

Unitarily Invariant Metrics on the Grassmann Space

Li Qiu

Department of Electrical & Electronic Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong, China

Yanxia Zhang

Department of Applied Mathematics
Beijing Institute of Technology
Beijing 100081 China

Chi-Kwong Li

Department of Mathematics
College of William and Mary
Williamsburg, Virginia 23187-8795 USA

Abstract

Let $\mathcal{G}_{m,n}$ be the Grassmann space of m -dimensional subspaces of \mathbb{F}^n . Denote by $\theta_1(\mathcal{X}, \mathcal{Y}), \dots, \theta_m(\mathcal{X}, \mathcal{Y})$ the canonical angles between subspaces $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$. It is shown that $\Phi(\theta_1(\mathcal{X}, \mathcal{Y}), \dots, \theta_m(\mathcal{X}, \mathcal{Y}))$ defines a unitarily invariant metric on $\mathcal{G}_{m,n}$ for every symmetric gauge function Φ . This provides a wide class of new metrics on $\mathcal{G}_{m,n}$. Some related results on perturbation and approximation of subspaces in $\mathcal{G}_{m,n}$, as well as the canonical angles between them, are also discussed. Furthermore, the equality cases of the triangle inequalities for several unitarily invariant metrics are analyzed.

1 Introduction

Let \mathbb{F} be the real field \mathbb{R} or the complex field \mathbb{C} . Denote by \mathbb{F}^n the linear space of $n \times 1$ vectors with entries in \mathbb{F} . Suppose \mathcal{X} and \mathcal{Y} are one dimensional subspaces of \mathbb{F}^n spanned by the unit vectors x and y , respectively. There are natural metrics to measure the distance between \mathcal{X} and \mathcal{Y} .

- (a) The acute angle between \mathcal{X} and \mathcal{Y} , i.e., $\theta(\mathcal{X}, \mathcal{Y}) = \arccos |x^H y| \in [0, \pi/2]$.
- (b) The gap metric between \mathcal{X} and \mathcal{Y} defined as $\sin \theta(\mathcal{X}, \mathcal{Y}) = \inf \{\|x - yq\| : q \in \mathbb{F}\}$.
- (c) The Hausdorff metric between \mathcal{X} and \mathcal{Y} defined as

$$2 \sin(\theta(\mathcal{X}, \mathcal{Y})/2) = \inf \{\|x - yu\| : u \in \mathbb{F}, |u| = 1\}.$$

To gain more geometrical insight, let us temporarily specialize these measures to real Euclidean space \mathbb{R}^n . It is clear that the quantity in (a) is in fact the acute angle between the lines \mathcal{X} and \mathcal{Y} . The quantity in (b) represents the distance of any unit vector in one line, either x or $-x$, to the other line \mathcal{Y} ; it also turns out to be the difference between the orthogonal projections onto \mathcal{X} and \mathcal{Y} respectively, i.e., $\|xx^H - yy^H\|$, measured by the operator norm. The quantity in (c) represents the Hausdorff distance between the set of unit vectors in one line, $\{x, -x\}$, and that in the other line, $\{y, -y\}$.

It is well-known and not hard to check that each of the measures in (a), (b) and (c) indeed defines a metric ρ on the set $\mathcal{G}_{1,n}$ of one dimensional subspaces of \mathbb{F}^n , i.e., for any one dimensional subspaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ of \mathbb{F}^n , we have

- (i) $\rho(\mathcal{X}, \mathcal{Y}) \geq 0$; the equality holds if and only if $\mathcal{X} = \mathcal{Y}$,
- (ii) $\rho(\mathcal{X}, \mathcal{Y}) = \rho(\mathcal{Y}, \mathcal{X})$,
- (iii) $\rho(\mathcal{X}, \mathcal{Z}) \leq \rho(\mathcal{X}, \mathcal{Y}) + \rho(\mathcal{Y}, \mathcal{Z})$.

Furthermore, each of these metrics is unitarily invariant in the sense that

- (iv) $\rho(U\mathcal{X}, U\mathcal{Y}) = \rho(\mathcal{X}, \mathcal{Y})$ for any unitary transformation U on \mathbb{F}^n .

In a certain sense, the metric defined in (a) is superior to those in (b) or (c) for the following reasons.

First, for any positive subadditive nondecreasing function $f : [0, \pi/2] \rightarrow [0, \infty)$, i.e., any function f satisfying $f(\alpha) > 0$ for $\alpha \in (0, \pi/2]$, $f(\alpha + \beta) \leq f(\alpha) + f(\beta)$ for any $\alpha, \beta \in [0, \pi/2]$ with $\alpha + \beta \leq \pi/2$, and $f(\alpha) \leq f(\beta)$ for $\alpha \leq \beta$, the function $(\mathcal{X}, \mathcal{Y}) \rightarrow f(\theta(\mathcal{X}, \mathcal{Y}))$ is again a metric on $\mathcal{G}_{1,n}$. Proper choice of such an f generates a wide class of metrics. Taking $f(\alpha) = \sin \alpha$ and $f(\alpha) = 2\sin(\alpha/2)$, one obtains the metrics in (b) and (c) from that of (a). One might also take, for example, $f(\alpha) = \ln(1 + \alpha)$ or $f(\alpha) = \tan(\alpha/2)$.

Second, for $\rho(\mathcal{X}, \mathcal{Y}) = \theta(\mathcal{X}, \mathcal{Y})$, the equality in (iii) holds for $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{1,n}$ if and only if there are unit vectors x, y, z in $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ such that $y = [tx + (1 - t)z] / \|tx + (1 - t)z\|$ for some $t \in [0, 1]$ and $x^*z \geq 0$. Geometrically, it means that $V = \text{span}\{x, y, z\}$ is two dimensional and the vector y lies on the circular arc with length not larger than $\pi/2$ joining the two vectors x and z in V . This is most useful when we consider problems involving small perturbation of subspaces. In contrast, for the metrics defined in (b) and (c), the equality in (iii) holds only in the trivial case when $\mathcal{X} = \mathcal{Y}$ or $\mathcal{Y} = \mathcal{Z}$.

