GEOMETRIC DISTANCE BETWEEN POSITIVE DEFINITE MATRICES OF DIFFERENT DIMENSIONS

LEK-HENG LIM, RODOLPHE SEPULCHRE, AND KE YE

ABSTRACT. We show how the Riemannian distance on $\mathbb{S}^n_{++}$, the cone of $n \times n$ real symmetric or complex Hermitian positive definite matrices, may be used to naturally define a distance between two such matrices of different dimensions. Given that $\mathbb{S}^n_{++}$ also parameterizes n-dimensional ellipsoids, and inner products on $\mathbb{R}^n$, $n \times n$ covariance matrices of nondegenerate probability distributions, this gives us a natural way to define a geometric distance between a pair of such objects of different dimensions.

1. Introduction

It is well-known that the cone of real symmetric positive definite or complex Hermitian positive definite matrices $\mathbb{S}^n_{++}$ has a natural Riemannian metric that gives it a *Riemannian distance*

$$\delta_2 : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \to \mathbb{R}_+, \quad \delta_2(A, B) = \Bigl[\sum\nolimits_{j=1}^n \log^2(\lambda_j(A^{-1}B))\Bigr]^{1/2}. \tag{1.1}$$

The Riemannian metric and distance endow $\mathbb{S}^n_{++}$ with rich geometric properties: in addition to being a Riemannian manifold, it is a symmetric space, a Bruhat–Tits space, a CAT(0) space, and a metric space of nonpositive curvature [2, Chapter 6].

The Riemannian distance δ_2 is arguably the most natural and useful distance on the positive definite cone $\mathbb{S}^n_{++}$ [3]. It may be thought as a generalization to $\mathbb{S}^n_{++}$ the geometric distance between two positive numbers $|\log(a/b)|$ [3]. It is invariant under any *congruence* transformation of the data: $\delta_2(XAX^{\mathsf{T}}, XBX^{\mathsf{T}}) = \delta_2(A, B)$ for any invertible matrix X. Because a positive definite matrix is congruent to identity, the distance is entirely characterized by the simple formula $\delta(A, I) = \|\log A\|_F$. It is also invariant under *inversion*, $\delta_2(A^{-1}, B^{-1}) = \delta_2(A, B)$, which again generalizes an important property of the geometric distance between positive scalars, as well as any *similarity* transformation: $\delta_2(XAX^{-1}, XBX^{-1}) = \delta_2(A, B)$ for any invertible matrix X. For comparison, all matrix norms are at best invariant under orthogonal or unitary transformations (e.g., Frobenius, spectral, nuclear, Schatten, Ky Fan norms) or otherwise only permutations and scaling (e.g., operator p-norms, Hölder p-norms, where $p \neq 2$).

From a practical perspective, δ_2 underlies important applications in computer vision [12], medical imaging [5, 9], radar signal processing [1], statistical inference [11], among other areas. In optimization, δ_2 has been shown [10] to be equivalent to the metric defined by the self-concordant log barrier in semidefinite programming, i.e., $\log\det : \mathbb{S}^n_{++} \to \mathbb{R}$. In statistics, it has been shown [13] to be equivalent to the Fisher information metric for Gaussian covariance matrix estimation problems. In numerical linear algebra, δ_2 gives rise to the matrix geometric mean [8], a topic that has been thoroughly studied and has many applications of its own.

We will show how δ_2 naturally gives a notion of geometric distance δ_2^+ between positive definite matrices of *different* dimensions, that is, we will define $\delta_2^+(A, B)$ for $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$ where $m \neq n$. Because of the ubiquity of positive definite matrices, this distance immediately extends to other objects. For example, real symmetric positive definite matrices $A \in \mathbb{S}^n_{++}$ are in one-to-one correspondence with:

2010 *Mathematics Subject Classification.* 15B48, 15A18, 51K99, 53C25.
Key words and phrases. Riemannian distance, positive definite matrices, covariance matrices, ellipsoids.

