Canad. J. Math. Vol. **56** (4), 2004 pp. 776–793

Best Approximation in Riemannian Geodesic Submanifolds of Positive Definite Matrices

Yongdo Lim

Abstract. We explicitly describe the best approximation in geodesic submanifolds of positive definite matrices obtained from involutive congruence transformations on the Cartan-Hadamard manifold $\mathrm{Sym}(n,\mathbb{R})^{++}$ of positive definite matrices. An explicit calculation for the minimal distance function from the geodesic submanifold $\mathrm{Sym}(p,\mathbb{R})^{++} \times \mathrm{Sym}(q,\mathbb{R})^{++}$ block diagonally embedded in $\mathrm{Sym}(n,\mathbb{R})^{++}$ is given in terms of metric and spectral geometric means, Cayley transform, and Schur complements of positive definite matrices when $p \le 2$ or $q \le 2$.

1 Introduction

In this paper, we consider the best approximation and the minimal distance function for geodesic submanifolds of the Riemannian symmetric space $\mathrm{Sym}(n,\mathbb{R})^{++}$ of positive definite matrices. The convex cone $\mathrm{Sym}(n,\mathbb{R})^{++}$ of $n \times n$ positive definite matrices is a typical example of Cartan-Hadamard manifold and hence a Bruhat-Tits space, complete metric space satisfying the "semiparallelogram law" ([6, 1]). The elementary linear algebra fact that the space $\mathrm{Sym}(n,\mathbb{R})^{++}$ is obtained by bending the inner product space $\mathrm{Sym}(n,\mathbb{R})$ equipped with the trace inner product via the exponential mapping and the nontrivial fact that the matrix exponential mapping is distance nondecreasing (seminegative curvature) are fundamental in the study of the cone as a Cartan-Hadamard or a Bruhat-Tits space in the geometric point of view (see, *e.g.*, [6, 7]).

If H is a finite dimensional inner product space, the parallelogram law holds with respect to the distance given by the inner product, and hence H becomes a Bruhat-Tits space. For a vector subspace K of H, the best approximation for K corresponds to the projection theorem, and the problems finding $x_K \in K$, the best approximation of x out of K, and computing the distance $\|x - x_K\|$ are completely determined in terms of the Gram determinant and matrix. The purpose of this paper is to study these problems in the *non-linear* space $\mathrm{Sym}(n,\mathbb{R})^{++}$ for the Riemannian metric distance δ replacing subspaces by geodesic submanifolds of $\mathrm{Sym}(n,\mathbb{R})^{++}$. Here, a set $\mathcal{S} = \exp(U)$ of $\mathrm{Sym}(n,\mathbb{R})^{++}$ for a vector subspace U of $\mathrm{Sym}(n,\mathbb{R})$ is said to be a geodesic submanifold if it contains the geodesic between any two of its points. The main

Received by the editors May 29, 2002; revised October 29, 2003.
This work was supported by the Korean Research Foundation Grant (KRF-2002-013-D00006)
AMS subject classification: Primary: 15A48; secondary: 49R50, 15A18, 53C3.
Keywords: Matrix approximation, positive definite matrix, geodesic submanifold, Cartan-Hadamard manifold, best approximation, minimal distance function, global tubular neighborhood theorem, Schur complement, metric and spectral geometric mean, Cayley transform.

https://doi.org/10.4153/CJM-2004-035-5 Published online by Cambridge University Press

problem we are considering related to the given geodesic submanifold $\mathcal{S}$ is to describe explicitly the best approximation $X_S \in S$ such that $\delta(X, X_{\mathcal{S}}) = \min_{A\in\mathcal{S}}\delta(X, A)$ and the minimal distance function $d_{\mathcal{S}}(X) = \delta(X, X_{\mathcal{S}})$ in terms of the give point X and the geodesic submanifold $\mathcal{S}$.

Although the projection theorem for arbitrary geodesic submanifolds (generally, for any closed and geodesically convex subsets) of positive definite matrices follows directly from the nonpositive curvature property of the Riemannian manifold $\mathrm{Sym}(n, \mathbb{R})^{++}$, the problem describing explicitly the best approximation of a point out of a geodesic submanifold $\mathcal{S}$ and its minimal distance is non-trivial because it is closely related to the minimization problem

$$\delta(X, A) \to \min, A \in \mathcal{S}.$$

(See Remark 3.1 for a complete version associated to eigenvalue functions on the space $\mathrm{Sym}(n, \mathbb{R})$.) In this paper, we restrict our attention to the geodesic submanifold of positive definite matrices fixed by an involutive congruence transformation of the form $\sigma_W(X) = WXW, W = W^T = W^{-1}$. In [10, 11], the authors have studied a similar type of matrix approximation via the dualistic nature of the manifold $\mathrm{Sym}(n, \mathbb{R})^{++}$ and its information theoretic implications. They considered two dual affine connections ∇ and ∇^* satisfying $Ag(B, C) = g(\nabla_A B, C) + g(B, \nabla^*{}_A C)$ for the Riemannian metric g on $\mathrm{Sym}(n, \mathbb{R})^{++}$ and considered the optimization problems defined on the intersection of the cone $\mathrm{Sym}(n, \mathbb{R})^{++}$ and an affine subspace for measures invariant under the congruent transformations $X \to MXM^T, M \in \mathrm{GL}(n, \mathbb{R})$. They obtained an algorithm for the optimization problem with respect to the ∇^*-divergence defined by $D^*(P_1, P_2) := \log\det(P_1 P_2^{-1}) + \mathrm{tr}(P_1^{-1} P_2) - n$ and they asked the optimization problem with respect to the Riemannian metric distance δ, the most natural measures on the cone. The geodesic submanifolds arising from Jordan involutions are domains of doubly autoparallel ([11, Theorem 4.6]).

This paper is organized as follows. In section 2, we describe the Riemannian structures of the cone $\mathrm{Sym}(n, \mathbb{R})^{++}$ together with basic properties of (metric and spectral) geometric means of positive definite matrices. In section 3, we describe the best approximation for geodesic submanifolds obtained from Jordan involutions in terms of metric and spectral geometric means, and in section 4, we give an explicit formula for the best approximation of the geodesic submanifold $\mathrm{Sym}(p, \mathbb{R})^{++} \times \mathrm{Sym}(q, \mathbb{R})^{++}$ block diagonally embedded in $\mathrm{Sym}(n, \mathbb{R})^{++}$. In section 5, we study the eigenvalues of matrices appearing in the best approximation theorem for $\mathrm{Sym}(p, \mathbb{R})^{++} \times \mathrm{Sym}(q, \mathbb{R})^{++}$ and give a formula for the distance function when either $p \le 2$ or $q \le 2$. Finally, we establish in section 6 a relation between the best approximation and the global tubular neighborhood theorem which produces a factorization of positive definite matrices with factors of geometric and spectral geometric means.

2 Riemannian Structures of the Positive Definite Cone

It is assumed that all matrices involved in this paper have real elements. The identity matrix and the null matrix are denoted by I and 0, respectively, and their sizes are determined by the context. A matrix X is symmetric if $X = X^T$, where X^T denotes the transpose of the matrix X. Let $\mathrm{Sym}(n,\mathbb{R})$ be the vector space of all $n\times n$ real symmetric matrices. For $X\in\mathrm{Sym}(n,\mathbb{R})$, we recall that A is positive semidefinite, denoted by $0\le X$, if $x^TXx\ge 0$ for all $x\in\mathbb{R}^n$. Similarly, X is positive definite, denoted by $X>0$, if it is positive semidefinite and invertible. We denote the set of positive definite (resp., semidefinite) matrices by $\mathrm{Sym}(n,\mathbb{R})^{++}$ (resp., $\mathrm{Sym}(n,\mathbb{R})^{+}$). We consider two useful relations on $\mathrm{Sym}(n,\mathbb{R})$, the Löwner partial order on $\mathrm{Sym}(n,\mathbb{R})$ defined by $X\le Y$ if and only if $Y-X$ is positive semidefinite, and $X<Y$ if and only if $Y-X$ is positive definite. The set $\mathrm{Sym}(n,\mathbb{R})^{++}$ is an open convex cone of $\mathrm{Sym}(n,\mathbb{R})$ and is a typical example of a Cartan-Hadamard manifold, complete simply connected Riemannian manifold with seminegative curvature. We shortly review the Riemannian structure of $\mathrm{Sym}(n,\mathbb{R})^{++}$. See, *e.g.,* [6, 7] for more details.