The main purpose of this paper is to extend the nice features of the metric $\theta(\mathcal{X}, \mathcal{Y})$ on $\mathcal{G}_{1,n}$ to the Grassmann space $\mathcal{G}_{m,n}$ of m -dimensional subspaces of \mathbb{F}^n for $1 \leq m \leq n$. It is well-known that between two members of $\mathcal{G}_{m,n}$, there are m angles, called canonical angles or principal angles. We will define a family of metrics using directly the canonical angles, instead of some trigonometrical functions of the canonical angles as in the current common practice. This strengthens the fundamental role played by the canonical angles in studying the relationship between subspaces. We will then establish some results in the perturbation analysis of the canonical angles and in the robustness of certain geometric relations, namely the nullity (dimension of the intersection) and deficiency (codimension of the sum), between a pair of subspaces using the new metrics defined. We will justify that the new family of metrics do have advantages over the existing ones.

The demand for a distance measure between linear subspaces arise in countless occasions, ranging from pure mathematics to engineering applications. Examples include perturbation theory of unbounded linear operators [16], perturbation analysis of invariant subspaces, deflating subspaces in generalized eigenvalue problems, and singular subspaces in singular value decompositions [3, 28, 27], optimization [6], robust control [9, 24, 25, 32], geometric approach to linear multivariable control [17], system identification [33], signal processing [21, 4], so on and so forth. Our study was mainly motivated from robust control of linear feedback systems. Roughly speaking a linear time-invariant system can be described by a subspace valued frequency function, and the description of an uncertain system calls for a suitable distance measure between subspaces. The stability of a feedback system can be restated as the complement property of two frequency dependent subspaces. Hence the robust stability deals with the robustness of subspace complement, motivating the study of robustness of subspace nullity and deficiency, which is also done in this paper. Not all the distance measures used in above works are metrics, i.e., some of them do not satisfy the triangle inequality, but metrics are certainly preferred from both theoretical and practical view points. Indeed most of the measures used are metrics and most of them are defined from certain trigonometrical functions of the canonical angles between the subspaces. Among the possible metrics, the ones with tight triangle inequalities are most desirable, since they give the tightest perturbation bounds, best estimation of errors, etc. The main ramification of this paper is that the newly defined family of metrics does offer the desired advantages.

In addition to being used to measure the distance between subspaces, the canonical angles also have other utilities. Notably, they can give all the singular values of an oblique projection ([10] related the largest singular value with the smallest canonical angle, but the relation can be extended to all singular values), they are used to study the volumes of matrices [23], they are used in statistics to define the so-called canonical correlations [13], which have attracted further mathematical treatment [11] and have wide applications (see those mentioned in [11]), and they are used in characterizing other geometric relations between subspaces such as nullity, deficiency, complementarity, and so on [16, 37]. Therefore, the perturbation analysis of the canonical angles themselves is of importance and has been conducted in works [29, 11, 18]. We will carry out the analysis using the newly defined family of metrics.

In the next section, we present the theory of canonical angles on a pair of subspaces in $\mathcal{G}_{m,n}$ and we develop some relationships among the canonical angles between three different subspaces, which constitute the key technical vehicle for the proof of our main results. In Section 3, we prove that any symmetric gauge function of the canonical angles defines a unitarily invariant metric; we also prove several results on approximation and perturbation of subspaces as well as canonical angles using the newly defined metrics. In Section 4, we analyze and compare the equality cases of the triangular inequalities for several unitarily invariant metrics. Further results and extensions will be discussed in Section 5.

We will use $\mathcal{M}_{m,n}(\mathbb{F})$ (respectively, $\mathcal{M}_n(\mathbb{F})$) to denote the linear space of $m \times n$ (respectively, $n \times n$) matrices over \mathbb{F} . The group of complex unitary or real orthogonal matrices is denoted by $\mathcal{U}_n(\mathbb{F})$. We will identify a matrix in $\mathcal{M}_n(\mathbb{F})$ and a linear transformation on \mathbb{F}^n . We often abbreviate the notations to $\mathcal{M}_{m,n}, \mathcal{M}_n$ and \mathcal{U}_n if the statements are valid for both $\mathbb{F} = \mathbb{C}$ or \mathbb{R} .

2 Canonical angles between subspaces

The *singular values* of $A \in \mathcal{M}_{m,n}$ are the nonnegative square roots of the l largest eigenvalues of positive semi-definite matrix $A^H A$, where $l = \min\{m, n\}$, and are denoted by $\sigma_1(A) \geq \dots \geq \sigma_l(A)$. We also use the more compact notation $\sigma(A)$ to denote the decreasingly ordered l -tuple of singular values of A .

Let $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$. Suppose the columns of $X_1, Y_1 \in \mathcal{M}_{n,m}$ form orthonormal bases for \mathcal{X} and \mathcal{Y} respectively. Then the singular values of $X_1^H Y_1$ lie in $[0, 1]$ and are independent of the particular choices of the orthonormal bases. We define the *canonical angles* between \mathcal{X} and \mathcal{Y} to be

$$\theta_i(\mathcal{X}, \mathcal{Y}) = \arccos \sigma_{m-i+1}(X_1^H Y_1) \in [0, \pi/2], \quad i = 1, 2, \dots, m.$$

In the following, we denote the decreasingly ordered m -tuple of canonical angles between \mathcal{X} and \mathcal{Y} by

$$\theta(\mathcal{X}, \mathcal{Y}) := (\theta_1(\mathcal{X}, \mathcal{Y}), \dots, \theta_m(\mathcal{X}, \mathcal{Y})).$$

