosine criterion)**.** *Let $c_{12}, c_{13}, c_{23} \in [0,1]$, and set*

$$C := \begin{bmatrix} 1 & c_{12} & c_{13} \\ c_{12} & 1 & c_{23} \\ c_{13} & c_{23} & 1 \end{bmatrix}.$$

Define $\alpha := \arccos c_{12}$, $\beta := \arccos c_{13}$, and $\gamma := \arccos c_{23}$. Then $C \succeq 0$ if and only if

$$|\alpha - \beta| \le \gamma \le \alpha + \beta. \tag{4.1}$$

Proof. The 1×1 and 2×2 principal minors of C are nonnegative because $0 \le c_{ij} \le 1$. Since a Hermitian matrix is positive semidefinite exactly when all its principal minors are nonnegative, $C \succeq 0$ is equivalent to $\det C \ge 0$. Directly,

$$\det C = 1 + 2c_{12}c_{13}c_{23} - c_{12}^2 - c_{13}^2 - c_{23}^2. \tag{4.2}$$

Regarded as a quadratic polynomial in c_{23} with leading coefficient -1, its roots are $c_{12}c_{13} \pm \sqrt{(1-c_{12}^2)(1-c_{13}^2)} = \cos(\alpha \mp \beta)$. Therefore

$$\det C = \big(\cos|\alpha-\beta| - \cos\gamma\big)\big(\cos\gamma - \cos(\alpha+\beta)\big). \tag{4.3}$$

The determinant is nonnegative exactly when $\cos\gamma$ lies between the two roots. Since $0 \le |\alpha-\beta| \le \alpha+\beta \le \pi$ and cosine is decreasing on $[0,\pi]$, this is equivalent to (4.1). □

Proof of Theorem 1.3. Choose a block Gram factorization $\mathbf{H} = X^*X$, with $X = [\,X_1 \;\; X_2 \;\; X_3\,]$, for instance by taking $X = \mathbf{H}^{1/2}$ and splitting its columns conformally with the block structure. Then $H_{ij} = X_i^*X_j$. All matrices X_i have a common row dimension, although their column dimensions may differ. Define positive semidefinite matrices on this common row space by $A_i := X_iX_i^*$ for $i = 1,2,3$. The block-norm entries are their affinities:

$$\begin{aligned} \|H_{ij}\|_p^p &= \operatorname{tr}|X_i^*X_j|^p \\ &= \operatorname{tr}\left(X_j^*X_iX_i^*X_j\right)^{p/2} \\ &= \operatorname{tr}\left(A_i^{1/2}A_jA_i^{1/2}\right)^{p/2} \\ &= \Phi_p(A_i,A_j). \end{aligned} \tag{4.4}$$

The third equality holds because, for $M = A_i^{1/2}X_j$, the positive matrices $M^*M = X_j^*A_iX_j$ and $MM^* = A_i^{1/2}X_jX_j^*A_i^{1/2} = A_i^{1/2}A_jA_i^{1/2}$ have the same nonzero eigenvalues, counted with multiplicity.

Set $r_i := r_p(A_i) = (\operatorname{tr}A_i^p)^{1/2}$. If $r_i = 0$, then $A_i = 0$, hence $X_i = 0$, and the ith row and column of the matrix in (1.5) vanish. We may therefore remove such indices and assume $r_i > 0$ for all i. Normalize as in (1.8), setting $\widehat{A}_i := r_i^{-2/p}A_i \in \mathbb{S}_{p,+}$. Radial homogeneity gives $\Phi_p(A_i,A_j) = r_ir_j\Phi_p(\widehat{A}_i,\widehat{A}_j)$. Consequently,

$$[\|H_{ij}\|_p^p]_{i,j=1}^3 = DCD, \qquad D := \operatorname{diag}(r_1,r_2,r_3), \tag{4.5}$$

where $C := [\Phi_p(\widehat{A}_i,\widehat{A}_j)]_{i,j=1}^3$. The diagonal entries are $\Phi_p(\widehat{A}_i,\widehat{A}_i) = \operatorname{tr}\widehat{A}_i^p = 1$, and its off-diagonal entries $c_{ij} := \Phi_p(\widehat{A}_i,\widehat{A}_j)$ lie in $[0,1]$ by (1.7).

Define $\theta_{ij} := \vartheta_p(\widehat{A}_i,\widehat{A}_j) = \arccos c_{ij}$. Applying Theorem 3.4 to all three orderings of the triple gives $|\theta_{12}-\theta_{13}| \le \theta_{23} \le \theta_{12}+\theta_{13}$. Therefore Lemma 4.1 yields $C \succeq 0$. Finally, (4.5) gives $[\|H_{ij}\|_p^p]_{i,j=1}^3 = DCD \succeq 0$, as required. □

Remark 4.2 (Semidefinite versus definite conclusions)**.** *If $\mathbf{H} \succ 0$, then $[\|H_{ij}\|_p^p] \succ 0$. Indeed, choose $\varepsilon > 0$ such that $\mathbf{H} - \varepsilon I \succeq 0$ and apply Theorem 1.3 to this matrix. Its off-diagonal blocks are unchanged, while each diagonal entry of the scalar matrix is strictly smaller. Thus $[\|H_{ij}\|_p^p]$ is a positive semidefinite matrix plus a strictly positive diagonal matrix.*

The following singular configuration also attains equality in the angular triangle inequality.

Example 4.3 (A tight configuration, for every p)**.** *Take $n = 2$, let $u = (1,1)^{\mathsf{T}}/\sqrt{2}$ and $v = (1,-1)^{\mathsf{T}}/\sqrt{2}$, and set $A_1 := uu^*$, $A_2 := vv^*$, and $A_3 := 2^{-1/p}I_2$. All three lie on $\mathbb{S}_{p,+}$, since $\operatorname{tr}A_1^p = \operatorname{tr}A_2^p = 1$ and $\operatorname{tr}A_3^p = 2\cdot 2^{-1} = 1$. For a rank-one $A = ww^*$ with $\|w\| = 1$ and any $B \succeq 0$ one has $A^{1/2} = A$ and $ABA = (w^*Bw)ww^*$, so $\Phi_p(A,B) = (w^*Bw)^{p/2}$. Hence*

$$\Phi_p(A_1,A_2) = |\langle u,\, v\rangle|^p = 0, \qquad \Phi_p(A_1,A_3) = \Phi_p(A_2,A_3) = \big(2^{-1/p}\big)^{p/2} = \tfrac{1}{\sqrt{2}},$$

that is, $\vartheta_p(A_1,A_2) = \pi/2$ and $\vartheta_p(A_1,A_3) = \vartheta_p(A_2,A_3) = \pi/4$. The angular triangle inequality (3.9) therefore holds with equality*, for every $1 \le p \le 2$; correspondingly the cosine matrix C has $\det C = 0$, so the matrix (1.5) built from $H_{ij} = A_i^{1/2}A_j^{1/2}$ is positive semidefinite and singular.*

Remark 4.4 (The range of p)**.** *For $n \geq 2$, the triangle inequality for D_p fails for every $p > 2$ (Vuong and Osaka, 2026, Theorem 1.3(ii)). The block-positivity statement also fails in this range: applied to $H_{ij} = A_i^{1/2} A_j^{1/2}$, it would make $[\Phi_p(A_i, A_j)]_{i,j=1}^3$ a Gram matrix, so the distances $D_p(A_i, A_j)$ would be Euclidean distances and would satisfy the triangle inequality.*

5. Barycenters

To study Fréchet means for D_p, fix $A_1, \ldots, A_m \in \mathbb{P}_n$, not all zero, and weights in $\Delta_m := \{w \in \mathbb{R}^m : w_i \geq 0,\ \sum_i w_i = 1\}$. The *Fréchet functional* is $\mathcal{F}_w(X) := \sum_{i=1}^m w_i D_p(X, A_i)^2$ for $X \in \mathbb{P}_n$; any minimizer is a *barycenter* of (A_i, w_i) for D_p. Expanding (1.2),

$$\mathcal{F}_w(X) = \operatorname{tr} X^p - 2\Psi_w(X) + \sum_{i=1}^m w_i \operatorname{tr} A_i^p, \qquad \Psi_w(X) := \sum_{i=1}^m w_i \Phi_p(X, A_i). \tag{5.1}$$

The affinity sum Ψ_w is positively homogeneous of degree $t = p/2$, while $\operatorname{tr} X^p$ is homogeneous of degree p. For $m = 2$ and $w = (a, b)$, Ψ_w is the objective in Proposition 3.3.