The inner product $\langle X|Y\rangle := \mathrm{tr}(XY)$ on the vector space $\mathrm{Sym}(n,\mathbb{R})$ which can be identified with the tangent space of $\mathrm{Sym}(n,\mathbb{R})^{++}$ at I gives rise to a natural Riemannian metric on $\mathrm{Sym}(n,\mathbb{R})^{++}$. The inner product on the tangent space of $\mathrm{Sym}(n,\mathbb{R})^{++}$ at $A>0$ is given by $\langle X|Y\rangle_A = \mathrm{tr}(A^{-1}XA^{-1}Y)$. The corresponding Riemannian metric distance is completely measured by

$$\delta(A,B)=\Big(\sum_{i=1}^{n}\log^2\lambda_i\Big)^{1/2},$$

where $\lambda_1,\dots,\lambda_n$ are eigenvalues of $A^{-1}B$. Since $A^{-1}B$ is similar to $A^{-1/2}BA^{-1/2}$, the eigenvalues of $A^{-1}B$ are all positive, and hence $\log\lambda_i$ is defined for each i. The Riemannian metric is $\mathrm{GL}(n,\mathbb{R})$-invariant and each member M of $\mathrm{GL}(n,\mathbb{R})$ acts as an isometry on $\mathrm{Sym}(n,\mathbb{R})^{++}$ via the congruence transformation, $X\mapsto MXM^T$. Let $F(X):=-\log\det(X)$ be the standard barrier function of the cone $\mathrm{Sym}(n,\mathbb{R})^{++}$ (see, *e.g.,* [2, Chapter 3]). Then the Riemannian metric coincides with the Hessian metric of F. Indeed, $\langle X|Y\rangle_A=\langle F''(A)X|Y\rangle_I=\mathrm{tr}(A^{-1}XA^{-1}Y)$, for $A\in\mathrm{Sym}(n,\mathbb{R})^{++}$ and $X,Y\in\mathrm{Sym}(n,\mathbb{R})$. The matrix inversion on $\mathrm{Sym}(n,\mathbb{R})^{++}$, $A\to A^{-1}$ is an involutive isometry on $\mathrm{Sym}(n,\mathbb{R})^{++}$ and hence $\mathrm{Sym}(n,\mathbb{R})^{++}$ becomes a Riemannian symmetric space. The unique geodesic curve joining A and B is given by

$$\gamma(t):=A^{1/2}(A^{-1/2}BA^{-1/2})^tA^{1/2}$$

and the geodesic middle $A\#B:=\gamma(1/2)=A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ is known as the *geometric mean* of A and B in matrix theory. The realization of the geometric mean $A\#B$ as the geometric middle of the invariant Riemannian metric gives the following well-known but non-trivial properties (the commutativity, inversion, and transformation properties) which will be useful for our purposes (see [5, 7] for more details):

$$A\#B=B\#A,\ A^{-1}\#B^{-1}=(A\#B)^{-1} \tag{2.1}$$

$$M(A\#B)M^T=(MAM^T)\#(MBM^T) \tag{2.2}$$

for all $M\in\mathrm{GL}(n,\mathbb{R})$. One remarkable and important property of the geometric mean is that $A\#B$ is a unique positive definite solution of the Riccati equation (Riccati Lemma)

$$XA^{-1}X=B. \tag{2.3}$$

Proposition 2.1 *Let $A \in \mathrm{Sym}(n,\mathbb{R})^{++}$. Then $(A\#B)^2 > B$ for all $B \in \mathrm{Sym}(n,\mathbb{R})^{++}$ if and only if $A > I$.*

Proof Suppose that $(A\#B)^2 > B$ for all $B \in \mathrm{Sym}(n,\mathbb{R})^{++}$. Then with $B = I$ we have $A = (A^{1/2})^2 = (A\#I)^2 > I$. Conversely, suppose that $A > I$. Then by the order decreasing property of the inversion, $A^{-1} < I$. Let $B \in \mathrm{Sym}(n,\mathbb{R})^{++}$. Then by (2.1) and by the Riccati Lemma, we have

$$(A\#B)^{-1}B(A\#B)^{-1} = (B^{-1}\#A^{-1})B(B^{-1}\#A^{-1}) = A^{-1} < I$$

and hence $B < (A\#B)I(A\#B) = (A\#B)^2$. ■

In [3] Fielder and Pták have introduced and developed a new positive definite geometric mean of two positive definite matrices; the *spectral geometric mean* $F(A,B)$ of positive definite matrices A and B which is defined by

$$F(A,B) := (A^{-1}\#B)^{1/2}A(A^{-1}\#B)^{1/2}$$

By the Riccati Lemma (2.3), we immediately see that the spectral geometric mean $F(A,B)$ is a unique positive definite solution of the equation

$$A^{-1}\#B = (A^{-1}\#X)^2. \tag{2.4}$$

See [3] for the following properties of the spectral geometric mean and for more details.

Proposition 2.2 *Let $A, B \in \mathrm{Sym}(n,\mathbb{R})^{++}$. Then*

(1) $F(A,B) = F(B,A)$,
(2) $F(A,B)^{-1} = F(A^{-1},B^{-1})$,
(3) $F(A,B) = A\#B$ *if and only if* $AB = BA$,
(4) *(Spectral mean)* $F(A,B)^2$ *is positively similar to* AB.

As they mentioned in [3], the geometric mean $A\#B$ and the spectral geometric mean $F(A,B)$ are in general not comparable in the Loewner ordering, but property (3) in Proposition 2.2 indicates that two means are equal if and only if A and B commute which is a consequence of Fuglede-Putnam theorem (Lemma 5.1 and Theorem 5.2 of [3]). These basic properties of geometric and spectral geometric means of positive definite matrices play a crucial role for our work.

3 Geodesic Submanifolds of Positive Definite Matrices

Throughout this paper, we denote $V := \mathrm{Sym}(n,\mathbb{R})$ and $\Omega := \mathrm{Sym}(n,\mathbb{R})^{++}$ for notational convenience. Let U be a vector subspace of V and let $\mathcal{S} = \exp(U) = \{\exp(X) \mid X \in U\}$. We say that $\mathcal{S}$ is a *symmetric submanifold* of Ω if $ABA \in \mathcal{S}$ whenever $A, B \in \mathcal{S}$. And $\mathcal{S}$ is said to be a *geodesic submanifold* of Ω if for all $A, B \in \mathcal{S}$ and $t \in \mathbb{R}$, $A\#_tB := A^{1/2}(A^{-1/2}BA^{-1/2})^tA^{1/2} \in \mathcal{S}$, that is, $\mathcal{S}$ contains the geodesic between

https://doi.org/10.4153/CJM-2004-035-5 Published online by Cambridge University Press

A and B. By Theorem XII.3.7 of [6], $\mathcal{S}$ is a symmetric submanifold if and only if $\mathcal{S}$ is a geodesic submanifold. Every geodesic submanifold $\mathcal{S}$ is a closed and (mid-point) convex in the Cartan-Hadamard manifold Ω with respect to the Riemannian distance δ and hence for each $X \in \Omega$ there is a unique point $X_{\mathcal{S}} \in \mathcal{S}$ of minimal distance to X (see Corollary I.5.6 of [1]):

$$\delta(X, X_{\mathcal{S}}) = \delta(X, \mathcal{S}) := \inf_{A \in \mathcal{S}} \delta(X, A).$$

The map $X \mapsto X_{\mathcal{S}}$ is called the projection onto $\mathcal{S}$ and it has Lipschitz constant 1. We denote $d_{\mathcal{S}}$ by the minimal distance function for the geodesic submanifold $\mathcal{S}$;

$$d_{\mathcal{S}} \colon \Omega \to [0, \infty), \; d_{\mathcal{S}}(X) = \delta(X, X_{\mathcal{S}}).$$

Then it is continuous and is a convex function, that is, $d_{\mathcal{S}} \circ \gamma$ is convex for any geodesic γ in Ω.

