

Article

Not peer-reviewed version

Chordal Metric Formula Between Generalized Singular Values of Grassmann Matrix Pairs by Riemannian Optimization Models

[Yujie Wang](#), Wen Kang, [Lei Zhu](#)*

Posted Date: 25 February 2025

doi: 10.20944/preprints202502.1991.v1

Keywords: riemannian optimization models; riemannian manifolds; grassman matrix pair; Generalized singular values



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Chordal Metric Formula Between Generalized Singular Values of Grassmann Matrix Pairs by Riemannian Optimization Models

Yujie Wang ¹ , Wen Kang ² and Lei Zhu ^{3,*}

¹ School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

² School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

³ College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

* Correspondence: zhulei@njau.edu.cn

Abstract: In this paper, we present a new explicit formula for the sum of the chordal distance between the generalized singular values of Grassmann matrix pairs, based on Riemannian optimization models. The new formulas involve small-size unitary matrices and real orthogonal matrices derived from Riemannian optimization models. We then apply Newton's method on Riemannian manifolds to efficiently solve the variable matrices involved. The new explicit formulas provide significant improvements over the existing theoretical and computational upper bounds.

Keywords: riemannian optimization models; riemannian manifolds; grassman matrix pair; Generalized singular values

0. Introduction

The generalized singular value decomposition (GSVD) is useful in matrix computation problems and practical applications, such as the general Gauss-Markov linear model, the Kronecker canonical form of a general matrix pencil, generalized total least squares problem, gene data analysis and so on [1–6,8–12]. In particular, GSVD and generalized singular values (GSVs) can provide a comparative mathematical framework for two genome-scale expression data sets as seen in [1–4,22].

0.1. GSVD and Its Applications

More precisely, GSVD induces a linear transformation of two data sets from the two genes $\times$ arrays spaces to two reduced and diagonalized “genelets” $\times$ “arraylets” spaces. A single microarray probes the relative expression levels of p genes in a single sample. A series of q_1 arrays probes the genome-scale expression levels in q_1 different samples, i.e., under q_1 different experimental conditions. Let matrix A of size p -genes $\times$ q_1 -arrays, denote the full expression data, whose k -th row is the expression of the k -th gene across the different samples that correspond to the different arrays. Let the matrix B of size t -genes $\times$ q_2 -arrays, denote the relative expression levels of t genes under $q_2 = q_1 = q < \max\{p, t\}$ experimental conditions that correspond one to one to the q_1 conditions underlying A . GSVD induces simultaneous linear transformation of two expression data sets A and B from two p -genes $\times$ q -arrays and t -genes $\times$ q -arrays spaces to two reduced q -genelets $\times$ q -arraylets spaces.

Let A and B have the following GSVD: $A = U\Sigma_A R^{-1}$, $B = V\Sigma_B R^{-1}$. In these spaces the data is represented by the diagonal non-negative matrices Σ_A and Σ_B and denote $\langle k|\Sigma_A|l \rangle = \varepsilon_{1,l}\delta_{kl}$, $\langle k|\Sigma_B|l \rangle = \varepsilon_{2,l}\delta_{kl}$ for all $1 \leq k, l \leq q$, where $\delta_{kl} = 1$ if $k = l$ and $\delta_{kl} = 0$ if $k \neq l$. By $\langle k|\Sigma_A$ we denote the k -th row of the matrix Σ_A , which lists the expression of the k -th gene across the different samples that correspond to the different arrays. By $\Sigma_A|l \rangle$ we denote the l -th column of the matrix Σ_A , which lists the genome-scale expression measured by the l -th array. The antisymmetric angular distance between the data sets [3], $\theta_l = \arctan(\varepsilon_{1,l}/\varepsilon_{2,l}) - \pi/4$ indicates the relative significance

of the l -th genelet, i.e., its significance in the first data set relative to that in the second in terms of the ratio of the expression information captured by this genelet in the first data set to that in the second. An angular distance of 0 indicates a genelet of equal significance in both data sets, with $\varepsilon_{1,l} = \varepsilon_{2,l}$.

Alter et al. [1] made comparative analysis of gene mRNA expression data sets by transforming those into GSVs and analyzed the degree of change for reduced and diagonalized “genelets” $\times$ “arraylets” spaces in gene expression levels from samples under different experimental conditions. This behavior is significant for guiding further experimental research. Motivated by practical applications proximity and range size between GSVs, Xu et al. [22] considered the upper bound of sum of chordal metric between generalized singular values of Grassmann matrix pairs and applied this result in comparative analysis of gene mRNA expression data.

0.2. Riemannian Optimization

Optimization on Riemannian manifolds is called Riemannian optimization. Unconstrained optimization methods on the Euclidean space, such as the steepest descent, the conjugate gradient, and Newton’s methods are generalized to those on a Riemannian manifold. The term “optimization on manifolds” is applied to algorithms that preserve manifold constraints during iterations. Generally speaking, preserving constraints has advantages in many situations, for example, when the cost functions are undefined or of little use outside the feasible region and when the algorithm is terminated prior to convergence yet a feasible solution is required. Theories and methods for optimization on manifolds date back to the 1970s, and algorithms for specific problems, e.g., various eigenvalue problems, appeared even earlier. Most of the general-purpose algorithms, however, did not appear until the 1990s [30]. While it is trivial to generate trial points in $\mathbb{R}^n$ along straight search lines, it is not as easy to do so in the curved manifold. A natural choice is the geodesic, which is the analog of straightline: it has the shortest length between two points in the manifold. For more details we may refer readers to [29–32].

In this paper we provide explicit formula of sum of chordal metric between generalized singular values of Grassmann matrix pairs with Riemannian optimization models, which absolutely improves the existing upper bounds in [22] in theory. The new formula involves two small-size unitary variable matrices from Riemannian optimization models. Then we use Newton methods on Riemannian manifolds to solve the involved Riemannian optimization models for deducing these two variable matrices solutions efficiently. The contributions of this paper are summarized as follows.

- On theoretic aspect we provide a new explicit formula of sum of chordal metric between generalized singular values of Grassmann matrix pairs with Riemannian optimization models instead of the existing upper bounds.
- On algorithm aspect the new expression formula involves only two small-size unitary variable matrices from Riemannian optimization models and Newton methods on Riemannian manifolds are given, which reduces the required computational cost.

0.3. Notations

Throughout this paper, by $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^n$, $\mathbb{C}^{m \times n}$, $\mathbb{U}_n$ and $\mathbb{O}\mathbb{R}_n$ we denote the sets of real numbers, complex numbers, real vectors of order n , $m \times n$ matrices with entries in $\mathbb{C}$, and $n \times n$ unitary matrices and real orthogonal matrices, respectively. The symbols $|\cdot|$ and $\Re(\cdot)$ stand for absolute value and real part of a complex number, respectively. The symbols I_n and $O_{m \times n}$ represent the identity matrix of order n and $m \times n$ zero matrix, respectively. For a square matrix $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, by A^T , A^H , A^{-1} , and $\text{tr}(A)$ we denote the transpose, conjugate transpose, inverse, and trace of matrix A , respectively. We use $\|\cdot\|_2$ to denote the spectral norm of a matrix. The singular value set of A is denoted by $\sigma(A)$. The complex number $a + bi$ ($a, b \in \mathbb{R}$) and the real number pair (a, b) can establish a one-to-one correspondence with a point on the coordinate plane, so that the coordinate plane that establishes a one-to-one correspondence with the whole complex number is called Gauss plane, which is denoted by $\mathfrak{P}(1, 1)$. For given matrices $A, B \in \mathbb{C}^{n \times n}$, $A < (\leq) B$ means $B - A$ is a positive (semi-)definite

matrix. As in [9,15–17], we also use the chordal metric on the Riemann sphere to measure the difference between two GSVs. The pair (α, β) can be regarded as a point in Gauss plane $\mathfrak{P}(1, 1)$. For two points $(\alpha, \beta) \neq (0, 0), (\gamma, \delta) \neq (0, 0)$ in Gauss plane $\mathfrak{P}(1, 1)$, their chordal distance is defined by

$$\rho((\alpha, \beta), (\gamma, \delta)) = \frac{|\alpha\delta - \beta\gamma|}{\sqrt{(|\alpha|^2 + |\beta|^2)(|\gamma|^2 + |\delta|^2)}} \quad (1)$$

to measure the difference between two points (α, β) and (γ, δ) . Next, we outline some definitions related to GSVD. Readers may refer to [7,9,16] for these definitions.

Definition 1. The matrix pair $\{A, B\}$ with $A, B \in \mathbb{C}^{n \times n}$ is said to be regular if there exists $\zeta \in \mathbb{C}$ such that $\det(A + \zeta B) \neq 0$. Denote by

$$\mathcal{G}_{1,2} = \{(\alpha, \beta) \neq (0, 0) : \alpha, \beta \in \mathbb{C}\}.$$

The pair $(\alpha, \beta) \in \mathcal{G}_{1,2}$ is called a generalized eigenvalue of a regular pair $\{A, B\}$, if $\det(\beta A - \alpha B) = 0$.

The set of generalized eigenvalues of the regular pair $\{A, B\}$ is denoted by $\lambda(A, B)$.