5.1. Convexity, existence, uniqueness.

Proposition 5.1 (Convexity)**.** *Let $1 \leq p \leq 2$.*

(i) *For every $A \in \mathbb{P}_n$, the map $X \mapsto \Phi_p(X, A)$ is concave and nondecreasing on $\mathbb{P}_n$.*
(ii) *The map $X \mapsto \operatorname{tr} X^p$ is convex on $\mathbb{P}_n$, and strictly convex when $p > 1$.*
(iii) *For every $w \in \Delta_m$, $\mathcal{F}_w$ is convex on $\mathbb{P}_n$, and strictly convex when $p > 1$.*

Proof. (i) By (1.1) and the coincidence of nonzero spectra, $\Phi_p(X, A) = \operatorname{tr}(A^{1/2} X A^{1/2})^t$ with $t = p/2 \in [\frac{1}{2}, 1]$. The map $X \mapsto A^{1/2} X A^{1/2}$ is linear and maps $\mathbb{P}_n$ into $\mathbb{P}_n$, and $M \mapsto \operatorname{tr} M^t$ is concave and nondecreasing on $\mathbb{P}_n$ for $t \in [0, 1]$, since $\tau \mapsto \tau^t$ is concave and nondecreasing on $[0, \infty)$ (Bhatia, 1997; Carlen, 2010). Composing these maps proves (i).

(ii) On $\mathbb{P}_n$ we have $\operatorname{tr} X^p = \|X\|_p^p$, and $\|\cdot\|_p$ is a norm for $p \geq 1$, so convexity follows from convexity and monotonicity of $\tau \mapsto \tau^p$ on $[0, \infty)$. For strictness let $p > 1$, let $X \neq Y$ in $\mathbb{P}_n$, let $\lambda \in (0, 1)$, and set $Z = \lambda X + (1 - \lambda) Y$. If $\|X\|_p \neq \|Y\|_p$, then

$$\|Z\|_p^p \leq \left(\lambda \|X\|_p + (1 - \lambda)\|Y\|_p\right)^p < \lambda \|X\|_p^p + (1 - \lambda)\|Y\|_p^p$$

by strict convexity of $\tau \mapsto \tau^p$. If $\|X\|_p = \|Y\|_p =: \rho$, then $\rho > 0$ because $X \neq Y$, and the Schatten class S_p is uniformly, hence strictly, convex for $1 < p < \infty$ (McCarthy, 1967; Simon, 2005); therefore $\|Z\|_p < \rho$ and $\|Z\|_p^p < \rho^p$.

(iii) By (5.1), $\mathcal{F}_w$ is the sum of the convex functions $\operatorname{tr} X^p$ and $-2\Psi_w$, and a constant. Strict convexity for $p > 1$ comes from the first summand. □

Theorem 5.2 (Existence and uniqueness)**.** *Let $1 \leq p \leq 2$ and $w \in \Delta_m$. Then $\mathcal{F}_w$ has a nonempty compact convex set of minimizers on $\mathbb{P}_n$, and the minimizer is unique if $1 < p \leq 2$.*

Proof. By (1.7), $D_p(X, A_i)^2 \geq (r_p(X) - r_p(A_i))^2$, so $\mathcal{F}_w(X) \to \infty$ as $r_p(X) \to \infty$. By continuity on the closed cone $\mathbb{P}_n$, its sublevel sets are compact, so it attains its minimum. The minimizer set is compact; by Proposition 5.1, it is convex and reduces to a point for $p > 1$. □

At $p = 1$ the term $\operatorname{tr} X$ is linear, so Proposition 5.1(iii) gives convexity but not strict convexity. Nevertheless, the barycenter is unique whenever some A_i with $w_i > 0$ is nonsingular (Agueh and Carlier, 2011; Álvarez-Esteban et al., 2016). For the Euler characterization, we next show that positive definite data give interior minimizers.

Lemma 5.3 (Positivity of the barycenter)**.** *Let $1 \leq p \leq 2$, $w \in \Delta_m$ with all $w_i > 0$, and suppose $A_i \succ 0$ for every i. Then every barycenter is positive definite.*

Proof. Suppose $X \in \mathbb{P}_n$ is singular, let E be the orthogonal projection onto $\ker X$, $k = \operatorname{rank} E > 0$, and let $m_0 > 0$ be the smallest positive eigenvalue of X (take $m_0 = 1$ if $X = 0$). For $0 < \varepsilon < m_0$ put $X_\varepsilon := X + \varepsilon E$, so that $X_\varepsilon \succeq \varepsilon I$. Since E is supported on $\ker X$, we have $\operatorname{tr} X_\varepsilon^p = \operatorname{tr} X^p + k\varepsilon^p$. As in (3.2), each affinity satisfies $\Phi_p(X_\varepsilon, A_i) \geq \Phi_p(X, A_i) + k(\varepsilon\lambda_{\min}(A_i))^t$. Hence, by (5.1),

$$\mathcal{F}_w(X_\varepsilon) - \mathcal{F}_w(X) \leq k\varepsilon^p - 2k\varepsilon^t \sum_i w_i \lambda_{\min}(A_i)^t.$$

Since $t = p/2 < p$, this is negative for sufficiently small $\varepsilon > 0$, so X cannot minimize $\mathcal{F}_w$. □

5.2. **The Euler equation.**

Proof of Theorem 1.4. By Lemma 3.1, each $\Phi_p(\cdot, A_i)$ is real analytic near $X \succ 0$, while $X \mapsto \operatorname{tr} X^p$ has differential $H \mapsto p \operatorname{tr}[X^{p-1}H]$. Thus (5.1), with $2t = p$, gives

$$\nabla \mathcal{F}_w(X) = p\left(X^{p-1} - X^{-1/2}\Big(\sum_i w_i (X^{1/2} A_i X^{1/2})^{p/2}\Big) X^{-1/2}\right). \tag{5.2}$$

Conjugating by $X^{1/2}$ shows that $\nabla \mathcal{F}_w(X) = 0$ is equivalent to (1.6).

A positive definite X is interior to $\mathbb{P}_n$ in the real vector space of Hermitian matrices, so any such minimizer has vanishing gradient. Conversely, convexity (Proposition 5.1(iii)) makes every stationary point a global minimizer. □

5.3. **Convergence of the fixed-point iteration.** Equation (1.6) suggests the iteration

$$\mathcal{T}(X) := X^{-1/2}\Big(\sum_{i=1}^m w_i (X^{1/2} A_i X^{1/2})^{p/2}\Big)^{2/p} X^{-1/2}, \tag{5.3}$$

whose fixed points in the positive definite cone are exactly the barycenters there, by Theorem 1.4. At $p = 2$ the map is constant, $\mathcal{T}(X) = \sum_i w_i A_i$, so it converges in one step. For real symmetric data at $p = 1$, it is the Gaussian Wasserstein barycenter iteration of Álvarez-Esteban et al. (2016); linear convergence rates were established by Chewi et al. (2020).

For the full range $1 \leq p \leq 2$, we combine sufficient decrease from the partition inequality (Lemma 5.4), confinement of the iterates to a spectral box (Lemma 5.6), and an error bound that converts the step length into control of the optimality gap (Lemma 5.7). Appendix D supplies the confinement and error-bound proofs, including $p = 1$; we prove sufficient decrease below.

Throughout the rest of this subsection, $A_1, \ldots, A_m \succ 0$ and $w \in \Delta_m$ has all $w_i > 0$. We abbreviate the right side of the Euler equation (1.6) by

$$S(X) := \sum_{i=1}^m w_i (X^{1/2} A_i X^{1/2})^{p/2}, \qquad \text{so that} \qquad \mathcal{T}(X) = X^{-1/2} S(X)^{1/t} X^{-1/2}. \tag{5.4}$$

Since $S(X) \succ 0$ whenever $X \succ 0$, the iteration is well defined:

$$X_{k+1} := \mathcal{T}(X_k), \qquad X_0 \succ 0. \tag{5.5}$$

Lemma 5.4 (Sufficient decrease)**.** *For every $X \succ 0$,*

$$\boxed{\mathcal{F}_w(\mathcal{T}(X)) + D_p(X, \mathcal{T}(X))^2 \;\leq\; \mathcal{F}_w(X).} \tag{5.6}$$

In particular $\mathcal{F}_w$ is nonincreasing along (5.5), strictly so unless $\mathcal{T}(X) = X$.