(i) ellipsoids centered at the origin in $\mathbb{R}^n$,

$$\mathcal{E}_A := \{x \in \mathbb{R} : x^{\mathsf{T}} A x \le 1\};$$

(ii) inner products on $\mathbb{R}^n$,

$$\langle \cdot, \cdot \rangle_A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x, y) \mapsto x^{\mathsf{T}} A y;$$

(iii) covariances of nondegenerate random variables $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$,

$$A = \operatorname{Cov}(X) = E[(X - \mu)(X - \mu)^{\mathsf{T}}];$$

as well as other objects such as diffusion tensors, mean-centered Gaussians, sums-of-squares polynomials, etc. In other words, our new notion of distance gives a way to measure separation between ellipsoids, inner products, covariances, etc, of different dimensions. Note that we may replace $\mathbb{R}$ by $\mathbb{C}$ and x^{T} by x^*, so these results also carry over to $\mathbb{C}$.

In fact, it is easiest to describe our approach in terms of ellipsoids, by virtue of (i). The result that forms the impetus behind our distance δ_2^+ is the following:

> *Given an m-dimensional ellipsoid $\mathcal{E}_A$ and an n-dimensional ellipsoid $\mathcal{E}_B$, say $m \le n$. The distance from $\mathcal{E}_A$ to the set of m-dimensional ellipsoids contained in $\mathcal{E}_B$ equals the distance from $\mathcal{E}_B$ to the set of n-dimensional ellipsoids containing $\mathcal{E}_A$, where both distances are measured via* (1.1). *Their common value gives a distance between $\mathcal{E}_A$ and $\mathcal{E}_B$ and therefore A and B.*

In addition, we show that this distance has an explicit, readily computable expression.

Notations and terminologies. All results in this article will apply to $\mathbb{R}$ and $\mathbb{C}$ alike. To avoid verbosity, we adopt the convention that the term 'Hermitian' will cover both 'complex Hermitian' and 'real symmetric.' $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$. For $X \in \mathbb{F}^{m \times n}$, X^* will mean the transpose of X if $\mathbb{F} = \mathbb{R}$ and the conjugate transpose of X if $\mathbb{F} = \mathbb{C}$.

We will adopt notations in [4]. Let n be a positive integer. $\mathbb{S}^n$ will denote the vector space of $n \times n$ Hermitian matrices, $\mathbb{S}^n_+$ the closed cone of of $n \times n$ Hermitian positive semidefinite matrices, and $\mathbb{S}^n_{++}$ the open cone of $n \times n$ Hermitian positive definite matrices. $\preceq$ will denote the partial order on $\mathbb{S}^n_+$ (and thus also on its subset $\mathbb{S}^n_{++}$) defined by

$$A \preceq B \qquad \text{if and only if} \qquad B - A \in \mathbb{S}^n_+.$$

For brevity, positive (semi)definite will henceforth mean[1] Hermitian positive (semi)definite.

2. Positive definite matrices

For the reader's easy reference, we will review some basic properties of positive definite matrices that we will need later: simultaneous diagonalizability, Cauchy interlacing, and majorization.

A pair of Hermitian matrices, one positive definite and the other nonsingular, may be simultaneously diagonalized. We state a version of this well-known result below [7, Theorem 12.19].

Theorem 2.1 (Simultaneous diagonalization)**.** *Let $A \in \mathbb{S}^n_{++}$ and $B \in \mathbb{S}^n$. Then there exists a nonsingular $X \in \mathbb{F}^{n \times n}$ such that*

$$XAX^* = I_n, \qquad XBX^* = D,$$

where I_n is the $n \times n$ identity matrix and D is the diagonal matrix whose diagonal entries are eigenvalues of $A^{-1}B$.

[1] Recall that while a complex positive (semi)definite matrix is necessarily Hermitian, a real positive (semi)definite matrix does not need to be symmetric.

As usual, we will order the eigenvalues of $X \in \mathbb{S}^n_{++}$ nonincreasingly:

$$\lambda_1(X) \le \lambda_2(X) \le \cdots \le \lambda_n(X).$$

The next two standard results may be found as [6, Theorem 4.3.28, Corollary 7.7.4].