Definition 2. Let $A \in \mathbb{C}^{m \times n}$ and $B \in \mathbb{C}^{p \times n}$. A matrix pair $\{A, B\}$ is an (m, p, n) -Grassman matrix pair (GMP) if $\text{rank}(A^T, B^T)^T = n$.

For GSVD, it has several formulations in the literature. In this paper we adopt the following form as in [9,16,19].

Definition 3. Let $\{A, B\}$ be an (m, p, n) -GMP. Then there exist unitary matrices $U \in \mathbb{C}^{m \times m}, V \in \mathbb{C}^{p \times p}$, and a nonsingular matrix $R \in \mathbb{C}^{n \times n}$ such that

$$U^H A R = \Sigma_A, \quad V^H B R = \Sigma_B, \quad (2)$$

$$\Sigma_A = \begin{pmatrix} \Lambda & \\ & O_{(m-r-s) \times (n-r-s)} \end{pmatrix}, \quad \Sigma_B = \begin{pmatrix} O_{(p+r-n) \times r} & \\ & \Omega \end{pmatrix}, \quad (3)$$

where $O_{(m-r-s) \times (n-r-s)}$ and $O_{(p+r-n) \times r}$ are zero matrices, and

$$\Lambda = \text{diag}(\alpha_1, \dots, \alpha_{r+s}), \quad \Omega = \text{diag}(\beta_{r+1}, \dots, \beta_n),$$

with

$$1 = \alpha_1 = \dots = \alpha_r > \alpha_{r+1} \geq \dots \geq \alpha_{r+s} > \alpha_{r+s+1} = \dots = \alpha_n = 0,$$

$$0 = \beta_1 = \dots = \beta_r < \beta_{r+1} \leq \dots \leq \beta_{r+s} < \beta_{r+s+1} = \dots = \beta_n = 1,$$

and

$$\alpha_i^2 + \beta_i^2 = 1, \quad 1 \leq i \leq n.$$

We note that by Definition 3, given another (m, p, n) -GMP $\{\tilde{A}, \tilde{B}\}$, there also exist unitary matrices $\tilde{U} \in \mathbb{C}^{m \times m}, \tilde{V} \in \mathbb{C}^{p \times p}$, nonsingular $\tilde{R} \in \mathbb{C}^{n \times n}$, and nonnegative diagonal matrices $\Sigma_{\tilde{A}} \in \mathbb{C}^{m \times n}, \Sigma_{\tilde{B}} \in \mathbb{C}^{p \times n}$ for the GSVD of $\{\tilde{A}, \tilde{B}\}$.

For GSVD of a real matrix pair $\{A, B\}$ we adopt the following definition form as in [9,16,19].

Definition 4. Let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times n}$ have $\text{rank}(A^T, B^T) = n$. Then there exist two orthogonal matrices $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{p \times p}$ and a nonsingular matrix $R \in \mathbb{R}^{n \times n}$ such that

$$U^T A R = \Sigma_A, \quad V^T B R = \Sigma_B, \quad (4)$$

$$\Sigma_A = \begin{pmatrix} \Lambda & \\ & O_{(m-r-s) \times (n-r-s)} \end{pmatrix}, \Sigma_B = \begin{pmatrix} O_{(p+r-n) \times r} & \\ & \Omega \end{pmatrix}, \quad (5)$$

where $\Lambda = \text{diag}(\alpha_1, \dots, \alpha_{r+s})$ and $\Omega = \text{diag}(\beta_{r+1}, \dots, \beta_n)$ with

$$\begin{aligned} 1 &= \alpha_1 = \dots = \alpha_r > \alpha_{r+1} \geq \dots \geq \alpha_{r+s} > \alpha_{r+s+1} = \dots = \alpha_n = 0, \\ 0 &= \beta_1 = \dots = \beta_r < \beta_{r+1} \leq \dots \leq \beta_{r+s} < \beta_{r+s+1} = \dots = \beta_n = 1, \end{aligned}$$

and $\alpha_i^2 + \beta_i^2 = 1$ for $1 \leq i \leq n$.

0.4. Overview of Existing Results about Chordal Metric Between GSVs of GMP

There have been a large amount of papers in various topics related to GSVD and its applications, e.g., see [5,6,8–12,16–19,23]. One of interesting and important topics is about chordal metric between GSVs of Grassmann matrix pairs (GMPs), which was first analyzed by Sun [16] in 1983. In [16], Sun presented the Hoffman-Wielandt theorem for GSVs of GMPs and gave bounds on perturbations of GSVs, which generalized several well-known results for the standard singular value problem. Since then, Paige and Saunders [15] in 1981 gave a bound for GSV variations. Li [9] in 1993 presented several perturbation bounds of GSVs and its associated subspace. Recently Xu et al. [22] considered sharp upper bounds of the chordal metric between generalized singular values of Grassmann matrix pairs. The following results provide perturbation bounds based on the chordal metric of GSVs.

Theorem 1 (The Hoffman-Wielandt type theorem for GSV of GMP[16]). *Let $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ be two (m, p, n) -GMPs, $\sigma\{A, B\} = \{(\alpha_i, \beta_i)\}_{i=1}^n$ and $\sigma\{\tilde{A}, \tilde{B}\} = \{(\tilde{\alpha}_i, \tilde{\beta}_i)\}_{i=1}^n$. Then*

$$\prod_{i=1}^n \left(1 - \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))\right) \geq 1 - d^2(Y, \tilde{Y}), \quad (6)$$

$$\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) \leq n \left(1 - \sqrt[n]{1 - d^2(Y, \tilde{Y})}\right), \quad (7)$$

where

$$Y = \begin{pmatrix} A \\ B \end{pmatrix}, \tilde{Y} = \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix},$$

$$\rho((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) \text{ is given by (1) and } d(Y, \tilde{Y}) = \left\{1 - \frac{|\det(Y^H \tilde{Y})|^2}{\det(Y^H Y) \det(\tilde{Y}^H \tilde{Y})}\right\}^{\frac{1}{2}}.$$

Theorem 2 (Explicit expression and sharp upper bound[22]). *Let $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ be two (m, p, n) -GMPs, $\sigma\{A, B\} = \{(\alpha_i, \beta_i)\}_{i=1}^n$ and $\sigma\{\tilde{A}, \tilde{B}\} = \{(\tilde{\alpha}_i, \tilde{\beta}_i)\}_{i=1}^n$. Then*

$$\begin{aligned} \prod_{i=1}^n \left(1 - \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))\right) &\equiv \max_{\Psi_1 \in \mathbb{U}_m, \Psi_2 \in \mathbb{U}_p} (1 - d^2(Y, \tilde{Y}_{\Psi})) \\ &= 1 - d^2(Y, \tilde{Y}_{\Psi_0}), \end{aligned} \quad (8)$$

$$\begin{aligned} \sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) &\leq n \left(1 - \max_{\Psi_1 \in \mathbb{U}_m, \Psi_2 \in \mathbb{U}_p} \sqrt[n]{1 - d^2(Y, \tilde{Y}_{\Psi})}\right) \\ &= n \left(1 - \sqrt[n]{1 - d^2(Y, \tilde{Y}_{\Psi_0})}\right), \end{aligned} \quad (9)$$

where

$$Y = \begin{pmatrix} A \\ B \end{pmatrix}, \tilde{Y} = \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix}, \tilde{Y}_{\Psi} = \Psi \tilde{Y}, \tilde{Y}_{\Psi_0} = \Psi_0 \tilde{Y},$$

$$\Psi = \begin{pmatrix} \Psi_1 & O_{m \times p} \\ O_{p \times m} & \Psi_2 \end{pmatrix}, \quad \Psi_0 = \begin{pmatrix} U\tilde{U}^H & O_{m \times p} \\ O_{p \times m} & V\tilde{V}^H \end{pmatrix},$$

and $\Psi_1 \in \mathbb{U}_m, \Psi_2 \in \mathbb{U}_p; U, \tilde{U}, V, \tilde{V}$ are the unitary matrices in GSVD of $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ (refer to (2)-(3)). The inequality (9) is an equality if and only if $\rho((\alpha_1, \beta_1), (\tilde{\alpha}_1, \tilde{\beta}_1)) = \rho((\alpha_2, \beta_2), (\tilde{\alpha}_2, \tilde{\beta}_2)) = \dots = \rho((\alpha_n, \beta_n), (\tilde{\alpha}_n, \tilde{\beta}_n))$.

In this paper we continue along this line and further consider explicit expression of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ instead of the existing upper bound in (9).

0.5. Organization

The rest of this paper is organized as follows. Section 1 is the preliminary. In Section 2, we present an explicit expression of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ by Riemannian optimization models and the corresponding numerical methods. Finally, in Section 3 concluding remarks are drawn.

1. Preliminaries

In this section we propose some useful lemmas in order to deduce the main results.