Proof. Put $K_i := (X^{1/2} A_i X^{1/2})^{1/2}$ and $K := (\sum_{i=1}^m w_i K_i^p)^{1/p}$, so that $K^p = S(X)$ and $\mathcal{T}(X) = X^{-1/2} K^2 X^{-1/2}$. Since $X^{1/2}\mathcal{T}(X)X^{1/2} = K^2$, we have the exact identity

$$\Phi_p(X, \mathcal{T}(X)) = \operatorname{tr} K^p = \sum_{i=1}^m w_i \Phi_p(X, A_i). \tag{5.7}$$

This cancels the affinity terms in the decrease estimate. For the remaining terms, use

$$\Phi_p(\mathcal{T}(X), A_i) = \|K_i X^{-1} K\|_p^p, \qquad \operatorname{tr} \mathcal{T}(X)^p = \|K X^{-1} K\|_p^p. \tag{5.8}$$

For the first, $A_i = X^{-1/2} K_i^2 X^{-1/2}$, so $\mathcal{T}(X) A_i$ is similar to $K^2 X^{-1} K_i^2 X^{-1}$, which has the same nonzero spectrum as $K X^{-1} K_i^2 X^{-1} K = (K_i X^{-1} K)^*(K_i X^{-1} K)$; taking the $(p/2)$th power trace gives the claim. For the second, $\mathcal{T}(X) = (X^{-1/2} K)(K X^{-1/2})$ and $K X^{-1} K = (K X^{-1/2})(X^{-1/2} K)$ have the same nonzero spectrum, and both are positive semidefinite.

Apply Theorem 1.2 with $H_i := w_i K_i^p$, $H := \sum_i H_i = K^p$, and $K X^{-1} K$ as the free argument. Here $H_i^{1/p} = w_i^{1/p} K_i$ and $H^{-1/p} = K^{-1}$, so (1.4) reads

$$\sum_{i=1}^m w_i \|K_i X^{-1} K\|_p^p \;\geq\; \|K X^{-1} K\|_p^p, \tag{5.9}$$

which by (5.8) is exactly

$$\sum_{i=1}^m w_i \Phi_p(\mathcal{T}(X), A_i) \;\geq\; \operatorname{tr} \mathcal{T}(X)^p. \tag{5.10}$$

Expanding by (1.2) and cancelling with (5.7) gives

$$\mathcal{F}_w(X) - \mathcal{F}_w(\mathcal{T}(X)) - D_p(X, \mathcal{T}(X))^2 = 2 \left[\sum_{i=1}^m w_i \Phi_p(\mathcal{T}(X), A_i) - \operatorname{tr} \mathcal{T}(X)^p \right], \tag{5.11}$$

whose right side is nonnegative by (5.10). If $\mathcal{T}(X) \neq X$ then $D_p(X, \mathcal{T}(X)) > 0$, because D_p is a metric (Theorem 1.1), so the decrease is strict. □

Remark 5.5 (What the proof uses)**.** *The proof also applies to semidefinite A_i provided $X \succ 0$ and $S(X) \succ 0$. By (5.11), the remainder in the decrease estimate is twice the surplus in the partition inequality (5.9). At $p = 2$ both inequalities hold with equality.*

Lemma 5.4 gives convergence of the objective values. To obtain a rate, we first bound the spectra of the iterates away from zero and infinity.

Lemma 5.6 (Uniform spectral confinement)**.** *Fix $X_0 \succ 0$, let $F_0 := \mathcal{F}_w(X_0)$, and put*

$$\bar{r} := \sum_{i=1}^m w_i r_p(A_i), \qquad \lambda_+ := \left(\bar{r} + \sqrt{F_0}\right)^{2/p}, \tag{5.12}$$

$$c_A := \sum_{i=1}^m w_i \lambda_{\min}(A_i)^t, \qquad \lambda_- := \frac{c_A^{n/t}}{\lambda_+^{n-1}}. \tag{5.13}$$

Then every iterate of (5.5) with $k \geq 1$, and every barycenter $X_\star$, lies in the spectral box $\mathcal{B} := \{X \in \mathbb{P}_n : \lambda_- I \preceq X \preceq \lambda_+ I\}$.

Within this box, the step length $D_p(X, \mathcal{T}(X))$ controls the optimality gap. This step uses $p > 1$; Proposition D.1 treats $p = 1$.

Lemma 5.7 (Residual error bound)**.** *Let $1 < p \leq 2$ and let $X \in \mathcal{B}$ with $\mathcal{T}(X) \in \mathcal{B}$. Then*

$$D_p(X, \mathcal{T}(X))^2 \;\geq\; \rho_p(\mathcal{F}_w(X) - \mathcal{F}_w(X_\star)), \qquad \rho_p := \frac{4(p-1)^2}{p^2} \left(\frac{\lambda_-}{\lambda_+}\right)^{6-2p}, \tag{5.14}$$

where $X_\star$ is the barycenter. Moreover $0 < \rho_p \leq 1$.

Appendix D proves both lemmas; combined with Lemma 5.4, they give linear convergence.

Theorem 5.8 (Global linear convergence)**.** *Let $1 < p \le 2$, let $A_1, \dots, A_m \succ 0$, let $w \in \Delta_m$ have positive coordinates, let $X_0 \succ 0$, and let $X_\star$ be the barycenter. With $\lambda_\pm$ as in (5.12)–(5.13) and ρ_p as in (5.14), the iteration (5.5) satisfies, for every $k \ge 1$,*

$$\boxed{\mathcal{F}_w(X_k) - \mathcal{F}_w(X_\star) \;\le\; (1-\rho_p)^{k-1}\big(\mathcal{F}_w(X_1) - \mathcal{F}_w(X_\star)\big),} \tag{5.15}$$

so the objective values converge Q-linearly, and the iterates converge R-linearly,

$$\|X_k - X_\star\|_F \;\le\; \left(\frac{2\big(\mathcal{F}_w(X_1) - \mathcal{F}_w(X_\star)\big)}{p(p-1)\lambda_+^{p-2}}\right)^{1/2} (1-\rho_p)^{(k-1)/2}. \tag{5.16}$$

Moreover the computable step length certifies the gap a posteriori:

$$\mathcal{F}_w(X_k) - \mathcal{F}_w(X_\star) \;\le\; \rho_p^{-1} D_p(X_k, X_{k+1})^2. \tag{5.17}$$

Proof. By Lemma 5.6, $X_k \in \mathcal{B}$ for every $k \ge 1$, and $X_\star \in \mathcal{B}$. For $k \ge 1$, Lemma 5.4 followed by Lemma 5.7 at $X = X_k$ gives

$$\mathcal{F}_w(X_{k+1}) - \mathcal{F}_w(X_\star) \le \mathcal{F}_w(X_k) - \mathcal{F}_w(X_\star) - D_p(X_k, X_{k+1})^2 \le (1-\rho_p)\big(\mathcal{F}_w(X_k) - \mathcal{F}_w(X_\star)\big),$$

and iterating from $k = 1$ proves (5.15). Strong convexity of $\mathcal{F}_w$ on $\mathcal{B}$ with modulus $\mu_F = p(p-1)\lambda_+^{p-2}$, together with $\nabla\mathcal{F}_w(X_\star) = 0$, gives $\mathcal{F}_w(X) - \mathcal{F}_w(X_\star) \ge \frac{1}{2}\mu_F\|X - X_\star\|_{\mathrm{F}}^2$ on $\mathcal{B}$, which with (5.15) gives (5.16). Estimate (5.17) is (5.14) at $X = X_k$. □

Remark 5.9 (What is sharp and what is not)**.** *The sufficient decrease estimate is sharp, but the rate ρ_p is uniform over the spectral box containing the iterates after the first step. The determinant bound for λ_- can make this rate deteriorate rapidly with the dimension. On random instances with $n \le 5$, the certified ρ_p goes down to 10^{-57}, against a per-step contraction of order one. Theorem 5.8 thus establishes global linear convergence with explicit constants depending on the data, weights, p, and X_0, but the bound can be too small to predict the observed rate. A local estimate from the derivative of $\mathcal{T}$ at $X_\star$ may be sharper. The a posteriori bound (5.17) inherits the same limitation; Proposition C.6 instead bounds $\min \mathcal{F}_w$ from below without using $\lambda_\pm$.*

AI USAGE

The results in this paper were developed through extensive use of an AI assistant (OpenAI's GPT Pro). None of the claims were obtained zero-shot: they emerged through sustained back-and-forth discussion, including failed attempts, counterexamples, audits, and repairs. The AI-driven side of this work began in October 2025; the first complete proof was obtained on 11 August 2026.

ACKNOWLEDGEMENTS

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

REFERENCES

A. Afham and Chris Ferrie. Riemannian-geometric generalizations of quantum fidelities and Bures–Wasserstein distance. *Journal of Mathematical Physics*, 66(8):082201, 2025. doi: 10.1063/5.0252803.

Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011. doi: 10.1137/100805741.

Pedro C. Álvarez-Esteban, Eustasio del Barrio, Juan A. Cuesta-Albertos, and Carlos Matrán. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2):744–762, 2016. doi: 10.1016/j.jmaa.2016.04.045.