Theorem 2.2 (Cauchy interlacing inequalities)**.** *Let $m \le n$ and $A \in \mathbb{S}^n$. If we partition A into*

$$A = \begin{bmatrix} A_1 & A_2 \\ A_2^* & A_3 \end{bmatrix}, \quad A_1 \in \mathbb{S}^m, \quad A_2 \in \mathbb{F}^{m\times(n-m)}, \quad A_3 \in \mathbb{S}^{n-m},$$

then

$$\lambda_j(A) \le \lambda_j(A_1) \le \lambda_{j+n-m}(A), \quad j = 1, \dots, m.$$

Proposition 2.3 (Majorization)**.** *If $A, B \in \mathbb{S}^n_{++}$ and $A \preceq B$, then $\lambda_j(A) \le \lambda_j(B)$, $j = 1, \dots, n$.*

3. Containment of ellipsoids of different dimensions

It helps to picture our construction with a concrete geometric object in mind and for this purpose we will exploit the one-to-one correspondence between positive definite matrices and ellipsoids mentioned in Section 1. For $A \in \mathbb{S}^n_{++}$, the n-dimensional *ellipsoid* $\mathcal{E}_A$ centered at the origin is

$$\mathcal{E}_A := \{x \in \mathbb{F}^n : x^* A x \le 1\}.$$

All ellipsoids in this article will be centered at the origin and henceforth we will drop the 'centered at the origin' for brevity. There is a simple equivalence between containment of ellipsoids and the partial order on positive definite matrices.

Lemma 3.1. *Let $A, B \in \mathbb{S}^n_{++}$. Then $\mathcal{E}_A \subseteq \mathcal{E}_B$ if and only if $B \preceq A$.*

Proof. If $\mathcal{E}_A \subseteq \mathcal{E}_B$, then for each $x \in \mathbb{F}^n$ satisfying

$$x^* A x \le 1 \tag{3.1}$$

we also have $x^* B x \le 1$. Thus we have $y^* B y \le y^* A y$ for any $y \in \mathbb{F}^n$ since $x = y/\sqrt{y^* A y}$ satisfies (3.1). Conversely, if $B \preceq A$, then whenever x satisfies (3.1), we have $x^* B x \le x^* A x \le 1$. □

Lemma 3.1 gives the one-to-one correspondence we have alluded to: $\mathcal{E}_A = \mathcal{E}_B$ if and only if $A = B \in \mathbb{S}^n_{++}$.

We extend this to the containment of ellipsoids of different dimensions. Let $m \le n$ be positive integers and $A \in \mathbb{S}^m_{++}$, $B \in \mathbb{S}^n_{++}$. Consider the embedding

$$\iota_{m,n} : \mathbb{F}^m \to \mathbb{F}^n, \quad (x_1, \dots, x_m) \mapsto (x_1, \dots, x_m, 0, \dots, 0).$$

Then we have

$$\iota_{m,n}(\mathcal{E}_A) = \{(x, 0) \in \mathbb{F}^n : x^* A x \le 1\},$$

where $x \in \mathbb{F}^m$ and $0 \in \mathbb{F}^{n-m}$ is the zero vector. Let B_{11} be the upper left $m \times m$ principal submatrix of $B \in \mathbb{S}^n_{++}$, i.e., $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix}$ for matrices B_{11}, B_{12}, B_{22} of appropriate dimensions. Then the same argument used in the proof of Lemma 3.1 gives the following.

Lemma 3.2. *Let $m \le n$ and $A \in \mathbb{S}^m_{++}$, $B \in \mathbb{S}^n_{++}$. Then $\iota_{m,n}(\mathcal{E}_A) \subseteq \mathcal{E}_B$ if and only if $B_{11} \preceq A$.*

4. Geometric distance between ellipsoids of different dimensions

Our method of defining a geometric distance δ_2^+ for pairs of positive definite matrices of different dimensions is inspired by a similar (at least in spirit) extension of the distance on a Grassmannian to subspaces of different dimensions proposed in [14]. The following convex sets will play the role of the Schubert varieties in [14].