Lemma 1. [21] Let $A_1, \dots, A_m \in \mathbb{C}^{n \times n}$ with $\sigma_1(A_j) \geq \sigma_2(A_j) \geq \dots \geq \sigma_n(A_j) \geq 0$ the singular values arranged in decreasing order and $c \in \mathbb{R}, j = 1, \dots, m$. We have the following extreme values:

$$\max_{U_1, \dots, U_m \in \mathbb{U}_n} \left| \text{tr}(cI_n \pm \prod_{j=1}^m U_j A_j) \right| = n |c| + \sum_{i=1}^n \prod_{j=1}^m \sigma_i(A_j).$$

Let (A, B) and $(\tilde{A}, \tilde{B})$ be two (m, p, n) GMPs and

$$\begin{aligned} \mathbf{S}_1 &= \begin{cases} (Z^H Z)^{-1/2} A^H, & m \leq n, m \leq p; \\ A(Z^H Z)^{-1/2}, & n \leq m, m \leq p; \\ (Z^H Z)^{-1/2} B^H, & p \leq n, p \leq m; \\ B(Z^H Z)^{-1/2}, & n \leq p, p \leq m; \end{cases} \\ \mathbf{S}_2 &= \begin{cases} (\tilde{Z}^H \tilde{Z})^{-1/2} \tilde{A}^H, & m \leq n, m \leq p; \\ \tilde{A}(\tilde{Z}^H \tilde{Z})^{-1/2}, & n \leq m, m \leq p; \\ (\tilde{Z}^H \tilde{Z})^{-1/2} \tilde{B}^H, & p \leq n, p \leq m; \\ \tilde{B}(\tilde{Z}^H \tilde{Z})^{-1/2}, & n \leq p, p \leq m \end{cases} \end{aligned} \quad (10)$$

and

$$\mathbf{M}(\Phi) = \Phi^H \mathbf{S}_2^H \mathbf{S}_2 \Phi \mathbf{S}_1^H \mathbf{S}_1, \forall \Phi \in \mathbb{U}_{\min\{m, p, n\}}. \quad (11)$$

Let (A, B) and $(\tilde{A}, \tilde{B})$ be two (m, p, n) real matrix pairs defined in (4) and

$$\begin{aligned} \mathcal{S}_1 &= \begin{cases} (Z^T Z)^{-1/2} A^T, & m \leq n, m \leq p; \\ A(Z^T Z)^{-1/2}, & n \leq m, m \leq p; \\ (Z^T Z)^{-1/2} B^T, & p \leq n, p \leq m; \\ B(Z^T Z)^{-1/2}, & n \leq p, p \leq m; \end{cases} \\ \mathcal{S}_2 &= \begin{cases} (\tilde{Z}^T \tilde{Z})^{-1/2} \tilde{A}^T, & m \leq n, m \leq p; \\ \tilde{A}(\tilde{Z}^T \tilde{Z})^{-1/2}, & n \leq m, m \leq p; \\ (\tilde{Z}^T \tilde{Z})^{-1/2} \tilde{B}^T, & p \leq n, p \leq m; \\ \tilde{B}(\tilde{Z}^T \tilde{Z})^{-1/2}, & n \leq p, p \leq m \end{cases} \end{aligned} \quad (12)$$

and

$$\mathcal{M}(\Phi) = \Phi^T \mathcal{S}_2^T \mathcal{S}_2 \Phi \mathcal{S}_1^T \mathcal{S}_1, \forall \Phi \in \mathbb{O}_{\min\{m,p,n\}}. \quad (13)$$

We set

$$\mathbf{Q}_i = \begin{cases} \text{diag}(I_i, O_{(\min\{m,p,n\}-i) \times (\min\{m,p,n\}-i)}), m \leq n, m \leq p; n \leq m, m \leq p; \\ \text{diag}(O_{(\min\{m,p,n\}-i) \times (\min\{m,p,n\}-i)}, I_i), p \leq n, p \leq m; n \leq p, p \leq m \end{cases} \quad (14)$$

for $1 \leq i \leq \min\{m, p, n\}$.

Lemma 2. [24] Let $f(X) : \mathbb{C}^{n \times n} \rightarrow \mathbb{R}$ be an analytical function of several complex variables on the domain $XX^H \leq I_n$. Then $f(X)$ attains its maximum modulus on the characteristic manifold $\{X \in \mathbb{C}^{n \times n} : XX^H = I_n\}$.

Lemma 3. Let $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ be two (m, p, n) -GMPs and let $\sigma\{A, B\} = \{(\alpha_i, \beta_i), i = 1, 2, \dots, n\}$ and $\sigma\{\tilde{A}, \tilde{B}\} = \{(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, 2, \dots, n\}$. Then

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{U}_{\min\{m,p,n\}}} \text{tr}(\mathbf{M}(\Phi)) - \text{tr}(\mathbf{S}_1^H \mathbf{S}_1 + \mathbf{S}_2^H \mathbf{S}_2) + n, \quad (15)$$

where $\mathbf{M}(\Phi)$ is given by (11).

Proof. Let $Z = \begin{pmatrix} A \\ B \end{pmatrix}$, $\tilde{Z} = \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix}$, then $Z^H Z = R^{-H} R^{-1}$, $\tilde{Z}^H \tilde{Z} = \tilde{R}^{-H} \tilde{R}^{-1}$ are nonsingular. Let

$$\begin{aligned} S_1 &= (Z^H Z)^{-1/2} A^H = (R^{-H} R^{-1})^{-1/2} R^{-H} \Sigma_A^H U^H, \\ S_2 &= (\tilde{Z}^H \tilde{Z})^{-1/2} \tilde{A}^H = (\tilde{R}^{-H} \tilde{R}^{-1})^{-1/2} \tilde{R}^{-H} \Sigma_{\tilde{A}}^H \tilde{U}^H \end{aligned} \quad (16)$$

$$\begin{aligned} S_3 &= (Z^H Z)^{-1/2} B^H = (R^{-H} R^{-1})^{-1/2} R^{-H} \Sigma_B^H V^H, \\ S_4 &= (\tilde{Z}^H \tilde{Z})^{-1/2} \tilde{B}^H = (\tilde{R}^{-H} \tilde{R}^{-1})^{-1/2} \tilde{R}^{-H} \Sigma_{\tilde{B}}^H \tilde{V}^H \end{aligned} \quad (17)$$

and

$$\begin{aligned} M_1(\Phi) &= S_1 \Phi^H S_2^H S_2 \Phi S_1^H, \forall \Phi \in \mathbb{U}_m; M_2(\Phi) = S_1^H \Phi^H S_2 S_2^H \Phi S_1, \forall \Phi \in \mathbb{U}_n \\ M_3(\Phi) &= S_3 \Phi^H S_4^H S_4 \Phi S_3^H, \forall \Phi \in \mathbb{U}_p; M_4(\Phi) = S_3^H \Phi^H S_4 S_4^H \Phi S_3, \forall \Phi \in \mathbb{U}_n. \end{aligned} \quad (18)$$

We first prove

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) - \text{tr}(\mathbf{S}_1^H \mathbf{S}_1 + \mathbf{S}_2^H \mathbf{S}_2) + n. \quad (19)$$

Using the GSVDs of $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ in Definition 3 we have

$$A = U \Sigma_A R, B = V \Sigma_B R, \tilde{A} = \tilde{U} \Sigma_{\tilde{A}} \tilde{R}, \tilde{B} = \tilde{V} \Sigma_{\tilde{B}} \tilde{R},$$

where $U, \tilde{U}, V, \tilde{V}, \Sigma_A, \Sigma_B, \Sigma_{\tilde{A}}$ and $\Sigma_{\tilde{B}}$ are given by (2) and (3). Let $M_1(\Phi)$ be given by (18). Then

$$\begin{aligned} \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) &= \max_{\Phi \in \mathbb{U}_m} \text{tr}(\Phi^H S_2^H S_2 \Phi S_1^H S_1) \\ &= \max_{\Phi \in \mathbb{U}_m} \text{tr}(\Phi \tilde{U} \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^H \tilde{U}^H \Phi^H U \Sigma_A \Sigma_A^H U^H). \end{aligned} \quad (20)$$

Next we consider two cases for different m and n . Case 1: if $n \leq m$, then we write

$$\Sigma_A = \begin{pmatrix} \hat{\Sigma}_A \\ O_{(m-n) \times n} \end{pmatrix}, \Sigma_{\tilde{A}} = \begin{pmatrix} \hat{\Sigma}_{\tilde{A}} \\ O_{(m-n) \times n} \end{pmatrix}, U^H \Phi \tilde{U} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \forall \Phi \in \mathbb{U}_m,$$

where $\hat{\Sigma}_A, \hat{\Sigma}_{\tilde{A}}, W_{11} \in \mathbb{C}^{n \times n}$ and $W_{22} \in \mathbb{C}^{(m-n) \times (m-n)}$. Substituting into (20) we have

$$\max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) = \max_{W_{11} W_{11}^H \leq I_n} \text{tr}(\hat{\Sigma}_A^H W_{11} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_{\tilde{A}}^H W_{11}^H \hat{\Sigma}_A),$$

where $W_{11} W_{11}^H \leq I_n$ means that $I_n - W_{11} W_{11}^H$ is positive semi-definite. As we know, $\text{tr}(\hat{\Sigma}_A^H W_{11} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_{\tilde{A}}^H W_{11}^H \hat{\Sigma}_A)$ as an analytical function of several complex variables on the domain $W_{11} W_{11}^H \leq I_n$ attains its maximum modulus on the characteristic manifold $\{W_{11} \in \mathbb{C}^{n \times n} : W_{11} W_{11}^H = I_n\}$ (see [21]), therefore by Lemma 2 we have

$$\max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) = \max_{W_{11} W_{11}^H = I_n} \text{tr}(\hat{\Sigma}_A^H W_{11} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_{\tilde{A}}^H W_{11}^H \hat{\Sigma}_A),$$

which together with Lemma 1 and (20) yields

$$\max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) = \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2. \quad (21)$$