William N. Anderson, Jr. and Richard J. Duffin. Series and parallel addition of matrices. *Journal of Mathematical Analysis and Applications*, 26(3):576–594, 1969. doi: 10.1016/0022-247X(69)90200-5.

Huzihiro Araki. On an inequality of Lieb and Thirring. *Letters in Mathematical Physics*, 19(2): 167–170, 1990. doi: 10.1007/BF01045887.

Jöran Bergh and Jörgen Löfström. *Interpolation Spaces: An Introduction*. Grundlehren der mathematischen Wissenschaften 223. Springer, Berlin, 1976. doi: 10.1007/978-3-642-66451-9.

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

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019. doi: 10.1016/j.exmath.2018.01.002.

Donald Bures. An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w^*-algebras. *Transactions of the American Mathematical Society*, 135: 199–212, 1969. doi: 10.2307/1995012.

Eric Carlen. Trace inequalities and quantum entropy: an introductory course. In *Entropy and the Quantum*, Contemporary Mathematics 529, pages 73–140. American Mathematical Society, Providence, RI, 2010. doi: 10.1090/conm/529/10428.

Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J. Stromme. Gradient descent algorithms for Bures–Wasserstein barycenters. In *Proceedings of the Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 1276–1304. PMLR, 2020. URL https://proceedings.mlr.press/v125/chewi20a.html.

Trung Hoa Dinh, Cong Trinh Le, Bich Khue Vo, and Trung Dung Vuong. The α-z-Bures Wasserstein divergence. *Linear Algebra and its Applications*, 624:267–280, 2021. doi: 10.1016/j.laa.2021.04.007.

Stephen W. Drury. Positive semidefiniteness of a 3×3 matrix related to partitioning. *Linear Algebra and its Applications*, 446:369–376, 2014. doi: 10.1016/j.laa.2013.12.037.

Mark Fannes, Fernando de Melo, Wojciech Roga, and Karol Życzkowski. Matrices of fidelities for ensembles of quantum states and the Holevo quantity. *Quantum Information and Computation*, 12(5–6):472–489, 2012. doi: 10.26421/QIC12.5-6-8.

Carl H. FitzGerald and Roger A. Horn. On fractional Hadamard powers of positive definite matrices. *Journal of Mathematical Analysis and Applications*, 61(3):633–642, 1977. doi: 10.1016/0022-247X(77)90167-6.

Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10(4):215–310, 1948.

Miran Jeong, Jinmi Hwang, and Sejong Kim. Right mean for the $\alpha - z$ Bures–Wasserstein quantum divergence. *Acta Mathematica Scientia*, 43(5):2320–2332, 2023. doi: 10.1007/s10473-023-0523-7.

Ábel Komálovics and Lajos Molnár. On a parametric family of distance measures that includes the Hellinger and the Bures distances. *Journal of Mathematical Analysis and Applications*, 529(2): Paper No. 127226, 2024. doi: 10.1016/j.jmaa.2023.127226.

Chi-Kwong Li and Fuzhen Zhang. Positivity of partitioned Hermitian matrices with unitarily invariant norms. *Positivity*, 19(3):439–444, 2015. doi: 10.1007/s11117-014-0306-4.

Minghua Lin and P. van den Driessche. Positive semidefinite 3×3 block matrices. *Electronic Journal of Linear Algebra*, 27:827–836, 2014. doi: 10.13001/1081-3810.2003.

Marvin Marcus and William Watkins. Partitioned Hermitian matrices. *Duke Mathematical Journal*, 38(2):237–249, 1971. doi: 10.1215/S0012-7094-71-03830-0.

Charles A. McCarthy. c_p. *Israel Journal of Mathematics*, 5:249–271, 1967. doi: 10.1007/BF02771613.

Hà Quang Minh. Alpha Procrustes metrics between positive definite operators: a unifying formulation for the Bures–Wasserstein and Log-Euclidean/Log-Hilbert–Schmidt metrics. *Linear Algebra and its Applications*, 636:25–68, 2022. doi: 10.1016/j.laa.2021.11.011.

Ingram Olkin and Friedrich Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982. doi: 10.1016/0024-3795(82)90112-4.

Barry Simon. *Trace Ideals and Their Applications*. Mathematical Surveys and Monographs 120. American Mathematical Society, Providence, RI, 2nd edition, 2005. doi: 10.1090/surv/120.

Asuka Takatsu. Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 48(4): 1005–1026, 2011.

Trung Dung Vuong and Hiroyuki Osaka. Three-point trace positivity: Schatten norm compression and interpolating metrics. arXiv:2610.11021 [math.OA], 2026. Posted 8 October 2026.

Lu Wei, Zhian Jia, Dagomir Kaszlikowski, and Sheng Tan. Antilinear superoperator, quantum geometric invariance, and antilinear symmetry for higher-dimensional quantum systems. *Quantum Information Processing*, 23:290, 2024. doi: 10.1007/s11128-024-04499-3.

Appendices

A A balanced trace inequality 13
B Endpoints and closed forms 14
C The variance inequality and three-point geometry 14
D Proofs of the convergence estimates, and the endpoint $p = 1$ 18
E A numerical comparison with first-order methods 20

Appendix A. A balanced trace inequality

For comparison with the support inequality, we record the trace estimate obtained by combining Theorem 1.2 with the Araki–Lieb–Thirring inequality. For positive definite P, Q, the parallel sum is $P : Q := (P^{-1} + Q^{-1})^{-1} = P(P + Q)^{-1}Q$. It extends to semidefinite matrices by positive regularization (Anderson and Duffin, 1969) and satisfies $(\alpha P) : (\alpha Q) = \alpha(P : Q)$ for $\alpha \geq 0$.

Theorem A.1 (Balanced trace inequality)**.** *Let $1 \leq p \leq 2$, $t = p/2$, and let $A, B, C \succ 0$. Define $Y := A^{1/2}BA^{1/2}$ and $Z := A^{1/2}CA^{1/2}$. If $Y^t + Z^t = \lambda A^p$ for some $\lambda > 0$, then*

$$\Phi_p(B, C) \geq \lambda \operatorname{tr}(Y^t : Z^t). \tag{A.1}$$

The conclusion also holds when A, B, C have a common support, by restricting to that support.

Proof. Set $P := Y^t$, $Q := Z^t$, and $T := P + Q = \lambda A^p$. Since $t/p = 1/2$, we have $P^{1/p} = Y^{1/2}$, $Q^{1/p} = Z^{1/2}$, and $T^{-1/p} = \lambda^{-1/p}A^{-1}$. Consequently,

$$\Phi_p(B, C) = \lambda \|Q^{1/p}T^{-1/p}P^{1/p}\|_p^p. \tag{A.2}$$

Indeed, $\|Z^{1/2}A^{-1}Y^{1/2}\|_p^p = \Phi_p(B, C)$ by the argument given for (3.5) in the proof of Proposition 3.3, with A in place of A_*, and (A.2) follows from the scalar factor $\lambda^{-1/p}$.

Apply Theorem 1.2 to the decomposition $T = Q + P$ and to $X = P^{1/p}$. We get

$$\|Q^{1/p}T^{-1/p}P^{1/p}\|_p^p + \|P^{1/p}T^{-1/p}P^{1/p}\|_p^p \geq \operatorname{tr} P. \tag{A.3}$$

For $p \geq 1$, the Araki–Lieb–Thirring trace inequality (Araki, 1990; Bhatia, 1997) gives

$$\operatorname{tr}\left(P^{1/p}T^{-1/p}P^{1/p}\right)^p \leq \operatorname{tr}(PT^{-1}P). \tag{A.4}$$

Indeed, apply the Araki–Lieb–Thirring inequality to $X = P^{2/p}$, $W = T^{-1/p}$, and exponent p. Combining (A.3) and (A.4),

$$\begin{aligned} \|Q^{1/p}T^{-1/p}P^{1/p}\|_p^p &\geq \operatorname{tr} P - \operatorname{tr}(PT^{-1}P) \\ &= \operatorname{tr}(PT^{-1}Q) = \operatorname{tr}(P : Q). \end{aligned} \tag{A.5}$$

Multiply by λ and use (A.2). □

Remark A.2. *No Araki–Lieb–Thirring inequality is used in Proposition 3.3. The partition inequality evaluates the cross term and the relevant self term simultaneously, which is stronger than first bounding the self term by a separate trace inequality.*

Appendix B. Endpoints and closed forms

Corollary B.1 ($p = 2$: the arithmetic mean)**.** *For $p = 2$ the unique barycenter is $X_\star = \sum_i w_i A_i$.*

Proof. At $p = 2$ the distance $D_2(A, B) = \|A - B\|_2$ is Euclidean, so $\mathcal{F}_w(X) = \|X - \sum_i w_i A_i\|_2^2 + \text{const}$. The arithmetic mean lies in $\mathbb{P}_n$ and is the unique minimizer by Theorem 5.2. □

Corollary B.2 ($p = 1$: the Bures–Wasserstein barycenter)**.** *For $p = 1$, a positive definite matrix X is a barycenter if and only if*

$$X = \sum_{i=1}^{m} w_i (X^{1/2} A_i X^{1/2})^{1/2}. \tag{B.1}$$

For real symmetric matrices, this is the covariance equation for a barycenter of the centred Gaussians $N(0, A_i)$ in 2-Wasserstein space (Agueh and Carlier, 2011; Álvarez-Esteban et al., 2016).