Definition 4.1. Let $m \le n$. For any $A \in \mathbb{S}^m_{++}$, we define the *convex set of n-dimensional ellipsoids containing $\mathcal{E}_A$* to be

$$\Omega_+(A) := \left\{ G = \begin{bmatrix} G_{11} & G_{12} \\ G_{12}^* & G_{22} \end{bmatrix} \in \mathbb{S}^n_{++} : G_{11} \preceq A \right\}. \tag{4.1}$$

For any $B \in \mathbb{S}^n_{++}$, we define the *convex set of m-dimensional ellipsoids contained in $\mathcal{E}_B$* to be

$$\Omega_-(B) := \{ H \in \mathbb{S}^m_{++} : B_{11} \preceq H \}, \tag{4.2}$$

where B_{11} is the upper left $m \times m$ principal submatrix of B.

Lemma 3.2 provides justification for the names: more precisely, $\Omega_+(A)$ parametrizes all n-dimensional ellipsoids containing $\iota_{m,n}(\mathcal{E}_A)$ whereas $\Omega_-(B)$ parametrizes all m-dimensional ellipsoids contained in $\mathcal{E}_{B_{11}}$.

Given $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$, a natural way to define the distance between A and B is to define it as the distance from A to the set $\Omega_-(B)$, i.e.,

$$\delta_2\bigl(A, \Omega_-(B)\bigr) := \inf_{H \in \Omega_-(B)} \delta_2(A, H) = \inf_{H \in \Omega_-(B)} \Bigl[\sum\nolimits_{j=1}^m \log^2 \lambda_j(AH^{-1})\Bigr]^{1/2}; \tag{4.3}$$

but another equally natural way is to define it as the distance from $B \in \mathbb{S}^n_{++}$ to the set $\Omega_+(A)$, i.e.,

$$\delta_2\bigl(B, \Omega_+(A)\bigr) := \inf_{G \in \Omega_+(A)} \delta_2(G, B) = \inf_{G \in \Omega_+(A)} \Bigl[\sum\nolimits_{j=1}^n \log^2 \lambda_j(GB^{-1})\Bigr]^{1/2}. \tag{4.4}$$

We will show that

$$\delta_2\bigl(A, \Omega_-(B)\bigr) = \delta_2\bigl(B, \Omega_+(A)\bigr)$$

and their common value gives the distance we seek between A and B.

Note that $\Omega_+(A) \subseteq \mathbb{S}^n_{++}$ and $\Omega_-(B) \subseteq \mathbb{S}^m_{++}$, (4.3) is the distance of a point A to a set $\Omega_-(B)$ within the Riemannian manifold $\mathbb{S}^m_{++}$, (4.4) is the distance of a point B to a set $\Omega_+(A)$ within the Riemannian manifold $\mathbb{S}^n_{++}$. There is no reason to expect that they are equal but in fact they are — this is our main result.

Theorem 4.2. *Let $m \le n$ be positive integers and let $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$. Let B_{11} be the upper left $m \times m$ principal submatrix of B. Then*

$$\delta_2\bigl(A, \Omega_-(B)\bigr) = \delta_2\bigl(B, \Omega_+(A)\bigr) \tag{4.5}$$

and their common value is given by

$$\delta_2^+(A, B) := \Bigl[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j(A^{-1}B_{11})\}^2\Bigr]^{1/2}, \tag{4.6}$$

or, alternatively,

$$\delta_2^+(A, B) = \Bigl[\sum\nolimits_{j=1}^k \log^2 \lambda_j(A^{-1}B_{11})\Bigr]^{1/2},$$

where k is such that $\lambda_j(A^{-1}B_{11}) \le 1$ for $j = k+1, \dots, m$.

We will defer the proof of Theorem 4.2 to Section 5 but first make a few immediate observations regarding this new distance.

An implicit assumption in Theorem 4.2 is that whenever we write $\delta^+(A, B)$, we will require that the dimension of the matrix in the first argument be not more than the dimension of the matrix in the second argument. In particular, $\delta^+(A, B) \ne \delta^+(B, A)$; in fact the latter is not meaningful except in the case when $m = n$. An immediate conclusion is that δ_2^+ does not define a *metric* on $\bigcup_{n=1}^\infty \mathbb{S}_{++}^n$, which is not surprising as δ_2^+ is a distance in the sense of a distance from a point to a set.

