Geometric Distance Between Positive Definite Matrices of Different Dimensions

Lek-Heng Lim, Rodolphe Sepulchre, *Fellow, IEEE*, and Ke Ye

***Abstract*— We show how the geodesic distance on $\mathbb{S}^n_{++}$, the cone of $n \times n$ real symmetric or complex Hermitian positive definite matrices regarded as a Riemannian manifold, 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, inner products on $\mathbb{R}^n$, and $n \times n$ covariances of nondegenerate probability distributions, this gives us a natural way to define a geometric distance between a pair of such objects of different dimensions.**

***Index Terms*— Riemannian manifold, geodesic distance, positive definite matrices, covariance matrices, ellipsoids.**

I. 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 tensor, and the infimum length of geodesics connecting two points gives a *geodesic distance* $\delta_2 : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \to \mathbb{R}_+$,

$$\delta_2(A, B) := \Big[\sum\nolimits_{j=1}^{n} \log^2(\lambda_j(A^{-1}B))\Big]^{1/2}. \tag{1}$$

The Riemannian metric tensor and geodesic 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 [1, Chapter 6].

The geodesic distance δ_2 is arguably the most natural and useful distance on the positive definite cone $\mathbb{S}^n_{++}$ [2]. It may be thought as a generalization of $|\log(a/b)|$, the geometric distance between two positive numbers, to $\mathbb{S}^n_{++}$ [2]. 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. For comparison, all common 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 [3], medical imaging [4], [5], radar signal processing [6], statistical inference [7], among other areas. In optimization, δ_2 has been shown [8] 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 [9] 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 [10], 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:

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

$$\mathcal{E}_A := \{x \in \mathbb{R}^n : x^{\mathsf{T}}Ax \leq 1\};$$

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

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

(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 $\mathbb{C}$.

Manuscript received June 4, 2018; revised December 17, 2018; accepted March 22, 2019. Date of publication April 30, 2019; date of current version August 16, 2019. L.-H. Lim was supported in part by the Defense Advanced Research Projects Agency (DARPA) under Grant D15AP00109, in part by the National Science Foundation under the NSF under Grant IIS 1546413, in part by the DARPA Director's Fellowship, and in part by the Eckhardt Faculty Fund through The University of Chicago. R. Sepulchre was supported by the European Research Council through the Advanced ERC Grant under Grant 670645. K. Ye was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11688101 and Grant 11801548, in part by the National Key R&D Program of China under Grant 2018YFA0306702, in part by the Hundred Talents Program of the Chinese Academy of Sciences, and in part by the Thousand Talents Plan of the State Council of China.

L.-H. Lim is with the Computational and Applied Mathematics Initiative, Department of Statistics, The University of Chicago, Chicago, IL 60637 USA (e-mail: lekheng@galton.uchicago.edu).

R. Sepulchre is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: r.sepulchre@eng.cam.ac.uk).

K. Ye is with the KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: keyk@amss.ac.cn).

Communicated by P. Grohs, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2019.2913874

0018-9448 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

In fact, it is easier 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)*. 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 [11]. 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 $n \times n$ Hermitian positive semidefinite matrices, and $\mathbb{S}^n_{++}$ the open cone of $n \times n$ Hermitian positive definite matrices. If $A \in \mathbb{S}^n$, we write $\dim A := n$ for its number of rows/columns. $\preceq$ 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.

II. 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 Hermitian matrix and a positive definite matrix may be simultaneously diagonalized. We state a version of this well-known result below [13, Theorem 12.19].

Theorem 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*

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

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

Note that the eigenvalues of $A^{-1}B$ are necessarily real since it is similar to the Hermitian matrix $A^{-1/2}BA^{-1/2}$. As usual, we will order the eigenvalues of $A \in \mathbb{S}^n_{++}$ nonincreasingly:

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

[1]While a complex positive (semi)definite matrix is necessarily Hermitian, a real positive (semi)definite matrix does not need to be symmetric [12, p. 80].

The next two standard results may be found as [14, Theorem 4.3.28, Corollary 7.7.4].