Proof. The characterization is Theorem 1.4 at $p = 1$. For real symmetric covariances, D_1 is the W_2 distance between the corresponding centred Gaussians (Olkin and Pukelsheim, 1982; Bhatia et al., 2019). □

For commuting inputs, the barycenter formula is explicit at every exponent.

Proposition B.3 (Commuting inputs)**.** *If $A_1, \ldots, A_m$ commute pairwise and $\sum_i w_i A_i^{p/2} \succ 0$, then*

$$X_\star = \Big(\sum_{i=1}^{m} w_i A_i^{p/2} \Big)^{2/p} \tag{B.2}$$

is a barycenter, the power mean of order $p/2$. On commuting matrices, the map $A \mapsto A^{p/2}$ converts D_p to the Hilbert–Schmidt distance:

$$D_p(A, B) = \|A^{p/2} - B^{p/2}\|_2. \tag{B.3}$$

Proof. Let X be the right side of (B.2); it commutes with every A_i and is positive definite. Then $(X^{1/2} A_i X^{1/2})^{p/2} = (X A_i)^{p/2} = X^{p/2} A_i^{p/2}$, whence

$$\sum_i w_i (X^{1/2} A_i X^{1/2})^{p/2} = X^{p/2} \sum_i w_i A_i^{p/2} = X^{p/2} X^{p/2} = X^p,$$

which is (1.6); apply Theorem 1.4. For (B.3), commutativity gives $\Phi_p(A, B) = \operatorname{tr}(AB)^{p/2} = \operatorname{tr}(A^{p/2} B^{p/2})$, so $D_p(A, B)^2 = \operatorname{tr}(A^{p/2} - B^{p/2})^2$. □

Remark B.4 (Equivariance)**.** *Since $\Phi_p(UAU^*, UBU^*) = \Phi_p(A, B)$ for unitary U, barycenters are unitarily equivariant: if $X_\star$ is a barycenter, then $U X_\star U^*$ is a barycenter of $(U A_i U^*, w_i)$. Since $D_p(sA, sB)^2 = s^p D_p(A, B)^2$ for $s \geq 0$, $sX_\star$ is likewise a barycenter of (sA_i, w_i). General invertible congruences need not preserve barycenters, except at $p = 2$, where the arithmetic mean is equivariant.*

Remark B.5 (Endpoint checks)**.** *For the matrices $A_i = X_i X_i^*$ in the proof of Theorem 1.3, at $p = 2$ we have $\|H_{ij}\|_2^2 = \operatorname{tr}(A_i A_j)$, so the matrix in (1.5) is the ordinary Hilbert–Schmidt Gram matrix of A_1, A_2, A_3. Its positivity for any number of blocks is the Marcus–Watkins case. At $p = 1$, three-point positivity for the Bures affinity recovers Drury's theorem.*

Appendix C. The variance inequality and three-point geometry

To bound $\min \mathcal{F}_w$, we optimize the radius and apply a support inequality on $\mathbb{S}_{p,+}$.

C.1. **Reduction to the sphere.**

Lemma C.1 (Radial reduction). *For $1 \le p \le 2$ and $w \in \Delta_m$, $\Psi_* := \max_{\widehat{X} \in \mathbb{S}_{p,+}} \Psi_w(\widehat{X}) \ge 0$, and*

$$\min_{X \in \mathbb{P}_n} \mathcal{F}_w(X) = \sum_{i=1}^m w_i \operatorname{tr} A_i^p - \Psi_*^2. \tag{C.1}$$

Moreover X is a barycenter if and only if $X = \Psi_*^{2/p} \widehat{X}$ for some maximizer $\widehat{X}$ of Ψ_w on $\mathbb{S}_{p,+}$.

Proof. Every nonzero $X \in \mathbb{P}_n$ is uniquely $X = s\widehat{X}$ with $\widehat{X} \in \mathbb{S}_{p,+}$ and $s = r_p(X)^{2/p} > 0$; the value $s = 0$ corresponds to $X = 0$. By (5.1) and homogeneity,

$$\mathcal{F}_w(s\widehat{X}) = s^p - 2s^{p/2}\Psi_w(\widehat{X}) + \sum_i w_i \operatorname{tr} A_i^p.$$

Substituting $u := s^{p/2} \in [0, \infty)$, the dependence on s is $u^2 - 2u\Psi_w(\widehat{X})$, which is minimized uniquely at $u = \Psi_w(\widehat{X})$ with value $-\Psi_w(\widehat{X})^2$. Since $\Psi \mapsto -\Psi^2$ is strictly decreasing on $[0, \infty)$, minimizing over $\widehat{X}$ selects exactly the maximizers of Ψ_w on $\mathbb{S}_{p,+}$, and gives (C.1).

The value Ψ_* is positive if $w_i > 0$ and $A_i \neq 0$ for some i: at $\widehat{X} = n^{-1/p} I$, the corresponding term is $\Phi_p(\widehat{X}, A_i) = n^{-1/2} \operatorname{tr} A_i^{p/2} > 0$. If no such index exists, $\Psi_w \equiv 0$. □

Corollary C.2 (Value and radius of the barycenter). *Let $X_\star \succ 0$ be a barycenter. Then $\operatorname{tr} X_\star^p = \Psi_w(X_\star) = \Psi_*^2$,*

$$\min_{X \in \mathbb{P}_n} \mathcal{F}_w(X) = \sum_{i=1}^m w_i \operatorname{tr} A_i^p - \operatorname{tr} X_\star^p, \tag{C.2}$$

$$r_p(X_\star) \le \sum_{i=1}^m w_i r_p(A_i). \tag{C.3}$$

Proof. Taking traces in (1.6) gives $\operatorname{tr} X_\star^p = \sum_i w_i \Phi_p(X_\star, A_i) = \Psi_w(X_\star)$; the value Ψ_*^2 and (C.2) then follow from Lemma C.1 and (5.1). For (C.3), combine $r_p(X_\star)^2 = \operatorname{tr} X_\star^p = \Psi_w(X_\star)$ with (1.7) to obtain $r_p(X_\star)^2 = \sum_i w_i \Phi_p(X_\star, A_i) \le r_p(X_\star) \sum_i w_i r_p(A_i)$, and divide by $r_p(X_\star) > 0$. □

C.2. **A multipoint support inequality and the variance inequality.** To control Ψ_* for all the data points at once, we extend Proposition 3.3 to a weighted sum of affinities.

Proposition C.3 (Multipoint support inequality). *Let $1 \le p \le 2$, let $A, A_1, \dots, A_m \in \mathbb{S}_{p,+}$, and let $a_1, \dots, a_m \ge 0$. Then*

$$\sum_{i=1}^m a_i \Phi_p(A, A_i) \le \Big(\sum_{i=1}^m a_i^2 + \sum_{i \neq j} a_i a_j \Phi_p(A_i, A_j) \Big)^{1/2}. \tag{C.4}$$

Proof. Assume first that every $a_i > 0$ and every $A_i \succ 0$. Let A_* maximize $F(X) = \sum_i a_i \Phi_p(X, A_i)$ over $\mathbb{S}_{p,+}$; by Lemma 3.2, $A_* \succ 0$. Put $Y_i := A_*^{1/2} A_i A_*^{1/2}$. By Lemma 3.1 the Lagrange condition is $\sum_{i=1}^m a_i Y_i^t = \lambda A_*^p$, and taking traces with $\operatorname{tr} A_*^p = 1$ identifies $\lambda = F(A_*)$. Set $P_i := a_i Y_i^t$ and $T := \sum_i P_i = \lambda A_*^p$. Since $t/p = 1/2$, we have $P_i^{1/p} = a_i^{1/p} Y_i^{1/2}$ and $T^{-1/p} = \lambda^{-1/p} A_*^{-1}$, so exactly as in (3.5)–(3.7),

$$\|P_j^{1/p} T^{-1/p} P_i^{1/p}\|_p^p = \begin{cases} \dfrac{a_i a_j}{\lambda} \Phi_p(A_i, A_j), & i \neq j, \\ \dfrac{a_i^2}{\lambda}, & i = j, \end{cases} \tag{C.5}$$

the diagonal case using $\operatorname{tr} A_i^p = 1$. Applying Theorem 1.2 to the decomposition $T = \sum_j P_j$ with $X = P_i^{1/p}$ gives

$$\sum_{j=1}^{m} \|P_j^{1/p} T^{-1/p} P_i^{1/p}\|_p^p \;\geq\; \operatorname{tr} P_i = a_i \Phi_p(A_*, A_i).$$