Remark 3.1 If $\mathcal{S}$ is a geodesic submanifold (*e.g.*, the geodesic submanifold of diagonal matrices with positive entries) of Ω then it is closed under the inversion for since $I \in \mathcal{S}$ and $A^t \in \mathcal{S}$ for all $t \in \mathbb{R}$ whenever $A \in \mathcal{S}$. Observing that for $A \in \mathcal{S}$, $\delta(X, A^{-1}) = \delta(I, X^{-1/2} A^{-1} X^{-1/2}) = \delta(I, X^{1/2} A X^{1/2})$ from the fact that the inversion is an isometry, the problem of finding the best approximation $X_{\mathcal{S}}$ in $\mathcal{S}$ turns to be the minimization problem

$$\min_{A \in \mathcal{S}} \| \log \lambda(X^{1/2} A X^{1/2}) \| = \min_{Y \in X^{1/2} \mathcal{S} X^{1/2}} \| \log \lambda(Y) \|$$

where $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^n$ and $\lambda \colon \operatorname{Sym}(n, \mathbb{R}) \to \mathbb{R}^n$ denotes the "eigenvalue map" of non-increasing order. The map

$$\| \log(\,\cdot\,) \| \colon \mathbb{R}^n \to [0, \infty], \; x = (x_1, x_2, \dots, x_n) \mapsto \| \log x \| \in [0, \infty]$$

is permutation-invariant and hence the function $Y \mapsto \| \log \lambda(Y) \|$ is an "eigenvalue function" on $\operatorname{Sym}(n, \mathbb{R})$ (see [8, 9] for analysis of eigenvalue functions).

A linear invertible transformation $\sigma \colon V \to V$ is called a *Jordan automorphism* if

$$\sigma(XY + YX) = \sigma(X)\sigma(Y) + \sigma(Y)\sigma(X),$$

for all $X, Y \in V$. Observe that $XYX = 2X \circ (X \circ Y) - X^2 \circ Y$ and every Jordan automorphism σ preserves the Jordan product $X \circ Y := \frac{1}{2}(XY + YX)$. This implies that $\sigma(XYX) = \sigma(X)\sigma(Y)\sigma(X)$ for all $X, Y \in V$ and $\sigma(A^{-1}) = \sigma(A)^{-1}$ for any invertible matrix A. Furthermore, every Jordan automorphism acts as an isometry on the Riemannian manifold Ω. Indeed, if $A, B \in \Omega$ then $\sigma(A)^{-1}\sigma(B)$ is similar to $\sigma(A)^{-1/2}\sigma(B)\sigma(A)^{-1/2} = \sigma(A^{-1/2} B A^{-1/2})$ which has the same spectrum of $A^{-1/2} B A^{-1/2}$ by spectral decomposition. In particular, each Jordan automorphism preserves the geometric mean: $\sigma(A \# B) = \sigma(A) \# \sigma(B)$ for all $A, B \in \Omega$.

Let σ be a Jordan automorphism of V. Consider the ± 1-*eigenspaces* of σ on V and Ω, respectively:

$$V_\sigma^+ := \{A \in V : \sigma(A) = A\},\ V_\sigma^- := \{A \in V : \sigma(A) = -A\}$$

and

$$\Omega_\sigma^+ := \{A \in \Omega : \sigma(A) = A\} = \Omega \cap V_\sigma^+, \Omega_\sigma^- := \{A \in \Omega : \sigma(A) = A^{-1}\}.$$

By observing that $\sigma(\exp X) = \exp(\sigma(X))$ for all $X \in V$ and that $\exp\colon V \to \Omega$ is bijective, we have that $\Omega_\sigma^+ = \exp(V_\sigma^+)$ and $\Omega_\sigma^- = \exp(V_\sigma^-)$. Since $\sigma(ABA) = \sigma(A)\sigma(B)\sigma(A)$, it follows that $ABA \in \Omega_\sigma^\pm$ for all $A, B \in \Omega_\sigma^\pm$ and hence $\Omega_\sigma^\pm$ are geodesic submanifolds of Ω.

Proposition 3.2 *If σ is an involutive Jordan automorphism then*

$$X_{\Omega_\sigma^\pm} = X\#\sigma(X^{\pm 1}), \tag{3.1}$$

$$F(X, \sigma(X^{\pm 1})) \in \Omega_\sigma^\pm. \tag{3.2}$$

Proof It is immediate seen from (2.1) and (2.2) that for $X \in \Omega$,

$$\sigma(X\#\sigma(X)) = \sigma(X)\#\sigma^2(X) = \sigma(X)\#X = X\#\sigma(X)$$

and

$$\sigma(X\#\sigma(X^{-1})) = \sigma(X)\#X^{-1} = X^{-1}\#\sigma(X) = (X\#\sigma(X^{-1}))^{-1}.$$

This implies that $X\#\sigma(X^{\pm 1}) \in \Omega_\sigma^\pm$. Since the Riemannian metric distance δ is invariant under the inversion $A \mapsto A^{-1}$ and under the Jordan involution σ, we find that $\delta(\sigma(A), \sigma(B)) = \delta(A, B) = \delta(A^{-1}, B^{-1})$. Let $X \in \Omega$. Then for any $A \in \Omega_\sigma^\pm$,

$$\begin{aligned}
\delta(X, X\#\sigma(X^{\pm 1})) &= \frac{1}{2}\delta\left(X, \sigma(X^{\pm 1})\right) \\
&\le \frac{1}{2}\left(\delta(X, A) + \delta(A, \sigma(X^{\pm 1}))\right) \\
&= \frac{1}{2}\left(\delta(X, A) + \delta(\sigma(A^{\pm 1}), \sigma(X^{\pm 1}))\right) \\
&= \frac{1}{2}\left(\delta(X, A) + \delta(A, X)\right) \\
&= \delta(X, A).
\end{aligned}$$

which implies that $X_{\Omega_\sigma^\pm} = X\#\sigma(X^{\pm 1})$.

Next, suppose that $X \in \Omega$. Then

$$\begin{aligned}
\sigma\left(F(X, \sigma(X^{\pm 1}))\right) &= \sigma\left((X^{-1}\#\sigma(X^{\pm 1}))^{1/2} X (X^{-1}\#\sigma(X^{\pm 1}))^{1/2}\right) \\
&= \sigma\left((X^{-1}\#\sigma(X^{\pm 1}))^{1/2}\right)\sigma(X)\sigma\left((X^{-1}\#\sigma(X^{\pm 1}))^{1/2}\right) \\
&= (\sigma(X^{-1})\#X^{\pm 1})^{1/2}\sigma(X)(\sigma(X^{-1})\#X^{\pm 1})^{1/2} \\
&= F(X^{\pm 1}, \sigma(X)).
\end{aligned}$$

It follows from Proposition 2.2(2) that $F(X, \sigma(X^{\pm 1})) \in \Omega_\sigma^\pm$. ■

If W is an involutive symmetric (orthogonal) matrix (that is, $W^2 = I$) then the map σ_W defined by $\sigma_W(X) = WXW$ is an involutive Jordan automorphism. Our main interest in this paper is to describe explicitly the best approximation and the distance function for the geodesic submanifold induced by the Jordan automorphism $\sigma_W, W^2 = I$. Note that every involutive symmetric matrix is orthogonally similar to $J_{p,q} = \left(\begin{smallmatrix} I_p & 0 \\ 0 & -I_q \end{smallmatrix}\right)$ for some p, and hence it is enough to restrict our interest to the Jordan involution $\sigma_{p,q} := \sigma_{J_{p,q}}$. We observe that the involution $\sigma_{p,q}\colon \operatorname{Sym}(n,\mathbb{R}) \to \operatorname{Sym}(n,\mathbb{R})$ is given by

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \mapsto \begin{pmatrix} A & -B \\ -B^T & C \end{pmatrix}.$$