Let $U^H \Psi U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$, $\tilde{U}^H \tilde{\Psi} \tilde{U} = \begin{pmatrix} \tilde{U}_{11} & \tilde{U}_{12} \\ \tilde{U}_{21} & \tilde{U}_{22} \end{pmatrix}$, $\forall \Psi, \tilde{\Psi} \in \mathbb{U}_m$ with $U_{11}, \tilde{U}_{11} \in \mathbb{C}^{n \times n}$ and $U_{22}, \tilde{U}_{22} \in \mathbb{C}^{(m-n) \times (m-n)}$, then by Lemma 1 and (33) we have

$$\begin{aligned} \text{tr}(S_1^H S_1 + S_2^H S_2) &= \text{tr}(\Sigma_A \Sigma_A^H + \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^H), \\ &= \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2). \end{aligned} \quad (22)$$

Case 2: if $m \leq n$, then we write

$$\Sigma_A = (\hat{\Sigma}_A, O_{m \times (n-m)}), \Sigma_{\tilde{A}} = (\hat{\Sigma}_{\tilde{A}}, O_{m \times (n-m)}).$$

Then by (20) and Lemma 1 we have

$$\begin{aligned} \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) &= \max_{\Phi \in \mathbb{U}_m} \text{tr}(\Phi \tilde{U} \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^H \tilde{U}^H \Phi^H U \Sigma_A \Sigma_A^H U^H) \\ &= \max_{\Phi \in \mathbb{U}_m} \text{tr} \begin{pmatrix} \Phi \tilde{U} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_{\tilde{A}}^H \tilde{U}^H \Phi^H U \hat{\Sigma}_A \hat{\Sigma}_A^H U^H & 0 \\ 0 & 0_{(n-m) \times (n-m)} \end{pmatrix} \\ &= \max_{\Phi \in \mathbb{U}_m} \text{tr}(\Phi \tilde{U} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_{\tilde{A}}^H \tilde{U}^H \Phi^H U \hat{\Sigma}_A \hat{\Sigma}_A^H U^H) \\ &= \sum_{i=1}^m \alpha_i^2 \tilde{\alpha}_i^2 = \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2 \end{aligned} \quad (23)$$

since $\alpha_{m+1} = \dots = \alpha_n = \tilde{\alpha}_{m+1} = \dots = \tilde{\alpha}_n = 0$ by the definition of GSVD. It follows from Lemma 1 and (33) that

$$\begin{aligned} \text{tr}(S_1^H S_1 + S_2^H S_2) &= \text{tr}(U \Sigma_A \Sigma_A^H U^H + \tilde{U} \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^H \tilde{U}^H), \\ &= \sum_{i=1}^m (\alpha_i^2 + \tilde{\alpha}_i^2) = \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2). \end{aligned} \quad (24)$$

By Definition 3 we have $\beta_i^2 = 1 - \alpha_i^2$, $\tilde{\beta}_i^2 = 1 - \tilde{\alpha}_i^2$, and together with (20)-(24) we have

$$\begin{aligned} \sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) &= \sum_{i=1}^n (2\alpha_i^2 \tilde{\alpha}_i^2 - (\alpha_i^2 + \tilde{\alpha}_i^2) + 1) \\ &= 2 \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2 - \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2) + n \\ &= 2 \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n. \end{aligned} \quad (25)$$

By Cases 1 and 2 we deduce (35) holds true. By the similar aforementioned methods we get

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{U}_n} \text{tr}(M_2(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n. \quad (26)$$

Note that in (41)

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \sum_{i=1}^n \beta_i^2 \tilde{\beta}_i^2 - \sum_{i=1}^n (\beta_i^2 + \tilde{\beta}_i^2) + n,$$

and hence by the similar aforementioned methods we have

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{U}_p} \text{tr}(M_3(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n, \quad (27)$$

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{U}_n} \text{tr}(M_4(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n. \quad (28)$$

In order to select the smallest dimensional variable unitary matrix by (35), (42)-(44) we have (31) holds. This completes the proof. $\square$

Lemma 4. Let $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\Delta = \text{diag}(\delta_1, \dots, \delta_n) \in \mathbb{R}^{n \times n}$ be diagonal matrices with $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ and $\delta_1 \geq \dots \geq \delta_n \geq 0$. Then we have

$$\max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(\Lambda \Phi^T \Delta \Phi) = \sum_{i=1}^n \lambda_i \delta_i,$$

Proof. It follows from (1) that

$$\begin{aligned} &\max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(\Lambda \Phi^T \Delta \Phi) \\ &= \max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(\Lambda \Phi^H \Delta \Phi) \\ &\leq \max_{\Phi \in \mathbb{U}_n} \text{tr}(\Lambda \Phi^H \Delta \Phi) \\ &= \sum_{i=1}^n \lambda_i \delta_i. \end{aligned} \quad (29)$$

Meanwhile,

$$\sum_{i=1}^n \lambda_i \delta_i = \text{tr}(\Lambda \Delta) \leq \max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(\Lambda \Phi^T \Delta \Phi). \quad (30)$$

Combing (29) and (30) yields

$$\max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(\Lambda \Phi^T \Delta \Phi) = \sum_{i=1}^n \lambda_i \delta_i.$$

This completed the proof. $\square$

Lemma 5. Let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{p \times n}$ have $\text{rank}(A^T, B^T) = n$ and the GSV (α_i, β_i) of $\{A, B\}$ be given in (4) and (5). Then

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{O}_{\min\{m,p,n\}}} \text{tr}(\mathcal{M}(\Phi)) - \text{tr}(S_1^T S_1 + S_2^T S_2) + n, \quad (31)$$

where $\mathcal{M}(\Phi)$ is given by (13).

Proof. Let $Z = \begin{pmatrix} A \\ B \end{pmatrix}$, $\tilde{Z} = \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix}$, then $Z^T Z = R^{-T} R^{-1}$, $\tilde{Z}^T \tilde{Z} = \tilde{R}^{-T} \tilde{R}^{-1}$ are nonsingular. Let

$$S_1 = (Z^T Z)^{-1/2} A^T = (R^{-T} R^{-1})^{-1/2} R^{-T} \Sigma_A^T U^T, \quad (32)$$

$$S_2 = (\tilde{Z}^T \tilde{Z})^{-1/2} \tilde{A}^T = (\tilde{R}^{-T} \tilde{R}^{-1})^{-1/2} \tilde{R}^{-T} \Sigma_{\tilde{A}}^T \tilde{U}^T$$

$$S_3 = (Z^T Z)^{-1/2} B^T = (R^{-T} R^{-1})^{-1/2} R^{-T} \Sigma_B^T V^T, \quad (33)$$

$$S_4 = (\tilde{Z}^T \tilde{Z})^{-1/2} \tilde{B}^T = (\tilde{R}^{-T} \tilde{R}^{-1})^{-1/2} \tilde{R}^{-T} \Sigma_{\tilde{B}}^T \tilde{V}^T$$

and

$$M_1(\Phi) = S_1 \Phi^T S_2^T S_2 \Phi S_1^T, \forall \Phi \in \mathbb{O}_{\mathbb{R}_m}; M_2(\Phi) = S_1^T \Phi^T S_2 S_2^T \Phi S_1, \forall \Phi \in \mathbb{O}_{\mathbb{R}_n} \quad (34)$$

$$M_3(\Phi) = S_3 \Phi^T S_4^T S_4 \Phi S_3^T, \forall \Phi \in \mathbb{O}_{\mathbb{R}_p}; M_4(\Phi) = S_3^T \Phi^T S_4 S_4^T \Phi S_3, \forall \Phi \in \mathbb{O}_{\mathbb{R}_n}.$$

We first prove

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{O}_{\mathbb{R}_m}} \text{tr}(M_1(\Phi)) - \text{tr}(S_1^T S_1 + S_2^T S_2) + n. \quad (35)$$

Then

$$\begin{aligned} \max_{\Phi \in \mathbb{O}_{\mathbb{R}_m}} \text{tr}(M_1(\Phi)) &= \max_{\Phi \in \mathbb{O}_{\mathbb{R}_m}} \text{tr}(\Phi^T S_2^H S_2 \Phi S_1^T S_1) \\ &= \max_{\Phi \in \mathbb{O}_{\mathbb{R}_m}} \text{tr}(\Phi \tilde{U} \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^T \tilde{U}^T \Phi^H U \Sigma_A \Sigma_A^H U^T), \end{aligned} \quad (36)$$

which together with Lemma 1 and (20) yields

$$\max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) = \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2. \quad (37)$$