Summing over i and inserting (C.5),

$$\frac{1}{\lambda}\Big(\sum_i a_i^2 + \sum_{i\neq j} a_i a_j \Phi_p(A_i, A_j)\Big) \;\geq\; \sum_i a_i \Phi_p(A_*, A_i) = \lambda.$$

Since $F(A) \leq \lambda$, this proves (C.4). Approximate singular A_i and omit terms with $a_i = 0$. □

Theorem C.4 (Variance inequality)**.** *Let $1 \leq p \leq 2$, $A_1, \ldots, A_m \in \mathbb{P}_n$ and $w \in \Delta_m$. Then*

$$\min_{X \in \mathbb{P}_n} \sum_{i=1}^{m} w_i D_p(X, A_i)^2 \;\geq\; \sum_{1 \leq i < j \leq m} w_i w_j D_p(A_i, A_j)^2. \tag{C.6}$$

Equality holds when $p = 2$, and on any commuting family.

Proof. Set $r_i := r_p(A_i)$. For $r_i > 0$, let $\widehat{A}_i$ be the normalization (1.8); for $r_i = 0$, choose any $\widehat{A}_i \in \mathbb{S}_{p,+}$. Put $z_{ij} := \Phi_p(\widehat{A}_i, \widehat{A}_j)$. By homogeneity (1.8), for $\widehat{X} \in \mathbb{S}_{p,+}$, $\Psi_w(\widehat{X}) = \sum_i w_i r_i \Phi_p(\widehat{X}, \widehat{A}_i)$, so Proposition C.3 applied with $a_i = w_i r_i$ bounds the maximum Ψ_*:

$$\Psi_*^2 \;\leq\; \sum_i w_i^2 r_i^2 + \sum_{i\neq j} w_i w_j r_i r_j z_{ij}.$$

Insert this into (C.1) and use $\operatorname{tr} A_i^p = r_i^2$:

$$\min_X \mathcal{F}_w(X) \;\geq\; \sum_i w_i r_i^2 - \sum_i w_i^2 r_i^2 - \sum_{i\neq j} w_i w_j r_i r_j z_{ij} = \sum_i w_i(1 - w_i) r_i^2 - \sum_{i\neq j} w_i w_j r_i r_j z_{ij}.$$

Since $1 - w_i = \sum_{j\neq i} w_j$, the first sum equals $\sum_{i\neq j} w_i w_j r_i^2 = \sum_{i<j} w_i w_j (r_i^2 + r_j^2)$, while the second equals $2\sum_{i<j} w_i w_j r_i r_j z_{ij}$. Hence

$$\min_X \mathcal{F}_w(X) \;\geq\; \sum_{i<j} w_i w_j (r_i^2 + r_j^2 - 2 r_i r_j z_{ij}) = \sum_{i<j} w_i w_j D_p(A_i, A_j)^2$$

by (3.13). At $p = 2$ the distance is Euclidean and (C.6) is the classical identity $\sum_i w_i \|X - A_i\|_2^2 = \|X - \sum_j w_j A_j\|_2^2 + \sum_{i<j} w_i w_j \|A_i - A_j\|_2^2$ evaluated at the mean; on a commuting family the same identity applies after the isometry (B.3). □

Remark C.5 (Two points and geodesicity)**.** *For $m = 2$ inequality (C.6) reads*

$$\min_{X \in \mathbb{P}_n} \left(w_1 D_p(X, A_1)^2 + w_2 D_p(X, A_2)^2\right) \;\geq\; w_1 w_2 D_p(A_1, A_2)^2. \tag{C.7}$$

The reverse inequality holds in any *geodesic metric space: the point of a geodesic from A_1 to A_2 at parameter w_2 attains the value $w_1 w_2 D_p(A_1, A_2)^2$. Strict inequality therefore rules out a geodesic joining that pair; for $w_1 = w_2 = 1/2$, it also rules out a midpoint.*

Equality holds at both endpoints: D_2 is Euclidean, and the Bures–Wasserstein space at $p = 1$ is geodesic (Takatsu, 2011; Bhatia et al., 2019). For $1 < p < 2$, (C.7) with $w_1 = w_2 = 1/2$ is strict for suitable pairs of rank-one projections in dimension two (we omit the proof). Consequently, D_p is not a length metric on $\mathbb{P}_n$ for any $n \geq 2$, even on the positive definite cone; in particular, $(\mathbb{P}_n, D_p)$ is not geodesic. Commuting pairs do have midpoints, by (B.3).

C.3. **An a posteriori optimality certificate.** The concavity of Ψ_w gives a lower bound on the minimum of $\mathcal{F}_w$. Once $\nabla\Psi_w$ has been computed, the bound requires one Schatten norm.

Proposition C.6 (Optimality certificate)**.** *Let $1 \le p \le 2$, $w \in \Delta_m$, and $X_0 \succ 0$. Put*

$$G_0 := \tfrac{p}{2}\, X_0^{-1/2}\Big(\sum_i w_i (X_0^{1/2} A_i X_0^{1/2})^{p/2}\Big) X_0^{-1/2} \succeq 0. \tag{C.8}$$

Then, with Ψ_* as in Lemma C.1,

$$\Psi_* \;\le\; \Big(1 - \tfrac{p}{2}\Big)\Psi_w(X_0) + \|G_0\|_{p'}, \tag{C.9}$$

and consequently $\min_X \mathcal{F}_w(X)$ is at least $\sum_i w_i \operatorname{tr} A_i^p$ minus the square of the right side of (C.9). The bound (C.9) holds with equality when X_0 is a maximizer of Ψ_w on $\mathbb{S}_{p,+}$.

Proof. Let $\mathcal{K} := \{X \in \mathbb{P}_n : \operatorname{tr} X^p \le 1\}$, a convex set by Proposition 5.1(ii). By homogeneity of Ψ_w, $\max_{\mathcal{K}} \Psi_w = \max_{\mathbb{S}_{p,+}} \Psi_w = \Psi_*$. Since Ψ_w is concave (Proposition 5.1(i)) and differentiable at $X_0 \succ 0$ with $\nabla\Psi_w(X_0) = G_0$ (Lemma 3.1),

$$\Psi_* = \max_{X\in\mathcal{K}} \Psi_w(X) \le \Psi_w(X_0) - \langle G_0,\, X_0\rangle + \max_{X\in\mathcal{K}} \langle G_0,\, X\rangle.$$

The inner product term simplifies: writing S_0 for the bracket in (C.8),

$$\langle G_0,\, X_0\rangle = \tfrac{p}{2}\operatorname{tr}(X_0^{-1/2} S_0 X_0^{-1/2} X_0) = \tfrac{p}{2}\operatorname{tr} S_0 = \tfrac{p}{2}\Psi_w(X_0).$$

For the last term, Hölder gives $\langle G_0,\, X\rangle \le \|G_0\|_{p'}\|X\|_p \le \|G_0\|_{p'}$ on $\mathcal{K}$, including $p = 1$, when $p' = \infty$. Substituting these bounds into the concavity inequality proves (C.9); combining with (C.1) gives the stated lower bound on $\min \mathcal{F}_w$.

For the equality case, let $X_0 \in \mathbb{S}_{p,+}$ maximize Ψ_w. Since $X_0 \succ 0$, Lemma 3.1 gives the Lagrange condition $\sum_i w_i (X_0^{1/2} A_i X_0^{1/2})^{p/2} = \Psi_* X_0^p$, so $G_0 = \frac{p}{2}\Psi_* X_0^{p-1}$. The normalization $\operatorname{tr} X_0^p = 1$ gives $\|X_0^{p-1}\|_{p'} = 1$, also at $p = 1$, where $X_0^0 = I$ and $p' = \infty$. Hence $\|G_0\|_{p'} = \frac{p}{2}\Psi_*$. Since $\Psi_w(X_0) = \Psi_*$, the right side of (C.9) equals $(1 - \frac{p}{2})\Psi_* + \frac{p}{2}\Psi_* = \Psi_*$. □

C.4. **Three-point consequences.** The block positivity theorem also controls quadratic forms in the squared distances of a triple.