Theorem 2 (Cauchy interlacing inequalities). *Let $m \le n$ and $A \in \mathbb{S}^n$. If we partition A into*

$$A = \begin{bmatrix} A_1 & A_2 \\ A_2^* & A_3 \end{bmatrix},$$

with $A_1 \in \mathbb{S}^m$, $A_2 \in \mathbb{F}^{m\times(n-m)}$, $A_3 \in \mathbb{S}^{n-m}$, then

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

Proposition 3 (Majorization). *If $A, B \in \mathbb{S}^n_{++}$ and $A \preceq B$, then $\lambda_j(A) \le \lambda_j(B)$, $j = 1, \dots, n$.*

III. 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 I. For $A \in \mathbb{S}^n_{++}$, the n-dimensional *ellipsoid* $\mathcal{E}_A$ centered at the origin is

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

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 4. *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

$$x^*Ax \le 1 \tag{2}$$

we also have $x^*Bx \le 1$. Thus we have $y^*By \le y^*Ay$ for any $y \in \mathbb{F}^n$ since $x = y/\sqrt{y^*Ay}$ satisfies (2). Conversely, if $B \preceq A$, then whenever x satisfies (2), we have $x^*Bx \le x^*Ax \le 1$. $\square$

Lemma 4 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

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

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 4 gives the following.

Lemma 5. *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 $B_{11} \preceq A$.*

IV. 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 geodesic distance on a Grassmannian to subspaces of different dimensions proposed in [15]. The following convex sets will play the role of the Schubert varieties in [15].

Definition 6. 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

$$\Omega_+(A) := \left\{ G = \begin{bmatrix} G_{11} & G_{12} \\ G_{12}^* & G_{22} \end{bmatrix} \in \mathbb{S}^n_{++} : G_{11} \preceq A \right\}. \quad (3)$$

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

$$\Omega_-(B) := \{ H \in \mathbb{S}^m_{++} : B_{11} \preceq H \}, \quad (4)$$

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

Lemma 5 provides justification for the names: more precisely, $\Omega_+(A)$ parameterizes all n-dimensional ellipsoids containing $\iota_{m,n}(\mathcal{E}_A)$ whereas $\Omega_-(B)$ parameterizes 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.,

$$\begin{aligned} \delta_2\big(A, \Omega_-(B)\big) &:= \inf_{H \in \Omega_-(B)} \delta_2(A, H) \\ &= \inf_{H \in \Omega_-(B)} \Big[\sum\nolimits_{j=1}^m \log^2 \lambda_j (A H^{-1}) \Big]^{1/2}; \end{aligned} \quad (5)$$

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.,

$$\begin{aligned} \delta_2\big(B, \Omega_+(A)\big) &:= \inf_{G \in \Omega_+(A)} \delta_2(G, B) \\ &= \inf_{G \in \Omega_+(A)} \Big[\sum\nolimits_{j=1}^n \log^2 \lambda_j (G B^{-1}) \Big]^{1/2}. \end{aligned} \quad (6)$$

We will show that

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

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

Note that $\Omega_+(A) \subseteq \mathbb{S}^n_{++}$ and $\Omega_-(B) \subseteq \mathbb{S}^m_{++}$, (5) is the distance of a point A to a set $\Omega_-(B)$ within the Riemannian manifold $\mathbb{S}^m_{++}$, (6) 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 7. *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*

$$\delta_2\big(A, \Omega_-(B)\big) = \delta_2\big(B, \Omega_+(A)\big) \quad (7)$$

and their common value is given by

$$\delta_2^+(A, B) := \Big[\sum\nolimits_{j=1}^m \max\{0, \log \lambda_j (A^{-1} B_{11})\}^2 \Big]^{1/2}, \quad (8)$$

or, alternatively,

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

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 7 to Section V but first make a few immediate observations regarding this new distance.

An implicit assumption in Theorem 7 is that whenever we write $\delta_2^+(A, B)$, 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, $\delta_2^+(A, B) \ne \delta_2^+(B, A)$; 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 $m = n$, (8) becomes

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

However, since $m = n$, we may swap the matrices A and B in (7) to get

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

and their common value is given by

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

Note that even in this case, $\delta^+(A, B) \ne \delta^+(B, A)$ in general. Nevertheless, this gives us the relation between our original geodesic distance δ_2 and the distance δ_2^+ defined in Theorem 7.