Let \mathcal{X}_\perp and \mathcal{Y}_\perp be the orthogonal complements of \mathcal{X} and \mathcal{Y} , respectively. Suppose $X_2, Y_2 \in \mathcal{M}_{n,n-m}$ have columns forming orthonormal bases for \mathcal{X}_\perp and \mathcal{Y}_\perp , respectively. Then $X = [X_1 \ X_2]$ and $Y = [Y_1 \ Y_2]$ are unitary matrices. By the CS decomposition of matrices in \mathcal{U}_n (e.g., see Theorem 5.2 in [27]), there are $E = \text{diag}(E_1, E_2), F = \text{diag}(F_1, F_2) \in \mathcal{U}_n$ with $E_1, F_1 \in \mathcal{U}_m$ such that

$$E^H X^H Y F = \text{diag} \left(I_{m-l}, \begin{bmatrix} \Gamma & -\Sigma \\ \Sigma & \Gamma \end{bmatrix}, I_{n-m-l} \right),$$

where $l = \min\{m, n-m\}$, Γ and Σ are diagonal matrices with diagonal entries in $[0, 1]$ satisfying $\Gamma^2 + \Sigma^2 = I_l$. Hence the diagonal entries of Γ are the cosines of the first l canonical angles between \mathcal{X} and \mathcal{Y} ; the diagonal entries of Σ are the sines of the first l canonical angles. The rest $m-l$ canonical angles between \mathcal{X} and \mathcal{Y} are always zero. This leads to several observations:

- (1) The nonzero canonical angles between \mathcal{X} and \mathcal{Y} are the same as those of \mathcal{X}_\perp and \mathcal{Y}_\perp .
- (2) The singular values of $X_2^H Y_1$ are $\sin \theta_1(\mathcal{X}, \mathcal{Y}) \geq \dots \geq \sin \theta_l(\mathcal{X}, \mathcal{Y})$, and so are those of $X_1^H Y_2$.
- (3) The canonical angles are unitary invariant, i.e., $\theta(U\mathcal{X}, U\mathcal{Y}) = \theta(\mathcal{X}, \mathcal{Y})$ for all $U \in \mathcal{U}_n$, since $(UX_1)^H(UY_1) = X_1^H Y_1$.

In view of the above observations, we can always assume $m \leq n/2$ in our discussion; otherwise, we can consider the canonical angles between \mathcal{X}_\perp and \mathcal{Y}_\perp instead.

Jordan [15] discovered the canonical angles shortly after he discovered the matrix singular values [14]. In statistics, Hotelling [13] used canonical angles to define the canonical correlations shortly after he used the singular values to define the principal components [12]. Also see, for example, [3], [35], [2].

The CS decomposition also yields a special unitary transformation mapping \mathcal{X} to \mathcal{Y} , called a direct rotation from \mathcal{X} to \mathcal{Y} [3]. Set $\hat{X} = XE$ and $\hat{Y} = YF$. We have

$$\hat{X}^H \hat{Y} = \begin{bmatrix} \Gamma & -\Sigma & 0 \\ \Sigma & \Gamma & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix}. \quad (1)$$

A direct rotation from \mathcal{X} to \mathcal{Y} is given by

$$R = \hat{X} \begin{bmatrix} \Gamma & -\Sigma & 0 \\ \Sigma & \Gamma & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix} \hat{X}^H = \hat{X} \exp \left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} \right) \hat{X}^H, \quad (2)$$

where $A = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Y}) & 0_{m, n-2m} \end{bmatrix}$. It is shown in [3] that among all unitary transformations mapping \mathcal{X} to \mathcal{Y} , the direct rotation is the ‘‘most economic’’ in some sense. We will revisit this point in Section 5.

Similarly, if we have another subspace $\mathcal{Z} \in \mathcal{G}_{m,n}$, then a direct rotation from \mathcal{Y} to \mathcal{Z} is given by

$$S = \bar{Y} \exp \left(\begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} \right) \bar{Y}^H$$

where $B = \begin{bmatrix} \text{diag } \theta(\mathcal{Y}, \mathcal{Z}) & 0_{m, n-2m} \end{bmatrix}$, $\bar{Y} = \begin{bmatrix} \bar{Y}_1 & \bar{Y}_2 \end{bmatrix} \in \mathcal{U}_n$, and the columns of \bar{Y}_1 and \bar{Y}_2 form special orthonormal bases of \mathcal{Y} and \mathcal{Y}_\perp respectively. Also a direct rotation mapping \mathcal{X} to \mathcal{Z} is given by

$$T = \tilde{X} \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) \tilde{X}^H$$

where $C = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Z}) & 0_{m, n-2m} \end{bmatrix}$, $\tilde{X} = \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 \end{bmatrix} \in \mathcal{U}_n$, and the columns of \tilde{X}_1 and \tilde{X}_2 form special orthonormal bases of \mathcal{X} and \mathcal{X}_\perp respectively.

Since SR also takes \mathcal{X} onto \mathcal{Z} , it can only differ from T locally in \mathcal{X} and \mathcal{X}_\perp [3], i.e., there exists unitary matrix $Q = \text{diag}(Q_1, Q_2)$ with $Q_1 \in \mathcal{U}_m$ and $Q_2 \in \mathcal{U}_{n-m}$ such that

$$\tilde{X} \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) Q \tilde{X}^H = SR,$$

which is equivalent to

$$\exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) Q = \tilde{X}^H \bar{Y} \exp \left(\begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} \right) \bar{Y}^H \tilde{X} \tilde{X}^H \hat{X} \exp \left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} \right) \hat{X}^H \tilde{X}.$$

Denote $\tilde{X}^H \tilde{Y}$ by M and $\tilde{X}^H \hat{X}$ by N . Notice that $N = \text{diag}(N_1, N_2)$ for some $N_1 \in \mathcal{U}_m$ and $N_2 \in \mathcal{U}_{n-m}$. Hence

$$\exp\left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix}\right) Q = M \exp\left(\begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix}\right) M^H N \exp\left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix}\right) N^H. \quad (3)$$