Let $U^T \Psi U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$, $\tilde{U}^T \tilde{\Psi} \tilde{U} = \begin{pmatrix} \tilde{U}_{11} & \tilde{U}_{12} \\ \tilde{U}_{21} & \tilde{U}_{22} \end{pmatrix}$, $\forall \Psi, \tilde{\Psi} \in \mathbb{O}_{\mathbb{R}_m}$ with $U_{11}, \tilde{U}_{11} \in \mathbb{C}^{n \times n}$ and $U_{22}, \tilde{U}_{22} \in \mathbb{C}^{(m-n) \times (m-n)}$, then by Lemma 1 and (33) we have

$$\begin{aligned} \text{tr}(S_1^H S_1 + S_2^H S_2) &= \text{tr}(\Sigma_A \Sigma_A^H + \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^H), \\ &= \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2). \end{aligned} \quad (38)$$

Case 2: if $m \leq n$, then we write

$$\Sigma_A = (\hat{\Sigma}_A, O_{m \times (n-m)}), \Sigma_{\tilde{A}} = (\hat{\Sigma}_{\tilde{A}}, O_{m \times (n-m)}).$$

Then by (20) and Lemma 1 we have

$$\begin{aligned}
 \max_{\Phi \in \mathbb{O}\mathbb{R}_m} \text{tr}(M_1(\Phi)) &= \max_{\Phi \in \mathbb{O}\mathbb{R}_m} \text{tr}(\Phi \tilde{U} \Sigma_{\tilde{A}} \Sigma_A^T \tilde{U}^T \Phi^T U \Sigma_A \Sigma_A^T U^T) \\
 &= \max_{\Phi \in \mathbb{O}\mathbb{R}_m} \text{tr} \begin{pmatrix} \Phi \tilde{U} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_A^T \tilde{U}^T \Phi^T U \hat{\Sigma}_A \hat{\Sigma}_A^T U^T & 0 \\ 0 & 0_{(n-m) \times (n-m)} \end{pmatrix} \\
 &= \max_{\Phi \in \mathbb{O}\mathbb{R}_m} \text{tr}(\Phi \tilde{U} \hat{\Sigma}_{\tilde{A}} \hat{\Sigma}_A^T \tilde{U}^T \Phi^T U \hat{\Sigma}_A \hat{\Sigma}_A^T U^T) \\
 &= \sum_{i=1}^m \alpha_i^2 \tilde{\alpha}_i^2 = \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2
 \end{aligned} \tag{39}$$

since $\alpha_{m+1} = \dots = \alpha_n = \tilde{\alpha}_{m+1} = \dots = \tilde{\alpha}_n = 0$ by the definition of GSVD. It follows from Lemma 1 and (33) that

$$\begin{aligned}
 \text{tr}(S_1^H S_1 + S_2^H S_2) &= \text{tr}(U \Sigma_A \Sigma_A^T U^T + \tilde{U} \Sigma_{\tilde{A}} \Sigma_{\tilde{A}}^T \tilde{U}^T), \\
 &= \sum_{i=1}^m (\alpha_i^2 + \tilde{\alpha}_i^2) = \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2).
 \end{aligned} \tag{40}$$

By Definition 3 we have $\beta_i^2 = 1 - \alpha_i^2$, $\tilde{\beta}_i^2 = 1 - \tilde{\alpha}_i^2$, and together with (20)-(24) we have

$$\begin{aligned}
 \sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) &= \sum_{i=1}^n (2\alpha_i^2 \tilde{\alpha}_i^2 - (\alpha_i^2 + \tilde{\alpha}_i^2) + 1) \\
 &= 2 \sum_{i=1}^n \alpha_i^2 \tilde{\alpha}_i^2 - \sum_{i=1}^n (\alpha_i^2 + \tilde{\alpha}_i^2) + n \\
 &= 2 \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n.
 \end{aligned} \tag{41}$$

By Cases 1 and 2 we deduce (35) holds true. By the similar aforementioned methods we get

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(M_2(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n. \tag{42}$$

Note that in (41)

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \sum_{i=1}^n \beta_i^2 \tilde{\beta}_i^2 - \sum_{i=1}^n (\beta_i^2 + \tilde{\beta}_i^2) + n,$$

and hence by the similar aforementioned methods we have

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{O}\mathbb{R}_p} \text{tr}(M_3(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n, \tag{43}$$

$$\sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) = 2 \max_{\Phi \in \mathbb{O}\mathbb{R}_n} \text{tr}(M_4(\Phi)) - \text{tr}(S_1^H S_1 + S_2^H S_2) + n. \tag{44}$$

In order to select the smallest dimensional variable unitary matrix by (35), (42)-(44) we have (31) holds. This completes the proof. $\square$

2. Explicit formula of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ with Riemannian optimization models

We first present a new explicit formula of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ for complex grassman matrix pair. Then we use the corresponding numerical methods to solve solutions of the involved Riemannian optimization models.

2.1. For complex grassman matrix pair

We present a new explicit formula of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ for complex grassman matrix pair.

Lemma 6. Let $\mathbf{M}(\Phi), \mathbf{Q}_i, \mathbf{S}_1, \mathbf{S}_2$ be given by (10), (11) and (14). Then

$$\max_{\Phi \in \mathbb{U}_{\min\{m,p,n\}}} \text{tr}(\mathbf{M}(\Phi)) = \text{tr}(\mathbf{M}(\mathbf{F}_2 \mathbf{F}_1^H)),$$

where $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{U}_{\min\{m,p,n\}}$ can be obtained from Π_1, Π_2 by solving the trace function optimization models

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1^H \mathbf{S}_1^H \mathbf{S}_1 \Pi_1 \mathbf{Q}_i)|, \quad (45)$$

$$\max_{\Pi_2 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_2^H \mathbf{S}_2^H \mathbf{S}_2 \Pi_2 \mathbf{Q}_i)| \quad (46)$$

for every $i = 1, \dots, n$, respectively.

Proof. We first prove the case for $m \leq n, m \leq p$ and for others cases when $n \leq m, m \leq p; p \leq n, p \leq m$; and $n \leq p, p \leq m$ the corresponding results can be proved similarly. By (10), (11) and (18) $\mathbf{S}_1, \mathbf{S}_2, \mathbf{M}, \mathbf{F}_1, \mathbf{F}_2$ are taken as $\mathbf{S}_1, \mathbf{S}_2, \mathbf{M}_1, \mathbf{F}_1, \mathbf{F}_2$, where $\mathbf{F}_1$ and $\mathbf{F}_2$ can be obtained by solving

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1^H \mathbf{S}_1^H \mathbf{S}_1 \Pi_1 \mathbf{Q}_i)|, \quad (47)$$

$$\max_{\Pi_2 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_2^H \mathbf{S}_2^H \mathbf{S}_2 \Pi_2 \mathbf{Q}_i)|$$

for every $i = 1, \dots, n$, respectively. Let $\mathbf{f}_i$ be the i -th column of unitary matrix $\mathbf{F}_1$. By Lemma 1 we have for $i = 1$

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1^H \mathbf{S}_1^H \mathbf{S}_1 \Pi_1 \mathbf{Q}_1)| = |\text{tr}(\mathbf{F}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{F}_1 \mathbf{Q}_1)| = \mathbf{f}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_1 = \lambda_1. \quad (48)$$

Here $\mathbf{f}_1$ has some solutions and is set to satisfy $\mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_1 = \lambda_1 \mathbf{f}_1$. Similarly, for $i > 1$

$$\begin{aligned} \max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1^H \mathbf{S}_1^H \mathbf{S}_1 \Pi_1 \mathbf{Q}_i)| &= |\text{tr}(\mathbf{F}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{F}_1 \mathbf{Q}_i)| \\ &= \mathbf{f}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_1 + \dots + \mathbf{f}_i^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_i \\ &= \lambda_1 + \dots + \lambda_i. \end{aligned} \quad (49)$$

Here $\mathbf{f}_i$ has some solutions and is set to satisfy $\mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_i = \lambda_i \mathbf{f}_i$. Since $\mathbf{F}_1$ is unitary, then $\mathbf{f}_k^H \mathbf{f}_j = 0$ and $\mathbf{f}_k^H \mathbf{f}_k = 1, k \neq j; k, j = 1, \dots, n$. It follows that $\mathbf{f}_k^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_j = \lambda_i \mathbf{f}_k^H \mathbf{f}_j = 0$ and $\mathbf{f}_k^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{f}_k = \lambda_k \mathbf{f}_k^H \mathbf{f}_k = \lambda_k, k \neq j; k, j = 1, \dots, n$. Then

$$\mathbf{F}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{F}_1 = \Lambda_1, \quad (50)$$

where $\Lambda_1 = \text{diag}(\lambda_i)$. Similarly,

$$\mathbf{F}_2^H \mathbf{S}_2^H \mathbf{S}_2 \mathbf{F}_2 = \Lambda_2,$$

where $\Lambda_2 = \text{diag}(\chi_i)$. By Lemma 1 and (47) we have for $i = 1$

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1^H \mathbf{S}_1^H \mathbf{S}_1 \Pi_1 \mathbf{Q}_1)| = |\text{tr}(\mathbf{F}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{F}_1 \mathbf{Q}_1)| = \lambda_1 = \max_{1 \leq i \leq n} \lambda_i. \quad (51)$$