In the following, we fix positive integers p and q such that $p + q = n$ and denote $\Omega_{p,q}$ by the space $\Omega = \operatorname{Sym}(n,\mathbb{R})^{++}$ together with the specified involution $\sigma_{p,q}$. By $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \in \Omega_{p,q}$ we shall mean that $X \in \Omega$, A and C are $p \times p$ and $q \times q$ matrices. We also denote $\Omega^+_{p,q}$ by the geodesic submanifold fixed by the involution $\sigma_{p,q}$ and similarly, $\Omega^-_{p,q}$. Obviously,

$$V^+_{\sigma_{p,q}} = \left\{ \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} : A \in \operatorname{Sym}(p,\mathbb{R}), C \in \operatorname{Sym}(q,\mathbb{R}) \right\},$$

$$V^-_{\sigma_{p,q}} = \left\{ \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} : B \text{ is a } p \times q \text{ matrix} \right\},$$

$$\Omega^+_{p,q} = \left\{ \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} : A \in \operatorname{Sym}(q,\mathbb{R})^{++}, C \in \operatorname{Sym}(p,\mathbb{R})^{++} \right\}.$$

Let $d_{p,q}\colon \Omega_{p,q} \to [0,\infty)$ be the minimal distance function associated to the involution $\sigma_{p,q}$:

$$d_{p,q}(X) = \min_{A \in \Omega^+_{p,q}} \delta(X, A) = \delta(X, X \# \sigma_{p,q}(X)).$$

We note that

$$\delta(X, X\#\sigma_{p,q}(X)) = \frac{1}{2}\delta(X, \sigma_{p,q}(X)) = \Big(\sum_{i=1}^n \log^2 \lambda_i \Big)^{1/2}$$

where $\lambda_1, \dots, \lambda_n$ are eigenvalues of $X^{-1}\sigma_{p,q}(X)$, and that the function $2d_{p,q}$ is the displacement function of the elliptic isometry $\sigma_{p,q}$ in the context of Hadamard manifolds (*cf.* [1]).

4 Best Approximation in the Geodesic Submanifold $\Omega^+_{p,q}$

The following result is well-known (*cf.* [4, Theorem 7.7.6]).

Proposition 4.1 *Let $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right)$. Then X is positive definite if and only if $A > 0, C > B^T A^{-1} B$ if and only if $C > 0, A > BC^{-1}B^T$.*

One of our main results is the following explicit description of the best approximation for the geodesic submanifold $\Omega^+_{p,q}$.

Theorem 4.2 *Let $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \in \Omega_{p,q}$. Then*

$$X_{\Omega^+_{p,q}} = X\#\sigma_{p,q}(X) = \begin{pmatrix} A\#(A - BC^{-1}B^T) & 0 \\ 0 & C\#(C - B^T A^{-1} B) \end{pmatrix}$$

and

$$F(X, \sigma_{p,q}(X^{-1})) = \begin{pmatrix} F(A, (A - BC^{-1}B^T)^{-1}) & G(X) \\ G(X)^T & F(C, (C - B^T A^{-1} B)^{-1}) \end{pmatrix},$$

where $G(X) = \big(A\#(A - BC^{-1}B^T)\big)^{-1/2} B \big(C\#(C - B^T A^{-1} B)\big)^{-1/2}$. In particular, $AB = BC$ if and only if $X_{\Omega^-_{p,q}} = F(X, \sigma_{p,q}(X^{-1}))$.

Proof Since $X\#\sigma_{p,q}(X) \in \Omega^+_{p,q}$, there exist positive definite matrices $U \in \mathrm{Sym}(p, \mathbb{R})^{++}$ and $W \in \mathrm{Sym}(q, \mathbb{R})^{++}$ such that

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \# \begin{pmatrix} A & -B \\ -B^T & C \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & W \end{pmatrix}.$$

It then follows by the Riccati Lemma (2.3) that

$$\begin{pmatrix} U & 0 \\ 0 & W \end{pmatrix} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}^{-1} \begin{pmatrix} U & 0 \\ 0 & W \end{pmatrix} = \begin{pmatrix} A & -B \\ -B^T & C \end{pmatrix}.$$

By a direct computation, we find that

$$\begin{aligned} AU^{-1}A &= U + BW^{-1}B^T, \\ CW^{-1}C &= W + B^T U^{-1} B, \\ CW^{-1}B^T &= B^T U^{-1} A, \end{aligned}$$

and hence

$$\begin{aligned} U &= U + \big(BW^{-1}B^T - BC^{-1}(CW^{-1}B^T)\big) \\ &= U + \big(BW^{-1}B^T - BC^{-1}(B^T U^{-1} A)\big) \\ &= (U + BW^{-1}B^T) - BC^{-1}(B^T U^{-1} A) \\ &= AU^{-1}A - BC^{-1}(B^T U^{-1} A) = (A - BC^{-1}B^T)U^{-1}A. \end{aligned}$$

This implies that $UA^{-1}U = A - BC^{-1}B^T$ and hence

$$U = A\#(A - BC^{-1}B^T) \tag{4.1}$$

https://doi.org/10.4153/CJM-2004-035-5 Published online by Cambridge University Press

by the Riccati Lemma (2.3). Similarly, we obtain

$$W = C\#(C - B^T A^{-1} B). \tag{4.2}$$

Therefore

$$X\#\sigma_{p,q}(X) = \begin{pmatrix} A\#(A - BC^{-1}B^T) & 0 \\ 0 & C\#(C - B^T A^{-1} B) \end{pmatrix}.$$

Next, by (4.1) and (4.2) we have that

$$U^{-1/2} A U^{-1/2} = F(A, (A - BC^{-1}B^T)^{-1}),$$
$$W^{-1/2} C W^{-1/2} = F(C, (C - B^T A^{-1} B)^{-1})$$

and we also have

$$\begin{aligned} F(X, \sigma_{p,q}(X^{-1})) &= (X\#\sigma_{p,q}(X))^{-1/2} X (X\#\sigma_{p,q}(X))^{-1/2} \\ &= \begin{pmatrix} U^{-1/2} & 0 \\ 0 & W^{-1/2} \end{pmatrix} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} U^{-1/2} & 0 \\ 0 & W^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} U^{-1/2} A U^{-1/2} & U^{-1/2} B W^{-1/2} \\ W^{-1/2} B^T U^{-1/2} & W^{-1/2} C W^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} F(A, (A - BC^{-1}B^T)^{-1}) & G(X) \\ G(X)^T & F(C, (C - B^T A^{-1} B)^{-1}) \end{pmatrix}, \end{aligned}$$

where $G(X) := \left(A\#(A - BC^{-1}B^T)\right)^{-1/2} B \left(C\#(C - B^T A^{-1} B)\right)^{-1/2}$.

Finally, one can see directly that $AB = BC$ if and only if $X\sigma_{p,q}(X) = \sigma_{p,q}(X)X$ if and only if $X\sigma_{p,q}(X)^{-1} = \sigma_{p,q}(X)^{-1}X$. By Proposition 3.2 and Proposition 2.2 (3), it is equivalent to $X_{\Omega^-_{p,q}} = X\#\sigma_{p,q}(X^{-1}) = F(X, \sigma_{p,q}(X^{-1}))$. ■

Remark 4.3 The matrices $A - BC^{-1}B^T$ and $C - B^T A^{-1} B$ are known as the Schur complements of C and A in $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right)$, respectively (*cf.* [4]).

5 The Distance Function $d_{p,q}$

Recall that the distance function $d_{p,q}$ from the geodesic submanifold $\Omega^+_{p,q}$ is given by

$$d_{p,q}(X) = \inf_{A \in \Omega^+_{p,q}} \delta(X, A) = \delta(X, X\#\sigma_{p,q}(X)) = \frac{1}{2}\delta(X, \sigma_{p,q}(X))$$

and it is determined by the eigenvalues of $X^{-1}\sigma_{p,q}(X)$.