Corollary C.7 (Three-point negative type)**.** *Let $1 \le p \le 2$. For every $A_1, A_2, A_3 \in \mathbb{P}_n$, the 3×3 affinity matrix $K := [\Phi_p(A_i, A_j)]_{i,j=1}^3$ is positive semidefinite. Consequently, the squared distance matrix $\Delta := [D_p(A_i, A_j)^2]_{i,j=1}^3$ is conditionally negative semidefinite: $c^{\mathsf{T}}\Delta c \le 0$ for every $c = (c_1, c_2, c_3)^{\mathsf{T}} \in \mathbb{R}^3$ with $c_1 + c_2 + c_3 = 0$.*

Proof. Apply Theorem 1.3 to the positive semidefinite block Gram matrix with $H_{ij} := A_i^{1/2} A_j^{1/2}$ to obtain $K \succeq 0$. This is also the matrix case of Vuong and Osaka (2026, Theorem 1.1). For $c_1 + c_2 + c_3 = 0$, the radial terms in the definition of D_p^2 cancel, and hence

$$\begin{aligned} c^{\mathsf{T}}\Delta c &= \sum_{i,j=1}^3 c_i c_j \left(\operatorname{tr} A_i^p + \operatorname{tr} A_j^p - 2\Phi_p(A_i, A_j)\right) \\ &= -2c^{\mathsf{T}} K c \le 0. \end{aligned} \tag{C.10}$$

□

Corollary C.8 (A cosine-type support law)**.** *For $A, B, C \in \mathbb{S}_{p,+}$, write $x = \Phi_p(A, B)$, $y = \Phi_p(A, C)$, and $z = \Phi_p(B, C)$. Then*

$$z \ge xy - \sqrt{(1 - x^2)(1 - y^2)}. \tag{C.11}$$

Proof. Inequality (C.11) is equivalent to the angular triangle inequality after writing $x = \cos\alpha$, $y = \cos\beta$, and $z = \cos\gamma$. □

Appendix D. Proofs of the convergence estimates, and the endpoint $p = 1$

Proof of Lemma 5.6. By Lemma 5.4, $\mathcal{F}_w(X_k) \le F_0$ for every k. By (1.7), $D_p(X, A_i)^2 \ge (r_p(X) - r_p(A_i))^2$, so Jensen gives

$$\mathcal{F}_w(X) \;\ge\; \sum_i w_i \big(r_p(X) - r_p(A_i)\big)^2 \;\ge\; \big(r_p(X) - \bar{r}\big)^2.$$

Hence $r_p(X_k) \le \bar{r} + \sqrt{F_0}$, and since $\lambda_{\max}(X_k)^p \le \operatorname{tr} X_k^p = r_p(X_k)^2$, also $\lambda_{\max}(X_k) \le \lambda_+$.

For the lower bound, $A_i \succeq \lambda_{\min}(A_i) I$ gives $X^{1/2} A_i X^{1/2} \succeq \lambda_{\min}(A_i) X$, and $0 < t \le 1$, so the Löwner–Heinz inequality (Bhatia, 1997) yields $\big(X^{1/2} A_i X^{1/2}\big)^t \succeq \lambda_{\min}(A_i)^t X^t$ and therefore $S(X) \succeq c_A X^t$. Taking determinants gives $\det S(X) \ge c_A^n (\det X)^t$, whence

$$\det \mathcal{T}(X) = (\det X)^{-1} \big(\det S(X)\big)^{1/t} \;\ge\; c_A^{n/t}. \tag{D.1}$$

So $\det X_k \ge c_A^{n/t}$ for $k \ge 1$, and combining with the upper bound,

$$\lambda_{\min}(X_k) \;\ge\; \frac{\det X_k}{\lambda_{\max}(X_k)^{n-1}} \;\ge\; \frac{c_A^{n/t}}{\lambda_+^{n-1}} = \lambda_-.$$

A barycenter $X_\star$ satisfies $\mathcal{F}_w(X_\star) \le F_0$, hence $X_\star \preceq \lambda_+ I$ by the first paragraph; it is positive definite by Lemma 5.3 and fixed by $\mathcal{T}$ by Theorem 1.4, so (D.1) gives $\det X_\star \ge c_A^{n/t}$ and hence $X_\star \succeq \lambda_- I$. □

Proof of Lemma 5.7. (i) Strong convexity. By (5.1) and the symmetry of Φ_p,

$$\mathcal{F}_w(X) = \operatorname{tr} X^p - 2 \sum_{i=1}^m w_i \operatorname{tr}\big(A_i^{1/2} X A_i^{1/2}\big)^t + \text{const}, \tag{D.2}$$

and the second term is convex by Proposition 5.1(i). For $g(X) = \operatorname{tr} f(X)$ with f smooth, the Daleckii–Krein formula (Bhatia, 1997) gives $D^2 g(X)[H, H] = \sum_{j,k} (f')^{[1]}(\mu_j, \mu_k) |H_{jk}|^2$ in an eigenbasis of X, where $(f')^{[1]}$ is the first divided difference of f' and μ_j are the eigenvalues. With $f(x) = x^p$ each divided difference is a value of $f''(x) = p(p-1)x^{p-2}$ on $[\lambda_-, \lambda_+]$, and $p \le 2$ makes x^{p-2} nonincreasing, so on $\mathcal{B}$

$$D^2 \operatorname{tr} X^p[H, H] \;\ge\; p(p-1)\lambda_+^{p-2} \|H\|_{\mathrm{F}}^2. \tag{D.3}$$

Thus $\mathcal{F}_w$ is μ_F-strongly convex on the convex set $\mathcal{B}$, with $\mu_F := p(p-1)\lambda_+^{p-2}$. Minimizing the quadratic lower bound at X and evaluating at $X_\star \in \mathcal{B}$ gives

$$\|\nabla \mathcal{F}_w(X)\|_{\mathrm{F}}^2 \;\ge\; 2\mu_F \big(\mathcal{F}_w(X) - \mathcal{F}_w(X_\star)\big). \tag{D.4}$$

The same computation applied to $U \mapsto D_p(U, V)^2 = \operatorname{tr} U^p - 2\Phi_p(U, V) + \operatorname{tr} V^p$, which vanishes together with its gradient at $U = V$, gives for $U, V \in \mathcal{B}$

$$D_p(U, V)^2 \;\ge\; \frac{\mu_D}{2} \|U - V\|_{\mathrm{F}}^2, \qquad \mu_D := p(p-1)\lambda_+^{p-2}. \tag{D.5}$$

(ii) The step controls the gradient. Writing $X^{p-1} = X^{-1/2} X^p X^{-1/2}$ in (5.2),

$$\nabla \mathcal{F}_w(X) = p X^{-1/2} \big(X^p - S(X)\big) X^{-1/2}. \tag{D.6}$$

Put $Q := S(X)^{1/t} = X^{1/2} \mathcal{T}(X) X^{1/2}$. Since $p = 2t$, $X^p - S(X) = (X^2)^t - Q^t$, and $X, \mathcal{T}(X) \succeq \lambda_- I$ force $X^2 \succeq \lambda_-^2 I$ and $Q \succeq \lambda_- X \succeq \lambda_-^2 I$. On $[\lambda_-^2, \infty)$ the divided differences of $x \mapsto x^t$ are bounded by $t\lambda_-^{2t-2}$, so the Daleckii–Krein formula gives the Frobenius Lipschitz bound $\|U^t - V^t\|_{\mathrm{F}} \le t\lambda_-^{2t-2} \|U - V\|_{\mathrm{F}}$. Applying it to $U = X^2$, $V = Q$ and using $X^2 - Q = X^{1/2}(X - \mathcal{T}(X)) X^{1/2}$,

$$\|X^p - S(X)\|_{\mathrm{F}} \;\le\; t\lambda_+ \lambda_-^{p-2} \|X - \mathcal{T}(X)\|_{\mathrm{F}}.$$

Feeding this into (D.6) and using $\|X^{-1/2}ZX^{-1/2}\|_{\mathrm{F}} \le \lambda_-^{-1}\|Z\|_{\mathrm{F}}$,

$$\|\nabla \mathcal{F}_w(X)\|_{\mathrm{F}} \;\le\; C_R\|X - \mathcal{T}(X)\|_{\mathrm{F}}, \qquad C_R := pt\,\lambda_+\lambda_-^{p-3} = \frac{p^2}{2}\lambda_+\lambda_-^{p-3}. \tag{D.7}$$

(iii) Combination. Chaining (D.5), (D.7) and (D.4),

$$D_p(X,\mathcal{T}(X))^2 \;\ge\; \frac{\mu_D}{2}\|X-\mathcal{T}(X)\|_{\mathrm{F}}^2 \;\ge\; \frac{\mu_D}{2C_R^2}\|\nabla\mathcal{F}_w(X)\|_{\mathrm{F}}^2 \;\ge\; \frac{\mu_D\mu_F}{C_R^2}(\mathcal{F}_w(X)-\mathcal{F}_w(X_\star)),$$

and substituting the three constants,

$$\frac{\mu_D\mu_F}{C_R^2} = \frac{p^2(p-1)^2\lambda_+^{2p-4}}{\frac{p^4}{4}\lambda_+^2\lambda_-^{2p-6}} = \frac{4(p-1)^2}{p^2}\left(\frac{\lambda_-}{\lambda_+}\right)^{6-2p} = \rho_p.$$

Finally $4(p-1)^2/p^2 \le 1$ exactly because $p \le 2$, and $\lambda_- \le \lambda_+$ with $6-2p > 0$, so $\rho_p \le 1$. □

At $p = 1$ the leading term $\operatorname{tr} X^p$ is linear, so (D.3) degenerates. For positive definite data, the fidelity terms supply the curvature needed for the error bound.