For the special case $m = n$, (4.6) becomes

$$\delta_2^+(A, B) = \Bigl[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j(A^{-1}B)\}^2\Bigr]^{1/2}.$$

However, since $m = n$, we may swap the matrices A and B in (4.5) to get

$$\delta_2\bigl(B, \Omega_-(A)\bigr) = \delta_2\bigl(A, \Omega_+(B)\bigr)$$

and their common value is given by

$$\delta_2^+(B, A) = \Bigl[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j(B^{-1}A)\}^2\Bigr]^{1/2}.$$

Note that even in this case, $\delta^+(A, B) \neq \delta^+(B, A)$ in general. Nevertheless, this gives us the relation between our original Riemannian distance δ_2 and the distance δ_2^+ defined in Theorem 4.2.

Proposition 4.3. *Let $m = n$. Then the distances δ_2 in* (1.1) *and δ_2^+ in* (4.6) *are related via*

$$\delta_2(A, B) = \delta_2^+(A, B) + \delta_2^+(B, A).$$

The domain of δ_2^+ may be further extended to positive semidefinite matrices in the following sense: Suppose $A \in \mathbb{S}_+^m$ and $B \in \mathbb{S}_+^n$ with $m \le n$. We may replace $\mathbb{S}_{++}^m$ by $\mathbb{S}_+^m$ in the (4.1) and $\mathbb{S}_{++}^n$ by $\mathbb{S}_+^n$ in (4.2). If A is singular, i.e., it is positive semidefinite but not positive definite, then we have

$$\delta_2\bigl(A, \Omega_-(B)\bigr) = \infty = \delta_2\bigl(B, \Omega_+(A)\bigr). \tag{4.7}$$

as $\delta_2(A, H) = \infty$ for any $H \in \Omega_-(B)$ and $\delta_2(B, G) = \infty$ for any $G \in \Omega_+(A)$. However, if B is singular, then (4.7) is not true unless A is also singular. In general we only have

$$\delta_2\bigl(A, \Omega_-(B)\bigr) \le \delta_2\bigl(B, \Omega_+(A)\bigr) = \infty,$$

where the inequality can be strict when A is positive definite. In short, (4.5) extends to positive semidefinite A and B except in the case where A is nonsingular and B is singular.

5. Proof of Theorem 4.2

Throughout this section, we will assume that $m \le n$, $A \in \mathbb{S}_{++}^m$, and $B \in \mathbb{S}_{++}^n$. We will prove Theorem 4.2 by showing that

$$\delta_2\bigl(A, \Omega_-(B)\bigr) = \Bigl[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j(A^{-1}B_{11})\}^2\Bigr]^{1/2} \tag{5.1}$$

in Lemma 5.3 and

$$\delta_2\bigl(B, \Omega_+(A)\bigr) = \Bigl[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j(A^{-1}B_{11})\}^2\Bigr]^{1/2} \tag{5.2}$$

in Lemma 5.4. The key to establishing these is to repeatedly use the following invariance of δ_2 under congruence action by nonsingular matrices.

Lemma 5.1 (Invariance of δ_2)**.** *Let $A, B \in \mathbb{S}_{++}^n$ and $X \in \mathbb{F}^{n\times n}$ be nonsingular. Then*

$$\delta_2(XAX^*, XBX^*) = \delta_2(A, B).$$

Proof. Observe that

$$(XAX^*)(XBX^*)^{-1} = X(AB^{-1})X^{-1}.$$

Thus $\lambda_j(AB^{-1}) = \lambda_j((XAX^*)(XBX^*)^{-1})$ and the invariance of δ_2 follows. □