Taking inverse, we get

$$Q^H \exp\left(\begin{bmatrix} 0 & C \\ -C^H & 0 \end{bmatrix}\right) = N \exp\left(\begin{bmatrix} 0 & A \\ -A^H & 0 \end{bmatrix}\right) N^H M \exp\left(\begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix}\right) M^H.$$

Multiplying $J = \text{diag}(I_m, -I_{n-m})$ from both sides and noticing that J commutes with both Q and N , we get

$$Q^H \exp\left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix}\right) = N \exp\left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix}\right) N^H J M \exp\left(\begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix}\right) M^H J. \quad (4)$$

Multiplying (3) and (4), we obtain

$$\begin{aligned} & \exp\left(\begin{bmatrix} 0 & -2C \\ 2C^H & 0 \end{bmatrix}\right) \\ &= M \exp\left(\begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix}\right) M^H N \exp\left(\begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix}\right) N^H J M \exp\left(\begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix}\right) M^H J \\ &= \exp\left(M \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} M^H\right) \exp\left(N \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} N^H\right) \exp\left(J M \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} M^H J\right). \end{aligned}$$

The product of matrix exponentials on the right hand side can be transformed to the exponential of a sum by using a matrix exponential formula of Thompson [31] (also see [1]), which asserts that for skew-Hermitian matrices $S, T \in \mathcal{M}_n$, there exist unitary matrices $U, V \in \mathcal{U}_n(\mathbb{C})$ such that

$$\exp(S) \exp(T) = \exp(USU^H + VTV^H).$$

This leads to the following relation of the canonical angles.

Proposition 1 *There exist $U_1, U_2, V \in \mathcal{U}_n(\mathbb{C})$ such that*

$$\exp\left(\begin{bmatrix} 0 & -2C \\ 2C^H & 0 \end{bmatrix}\right) = \exp\left(V \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} V^H + U_1 \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} U_1^H + U_2 \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} U_2^H\right). \quad (5)$$

Note that even when $\mathbb{F} = \mathbb{R}$, we have the above relations of canonical angles among three subspaces in terms of complex skew-Hermitian and unitary matrices.

3 Unitarily invariant metrics

A norm $\|\cdot\|$ on $\mathcal{M}_{m,n}$ is *unitarily invariant* if $\|UAV\| = \|A\|$ for any $U \in \mathcal{U}_m$, $V \in \mathcal{U}_n$ and $A \in \mathcal{M}_{m,n}$. It is immediately seen from the singular value decomposition that a unitary invariant norm on $\mathcal{M}_{m,n}$ depends only on the singular values of matrices.

Unitarily invariant norms are intimately related to symmetric gauge functions. A *symmetric gauge function* $\Phi: \mathbb{R}^l \rightarrow \mathbb{R}$ is a norm function satisfying additional properties that it is symmetric, i.e.,

$$\Phi(P\xi) = \Phi(\xi)$$

for any $\xi \in \mathbb{R}^l$ and permutation matrix P and that it is absolute, i.e., $\Phi(|\xi|) = \Phi(\xi)$. The absolute property can be replaced by the monotone property, i.e., $\Phi(|\xi|) \leq \Phi(|\zeta|)$ if $|\xi| \leq |\zeta|$ in an elementwise sense. This follows from the well-known fact that a norm function is absolute if and only if it is monotone. A particular class of symmetric gauge functions, called Ky Fan functions [7], plays an important role. Let $1 \leq k \leq l$. The Ky Fan k -function is defined as

$$\Phi_k(\xi_1, \dots, \xi_l) = \max_{1 \leq i_1 \leq \dots \leq i_k \leq l} \{|\xi_{i_1}| + |\xi_{i_2}| + \dots + |\xi_{i_k}|\}.$$

The reason why the class of Ky Fan k -functions is special lies in the fact that for a pair of vectors $\xi, \zeta \in \mathbb{R}^l$, $\Phi_k(\xi) \leq \Phi_k(\zeta)$ for all k implies that $\Phi(\xi) \leq \Phi(\zeta)$ for all symmetric gauge functions Φ . When $\Phi_k(\xi) \leq \Phi_k(\zeta)$ for all $1 \leq k \leq l$, we say that ζ weakly majorizes ξ .

Unitarily invariant norms as well as their interplay with symmetric gauge functions have been very well understood. We list a few results that serve as models for our development of unitarily invariant metrics. For a more complete coverage, see [27].

- (I) [34, von Neumann] There is a one-one correspondence between unitarily invariant norms $\|\cdot\|$ on $\mathcal{M}_{m,n}$ and symmetric gauge functions $\Phi : \mathbb{R}^l \rightarrow \mathbb{R}$, where $l = \min\{m, n\}$, given by

$$\|A\| = \Phi(\sigma(A)).$$

- (II) [22, Mirsky] For the unitarily invariant norm $\|\cdot\|$ on $\mathcal{M}_{m,n}$ corresponding to the symmetric gauge function Φ ,

$$\Phi(\sigma(A) - \sigma(B)) \leq \|A - B\|$$

for all $A, B \in \mathcal{M}_{m,n}$.