Proposition D.1 (The Bures endpoint)**.** *Let $p = 1$, $A_1, \dots, A_m \succ 0$, and $w \in \Delta_m$ with all $w_i > 0$. Write $\alpha_i := \lambda_{\min}(A_i)$ and $\beta_i := \lambda_{\max}(A_i)$, and set*

$$\mu_F^{(1)} := \sum_{i=1}^m w_i \frac{\alpha_i^2}{2(\beta_i\lambda_+)^{3/2}}, \qquad \mu_D^{(1)} := \frac{\lambda_-^2}{2\lambda_+^3}, \qquad C_R^{(1)} := \frac{\lambda_+}{2\lambda_-^2}. \tag{D.8}$$

Then Lemma 5.7 and Theorem 5.8 hold verbatim at $p = 1$ with

$$\rho_1 := \min\left\{1,\; \frac{\mu_F^{(1)}\mu_D^{(1)}}{\left(C_R^{(1)}\right)^2}\right\} > 0 \tag{D.9}$$

in place of ρ_p, and with $\mu_F^{(1)}$ in place of $p(p-1)\lambda_+^{p-2}$ in (5.16).

Proof. Only part (i) of the proof of Lemma 5.7 changes. For $f(x) = \sqrt{x}$,

$$-\frac{f'(x)-f'(y)}{x-y} = \frac{1}{2\sqrt{x}\sqrt{y}(\sqrt{x}+\sqrt{y})} \;\ge\; \frac{1}{4(\beta_i\lambda_+)^{3/2}}, \qquad 0 < x, y \le \beta_i\lambda_+,$$

and the eigenvalues of $A_i^{1/2}XA_i^{1/2}$ do not exceed $\beta_i\lambda_+$ for $X \in \mathcal{B}$. Hence, by Daleckii–Krein as above and $\|A_i^{1/2}HA_i^{1/2}\|_{\mathrm{F}} \ge \alpha_i\|H\|_{\mathrm{F}}$,

$$D^2\Big[-2\operatorname{tr}\big(A_i^{1/2}XA_i^{1/2}\big)^{1/2}\Big][H,H] \;\ge\; \frac{\alpha_i^2}{2(\beta_i\lambda_+)^{3/2}}\|H\|_{\mathrm{F}}^2,$$

so $\mathcal{F}_w$ is $\mu_F^{(1)}$-strongly convex on $\mathcal{B}$. Applying the same bound with the fixed second argument $V \in \mathcal{B}$ in place of A_i, so that λ_- and λ_+ replace α_i and β_i, gives (D.5) with $\mu_D^{(1)}$. Part (ii) is unchanged and yields $C_R^{(1)} = pt\lambda_+\lambda_-^{p-3}$ at $p = 1$, $t = \frac{1}{2}$. Part (iii) then gives (5.14) with constant $\mu_F^{(1)}\mu_D^{(1)}/(C_R^{(1)})^2$; replacing it by ρ_1 only weakens the bound, and guarantees $1 - \rho_1 \ge 0$. The proof of Theorem 5.8 then applies verbatim. □

Remark D.2 (Uniqueness at $p = 1$)**.** *Strong convexity of $\mathcal{F}_w$ on $\mathcal{B}$ and the fact that every barycenter lies in $\mathcal{B}$ (Lemma 5.6) together give uniqueness of the barycenter at $p = 1$ whenever all $A_i \succ 0$. This consequence of the rate analysis uses a stronger hypothesis than the criterion recalled after Theorem 5.2, which needs only one nonsingular A_i with $w_i > 0$.*

Remark D.3 (Endpoints)**.** *At $p = 2$ the map $\mathcal{T}$ is constant and $X_1 = X_\star$, so (5.15) holds trivially, and (5.6) becomes the exact bias–variance decomposition $\mathcal{F}_w(X) - \mathcal{F}_w(\bar{A}) = \|X - \bar{A}\|_2^2$ for $\bar{A} = \sum_i w_iA_i$. At $p = 1$, Lemma 5.4 recovers the sufficient decrease estimate for the Gaussian Wasserstein barycenter iteration in Álvarez-Esteban et al. (2016, Proposition 3.3).*

Appendix E. A numerical comparison with first-order methods

Convexity of $\mathcal{F}_w$ on the cone (Proposition 5.1) allows a comparison of $\mathcal{T}$ with gradient descent (GD) and an accelerated FISTA variant with adaptive restart (AGD), both using backtracking. Work is measured in *passes*: one sweep over the m data matrices requiring an eigendecomposition of each. A step of $\mathcal{T}$, one gradient, and one evaluation of $\mathcal{F}_w$ each cost one pass. We count all objective evaluations required by GD and AGD, including those at extrapolated points; evaluations used only to monitor $\mathcal{T}$ are not charged.

Table 1 reports the medians and Figure 1 a single trajectory. For each instance, the relative objective gap is $(\mathcal{F}_w(X) - F_{\mathrm{ref}})/\max\{1, |F_{\mathrm{ref}}|\}$, where F_{ref} is the lowest final objective value obtained by the three methods. At the tested exponents below 2, the median pass counts are 4–12 for $\mathcal{T}$, 41–193 for GD, and 45–178 for AGD. Acceleration does not consistently improve on GD. These comparisons concern the tested implementations, without alternative parametrizations for GD or AGD. The convergence guarantee in Theorem 5.8 does not explain the observed advantage of $\mathcal{T}$.

Table 1: Median passes to a relative objective gap of 10^{-13}, among the runs reaching that target out of 15 random instances with $m = 4$. All three methods use the same random positive definite start on each instance. At $p = 2$, the map $\mathcal{T}$ is exact in a single step.

	$n = 2$			$n = 5$			$n = 10$		
p	$\mathcal{T}$	GD	AGD	$\mathcal{T}$	GD	AGD	$\mathcal{T}$	GD	AGD
1.00	6	75	81	10	115	151	12	193	178
1.25	6	92	60	8	79	98	9	106	129
1.50	6	57	55	7	65	65	8	59	80
1.75	4	48	46	5	42	45	6	41	48
2.00	1	4	4	1	3	36	1	3	4

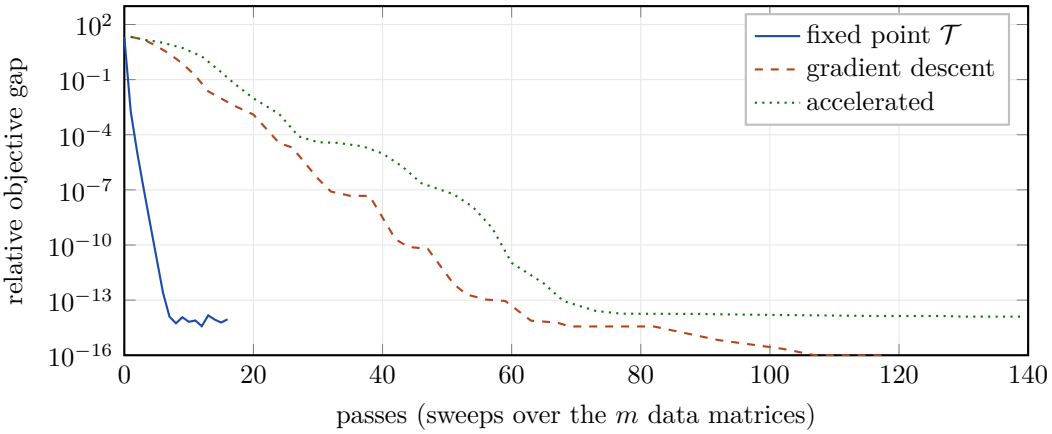


Figure 1. One instance, $n = 5$, $m = 4$, $p = 1.5$, from a random positive definite start. The fixed point reaches a gap of 10^{-13} in 7 passes; gradient descent needs 59 and the accelerated variant 68. Each trace ends at that method's stopping test. Values below 10^{-16} are floored; the tails lie at double-precision accuracy.