Proposition 5.1 *Let* $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \in \Omega_{p,q}$. *Then*

https://doi.org/10.4153/CJM-2004-035-5 Published online by Cambridge University Press

(1) $\det(X) = \det(C)\ \det(A - BC^{-1}B^T) = \det(A)\ \det(C - B^T A^{-1} B)$.
(2) *the eigenvalues of* $X^{-1}\sigma_{p,q}(X)$ *occur in reciprocal pairs: If* λ *is an eigenvalue of* $X^{-1}\sigma_{p,q}(X)$ *then so is* $1/\lambda$.
(3) *($p = 1$ or $q = 1$ cases):*

$$X = \sigma_{1,n-1}(X) \text{ if and only if } \det(X) = \Big(\sqrt{A\ \det(C)} \pm \sqrt{\det(C)\cdot BC^{-1}B^T}\Big)^2,$$
$$X = \sigma_{n-1,1}(X) \text{ if and only if } \det(X) = \Big(\sqrt{C\ \det(A)} \pm \sqrt{\det(A)\cdot B^T A^{-1}B}\Big)^2.$$

Proof (1) It follows from the fact that

$$\begin{pmatrix} I & -BC^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \begin{pmatrix} A - BC^{-1}B^T & 0 \\ B^T & C \end{pmatrix},$$
$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ B^T & C - B^T A^{-1} B \end{pmatrix}.$$

(2) Note that $X^{-1}\sigma_{p,q}(X)x = \lambda x$ if and only if $X^{-1}J_{p,q}XJ_{p,q}x = \lambda x$ if and only if $J_{p,q}x = \lambda X^{-1}J_{p,q}Xx$ if and only if $\frac{1}{\lambda}J_{p,q}x = (X^{-1}J_{p,q}XJ_{p,q})J_{p,q}x$. Therefore, λ is an eigenvalue of $X^{-1}\sigma_{p,q}(X)$ with eigenvector x if and only if $1/\lambda$ is an eigenvalue of $X^{-1}\sigma_{p,q}(X)$ with eigenvector $J_{p,q}x$.

(3) Suppose that $p = 1$ ($q = 1$ case is similar). In this case A is a positive real number. By (1), $\det(X) = A\ \det(C) - \det(C)BC^{-1}B^T$ and hence we immediately have that $X = \sigma_{1,n-1}(X)$ if and only if $B = 0$ if and only if $BC^{-1}B^T = 0$ if and only if $A\ \det(C) = \det(X)$. ■

Proposition 5.2 *Suppose that* $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \in \Omega_{p,q}$ *such that* $X \neq \sigma_{p,q}(X)$ *(that is,* $B \neq 0$*). Then*

(1) 1 *is not an eigenvalue of* $X^{-1}\sigma_{p,q}(X)$ *if and only if* $\operatorname{rank}(B) = \frac{n}{2}$. *This occurs only when* $p = q$ *and* B *is invertible. If either* $p \neq q$ *or* $p = q$ *and* B *is singular, then* 1 *is always an eigenvalue of* $X^{-1}\sigma_{p,q}(X)$ *with its algebraic multiplicity* $n - 2\operatorname{rank}(B)$.
(2) *For* $0 < \lambda \neq 1$, λ *is an eigenvalue of* $X^{-1}\sigma_{p,q}(X)$ *if and only if* $(1-\lambda/1+\lambda)^2$ *is an eigenvalue of the* $p \times p$ *matrix* $A^{-1}BC^{-1}B^T$ *if and only if* $(1-\lambda/1+\lambda)^2$ *is an eigenvalue of the* $q \times q$ *matrix* $C^{-1}B^TA^{-1}B$.
(3) *If* $p = 1$ *or* $q = 1$ *then the algebraic multiplicity of the eigenvalue* 1 *is* $n - 2$ *and the remaining two distinct eigenvalues are*

$$\frac{1}{\det(X)}\Big(\sqrt{A\ \det(C)} \pm \sqrt{\det(C)\cdot BC^{-1}B^T}\Big)^2 \quad (\text{if } p = 1)$$

and

$$\frac{1}{\det(X)}\Big(\sqrt{C\ \det(A)} \pm \sqrt{\det(A)\cdot B^TA^{-1}B}\Big)^2 \quad (\text{if } q = 1).$$

Proof Let $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \in \Omega_{p,q}$. First, we observe that

$$\begin{pmatrix} A & -B \\ -B^T & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

if and only if

$$(1-\lambda)Ax = (1+\lambda)By,\ (1+\lambda)B^Tx = (1-\lambda)Cy. \tag{5.1}$$

In particular, $\left(\begin{smallmatrix} A & -B \\ -B^T & C \end{smallmatrix}\right)\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right)\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ if and only if $By = 0$ and $B^Tx = 0$. Furthermore, the dimension of the subspace $\{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q : By = 0,\ B^Tx = 0\}$ is

$$(p - \operatorname{rank}(B)) + (q - \operatorname{rank}(B)) = n - 2\operatorname{rank}(B).$$

Therefore (1) follows.

(2) Let $0 < \lambda \neq 1$. Suppose that λ is an eigenvalue of $X^{-1}\sigma_{p,q}(X)$ and $(x, y)^T$ is an eigenvector of λ. Then since $\lambda \neq 1$ and since A, C are invertible, x and y are non-zero vectors by (5.1). Furthermore, we have that

$$x = \frac{1+\lambda}{1-\lambda}A^{-1}By = \Big(\frac{1+\lambda}{1-\lambda}\Big)^2 A^{-1}BC^{-1}B^Tx,$$
$$y = \frac{1+\lambda}{1-\lambda}C^{-1}B^Tx = \Big(\frac{1+\lambda}{1-\lambda}\Big)^2 C^{-1}B^TA^{-1}By$$

and hence $(1-\lambda/1+\lambda)^2$ is a common eigenvalue of $A^{-1}BC^{-1}B^T$ and $C^{-1}B^TA^{-1}B$. Conversely, suppose that $(1-\lambda/1+\lambda)^2$ is an eigenvalue of $A^{-1}BC^{-1}B^T$ and $x \in \mathbb{R}^p$ is an eigenvector. Set $y := (1+\lambda/1-\lambda)C^{-1}B^Tx \in \mathbb{R}^q$. Then $(x, y)^T$ satisfies (5.1) and hence λ is an eigenvalue of $X^{-1}\sigma_{p,q}(X)$. By the same argument, one can show that if $(1-\lambda/1+\lambda)^2$ is an eigenvalue of $C^{-1}B^TA^{-1}B$ then it is an eigenvalue of $X^{-1}\sigma_{p,q}(X)$.

(3) By (1), we may assume that $n > 2$. Suppose that $p = 1$ (and hence $q = n-1$). In this case, A is a positive real number. Then the algebraic multiplicity of the eigenvalue 1 is $n - 2\operatorname{rank}(B) = n - 2$ by (1). For positive real number k,

$$\Big(\frac{1-k}{1+k}\Big)^2 = \frac{BC^{-1}B^T}{A}$$

if and only if

$$(A - BC^{-1}B^T)k^2 - 2(A + BC^{-1}B^T)k + (A - BC^{-1}B^T) = 0.$$

Solving the quadratic equation for k, we find that k is one of the following

$$\frac{\Big(\sqrt{A\ \det(C)} \pm \sqrt{\det(C)\cdot BC^{-1}B^T}\Big)^2}{A\det(C) - \det(C)BC^{-1}B^T} = \frac{\Big(\sqrt{A\ \det(C)} \pm \sqrt{\det(C)\cdot BC^{-1}B^T}\Big)^2}{\det(X)}$$

and they are distinct and different from 1 by Proposition 5.1. Let k be one of the solutions (the other solution is $1/k$). Set $v = \left(\begin{smallmatrix} 1 \\ \frac{1+k}{1-k}C^{-1}B^T \end{smallmatrix}\right) \in \mathbb{R} \times \mathbb{R}^{n-1}$. Then

$$\begin{aligned}
\sigma_{1,n-1}(X)v &= \begin{pmatrix} A & -B \\ -B^t & C \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1+k}{1-k}C^{-1}B^T \end{pmatrix} = \begin{pmatrix} A - \frac{1+k}{1-k}BC^{-1}B^T \\ -B^t + \frac{1+k}{1-k}B^T \end{pmatrix} \\
&= \begin{pmatrix} k(A + \frac{1+k}{1-k}BC^{-1}B^T) \\ k(B^t + \frac{1+k}{1-k}B^T) \end{pmatrix} = k\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1+k}{1-k}C^{-1}B^T \end{pmatrix} \\
&= k(Xv),
\end{aligned}$$

and hence v is an eigenvector of $X^{-1}\sigma_{1,n-1}(X)$ with eigenvalue k. The proof for the case $q = 1$ is similar. ■