- (III) [5, Eckart and Young] [22, Mirsky] [26, Schmidt] Let $\|\cdot\|$ be a unitarily invariant norm on $\mathcal{M}_{m,n}$, and $1 \leq k \leq l = \min\{m, n\}$. Suppose $A \in \mathcal{M}_{m,n}$. Then

$$\min\{\|A - B\| : B \in \mathcal{M}_{m,n}, \text{rank}(B) \leq k\} = \Phi(0, \dots, 0, \sigma_{k+1}(A), \dots, \sigma_l(A)).$$

Specializing the triangle inequality of the norms in (I) to Ky Fan k -functions yields that for any $A, B, C \in \mathcal{M}_{m,n}$ with $C = A + B$,

$$\sum_{j=1}^k \sigma_j(C) \leq \sum_{j=1}^k \sigma_j(A) + \sum_{j=1}^k \sigma_j(B). \quad (6)$$

Specializing the Mirsky result (II) to Ky Fan k -functions yields that for any $A, B, C \in \mathcal{M}_{m,n}$ with $C = A + B$,

$$\sum_{j=1}^k \sigma_{i_j}(C) \leq \sum_{j=1}^k \sigma_{i_j}(A) + \sum_{j=1}^k \sigma_j(B) \quad (7)$$

for any $1 \leq i_1 < \dots < i_k \leq l = \min\{m, n\}$. In this sense, the inequalities in (II) are more general than the triangle inequalities of the norms in (I). Recently, the complete set of inequalities relating the singular values of matrices $A, B, C \in \mathcal{M}_{m,n}$ with $C = A + B$ has been determined; e.g., see [8].

One can use symmetric gauge functions to define unitarily invariant metrics on $\mathcal{G}_{m,n}$. Suppose $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$ is a symmetric gauge function and $\|\cdot\|$ is the corresponding unitarily invariant norm on $\mathcal{M}_{n,m}$. For $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$, let $X_1, Y_1 \in \mathcal{M}_{n,m}$ have columns forming orthonormal bases for \mathcal{X} and \mathcal{Y} . Two families of frequently used unitarily invariant metrics between \mathcal{X} and \mathcal{Y} are *the gap metrics*

$$\Phi(\sin \theta_1(\mathcal{X}, \mathcal{Y}), \dots, \sin \theta_m(\mathcal{X}, \mathcal{Y})) = \inf\{\|X_1 - Y_1 Q\| : Q \in \mathcal{M}_m\}, \quad (8)$$

and the Hausdorff metrics

$$\Phi(2\sin(\theta_1(\mathcal{X}, \mathcal{Y})/2), \dots, 2\sin(\theta_m(\mathcal{X}, \mathcal{Y})/2)) = \inf\{\|X_1 - Y_1 V\| : V \in \mathcal{U}_m\}, \quad (9)$$

which extend the metrics (b) and (c) for the case when $m = 1$ introduced in Section 1. As we have motivated in the case when $m = 1$, it is desirable to have an extension of metric (a). Potentially, the metrics defined by the canonical angles, instead of their sines or other functions, can be used to generate other metrics and have tighter triangle inequalities. It appears to be a long-established fact in the differential geometry literature [38, 6] that

$$\sqrt{\theta_1^2(\mathcal{X}, \mathcal{Y}) + \dots + \theta_m^2(\mathcal{X}, \mathcal{Y})}$$

is a metric. In [35], Wedin showed that the largest canonical angle $\theta_1(\mathcal{X}, \mathcal{Y})$ is a metric. Qiu and Davison [25] and Vinnicombe [32] also gave alternative proofs for this result in the engineering literature and demonstrated that the use of $\theta_1(\mathcal{X}, \mathcal{Y})$ as a metric is crucial in engineering applications. Recently, this fact was connected to a Finsler geometry of the Grassmann manifold [36]. It has been conjectured that any symmetric gauge function of the canonical angles is a metric. This conjecture was explicitly announced by the first author at the 10-th ILAS Conference in 2002. Using the results in the last section, we are now able to prove the conjecture.

Theorem 1 *Let $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$ be a symmetric gauge function. Define $\rho : \mathcal{G}_{m,n} \times \mathcal{G}_{m,n} \rightarrow \mathbb{R}$ by*

$$\rho(\mathcal{X}, \mathcal{Y}) = \Phi(\theta(\mathcal{X}, \mathcal{Y})).$$

Then ρ is a unitarily invariant metric.

Proof: Evidently, $\rho(\mathcal{X}, \mathcal{Y})$ is unitarily invariant and satisfies the positivity condition (i) and the symmetry condition (ii) in Section 1. It remains to show that it satisfies the triangle inequality (iii). As shown in Section 2, the canonical angles between subspaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ satisfy the relationship (5). Hence the eigenvalues of

$$\begin{bmatrix} 0 & -2C \\ 2C^H & 0 \end{bmatrix} \quad (10)$$

and those of

$$V \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} V^H + U_1 \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} U_1^H + U_2 \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} U_2^H \quad (11)$$

are equal modulo $\pm i2\pi$, where $i = \sqrt{-1}$. Since the eigenvalues values of (10) belong to $i[-\pi, \pi]$ and those of (11) are known to belong to $i[-2\pi, 2\pi]$, each eigenvalue of (11) either is equal to its corresponding eigenvalue of (10) or belongs to $i[-2\pi, -\pi] \cup i[\pi, 2\pi]$, hence having no less absolute value than its corresponding eigenvalue of (10). Since both matrix (10) and matrix (11) are skew-Hermitian, their singular values are the absolute values of their eigenvalues. Therefore, by the monotone property of the symmetric gauge function,

$$\begin{aligned} \left\| \begin{bmatrix} 0 & -2C \\ 2C^H & 0 \end{bmatrix} \right\| &\leq \left\| V \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} V^H + U_1 \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} U_1^H + U_2 \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} U_2^H \right\| \\ &\leq \left\| \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} \right\| + 2 \left\| \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} \right\| \end{aligned}$$

for all unitarily invariant norms $\|\cdot\|$ on \mathcal{M}_n . Hence

$$\|C\| \leq \|A\| + \|B\|$$

for all unitarily invariant norms $\|\cdot\|$ on $\mathcal{M}_{m,n-m}$. This is equivalent to

$$\rho(\mathcal{X}, \mathcal{Z}) \leq \rho(\mathcal{X}, \mathcal{Y}) + \rho(\mathcal{Y}, \mathcal{Z})$$

for all symmetric gauge functions Φ on \mathbb{R}^m . \square

The metrics defined by the symmetric gauge functions of the canonical angles as in Theorem 1 will be called the angular metrics. We designate the notation ρ for such a metric.