For other $i = 2, \dots, n$, we have

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |tr(\Pi_1^H S_1^H S_1 \Pi_1 Q_i)| = |tr(F_1^H S_1^H S_1 F_1 Q_i)| = |tr(\Lambda_1 Q_i)| = \lambda_1 + \dots + \lambda_i. \quad (52)$$

Then $\Lambda_1 = diag(\lambda_i)$, $\Lambda_2 = diag(\chi_i)$ have descending sorts with $\lambda_i \geq 0, \chi_i \geq 0$, otherwise, there will be conflicts with (51) and (52). Then by Lemma 1 we have

$$\begin{aligned} \max_{\Phi \in \mathbb{U}_m} tr(M_1(\Phi)) &= \max_{\Phi \in \mathbb{U}_m} tr(\Phi^H S_2^H S_2 \Phi S_1^H S_1) \\ &= \max_{\Phi \in \mathbb{U}_m} tr(\Phi^H F_2 \Lambda_2 F_2^H \Phi F_1 \Lambda_1 F_1^H) = tr(\Lambda_2 \Lambda_1). \end{aligned} \quad (53)$$

It is easy to check that $tr(M_1(F_2 F_1^H)) = tr(\Lambda_2 \Lambda_1)$, which together with (53) implies $\max_{\Phi \in \mathbb{U}_m} tr(M_1(\Phi)) = tr(M_1(F_2 F_1^H))$. For other cases for $n \leq m, m \leq p; p \leq n, p \leq m$; and $n \leq p, p \leq m$; the desired corresponding results can be proved similarly. The proof is complete. $\square$

Theorem 3. Let $\mathbf{M}(\Phi), \mathbf{Q}_i, \mathbf{S}_1, \mathbf{S}_2$ be given by (10), (11) and (14). Let $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ be two (m, p, n) -GMPs and $\sigma\{A, B\} = \{(\alpha_i, \beta_i), i = 1, 2, \dots, n\}$ and $\sigma\{\tilde{A}, \tilde{B}\} = \{(\tilde{\alpha}_j, \tilde{\beta}_j), j = 1, 2, \dots, n\}$. Then

$$\begin{aligned} &\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) \\ &= tr(\mathbf{S}_1^H \mathbf{S}_1 + \mathbf{S}_2^H \mathbf{S}_2) - 2 \max_{\Phi \in \mathbb{U}_{\min\{m,p,n\}}} tr(\mathbf{M}(\Phi)) - 2tr(\mathbf{L}_1^{\frac{1}{2}}(I_{\min\{m,p,n\}} - \mathbf{L}_1)^{\frac{1}{2}} \mathbf{L}_2^{\frac{1}{2}}(I_{\min\{m,p,n\}} - \mathbf{L}_2)^{\frac{1}{2}}) \\ &= tr(\mathbf{S}_1^H \mathbf{S}_1 + \mathbf{S}_2^H \mathbf{S}_2) - 2tr(\mathbf{M}(F_2 F_1^H)) - 2tr(\mathbf{L}_1^{\frac{1}{2}}(I_{\min\{m,p,n\}} - \mathbf{L}_1)^{\frac{1}{2}} \mathbf{L}_2^{\frac{1}{2}}(I_{\min\{m,p,n\}} - \mathbf{L}_2)^{\frac{1}{2}}), \end{aligned}$$

where

$$\mathbf{L}_1 = \mathbf{F}_1^H \mathbf{S}_1^H \mathbf{S}_1 \mathbf{F}_1, \quad \mathbf{L}_2 = \mathbf{F}_2^H \mathbf{S}_2^H \mathbf{S}_2 \mathbf{F}_2$$

and $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{U}_{\min\{m,p,n\}}$ can be obtained from Π_1, Π_2 by solving the trace function optimization models (45) and (46).

Proof. We first prove the case when $m \leq n, m \leq p$ and for others cases when $n \leq m, m \leq p; p \leq n, p \leq m$; and $n \leq p, p \leq m$ the corresponding results can be proved similarly. By (10), (11) and (18) $\mathbf{S}_1, \mathbf{S}_2, \mathbf{M}, \mathbf{F}_1, \mathbf{F}_2$ can be taken as S_1, S_2, M_1, F_1, F_2 . Since F_1 and F_2 are solutions of (45) and (46) for every $i = 1, \dots, n$, respectively, then $S_1^H S_1 = A(A^H A + B^H B)^{-1} A^H = U \Sigma_A \Sigma_A^H U^H$. For $i = 1$ by (45) and Lemma 1 we have

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |tr(\Pi_1^H A(A^H A + B^H B)^{-1} A^H \Pi_1 Q_1)| = |tr(F_1^H A(A^H A + B^H B)^{-1} A^H F_1 Q_1)| = \alpha_1,$$

For $i = 2$ we have

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |tr(\Pi_1^H A(A^H A + B^H B)^{-1} A^H \Pi_1 Q_i)| = |tr(F_1^H A(A^H A + B^H B)^{-1} A^H F_1 Q_2)| = \alpha_1 + \alpha_2.$$

For $i > 2$, we have

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |tr(\Pi_1^H A(A^H A + B^H B)^{-1} A^H \Pi_1 Q_i)| = \alpha_1 + \dots + \alpha_i.$$

Hence, we have $L_1 = F_1^H S_1^H S_1 F_1 = diag(\alpha_i)$, $L_2 = F_2^H S_2^H S_2 F_2 = diag(\tilde{\alpha}_i)$. It follows that

$$\sum_{i=1}^n \alpha_i \beta_i \tilde{\alpha}_i \tilde{\beta}_i = tr(L_1^{\frac{1}{2}}(I_m - L_1)^{\frac{1}{2}} L_2^{\frac{1}{2}}(I_m - L_2)^{\frac{1}{2}}),$$

which together with (41) and Lemma 6 yields

$$\begin{aligned} \sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) &= n - \sum_{i=1}^n (\alpha_i^2 \tilde{\alpha}_i^2 + \beta_i^2 \tilde{\beta}_i^2) - 2 \sum_{i=1}^n \alpha_i \beta_i \tilde{\alpha}_i \tilde{\beta}_i, \\ &= \text{tr}(S_1^H S_1 + S_2^H S_2) - 2 \max_{\Phi \in \mathbb{U}_m} \text{tr}(M_1(\Phi)) - 2 \text{tr}(L_1^{\frac{1}{2}}(I_m - L_1)^{\frac{1}{2}} L_2^{\frac{1}{2}}(I_m - L_2)^{\frac{1}{2}}), \\ &= \text{tr}(S_1^H S_1 + S_2^H S_2) - 2 \text{tr}(M_1(F_2 F_1^H)) - 2 \text{tr}(L_1^{\frac{1}{2}}(I_m - L_1)^{\frac{1}{2}} L_2^{\frac{1}{2}}(I_m - L_2)^{\frac{1}{2}}). \end{aligned}$$

For the cases when $n \leq m, m \leq p; p \leq n, p \leq m$; and $n \leq p, p \leq m$; by the similar aforementioned methods the desired results can be proved. $\square$

Remark 1. From Theorem 3 we see that the formula of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ in (3) only refers to two small size unitary matrices $\mathbf{F}_1 \in \mathbb{U}_{\min\{m,p,n\}}, \mathbf{F}_2 \in \mathbb{U}_{\min\{m,p,n\}}$, which may require less computational cost for computing the explicit expression in Theorem 3.

When using Theorem 3 we first determine $\mathbf{S}_1, \mathbf{S}_2$ by comparing m, p, n . Then by using (45) and (46) $\mathbf{F}_1, \mathbf{F}_2$ can be computed.