Recall that the "real" Cayley transformation $c\colon (0,\infty) \to (-1,1)$ is given by $c(x) = 1 - x/1 + x$. The map $d(x) = c(x)^2$ maps $(0,\infty)$ onto $[0,1)$. Note that $c(x)^2 = c(1/x)^2$. By Proposition 5.2 we have that for $X = \left(\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right) \neq \sigma_{p,q}(X)$,

$$\begin{aligned}
&d\big(\{0 < \lambda \neq 1 : \lambda \text{ is an eigenvalue of } X^{-1}\sigma_{p,q}(X)\}\big) \\
&\qquad\qquad = \{0 < \lambda < 1 : \lambda \text{ is an eigenvalue of } A^{-1}BC^{-1}B^T\}.
\end{aligned}$$

In fact, the matrix $A^{-1}BC^{-1}B^T$ is similar to the symmetric $p \times p$ matrix

$$A^{-1/2}BC^{-1}B^TA^{-1/2} = (A^{-1/2}BC^{-1/2})(A^{-1/2}BC^{-1/2})^T.$$

By Proposition 4.1, we have $A^{-1/2}BC^{-1}B^TA^{-1/2} < I$ and hence the eigenvalues of $A^{-1/2}BC^{-1}B^TA^{-1/2}$ lie in the interval $[0,1)$. If the rank of B is p then $BC^{-1}B^T$ is invertible and hence the eigenvalues of $A^{-1/2}BC^{-1}B^TA^{-1/2}$ must be in the interval $(0,1)$. Similar argument goes to the matrix $C^{-1/2}B^TA^{-1}BC^{-1/2}$ and hence we get natural functions Φ_p and Φ_q defined by

$$\begin{aligned}
\Phi_p\colon \Omega_{p,q} \to [0,I)_p, \quad &\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \mapsto A^{-1/2}BC^{-1}B^TA^{-1/2}, \\
\Psi_q\colon \Omega_{p,q} \to [0,I)_q, \quad &\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \mapsto C^{-1/2}B^TA^{-1}BC^{-1/2}
\end{aligned}$$

where $[0,I)_p$ denotes the Löwner order interval of $\mathrm{Sym}(p,\mathbb{R})^{++}$:

$$[0,I)_p = \{X \in \mathrm{Sym}(p,\mathbb{R}) : 0 \leq X < I\}.$$

Let $\mathbf{c}_l(W) = (I-W)(I+W)^{-1}$ be the Cayley transform on the cone $\mathrm{Sym}(l,\mathbb{R})^{++}$. Then it maps the open order interval $(0,I)_l := \{W \in \mathrm{Sym}(l,\mathbb{R}) : 0 < W < I\}$ onto itself. For $l \leq \min\{p,q\}$, we set

$$\mathcal{U}_l := \left\{ \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \Omega_{p,q} : \mathrm{rank}(B) = l \right\}.$$

Theorem 5.3 *We have*

(1) $\Phi_p^{-1}(0_p) = \Psi_q^{-1}(0_q) = \Omega_{p,q}^+$ *and if* $p \le q$ *(resp.* $q \le p$*), then* $\Phi_p^{-1}((0,I)_p) = \mathfrak{U}_p$ *and* $\Psi_q^{-1}((0,I)_q) = \mathfrak{U}_q$*, respectively.*

(2) *If* $p \le q$ *then* $\mathfrak{U}_p$ *is diffeomorphic to* $(0,I)_p \times F(p,q) \times \mathrm{Sym}(q,\mathbb{R})^{++}$*, where* $F(p,q)$ *denotes the Stiefel manifold of* p*-frames in* $\mathbb{R}^q$*, and the map* Φ_p *restricted to* $\mathfrak{U}_p$ *is surjective onto the open order interval* $(0,I)_p$*. The pre-image* $\Phi_p^{-1}(W)$ *of* $W \in (0,I)_p$ *is given by*

$$\Phi_p^{-1}(W) = \left\{ \begin{pmatrix} \left(W^{-1}\#(BC^{-1}B^T)\right)^2 & B \\ B^T & C \end{pmatrix} : \mathrm{rank}(B) = p,\ C \in \mathrm{Sym}(q,\mathbb{R})^{++} \right\}$$

which is diffeomorphic to $F(p,q) \times \mathrm{Sym}(q,\mathbb{R})^{++}$*. Furthermore,* λ *is an eigenvalue of* $X^{-1}\sigma_{p,q}(X)$ *if and only if* λ *is an eigenvalue of* $\mathbf{c}_p(\Phi_p(X)^{1/2})$ *if and only if* $\lambda = \frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}$ *for some eigenvalue* μ *of* $\Phi_p(X) = A^{-1/2}BC^{-1}B^TA^{-1/2}$*.*

(3) *If* $p = q$ *and* $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \mathfrak{U}_p$*, then* $\Phi_p(X)$ *and* $\Psi_p(X)$ *are similar.*

Proof (1) Straightforward.

(2) Suppose that $p \le q$. Then one can find a $p \times q$ matrix B such that $BB^T = I_p$. Thus for any given $p \times p$ positive definite matrix W such that $W < I$, the matrix $\begin{pmatrix} W^{-1} & B \\ B^T & I \end{pmatrix}$ is positive definite matrix by Proposition 4.1 and by the order reverting property of the inversion, and is mapped to W by Φ_p. This shows that Φ_p restricted to $\left\{ \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \Omega_{p,q} : \mathrm{rank}(B) = p \right\}$ is surjective onto the open order interval $(0,I)_p$. Let $W \in (0,I)_p$ and let $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \Omega_{p,q}$. Then $\Phi_p(X) = W$ if and only if $A^{-1/2}(BC^{-1}B^T)A^{-1/2} = W$ if and only if $A^{1/2} = W^{-1}\#(BC^{-1}B^T)$ by the Riccati Lemma (2.3). Therefore,

$$\Phi_p^{-1}(W) = \left\{ \begin{pmatrix} \left(W^{-1}\#(BC^{-1}B^T)\right)^2 & B \\ B^T & C \end{pmatrix} \in \Omega_{p,q} : \mathrm{rank}(B) = p \right\}$$

and is equal to

$$\left\{ \begin{pmatrix} \left(W^{-1}\#(BC^{-1}B^T)\right)^2 & B \\ B^T & C \end{pmatrix} : \mathrm{rank}(B) = p,\ C \in \mathrm{Sym}(q,\mathbb{R})^{++} \right\}$$

by Proposition 2.1 and Proposition 4.1. Therefore, $\Phi_p^{-1}(W)$ is diffeomorphic to $F(p,q) \times \mathrm{Sym}(q,\mathbb{R})^{++}$, where $F(p,q)$ is the Stiefel manifold of p-frames in $\mathbb{R}^q$. It is easy to check from (2.3) and Proposition 2.1 that the function

$$(0,I)_p \times F(p,q) \times \mathrm{Sym}(q,\mathbb{R})^{++} \to \mathfrak{U}_p$$

given by

$$(W,B,C) \mapsto \begin{pmatrix} \left(W^{-1}\#(BC^{-1}B^T)\right)^2 & B \\ B^T & C \end{pmatrix}$$

https://doi.org/10.4153/CJM-2004-035-5 Published online by Cambridge University Press

is a diffeomorphism with its inverse

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \mapsto (\Phi_p(X), B, C).$$

The remaining part of proof follows from Proposition 5.2 (2).