By continuity, slight perturbations in subspaces will result in slight perturbations in the canonical angles in between. How are the perturbations in the canonical angles bounded by the perturbations in the subspaces involved? The same question was asked in [29] and the following ‘‘Mirsky type result’’ gives an answer.

Theorem 2 *Let ρ be the angular metric corresponding to symmetric gauge function Φ . Then for $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m,n}$,*

$$\Phi(\theta(\mathcal{X}, \mathcal{Z}) - \theta(\mathcal{X}, \mathcal{Y})) \leq \rho(\mathcal{Y}, \mathcal{Z}) = \Phi(\theta(\mathcal{Y}, \mathcal{Z})).$$

Proof: Let $\sigma_1 \geq \dots \geq \sigma_n$ be singular values of

$$\Sigma := V \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} V^H + U_1 \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} U_1^H + U_2 \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} U_2^H.$$

Then it follows from the argument in the proof of Theorem 1 that there are singular values $c_i, i = 1, 2, \dots, n$, ordered properly, of

$$\begin{bmatrix} 0 & -2C \\ 2C^H & 0 \end{bmatrix}$$

such that either $\sigma_i = c_i$ or $\sigma_i = 2\pi - c_i$. Rewrite $\alpha_i = \theta_i(\mathcal{X}, \mathcal{Y})$, $\beta_i = \theta_i(\mathcal{Y}, \mathcal{Z})$ and $\gamma_i = \theta_i(\mathcal{X}, \mathcal{Z})$ for $i = 1, 2, \dots, m$. Since c_i is a reordering of $2\gamma_1, 2\gamma_1, \dots, 2\gamma_m, 2\gamma_m, 0, \dots, 0$, we have for any symmetric gauge function $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\begin{aligned} & \Psi(2\gamma_1 - 2\alpha_1, 2\gamma_1 - 2\alpha_1, \dots, 2\gamma_m - 2\alpha_m, 2\gamma_m - 2\alpha_m, 0, \dots, 0) \\ & \leq \Psi(c_1 - 2\alpha_1, c_2 - 2\alpha_1, \dots, c_{2m-1} - 2\alpha_m, c_{2m} - 2\alpha_m, c_{2m+1}, \dots, c_n) \\ & \leq \Psi(\sigma_1 - 2\alpha_1, \sigma_2 - 2\alpha_1, \dots, \sigma_{2m-1} - 2\alpha_m, \sigma_{2m} - 2\alpha_m, \sigma_{2m+1}, \dots, \sigma_n). \end{aligned}$$

Here we used the fact that $\sigma_i - \eta$ either is equal to $c_i - \eta$ or has no less absolute value than $c_i - \eta$ for any $\eta \in [-\pi, \pi]$. By (II), we obtain

$$\begin{aligned} & \Psi(2\gamma_1 - 2\alpha_1, 2\gamma_1 - 2\alpha_1, \dots, 2\gamma_m - 2\alpha_m, 2\gamma_m - 2\alpha_m, 0, \dots, 0) \\ & \leq \left\| \Sigma - V \begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} V^H \right\|_{\Psi} \\ & = \left\| U_1 \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} U_1^H + U_2 \begin{bmatrix} 0 & B \\ -B^H & 0 \end{bmatrix} U_2^H \right\|_{\Psi} \\ & \leq 2 \left\| \begin{bmatrix} 0 & -B \\ B^H & 0 \end{bmatrix} \right\|_{\Psi}. \end{aligned}$$

Hence

$$\Phi(\gamma_1 - \alpha_1, \dots, \gamma_m - \alpha_m) \leq \Phi(\beta_1, \dots, \beta_m)$$

for all symmetric gauge function Φ on \mathbb{R}^m , which immediately gives

$$\Phi(\theta(\mathcal{X}, \mathcal{Z}) - \theta(\mathcal{X}, \mathcal{Y})) \leq \rho(\mathcal{Y}, \mathcal{Z}). \quad \square$$

If we specialize the metrics in Theorem 1 to the ones corresponding to the Ky Fan k -functions, the triangle inequality gives us

$$\sum_{j=1}^k \theta_j(\mathcal{X}, \mathcal{Z}) \leq \sum_{j=1}^k \theta_j(\mathcal{X}, \mathcal{Y}) + \sum_{j=1}^k \theta_j(\mathcal{Y}, \mathcal{Z}). \quad (12)$$

On the other hand, specializing Theorem 2 to the Ky Fan functions gives us

$$\sum_{j=1}^k \theta_{i_j}(\mathcal{X}, \mathcal{Z}) \leq \sum_{j=1}^k \theta_{i_j}(\mathcal{X}, \mathcal{Y}) + \sum_{j=1}^k \theta_j(\mathcal{Y}, \mathcal{Z}) \quad (13)$$

for any $1 \leq i_1 < \dots < i_k \leq m$. In this sense, the inequalities in Theorem 2 cover the triangle inequalities in Theorem 1 as special cases.

Sometimes we are concerned with the perturbation of canonical angles when both subspaces involved are perturbed. In this case, the following version of Theorem 2 might be more convenient.

Corollary 1 *Let ρ be the angular metric corresponding to symmetric gauge function Φ . Then for $\mathcal{X}, \mathcal{Y}, \tilde{\mathcal{X}}, \tilde{\mathcal{Y}} \in \mathcal{G}_{m,n}$,*

$$\Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y})) \leq \rho(\tilde{\mathcal{X}}, \mathcal{X}) + \rho(\tilde{\mathcal{Y}}, \mathcal{Y}). \quad (14)$$

Proof: For any subspaces \mathcal{X}, \mathcal{Y} and the perturbations of them $\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}$, it is easy to show that

$$\begin{aligned} \Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y})) &= \Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \tilde{\mathcal{Y}}) + \theta(\mathcal{X}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y})) \\ &\leq \Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \tilde{\mathcal{Y}})) + \Phi(\theta(\mathcal{X}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y})) \\ &\leq \rho(\tilde{\mathcal{X}}, \mathcal{X}) + \rho(\tilde{\mathcal{Y}}, \mathcal{Y}). \quad \square \end{aligned}$$

Sun in [29] established certain bounds on $\Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y}))$ for the case when Φ is the Ky Fan 1-function (Hölder ∞ -norm) and when it is the Hölder 2-norm. Corollary 1 extends and improves the bounds in [29].