2.2. Solving Riemannian optimization models in (45) and (46) for complex grassman matrix pair

We will use Newton methods on Riemannian manifolds to solve Riemannian optimization models in (45) and (46) of Theorem 3.

$$\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1 \mathbf{S}_1^H \mathbf{S}_1 \Pi_1^H \mathbf{Q}_i)|, \quad (54)$$

for $n \leq m, n \leq p$ and $m \leq n, m \leq p$ and

$$\max_{\Pi_2 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_2 \mathbf{S}_1^H \mathbf{S}_1 \Pi_2^H \mathbf{Q}_i)|, \quad (55)$$

for $p \leq n, p \leq m$ and $n \leq p, p \leq m$. Solving $\max_{\Pi_1 \in \mathbb{U}_{\min\{m,p,n\}}} |\text{tr}(\Pi_1 \mathbf{S}_2^H \mathbf{S}_2 \Pi_1^H \mathbf{Q}_i)|$ can be similarly discussed. For simplicity, we assume that $n \leq m$ and $n \leq p$ for (54) and $n \leq p, p \leq m$ for (55) and other cases for different relationships between m, n, p can be similarly discussed. To compute (54) and (55), we consider the following matrix optimization problems:

$$\begin{aligned} \max \quad & f_1(\Phi_1) := \frac{1}{2} \text{tr}(\Phi_1^H \mathbf{Q}_i \Phi_1 C) \\ \text{s.t.} \quad & \Phi_1 \in \mathbb{U}_n \end{aligned} \quad (56)$$

and

$$\begin{aligned} \max \quad & f_2(\Phi_2) := \frac{1}{2} \text{tr}(\Phi_2^H \mathbf{Q}_i \Phi_2 D) \\ \text{s.t.} \quad & \Phi_2 \in \mathbb{U}_n \end{aligned} \quad (57)$$

for $1 \leq i \leq n$. We only need to focus on the following matrix optimization problems:

$$\begin{aligned} \max \quad & f_1(\Phi_{11}) := \frac{1}{2} \text{tr}(\Phi_{11}^H C \Phi_{11}) \\ \text{s.t.} \quad & \Phi_{11} \in St(n, i) := \{\Phi_{11} \in \mathbb{C}^{n \times i} \mid \Phi_{11}^H \Phi_{11} = I_i\}, \end{aligned} \quad (58)$$

$$\begin{aligned} \max \quad & f_2(\Phi_{22}) := \frac{1}{2} \text{tr}(\Phi_{22}^H D \Phi_{22}) \\ \text{s.t.} \quad & \Phi_{22} \in St(n, n - i + 1), \end{aligned} \quad (59)$$

where $C = (A^H A + B^H B)^{-1/2} A^H A (A^H A + B^H B)^{-1/2}$, $D = (A^H A + B^H B)^{-1/2} B^H B (A^H A + B^H B)^{-1/2}$, Φ_{11} is the first i columns of Φ_1^H and Φ_{22} is the last $n - i + 1$ columns of Φ_2^H . In the following, we focus on the solution of the matrix optimization problem (58) and the matrix optimization problem (59) can be solved in a similar way.

We note that $f_1(\Phi_{11}Q) = f_1(\Phi_{11})$ for all $\Phi_{11} \in St(n, i)$ and $Q \in \mathbb{U}_i$. Let the complex Grassmann manifold $\text{Grass}(n, i)$ be the set of all i -dimensional complex subspace of $\mathbb{C}^n$. If $[X]$ means the subspace spanned by the columns of $X \in St(n, i)$, then we have $[X] \in \text{Grass}(n, i)$. In particular, for any $X \in St(n, i)$, the natural projection $[X] \in \text{Grass}(n, i)$ corresponds the equivalent class $\{XQ \in St(n, i) \mid Q \in \mathbb{U}_i\}$ of $St(n, i)$. Thus, instead of problem (58), we consider the following optimization problem:

$$\begin{aligned} \max \quad & \tilde{f}_1([\Phi_{11}]) := f_1(\Phi_{11}) \\ \text{s.t.} \quad & [\Phi_{11}] \in \text{Grass}(n, i). \end{aligned} \quad (60)$$

Next, we present Newton's method for solving the optimization problem (60). Let $\Phi_{11} \in St(n, i)$. The tangent space of $St(n, i)$ at Φ_{11} is given by [29]

$$T_{\Phi_{11}} St(n, i) = \{Z \in \mathbb{C}^{n \times i} \mid Z = \Phi_{11}\Omega + (\Phi_{11})_{\perp} K, \Omega^H = -\Omega, \Omega \in \mathbb{C}^{i,i}, K \in \mathbb{C}^{n-i,i}\},$$

which can be endowed with the inner product

$$\langle Z_1, Z_2 \rangle = \Re[\text{tr}(Z_2^H (I - \frac{1}{2}\Phi_{11}\Phi_{11}^H)Z_1)], \quad Z_1, Z_2 \in T_{\Phi_{11}} St(n, i), \Phi_{11} \in St(m, i).$$

Here, $(\Phi_{11})_{\perp}$ means that $\text{span}((\Phi_{11})_{\perp})$ is the orthogonal complement of $\text{span}(\Phi_{11})$, where $\text{span}(\Phi_{11})$ denotes a linear space spanned by the column vectors of Φ_{11} . We note that the tangent space of $\text{Grass}(n, i)$ at $[\Phi_{11}]$ is given by [29]

$$T_{[\Phi_{11}]} \text{Grass}(n, i) = \{Z \in \mathbb{C}^{n \times i} \mid Z = (\Phi_{11})_{\perp} K, K \in \mathbb{C}^{n-i,i}\} \subset T_{\Phi_{11}} St(n, i).$$

Hence, we can define a Riemannian metric on $\text{Grass}(n, i)$ by

$$\langle Z_1, Z_2 \rangle = \Re[\text{tr}(Z_2^H Z_1)], \quad Z_1, Z_2 \in T_{[\Phi_{11}]} \text{Grass}(n, i), \Phi_{11} \in St(n, i)$$

with the induced norm $\|\cdot\|$. Then the orthogonal projection onto $T_{[\Phi_{11}]} \text{Grass}(m, i)$ is given by

$$P_{\Phi_{11}} Z = (I - \Phi_{11}\Phi_{11}^H)Z, \quad \forall Z \in \mathbb{C}^{n \times i}.$$

We define the local cost function $g : T_{[\Phi_{11}]} \text{Grass}(n, i) \rightarrow \mathbb{R}$ by

$$g(Z) = \tilde{f}_1([\Phi_{11} + Z]).$$

It is easy to check that, for any $Z \in T_{[\Phi_{11}]} \text{Grass}(n, i)$,

$$\begin{aligned} g(Z) &= f_1(\Phi_{11}) + \Re \text{tr}(Z^H D_{\Phi_{11}}) \\ &\quad + \frac{1}{2} \text{vec}(Z)^H \left(H_{\Phi_{11}} - (\Phi_{11}^H D_{\Phi_{11}})^T \otimes I_n \right) \text{vec}(Z), \end{aligned}$$

where $D_{\Phi_{11}} = C\Phi_{11}$ and $H_{\Phi_{11}} = I_i \otimes C$ are the derivative and Hessian of f_1 at Φ_{11} , respectively.

Based on the above analysis, Newton's method for solving the optimization problem (60) can be described as follows [29].

Algorithm 4. Step 0. Choose $\Phi_{11}^0 \in St(n, i)$, $\rho, \eta \in (0, 1)$, $\sigma \in (0, 1/2]$ and let $k := 1$.

Step 1. Apply the conjugate gradient (CG) method to solving

$$P_{\Phi_{11}^k} (CZ^k - Z^k (\Phi_{11}^k)^H D_{\Phi_{11}^k}) = -P_{\Phi_{11}^k} D_{\Phi_{11}^k}$$

for $Z^k \in T_{[\Phi_{11}^k]} \text{Grass}(n, i)$ such that

$$\|P_{\Phi_{11}^k} (CZ^k - Z^k (\Phi_{11}^k)^H D_{\Phi_{11}^k}) + P_{\Phi_{11}^k} D_{\Phi_{11}^k}\| \leq \eta_k \|P_{\Phi_{11}^k} D_{\Phi_{11}^k}\| \quad (61)$$

and

$$\langle P_{\Phi_{11}^k} D_{\Phi_{11}^k}, Z^k \rangle \leq -\eta_k \langle Z^k, Z^k \rangle, \quad (62)$$

where $\eta_k = \min\{\eta, \|P_{\Phi_{11}^k} D_{\Phi_{11}^k}\|\}$. If (61) and (62) are not attainable, then let

$$Z^k = -P_{\Phi_{11}^k} D_{\Phi_{11}^k}.$$

Step 2. Let $l_k > 0$ be the smallest integer m such that

$$f_1(\pi(\Phi_{11}^k + \beta^l Z^k)) \leq f_1(\Phi_{11}^k) + \sigma \rho^l \langle P_{\Phi_{11}^k} D_{\Phi_{11}^k}, Z^k \rangle$$

Set

$$Z^{k+1} = \pi(\Phi_{11}^k + \beta^l Z^k).$$

Step 3. Replace k by $k + 1$ and go to Step 1.

In Step 2 of Algorithm 4, $\pi : \mathbb{C}^{m \times i} \rightarrow \text{Grass}(n, i)$ is the projection onto $\text{Grass}(n, i)$, which is defined as follows: Let $X \in \mathbb{C}^{n \times i}$ be full column rank. Then

$$\pi(X) = \left[\underset{Y \in \text{St}(n, i)}{\text{argmin}} \|X - Y\|^2 \right].$$

As noted in [29], if the SVD of X is given by $X = U \Sigma V^H$, then $\pi(X) = U I_{n,i} V^H$. If the QR decomposition of X is $X = QR$, then $\pi(X) = Q_{n,i}$.

Remark 2. (i) An explicit expression of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ is given in Theorem 3. As a comparison, Theorem 2 only provides an upper bound. Obviously, Theorem 3 improves the existing results in [22] in theory, see Example below. We adopt trace function optimization formula with orthogonal constraints in Theorem 3 to evaluate the explicit expression of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$, which costs less complexity. While Theorem 3.2 in [22] needs unitary matrices $U, V, \tilde{U}, \tilde{V}$ of the GSVDs of two matrix pairs $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$. It is known that computing GSVDs may cost much higher amount of calculations than computing small-size unitary matrix. Hence, on algorithm aspect Theorem 3 also improves Theorem 3.2 in [22].