(3) It follows from that $(C^{-1}B^T)(A^{-1}BC^{-1}B^T)(C^{-1}B^T)^{-1} = C^{-1}B^TA^{-1}B$. ■

Remark 5.4 If $q \le p$ then $\mathcal{U}_q$ is diffeomorphic to $(0, I)_q \times F(q, p) \times \mathrm{Sym}(p, \mathbb{R})^{++}$ and the map Φ_q restricted to $\mathcal{U}_q$ is surjective onto $(0, I)_q$. The pre-image $\Psi_q^{-1}(W)$ of $W \in (0, I)_q$ is given by

$$\Psi_q^{-1}(W) = \left\{ \begin{pmatrix} A & B \\ B^T & \left(W^{-1}\#(B^TA^{-1}B)\right)^2 \end{pmatrix} : \mathrm{rank}(B) = q,\ A \in \mathrm{Sym}(p, \mathbb{R})^{++} \right\}$$

which is diffeomorphic to $\mathrm{Sym}(p, \mathbb{R})^{++} \times F(q, p)$.

Theorem 5.5 *If either p or q is less than equal to 2 then the distance function $d_{p,q}$ induced by the Jordan automorphism $\sigma_{p,q}$ is completely determined: If $p \le 2$ (resp., $q \le 2$) then*

$$d_{p,q}(X) = \frac{1}{\sqrt{2}}\delta_p(I, \mathbf{c}_p(\Phi_p(X)^{1/2})) = \frac{1}{\sqrt{2}}\delta_p(I_p + \Phi_p(X)^{1/2}, I_p - \Phi_p(X)^{1/2})$$

and

$$d_{p,q}(X) = \frac{1}{\sqrt{2}}\delta_q(I, \mathbf{c}_q(\Phi_q(X)^{1/2})) = \frac{1}{\sqrt{2}}\delta_q(I_q + \Psi_q(X)^{1/2}, I_q - \Psi_q(X)^{1/2})$$

respectively. Here δ_p and δ_q stand for the Riemannian metric distances on $\mathrm{Sym}(p, \mathbb{R})^{++}$ *and* $\mathrm{Sym}(q, \mathbb{R})^{++}$ *respectively.*

Proof If $X = \sigma_{p,q}(X)$ then $\Phi_p(X)$ and $\Psi_q(X)$ are zero matrices and $d_{p,q}(X) = 0$, and hence the statement is true. We assume that $X \ne \sigma_{p,q}(X)$. Note that $d_{p,q}(X) = \delta(X, X\#\sigma_{p,q}(X)) = \frac{1}{2}\delta(X, \sigma(X))$ and that for $W \in \mathrm{Sym}(l, \mathbb{R})^{++}$,

$$\delta_l(I, \mathbf{c}_l(W)) = \delta_l(I, (I - W)(I + W)^{-1}) = \delta_l(I + W, I - W)$$

by the invariance of the metric under the congruence transformations. Therefore

$$\delta_p(I, \mathbf{c}_p(\Phi_p(X)^{1/2})) = \delta_p(I + \Phi_p(X)^{1/2}, I - \Phi_p(X)^{1/2}) = \Big(\sum_{i=1}^{p} \log^2 \frac{1 + \sqrt{\mu_i}}{1 - \sqrt{\mu_i}}\Big)^{1/2}$$

where μ_i are eigenvalues of $\Phi_p(X)$. By the same argument, we have

$$\delta_q(I, \mathbf{c}_q(\Psi_q(X)^{1/2})) = \delta_q(I + \Psi_q(X)^{1/2}, I - \Psi_q(X)^{1/2}) = \Big(\sum_{i=1}^{q} \log^2 \frac{1 + \sqrt{\mu_i}}{1 - \sqrt{\mu_i}}\Big)^{1/2}$$

where μ_i are eigenvalues of $\Psi_q(X)$.

Case $p = 1$ ***or*** $q = 1$ The eigenvalues of $X^{-1}\sigma_{1,n-1}(X)$ are

$$\Big\{\underbrace{1,\cdots,1}_{n-2},\ \frac{\big(\sqrt{A\ \det(C)} \pm \sqrt{\det(C)\cdot BC^{-1}B^T}\big)^2}{\det(X)}\Big\},$$

by Proposition 5.2. Therefore

$$\begin{aligned} d_{1,n-1}(X) &= \frac{1}{\sqrt{2}}\Big(\log\frac{(\sqrt{A\ \det(C)} + \sqrt{\det(C)\cdot BC^{-1}B^T})^2}{\det(X)}\Big) \\ &= \frac{1}{\sqrt{2}}\Big(\log\frac{\sqrt{A} + \sqrt{BC^{-1}B^T}}{\sqrt{A} - \sqrt{BC^{-1}B^T}}\Big) = \frac{1}{\sqrt{2}}\Big(\log\frac{1 + \sqrt{A^{-1}BC^{-1}B^T}}{1 - \sqrt{A^{-1}BC^{-1}B^T}}\Big) \\ &= \frac{1}{\sqrt{2}}\delta_1\Big(1 + \Phi_1(X)^{1/2}, 1 - \Phi_1(X)^{1/2}\Big). \end{aligned}$$

Similarly, we have $d_{n-1,1}(X) = \frac{1}{\sqrt{2}}\delta_1\big(1 + \Psi_1(X)^{1/2}, 1 - \Psi_1(X)^{1/2}\big)$.

Case $p = 2$ We consider the rank of B. Suppose that $\operatorname{rank}(B) = 1$. Then by Proposition 5.2 (1) there exist only two distinct eigenvalues of $X^{-1}\sigma_{p,q}(X)$, say λ and β which are different from 1. By Proposition 5.1, $\beta = 1/\lambda$. Hence

$$d_{p,q}(X) = \frac{1}{2}\big(\log^2\lambda + \log^2\beta\big)^{1/2} = \frac{1}{\sqrt{2}}|\log\lambda|$$

which (by Proposition 5.2 (2)) is equal to

$$\frac{1}{\sqrt{2}}\Big(\log\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}}\Big) = \frac{1}{\sqrt{2}}\delta_p(I_p + \Phi_p(X)^{1/2}, I_p - \Phi_p(X)^{1/2})$$

where μ denotes the unique non-zero eigenvalue of $A^{-1/2}BC^{-1}B^TA^{-1/2} = \Phi_p(X)$.

Next, suppose that $\operatorname{rank}(B) = 2$. Let μ and ν be the eigenvalues of $\Phi_p(X)$. Then they must be positive. If $\mu = \nu$ then the four eigenvalues of $X^{-1}\sigma_{p,q}(X)$ distinct from 1 are one of the form $\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}}, \frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}$ by Proposition 5.2 (2) and hence

$$\begin{aligned} d_{p,q}(X) &= \frac{1}{2}\delta(X, \sigma_{p,q}(X)) = \frac{1}{2}\Big(4\log^2\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}}\Big)^{1/2} = \frac{1}{\sqrt{2}}\Big(2\log^2\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}}\Big)^{1/2} \\ &= \frac{1}{\sqrt{2}}\delta_p(I_p + \Phi_p(X)^{1/2}, I_p - \Phi_p(X)^{1/2}). \end{aligned}$$

If $\mu \neq \nu$, then the eigenvalues of $A^{-1/2}BC^{-1}B^TA^{-1/2}$ distinct from 1 are

$$\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}}, \frac{1-\sqrt{\mu}}{1+\sqrt{\mu}}, \frac{1+\sqrt{\nu}}{1-\sqrt{\nu}}, \frac{1-\sqrt{\nu}}{1+\sqrt{\nu}}$$

and thus

$$\begin{aligned} d_{p,q}(X) &= \frac{1}{2}\Big(2\log^2\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}} + 2\log^2\frac{1+\sqrt{\nu}}{1-\sqrt{\nu}}\Big)^{1/2} \\ &= \frac{1}{\sqrt{2}}\Big(\log^2\frac{1+\sqrt{\mu}}{1-\sqrt{\mu}} + \log^2\frac{1+\sqrt{\nu}}{1-\sqrt{\nu}}\Big)^{1/2} \\ &= \frac{1}{\sqrt{2}}\delta_p(I_p + \Phi_p(X)^{1/2}, I_p - \Phi_p(X)^{1/2}). \end{aligned}$$

Case $q = 2$ Similar to the case $p = 2$ from Proposition 5.2 and Theorem 5.3. ∎

Remark 5.6 It remains unanswered here whether the formula given in Theorem 5.5 holds true for arbitrary $1 \leq p < n$ or not. Definitely it depends on the structure of eigenvalues of $X^{-1}\sigma_{p,q}(X)$ or

$$X^{-1/2}(UXU)X^{-1/2} = (X^{-1/2}UX^{1/2})(X^{-1/2}UX^{1/2})^T,$$

where $U = U^T = U^{-1}$ involutive orthogonal matrix.