5.1. **Calculating $\delta_2\big(A, \Omega_-(B)\big)$.** Recall that we partition $B \in \mathbb{S}^n_{++}$ into $B = \left[\begin{smallmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{smallmatrix}\right]$. Note that $B_{11} \in \mathbb{S}^m_{++}$, $B_{12} \in \mathbb{F}^{m\times(n-m)}$, and $B_{22} \in \mathbb{S}^{n-m}_{++}$. By Theorem 2.1, there is a nonsingular $X \in \mathbb{F}^{m\times m}$ such that

$$XAX^* = I_m, \qquad XB_{11}X^* = D,$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ with $\lambda_j := \lambda_j(A^{-1}B_{11})$, $j = 1, \dots, m$. Since B is positive definite, so is B_{22}, and thus there is a nonsingular $Y \in \mathbb{F}^{(n-m)\times(n-m)}$ such that

$$YB_{22}Y^* = I_{n-m}.$$

Therefore, we have

$$\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix} \begin{bmatrix} X^* & 0 \\ 0 & Y^* \end{bmatrix} = \begin{bmatrix} D & XB_{12}Y^* \\ YB_{12}^*X^* & I_{n-m} \end{bmatrix}.$$

Set $Z := \left[\begin{smallmatrix} X & 0 \\ 0 & Y \end{smallmatrix}\right]$. Then, by Lemma 5.1,

$$\delta_2\big(A, \Omega_-(B)\big) = \delta_2\big(XAX^*, X\Omega_-(B)X^*\big) = \delta_2\big(I_m, \Omega_-(ZBZ^*)\big).$$

Hence we may assume without loss of generality that

$$A = I_m, \qquad B = \begin{bmatrix} D & B_{12} \\ B_{12}^* & I_{n-m} \end{bmatrix}, \tag{5.3}$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ and $B_{12} \in \mathbb{F}^{m\times(n-m)}$ is such that B is positive definite.

We will need a small observation regarding the eigenvalues of B.

Lemma 5.2. *Let $D = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$. Let $\mu_{m+1}, \dots, \mu_n$ be the eigenvalues of $B_{12}^*D^{-1}B_{12}$. Then $0 < \mu_{m+j} < 1$ for all $j = 1, \dots, n-m$ and the eigenvalues of $B = \left[\begin{smallmatrix} D & B_{12} \\ B_{12}^* & I_{n-m} \end{smallmatrix}\right]$ are $\lambda_1, \dots, \lambda_m, 1 - \mu_{m+1}, \dots, 1 - \mu_n$.*

Proof. Since $I_{n-m} - B_{12}^*D^{-1}B_{12}$ is the Schur complement of D in the positive definite matrix B, it follows that $0 < \mu_{m+j} < 1$ for all $j = 1, \dots, n-m$. The eigenvalues of B are obvious from

$$\begin{bmatrix} I_m & 0 \\ -B_{12}^*D^{-1} & I_{m-n} \end{bmatrix} \begin{bmatrix} D & B_{12} \\ B_{12}^* & I_{n-m} \end{bmatrix} \begin{bmatrix} I_m & 0 \\ -B_{12}^*D^{-1} & I_{m-n} \end{bmatrix}^{-1} = \begin{bmatrix} D & 0 \\ 0 & I_{n-m} - B_{12}^*D^{-1}B_{12} \end{bmatrix}. \qquad \square$$

We are now ready to prove (5.4).

Lemma 5.3. *Let $m \le n$ be positive integers and let $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$. Then there exists an $H_0 \in \mathbb{S}^m_{++}$ such that*

$$\delta_2\big(A, \Omega_-(B)\big) = \delta_2(A, H_0) = \Big[\sum\nolimits_{j=1}^m \min\{0, \log \lambda_j\}^2\Big]^{1/2}.$$

Proof. By the preceding discussions, we may assume that A and B are as in (5.3). So we must have

$$\delta_2\big(A, \Omega_-(B)\big) = \inf_{D \preceq H} \Big[\sum\nolimits_{j=1}^m \log^2 \lambda_j(H)\Big]^{1/2}.$$

The condition $D \preceq H$ implies that $\lambda_j \le \lambda_j(H)$, $j = 1, \dots, m$, by Proposition 2.3. Hence

$$\inf_{D \preceq H} \log^2 \lambda_j(H) = \begin{cases} \log^2 \lambda_j & \text{if } \lambda_j > 1, \\ 0 & \text{if } \lambda_j \le 1. \end{cases} \tag{5.4}$$