Proposition 8. *Let $m = n$. Then the distances δ_2 in* (1) *and δ_2^+ in* (8) *are related via*

$$\delta_2(A, B)^2 = \delta_2^+(A, B)^2 + \delta_2^+(B, A)^2.$$

Proposition 8 suggests that we may extend δ_2 in (1) to any $A, B \in \bigcup_{n=1}^\infty \mathbb{S}^n_{++}$ by defining

$$\delta_2(A, B) := \sqrt{\delta_2^+(A, B)^2 + \delta_2^+(B, A)^2}, \quad (9)$$

and setting $\delta_2^+(A, B) := 0$ whenever $\dim A > \dim B$. Then

$$\delta_2(A, B) = \begin{cases} \delta_2^+(A, B) & \text{if } \dim A \le \dim B, \\ \delta_2^+(B, A) & \text{if } \dim B \le \dim A. \end{cases}$$

Nevertheless δ_2 as defined in (9) is not a metric on $\bigcup_{n=1}^\infty \mathbb{S}^n_{++}$ either; indeed $\delta_2(I_m, I_n) = 0$ even when $m \ne n$.

V. Proof of Theorem 7

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 7 by showing that

$$\delta_2\big(A, \Omega_-(B)\big) = \Big[\sum_{j=1}^m \max\{0, \log \lambda_j (A^{-1} B_{11})\}^2 \Big]^{1/2} \quad (10)$$

in Lemma 10 and

$$\delta_2\big(B, \Omega_+(A)\big) = \Big[\sum_{j=1}^m \max\{0, \log \lambda_j (A^{-1} B_{11})\}^2 \Big]^{1/2} \quad (11)$$

in Lemma 11. The key to establishing these is to repeatedly use the following invariance of δ_2 under congruence action by nonsingular matrices.

Lemma 9 (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. □

A. Calculating $\delta_2\big(A, \Omega_-(B)\big)$

Recall that we partition $B \in \mathbb{S}^n_{++}$ into

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^* & B_{22} \end{bmatrix}.$$

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 1, there is a nonsingular $X \in \mathbb{F}^{m\times m}$ such that

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

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 9,

$$\begin{aligned}\delta_2\big(A, \Omega_-(B)\big) &= \delta_2\big(XAX^*, X\Omega_-(B)X^*\big) \\ &= \delta_2\big(I_m, \Omega_-(ZBZ^*)\big).\end{aligned}$$

Hence we may assume without loss of generality that

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

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 are now ready to prove (10).

Lemma 10. *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*

$$\begin{aligned}\delta_2\big(A, \Omega_-(B)\big) &= \delta_2(A, H_0) \\ &= \Big[\sum_{j=1}^m \max\{0, \log\lambda_j(A^{-1}B_{11})\}^2\Big]^{1/2}.\end{aligned}$$

Proof. By the preceding discussions, we may assume that A and B are as in (12). So we have

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

The condition $D \preceq H$ implies that $\lambda_j \le \lambda_j(H)$, $j = 1, \dots, m$, by Proposition 3. Hence

$$\inf_{D\preceq H}\log^2\lambda_j(H) = \begin{cases} \log^2\lambda_j & \text{if } \lambda_j > 1, \\ 0 & \text{if } \lambda_j \le 1. \end{cases} \tag{13}$$

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 (13). □

B. 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 V-A. Let L be the upper triangular matrix

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

Then

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

and

$$L\Omega_+(A)L^* = \Omega_+(A).$$

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 9, we have

$$\begin{aligned}\delta_2\big(B, \Omega_+(A)\big) &= \delta_2\big(LBL^*, L\Omega_+(A)L^*\big) \\ &= \delta_2\big(LBL^*, \Omega_+(A)\big).\end{aligned} \tag{14}$$

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

$$\begin{gathered} X_1\ AX_1^* = D^{-1}, \quad X_1\ B_{11}X_1^* = I_m, \\ Y_1(I_{n-m} - B_{12}^*B_{11}^{-1}B_{12})Y_1^* = I_{n-m}, \end{gathered}$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ with $\lambda_j := \lambda_j(A^{-1}B_{11})$, $j = 1, \dots, m$. Let

$$Z_1 = \begin{bmatrix} X_1 & 0 \\ 0 & Y_1 \end{bmatrix}.$$

Then

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

Hence, by (14) and Lemma 9,

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

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

$$A = D^{-1} = \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_m^{-1}), \quad B = I_n. \tag{15}$$

We are now ready to prove (11).

Lemma 11. *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*

$$\begin{aligned}\delta_2\big(B, \Omega_+(A)\big) &= \delta_2(G_0, B) \\ &= \Big[\sum_{j=1}^m \max\{0, \log\lambda_j(A^{-1}B_{11})\}^2\Big]^{1/2}.\end{aligned}$$

[2]We may take $X_1 = D^{-1/2}X$ where X and D are as in the beginning of Section V-A. X_1 exists by Theorem 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.