6 Global Tubular Neighborhood Theorem and $AB = BC$ Criterion

The following result shows in particular that the positive definite cone Ω admits the geometric and spectral geometric mean coordinates depending on Jordan involutions.

Theorem 6.1 *Let σ be an involutive Jordan automorphism of V. Then the map $T\colon \Omega_\sigma^+ \times \Omega_\sigma^- \to \Omega$, defined by $(A, B) \mapsto A^{1/2}BA^{1/2}$ is a differential diffeomorphism with its inverse $T^{-1}(X) = \big(X\#\sigma(X), F(X, \sigma(X^{-1}))\big)$.*

Proof Let $A_1, A_2 \in \Omega_\sigma^+$ and let $B_1, B_2 \in \Omega_\sigma^-$ such that $A_1B_1A_1 = A_2B_2A_2$. Then $B_2 = A_2^{-1}A_1B_1A_1A_2^{-1}$ or $B_2^{-1} = A_2A_1^{-1}B_1^{-1}A_1^{-1}A_2$. Applying the map σ, we have

$$\begin{aligned} B_2^{-1} &= \sigma(B_2) \\ &= \sigma(A_2^{-1}A_1B_1A_1A_2^{-1}) = A_2^{-1}\sigma(A_1B_1^{-1}A_1)A_2^{-1} \\ &= A_2^{-1}A_1\sigma(B_1)A_1A_2^{-1} = A_2^{-1}A_1B_1^{-1}A_1A_2^{-1}. \end{aligned}$$

Thus $A_2A_1^{-1}B_1^{-1}A_1^{-1}A_2 = B_2^{-1} = A_2^{-1}A_1B_1^{-1}A_1A_2^{-1}$ and hence

$$B_1^{-1} = (A_1A_2^{-2}A_1)B_1^{-1}(A_1A_2^{-2}A_1).$$

This implies that $I = B\#B^{-1} = A_1A_2^{-2}A_1$ by the Riccati Lemma (2.3) and hence we have $A_1 = A_2$ and so $B_1 = B_2$. This shows that the map T is injective.

Let $X \in \Omega$, and let $A := X\#\sigma(X)$ and $B := A^{-1/2}XA^{-1/2} = F(X, \sigma(X^{-1}))$. Then $A \in \Omega_\sigma^+$ and $B \in \Omega_\sigma^-$ by Proposition 3.2. By Proposition 2.2,

$$B = A^{-1/2}XA^{-1/2} = (X\#\sigma(X))^{-1/2}X(X\#\sigma(X))^{-1/2} = F(X, \sigma(X^{-1}))$$

and therefore $X = A^{1/2}BA^{1/2} = T(A, B)$. This implies that the map T is surjective. ■

Remark 6.2 One may see in similar way that the map $S\colon \Omega_\sigma^- \times \Omega_\sigma^+ \to \Omega$ defined by $(A, B) \mapsto A^{1/2}BA^{1/2}$ is a differential diffeomorphism with its inverse given by

$$S^{-1}(X) = (X\#\sigma(X^{-1}), F(X, \sigma(X))).$$

Indeed, it is well-defined by Proposition 3.2.

We explicitly describe the global tubular coordinates for $n = 2$. In this case the Jordan involution σ is given by $\left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} a & -b \\ -b & c \end{smallmatrix}\right)$ and the associated geodesic submanifolds are

$$\Omega_\sigma^+ = \left\{ \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix} : a, c > 0 \right\},$$
$$\Omega_\sigma^- = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a > 0, a^2 - b^2 = 1 \right\}.$$

Let $X = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right) \in \mathrm{Sym}(2, \mathbb{R})^+$. From Theorem 4.2, one sees that

$$T^{-1}(X) = \left(\sqrt{\det(X)} \begin{pmatrix} \sqrt{a/c} & 0 \\ 0 & \sqrt{c/a} \end{pmatrix}, \ \frac{1}{\sqrt{\det(X)}} \begin{pmatrix} \sqrt{ac} & \sqrt{b^2} \\ \sqrt{b^2} & \sqrt{ac} \end{pmatrix} \right)$$

and

$$\begin{aligned} S^{-1}(X) &= \left(\begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}, \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \right) \\ &= \left(\frac{1}{\sqrt{(a+c)^2 - 4b^2}} \begin{pmatrix} a+c & 2b \\ 2b & a+c \end{pmatrix}, \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \right) \end{aligned}$$

where

$$t = \tanh^{-1}(\frac{2b}{a+c}) = \log \sqrt{\frac{a+c+2b}{a+c-2b}},$$
$$x = \frac{a - c + \sqrt{(a+c)^2 - 4b^2}}{2},$$
$$y = \frac{c - a + \sqrt{(a+c)^2 - 4b^2}}{2}.$$

The distance function is given by

$$d(X) = \frac{1}{\sqrt{2}} \log \frac{ac + b^2}{ac - b^2}.$$

It is worth noting that $\frac{ac+b^2}{ac-b^2} = \frac{1}{2}\mathrm{tr}(X^{-1}\sigma(X))$.

By applying Theorem 6.1 for the involution $\sigma_{p,q}$ and Theorem 4.2, we have the following.

Corollary 6.3 *Let* $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \in \Omega_{p,q}$. *Then the following statements are equivalent:*

(1) $AB = BC$,
(2) $X_{\Omega^-_{p,q}} = F(X, \sigma_{p,q}(X^{-1}))$,
(3) $T^{-1}(X) = (X_{\Omega^+_{p,q}}, X_{\Omega^-_{p,q}})$,
(4) $S^{-1}(X) = (X_{\Omega^-_{p,q}}, X_{\Omega^+_{p,q}})$; *X has the coordinates of best approximants,*
(5) $X\sigma_{p,q}(X) = \sigma_{p,q}(X)X$; *X and* $\sigma_{p,q}(X)$ *commute,*
(6) $X\#\sigma_{p,q}(X^{-1}) = F(X, \sigma_{p,q}(X^{-1}))$,
(7) $X\#\sigma_{p,q}(X) = F(X, \sigma_{p,q}(X))$.

References

[1] W. Ballmann, *Lectures on spaces of nonpositive curvature*. Birkhäuser, Berlin, 1995.
[2] J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization*. CMS Books in Mathematics, Springer-Verlag, New York, 2000.
[3] M. Fiedler and V. Pták, *A new positive definite geometric mean of two positive definite matrices*. Linear Algebra Appl, **251**(1997), 1–20.
[4] R. Horn and C. Johnson, *Matrix analysis*. Cambridge University Press, Cambridge, 1985.
[5] F. Kubo and T. Ando, *Means of positive linear operators*. Math. Ann. **246**(1980), 205–224.
[6] S. Lang, *Fundamentals of differential geometry*. Graduate Texts in Mathematics 191, Springer-Verlag, New York, 1999.
[7] J. D. Lawson and Y. Lim, *The geometric mean, matrices, metrics, and more*. Amer. Math. Monthly **108**(2001), 797–812.
[8] A. S. Lewis, *Group invariance and convex matrix analysis*. SIAM J. Matrix Anal. Appl **17**(1996), 927–949.
[9] ———, *Nonsmooth analysis of eigenvalues*. Math. Program. **84**(1999), 1–24.
[10] A. Ohara, N. Suda and S. Amari, *Dualistic differential geometry of positive definite matrices and its applications to related problems*. Linear Algebra Appl. **247**(1996), 31–53.
[11] A. Ohara, *Information geometric analysis of an interior-point method for semidefinite programming*. Proceedings of Geometry in Present Day Science. World Scientific, 1999, pp. 49–74.

Department of Mathematics
Kyungpook National University
Taegu 702-701
Korea
e-mail: ylim@knu.ac.kr