Let $H_0 = \operatorname{diag}(h_1, \dots, h_m)$ where

$$h_j = \begin{cases} \lambda_j & \text{if } \lambda_j > 1, \\ 1 & \text{if } \lambda_j \le 1. \end{cases}$$

Then it is clear that $D \preceq H_0$ and H_0 is our desired matrix by (5.4). $\square$

5.2. **Calculating** $\delta_2\big(B,\Omega_+(A)\big)$**.** Let $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$. Again, we partition B as in Section 5.1. Let L be the upper triangular matrix

$$L = \begin{bmatrix} I_m & 0 \\ -B_{12}^* B_{11}^{-1} & I_{n-m} \end{bmatrix}.$$

Then

$$LBL^* = \begin{bmatrix} B_{11} & 0 \\ 0 & I_{n-m} - B_{12}^* B_{11}^{-1} B_{12} \end{bmatrix} \qquad \text{and} \qquad L\Omega_+(A)L^* = \Omega_+(A).$$

For the second equality, observe that $L\Omega_+(A)L^* \subseteq \Omega_+(A)$ and check that $L^{-1}\Omega_+(A)(L^{-1})^* \subseteq \Omega_+(A)$, which implies that $\Omega_+(A) \subseteq L\Omega_+(A)L^*$. Therefore, by Lemma 5.1, we have

$$\delta_2\big(B,\Omega_+(A)\big) = \delta_2\big(LBL^*, L\Omega_+(A)L^*\big) = \delta_2\big(LBL^*,\Omega_+(A)\big). \tag{5.5}$$

Let $X_1 \in \mathbb{F}^{m\times m}$ and $Y_1 \in \mathbb{F}^{(n-m)\times(n-m)}$ be nonsingular matrices[2] such that

$$X_1 A X_1^* = D^{-1}, \quad X_1 B_{11} X_1^* = I_{n-m}, \quad Y_1(I_{n-m} - B_{12}^* B_{11}^{-1} B_{12})Y_1^* = I_{n-m},$$

where $D = \operatorname{diag}(\lambda_1,\dots,\lambda_m)$ with $\lambda_j \coloneqq \lambda_j(A^{-1}B_{11})$, $j = 1,\dots,m$. Let $Z_1 = \left[\begin{smallmatrix} X_1 & 0 \\ 0 & Y_1 \end{smallmatrix}\right]$. Then

$$Z_1 LBL^* Z_1^* = I_n \qquad \text{and} \qquad Z_1\Omega_+(A)Z_1^* = \Omega_+(D^{-1}).$$

Hence, by (5.5) and Lemma 5.1,

$$\delta_2\big(B,\Omega_+(A)\big) = \delta_2\big(LBL^*,\Omega_+(A)\big) = \delta_2\big(Z_1LBL^*Z_1^*, Z_1\Omega_+(A)Z_1^*\big) = \delta_2\big(I_n,\Omega_+(D^{-1})\big),$$

So to calculate $\delta_2\big(B,\Omega_+(A)\big)$, it suffices to assume that

$$A = D^{-1} = \operatorname{diag}(\lambda_1^{-1},\dots,\lambda_m^{-1}), \qquad B = I_n. \tag{5.6}$$

We are now ready to prove (5.7).

Lemma 5.4. *Let $m \le n$ be positive integers and let $A \in \mathbb{S}^m_{++}$ and $B \in \mathbb{S}^n_{++}$. Then there exists some $G_0 \in \mathbb{S}^n_{++}$ such that*

$$\delta_2\big(B,\Omega_+(A)\big) = \delta_2(G_0,B) = \Big[\sum\nolimits_{j=1}^m \min\{0,\log\lambda_j(A^{-1}B_{11})\}^2\Big]^{1/2}.$$