(ii) The equality in Theorem 3 refers to matrix optimization with orthogonal constraints. Based on matrix optimization formula with orthogonal constraints we use classical Golub-Kahan bidiagonalization method to solve it efficiently. The proposed algorithms by trace function optimization under two variables are efficient for computing for $A^H A + B^H B$ being not ill-posed. For $A^H A + B^H B = W$ being ill-posed, by matrix decompositions (eg., Schur decomposition or QR decomposition) we compute W_1, W_2 with $W = W_1 W_2$, where W_1, W_2 have small condition numbers. We calculate W_1^{-1}, W_2^{-1} and then compute multiplications for $W^{-1} = W_2^{-1} W_1^{-1}$, see [27,28]. In general we first compute the inverse of the matrix with small condition numbers and then multiply these two inverse matrices, which can well get the inverse of the original matrix with large condition numbers. The proposed algorithms by trace function optimization under two variables are efficient for computing for condition numbers of

$A^H A + B^H B$ being large, see Table 1. In practical applications, obviously, by the new explicit expression formula getting exact values of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ are more precise than its upper bound.

Example. To test the comparison of the explicit formula in Theorem 3, and the upper bound of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ in [22], we randomly generate grassman matrix pencils $\{A, B\}$ and $\{\tilde{A}, \tilde{B}\}$ by MATLAB command `randn(m,n)+i*randn(m,n)` for $A, \tilde{A}$ and `randn(p,n)+i*randn(p,n)` for $B, \tilde{B}$ with different random m, n, p . We first assess condition numbers of $A^H A + B^H B$ by using two-norm and denoted by $\kappa_2(\cdot)$. By simple computations we present the table 1.

Table 1. Comparisons of exact values, explicit formula by Alg. 4 and the upper bound with different (m, p, n)

(m, p, n)	$\kappa_2(\cdot)$	Exact values	formula by Alg. 4	Upper Bound
(80,40,60)	30.2	27.24709031	27.24698583	41.3510
(200,500,450)	218.6	189.26904380	189.26910247	380.4629
(900,800,700)	359.4	458.42617335	458.42618210	573.0744
(60,120,140)	600.3	103.56173621	103.56172847	126.0846
(100,200,150)	400.6	79.35162440	79.35157629	118.5329
(250,500,450)	$O(10^3)$	268.49009138	268.49008938	376.4290
(800,600,500)	$O(10^3)$	409.05215925	409.05215570	454.7393
(2000,1900,1800)	$O(10^4)$	1516.38636811	1516.38632996	1704.1106
(500,600,800)	$O(10^5)$	379.23164800	379.23164655	634.2572

Here ‘Explicit expression’ denotes the equality in Theorem 3, ‘Upper Bound’ denotes the upper bound in [22]. This table obviously shows the equality in Theorem 3 is almost equal to exact value of $\sum_{i=1}^n \rho^2((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i))$ and may be much sharper than the upper bounds in [22]. This again verifies Remark (i).

3. Concluding Remarks

In this paper we give explicit expressions of sum of chordal metric between generalized singular values of Grassmann matrix pairs, which absolutely improves the existing upper bounds in [22]. For the new expression formula only two small-size singular value decompositions are required in the algorithm which dominates the computational cost. As a byproduct, we also present an interval estimate without using the singular values.

1. Alter, O., Brown, P. O., and Botstein, D. 2000. "Singular value decomposition for genome-wide expression data processing and modelling." *Proceedings of the National Academy of Sciences* 97: 10101–10106.
2. Alter, O., Brown, P. O., and Botstein, D. 2001. "Processing and modelling gene expression data using singular value decomposition." *Proceedings of SPIE* 4266: 171–186.
3. Alter, O., Brown, P. O., and Botstein, D. 2003. "Generalized singular decomposition for comparative analysis of genome-scale expression data sets of two different organisms." *Proceedings of the National Academy of Sciences* 100: 3351–3356.
4. Alter, O., Golub, G. H., Brown, P. O., and Botstein, D. 2004. "Novel genome-scale correlation between DNA replication and RNA transcription during the cell cycle in yeast is predicted by data-driven models." In *Proceedings of the Miami Nature Biotechnology Winter Symposium on the Cell Cycle, Chromosomes and Cancer*, vol. 15, edited by M. P. Deutscher et al., 15. University of Miami School of Medicine, Miami.
5. Bai, Z., and Demmel, J. W. 1993. "Computing the generalized singular value decomposition." *SIAM Journal on Scientific Computing* 14: 1464–1486.
6. Bai, Z., and Zha, H. 1993. "A new preprocessing algorithm for the computation of the generalized singular value decomposition." *SIAM Journal on Scientific Computing* 14: 1007–1012.
7. Bhatia, R. 1997. *Matrix Analysis*. New York: Springer.

8. Sun, J.-G. 1983. "Perturbation analysis for the generalized eigenvalue problem and the generalized singular value problem." In *Matrix Pencils*, Lecture Notes in Mathematics, vol. 973, edited by Springer Verlag, 221–244.
9. Li, R.-C. 1993. "Bounds on perturbations of generalized singular values and of associated subspaces." *SIAM Journal on Matrix Analysis and Applications* 14: 195–234.
10. Li, R.-C. 1993. "A perturbation bound for definite pencils." *Linear Algebra and its Applications* 179: 191–202.
11. Li, R.-C. 1994. "On perturbations of matrix pencils with real spectra." *Mathematics of Computation* 62: 231–265.
12. Li, R.-C. 2002. "On perturbations of matrix pencils with real spectra, a revisit." *Mathematics of Computation* 72: 715–728.
13. Mahony, R. E. 1996. "The constrained Newton method on a Lie group and the symmetric eigenvalue problem." *Linear Algebra and its Applications* 248(15): 67–89.
14. Owren, B., and Welfert, B. 2000. "The Newton iteration on Lie groups." *BIT* 40(1): 121–145.
15. Paige, C. C., and Saunders, M. A. 1981. "Towards a generalized singular value decomposition." *SIAM Journal on Numerical Analysis* 18: 398–405.
16. Sun, J.-G. 1983. "Perturbation analysis for the generalized singular value problem." *SIAM Journal on Matrix Analysis and Applications* 20: 611–625.
17. Sun, J.-G. 1998. "Perturbation analysis of generalized singular subspaces." *Numerische Mathematik* 79: 615–641.
18. Sun, J.-G. 2000. "Condition number and backward error for the generalized singular value decomposition." *SIAM Journal on Matrix Analysis and Applications* 22: 323–341.
19. Van Loan, C. F. 1976. "Generalizing the singular value decomposition." *SIAM Journal on Numerical Analysis* 13: 76–83.
20. Wang, J. H., and Li, C. 2011. "Kantorovich's theorems for Newton's method for mappings and optimization problems on Lie groups." *IMA Journal of Numerical Analysis* 31: ...
21. Xu, W. W., Li, W., Zhu, L., and Huang, X. P. 2019. "The analytic solutions of a class of constrained matrix minimization and maximization problems with applications." *SIAM Journal on Optimization* 29: 1657–1686.
22. Xu, W. W., Pang, H. K., Li, W., Huang, X. P., and Guo, W. J. 2018. "On the explicit expression of chordal metric between generalized singular values of Grassmann matrix pairs with applications." *SIAM Journal on Matrix Analysis and Applications* 39(4): 1547–1563.
23. Zha, H. 1992. "A numerical algorithm for computing restricted singular value decomposition of matrix triplets." *Linear Algebra and Its Applications* 168: 1–26.
24. Hua, L. K. 1963. *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*. Providence, RI: American Mathematical Society.
25. Golub, G. H., and Van Loan, C. F. 2013. *Matrix Computations*, 4th ed. Baltimore: Johns Hopkins University Press.
26. Golub, G. H., and Kahan, W. 1965. "Calculating the singular values and pseudo-inverse of a matrix." *SIAM Journal on Numerical Analysis* 2: 205–224.
27. Ben-Israel, A., and Greville, T. N. E. 1977. *Generalized Inverses: Theory and Applications*. New York: Wiley.
28. Jodár, L., Law, A. G., Rezazadeh, A., Watson, J. H., and Wu, G. 1991. "Computations for the Moore-Penrose and Other Generalized Inverses." *Congressus Numerantium* 80: 57–64.
29. Manton, J. H. 2002. "Optimization algorithms exploiting unitary constraints." *IEEE Transactions on Signal Processing* 50: 635–650.
30. Absil, P.-A., Mahony, R., and Sepulchre, R. 2008. *Optimization Algorithms on Matrix Manifolds*. Princeton: Princeton University Press.
31. Adler, R. L., Dedieu, J. P., Margulies, J. Y., Martens, M., and Shub, M. 2002. "Newton's method on Riemannian manifolds and a geometric model for the human spine." *IMA Journal of Numerical Analysis* 22: 359–390.
32. Edelman, A., Arias, T. A., and Smith, S. T. 1998. "The geometry of algorithms with orthogonality constraints." *SIAM Journal on Matrix Analysis and Applications* 20: 303–353.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.