Proof. By the preceding discussions, we may assume that A and B are as in (15). So we have

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

where G_{11} is the upper left $m \times m$ principal submatrix of $G \in \Omega_+(D^{-1})$. By Proposition 3, we have $\lambda_j(G_{11}) \le \lambda_j^{-1}$, $j = 1, \dots, m$. Moreover, by Theorem 2

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

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

$$\inf_{G_{11} \preceq D^{-1}} \log^2 \lambda_j(G) = \begin{cases} \log^2 \lambda_j & \text{if } \lambda_j > 1, \\ 0 & \text{if } \lambda_j \le 1, \end{cases} \tag{16}$$

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

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

as G_{22} can be chosen to be I_{n-m}. 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 (16). □

Acknowledgment

We thank the two anonymous referees for their very careful reviews and helpful comments. 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.

References

[1] R. Bhatia, *Positive Definite Matrices* (Princeton Series in Applied Mathematics). Princeton, NJ, USA: Princeton Univ. Press, 2007.

[2] S. Bonnabel and R. Sepulchre, "Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank" *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 3, pp. 1055–1070, Aug. 2009.

[3] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.

[4] P. T. Fletcher and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Process.*, vol. 87, no. 2, pp. 250–262, Feb. 2007.

[5] 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 Vis.*, vol. 40, no. 2, pp. 171–187, Jun. 2011.

[6] F. Barbaresco, "Innovative tools for radar signal processing based on cartan's geometry of SPD matrices & information geometry," in *Proc. IEEE Radar Conf.*, May 2008, pp. 1–6.

[7] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imaging Vis.*, vol. 25, no. 1, p. 127, Jul. 2006.

[8] T. Nesterov, "On the Riemannian geometry defined by self-concordant barriers and interior-point methods," *Found. Comput. Math.*, vol. 2, no. 4, pp. 333–361, Oct. 2002.

[9] S. T. Smith, "Covariance, subspace, and intrinsic Cramér–Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.

[10] J. D. Lawson and Y. Lim, "The geometric mean, matrices, metrics, and more," *Amer. Math. Monthly*, vol. 108, no. 9, pp. 797–812, Nov. 2001.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[12] F. Zhang, *Matrix Theory*, 2nd ed. New York, NY, USA: Springer, 2011.

[13] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2005.

[14] R. A. Horn, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[15] K. Ye and L.-H. Lim, "Schubert varieties and distances between subspaces of different dimensions," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 3, pp. 1176–1197, Sep. 2016.

Lek-Heng Lim was educated at the National University of Singapore (BS) and at Cornell (MS), Cambridge (Clare Hall Fellow), and Stanford (PhD) Universities. He was a Charles Morrey Assistant Professor at the University of California, Berkeley, and then an Assistant and Associate Professor at The University of Chicago. He is a winner of the 2019 ILAS Hans Schneider Prize, the 2017 SIAM James Wilkinson Prize, the 2017 FoCM Stephen Smale Prize, and a 2017 DARPA Director's Fellowship.

Rodolphe Sepulchre (M'96–SM'08–F'10) received the engineering degree and the Ph.D. degree from the Université catholique de Louvain in 1990 and in 1994, respectively. After a Postdoctoral stay at the University of California in Santa Barbara (1994–1996) and at the Université catholique de Louvain (1995–1997), he became a Faculty at the Université de Liège, Belgium. He moved to Cambridge University in 2013, where he is currently Professor of Engineering and a Professioral Fellow of Sidney Sussex College. He held visiting positions at Princeton University (2002–2003), the Ecole des Mines de Paris (2009–2010), Caltech (2018), and part-time positions at the University of Louvain (2000–2011) and at INRIA Lille Europe (2012–2013). In 2008, he was awarded the IEEE Control Systems Society Antonio Ruberti Young Researcher Prize. He is a fellow of IEEE and SIAM. He has been IEEE CSS distinguished lecturer between 2010 and 2015. In 2013, he was elected at the Royal Academy of Belgium.

Ke Ye received his Ph.D. from Texas A&M University at College Station in 2012. In 2012–2017, he served as an L. E. Dickson Instructor in the Mathematics Department and a PostDoctoral Scholar in the Statistics Department at The University of Chicago. He became a Tenured Associate Professor at the Chinese Academy of Sciences in 2017. He was recently selected for the Thousand Talent Program for Young Outstanding Scientists of China (2018) and the Hundred Talent Program of Chinese Academy of Sciences (2017).