Proof. By the preceding discussions, we may assume that A and B are as in (5.6). So we must have

$$\delta_2\big(I_n,\Omega_+(D^{-1})\big) = \inf_{G_{11}\preceq D^{-1}} \Big[\sum\nolimits_{j=1}^n \log^2\lambda_j(G)\Big]^{1/2},$$

where G_{11} is the upper left $m\times m$ principal submatrix of $G \in \Omega_+(D^{-1})$. By Proposition 2.3, we have $\lambda_j(G_{11}) \le \lambda_j^{-1}$, $j = 1,\dots,m$. Moreover, by Theorem 2.2,

$$\lambda_j(G) \le \lambda_j(G_{11}) \le \lambda_j^{-1}, \quad j = 1,\dots,m.$$

Therefore, for each $j = 1,\dots,m$,

$$\inf_{G_{11}\preceq D^{-1}} \log^2\lambda_j(G) = \begin{cases} \log^2\lambda_j & \text{if } \lambda_j > 1, \\ 0 & \text{if } \lambda_j \le 1, \end{cases} \tag{5.7}$$

and for each $j = m+1,\dots,n$,

$$\inf_{G_{11}\preceq D^{-1}} \log^2\lambda_j(G) = 0.$$

[2] We may take $X_1 = D^{-1/2}X$ where X and D are as in the beginning of Section 5.1. X_1 exists by Theorem 2.1 and Y_1 exists as $I_{n-m} - B_{12}^*B_{11}^{-1}B_{12}$ is the Schur complement of B_{11} in B, which is positive definite.

Let $G_0 = \operatorname{diag}(g_1, \dots, g_n)$ where

$$g_j = \begin{cases} \lambda_j^{-1} & \text{if } \lambda_j > 1 \text{ and } j = 1, \dots, m, \\ 1 & \text{otherwise.} \end{cases}$$

Then it is clear that $(G_0)_{11} \preceq D^{-1}$ and G_0 is our desired matrix by (5.7). □

Acknowledgment

This work came from an impromptu discussion of the first two authors at the workshop on "Nonlinear Data: Theory and Algorithms," where they met for the first time. LHL and RS gratefully acknowledge the workshop organizers and the Mathematical Research Institute of Oberwolfach for hosting the wonderful event.

The research leading to these results has received funding from the Defense Advanced Research Projects Agency under the DARPA Grant D15AP00109, the European Research Council under the Advanced ERC Grant Agreement Switchlet n.670645, the National Natural Science Foundation of China under the NSFC Grant no. 11688101, and the National Science Foundation under the NSF Grant IIS 1546413. In addition, LHL's work is supported by a DARPA Director's Fellowship and the Eckhardt Faculty Fund from the University of Chicago; KY's work is supported by the Hundred Talents Program of the Chinese Academy of Sciences as well as the Thousand Talents Plan of the State Council of China.

References

[1] F. Barbaresco. Innovative tools for radar signal processing based on Cartan's geometry of spd matrices and information geometry. In *2008 IEEE Radar Conference*, pages 1–6, May 2008.
[2] R. Bhatia. *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.
[3] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.*, 31(3):1055–1070, 2009.
[4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.
[5] P. T. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.*, 87(2):250–262, Feb. 2007.
[6] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.
[7] A. J. Laub. *Matrix analysis for scientists and engineers*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
[8] J. D. Lawson and Y. Lim. The geometric mean, matrices, metrics, and more. *Amer. Math. Monthly*, 108(9):797–812, 2001.
[9] M. Moakher and M. Zéraï. The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *J. Math. Imaging Vision*, 40(2):171–187, 2011.
[10] Y. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2(4):333–361, 2002.
[11] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Int. J. Comput. Vis.*, 25(1):127, Jul 2006.
[12] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *Int. J. Comput. Vis.*, 66(1):41–66, Jan 2006.
[13] S. T. Smith. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.*, 53(5):1610–1630, 2005.
[14] K. Ye and L.-H. Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM J. Matrix Anal. Appl.*, 37(3):1176–1197, 2016.

Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, Chicago, IL 60637, USA
E-mail address: `lekheng@galton.uchicago.edu`

Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK
E-mail address: `r.sepulchre@eng.cam.ac.uk`

KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China
E-mail address: `keyk@amss.ac.cn`