However, the analysis in [29] is more general in one direction: it applies to canonical angles between two subspaces with possibly different dimensions, which are also important in applications. Let us first extend the notion of canonical angles to this more general setting. For $\mathcal{X} \in \mathcal{G}_{m,n}$ and $\mathcal{Y} \in \mathcal{G}_{l,n}$, again let $X_1 \in \mathcal{M}_{n,m}$ and $Y_1 \in \mathcal{M}_{n,l}$ be matrices whose columns form orthonormal bases of \mathcal{X} and \mathcal{Y} respectively. The canonical angles between \mathcal{X} and \mathcal{Y} are defined to be

$$\theta_i(\mathcal{X}, \mathcal{Y}) = \arccos \sigma_{\min\{m,l\}-i+1}(X_1^H Y_1), \quad i = 1, 2, \dots, \min\{m, l\}.$$

Let us take a closer look at this definition for the case when $m < l$. By the singular value decomposition, we know that there exist $E_1 \in \mathcal{U}_m$ and $F_1 \in \mathcal{U}_l$ such that

$$E_1^H X_1^H Y_1 F_1 = \begin{bmatrix} \text{diag}(\cos \theta_1(\mathcal{X}, \mathcal{Y}), \dots, \cos \theta_m(\mathcal{X}, \mathcal{Y})) & 0_{m, l-m} \end{bmatrix}.$$

This shows that \mathcal{Y} has an m -dimensional subspace \mathcal{Y}_0 such that $\theta(\mathcal{X}, \mathcal{Y}_0) = \theta(\mathcal{X}, \mathcal{Y})$ and $\mathcal{X} \perp (\mathcal{Y} \ominus \mathcal{Y}_0)$. A similar interpretation can be made for the case when $l < m$. Now we would like to ask whether Corollary 1 can be extended to the case when $\mathcal{X}, \tilde{\mathcal{X}} \in \mathcal{G}_{m,n}$ and $\mathcal{Y}, \tilde{\mathcal{Y}} \in \mathcal{G}_{l,n}$ after reconciling the two metrics in the right hand side of (14) properly. We are not able to give an affirmative answer at this moment, though all indications show that this is possible. We nevertheless are able to treat another important special case when $l = n - m$. It is this case that is directly relevant to robust control [24, 37]. In this case, there are at most $\min\{m, n - m\}$ nonzero canonical angles between any two members of $\mathcal{G}_{\max\{m, n-m\}, n}$. Hence a symmetric gauge function Φ on $\mathbb{R}^{\min\{m, n-m\}}$ also defines a metric on $\mathcal{G}_{\max\{m, n-m\}, n}$ by applying to the $\min\{m, n - m\}$ largest canonical angles. We will denote the metric on $\mathcal{G}_{m,n}$ and $\mathcal{G}_{n-m,n}$ defined by Φ using the same symbol ρ . We then have $\rho(\mathcal{X}, \mathcal{Y}) = \rho(\mathcal{X}_\perp, \mathcal{Y}_\perp)$ for $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$.

Corollary 2 Let ρ be the angular metric on $\mathcal{G}_{m,n}$ and that $\mathcal{G}_{n-m,n}$ corresponding to symmetric gauge function Φ on $\mathbb{R}^{\min\{m,n-m\}}$. Then for $\mathcal{X}, \tilde{\mathcal{X}} \in \mathcal{G}_{m,n}$ and $\mathcal{Y}, \tilde{\mathcal{Y}} \in \mathcal{G}_{n-m,n}$

$$\Phi(\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) - \theta(\mathcal{X}, \mathcal{Y})) \leq \rho(\tilde{\mathcal{X}}, \mathcal{X}) + \rho(\tilde{\mathcal{Y}}, \mathcal{Y}).$$

Proof: It follows immediately from the fact that

$$\theta_i(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) = \pi/2 - \theta_{\min\{m,n-m\}-i+1}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}_\perp) \text{ and } \theta_i(\mathcal{X}, \mathcal{Y}) = \pi/2 - \theta_{\min\{m,n-m\}-i+1}(\mathcal{X}, \mathcal{Y}_\perp). \quad \square$$

In the rest of this section, we work towards a metric counterpart of the result in (III) on the robustness of the matrix rank. Following [16], we define the nullity and deficiency of subspaces \mathcal{X} and \mathcal{Y} to be

$$\text{nul}(\mathcal{X}, \mathcal{Y}) := \dim(\mathcal{X} \cap \mathcal{Y}) \quad \text{and} \quad \text{def}(\mathcal{X}, \mathcal{Y}) := \text{codim}(\mathcal{X} + \mathcal{Y}), \text{ respectively.}$$

Clearly, $\text{def}(\mathcal{X}, \mathcal{Y}) = \text{nul}(\mathcal{X}_\perp, \mathcal{Y}_\perp)$ and in the case when $\dim \mathcal{X} + \dim \mathcal{Y} = n$, we have $\text{def}(\mathcal{X}, \mathcal{Y}) = \text{nul}(\mathcal{X}, \mathcal{Y})$. Also, $\mathcal{X} \oplus \mathcal{Y} = \mathbb{F}^n$ if and only if $\text{nul}(\mathcal{X}, \mathcal{Y}) = 0$ and $\text{def}(\mathcal{X}, \mathcal{Y}) = 0$.

The robustness of the nullity and deficiency is of great interest in mathematics [16], statistics [13], and control theory [9, 25, 32, 37]. In particular, if we have subspaces $\mathcal{X} \in \mathcal{G}_{m,n}$ and $\mathcal{Y} \in \mathcal{G}_{l,n}$ with $\text{nul}(\mathcal{X}, \mathcal{Y}) < k$ for some $1 \leq k \leq \min\{m, l\}$ and if we also know the perturbed versions $\tilde{\mathcal{X}} \in \mathcal{G}_{m,n}$ and $\tilde{\mathcal{Y}} \in \mathcal{G}_{l,n}$ satisfy $\rho_1(\mathcal{X}, \tilde{\mathcal{X}}) \leq \alpha$ and $\rho_2(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \beta$, we wish to obtain the tightest condition on α and β to ensure $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$. The same problem can be considered for the deficiency. The following theorem solves these problems for the special case when $m = l$.

Theorem 3 Let ρ be the angular metric corresponding to symmetric gauge function Φ . Let $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$. Then for $\alpha \geq 0$, $\beta \geq 0$, and $1 \leq k \leq m$,

1. $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$ for all $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ satisfying $\rho(\mathcal{X}, \tilde{\mathcal{X}}) \leq \alpha$ and $\rho(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \beta$ if and only if

$$\alpha + \beta < \Phi(0, \dots, 0, \theta_{m-k+1}(\mathcal{X}, \mathcal{Y}), \dots, \theta_m(\mathcal{X}, \mathcal{Y}));$$

2. $\text{def}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$ for all $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ satisfying $\rho(\mathcal{X}, \tilde{\mathcal{X}}) \leq \alpha$ and $\rho(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \beta$ if and only if

$$\alpha + \beta < \Phi(0, \dots, 0, \theta_{m-k+1}(\mathcal{X}_\perp, \mathcal{Y}_\perp), \dots, \theta_m(\mathcal{X}_\perp, \mathcal{Y}_\perp)).$$

Proof: We only need to prove Statement 1. Statement 2 follows from $\text{def}(\mathcal{X}, \mathcal{Y}) = \text{nul}(\mathcal{X}_\perp, \mathcal{Y}_\perp)$. Suppose $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) \geq k$. Then $\theta_j(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) = 0, j = m - k + 1, \dots, m$. By Corollary 1,

$$\begin{aligned} \delta &:= \Phi(0, \dots, 0, \theta_{m-k+1}(\mathcal{X}, \mathcal{Y}), \dots, \theta_m(\mathcal{X}, \mathcal{Y})) \\ &\leq \Phi(\theta(\mathcal{X}, \mathcal{Y}) - \theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})) \leq \rho(\mathcal{X}, \tilde{\mathcal{X}}) + \rho(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \alpha + \beta. \end{aligned}$$

This shows that if

$$\alpha + \beta < \delta,$$

then $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$.

Now assume that $\alpha + \beta \geq \delta$. Then there exist $\alpha_1 \in [0, \alpha]$ and $\beta_1 \in [0, \beta]$ such that $\alpha_1 + \beta_1 = \delta$. By the CS decomposition, there exist $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix}$, where the columns of $X_1 = [x_1 \dots x_m]$ and $X_2 = [x_{m+1} \dots x_n]$ form orthonormal bases of \mathcal{X} and \mathcal{X}_\perp respectively, and the columns of $Y_1 = [y_1 \dots y_m]$ and $Y_2 = [y_{m+1} \dots y_n]$ form orthonormal bases for \mathcal{Y} and \mathcal{Y}_\perp respectively, such that

$$Y = X \begin{bmatrix} \cos(\text{diag } \theta(\mathcal{X}, \mathcal{Y})) & -\sin(\text{diag } \theta(\mathcal{X}, \mathcal{Y})) & 0 \\ \sin(\text{diag } \theta(\mathcal{X}, \mathcal{Y})) & \cos(\text{diag } \theta(\mathcal{X}, \mathcal{Y})) & 0 \\ 0 & 0 & I \end{bmatrix}. \quad (15)$$

Let $\lambda = \alpha_1/\delta$, then $1 - \lambda = \beta_1/\delta$. Also abbreviate $\theta_j(\mathcal{X}, \mathcal{Y})$ by θ_j , $j = 1, 2, \dots, m$. Define

$$\tilde{x}_j = x_j \cos(\lambda\theta_j) + x_{m+j} \sin(\lambda\theta_j), \quad j = m - k + 1, \dots, m,$$

$$\tilde{\mathcal{X}} = \text{span}(x_1, \dots, x_{m-k}, \tilde{x}_{m-k+1}, \dots, \tilde{x}_m) \quad \text{and} \quad \tilde{\mathcal{Y}} = \text{span}(y_1, \dots, y_{m-k}, \tilde{x}_{m-k+1}, \dots, \tilde{x}_m).$$

Also define

$$\tilde{X}_1 = \begin{bmatrix} x_1 & \cdots & x_{m-k} & \tilde{x}_{m-k+1} & \cdots & \tilde{x}_m \end{bmatrix} \quad \text{and} \quad \tilde{Y}_1 = \begin{bmatrix} y_1 & \cdots & y_{m-k} & \tilde{x}_{m-k+1} & \cdots & \tilde{x}_m \end{bmatrix}.$$

Then the columns of \tilde{X}_1 and \tilde{Y}_1 form orthonormal bases of $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ respectively. Furthermore,

$$X_1^H \tilde{X}_1 = \text{diag}(1, \dots, 1, \cos(\lambda\theta_{m-k+1}), \dots, \cos(\lambda\theta_m)).$$

It follows that

$$\rho(\mathcal{X}, \tilde{\mathcal{X}}) = \Phi(0, \dots, 0, \lambda\theta_{m-k+1}, \dots, \lambda\theta_m) = \alpha_1 \leq \alpha.$$

From (15), we see that $y_j = x_j \cos \theta_j + x_{m+j} \sin \theta_j$, $j = m - k + 1, \dots, m$. Hence

$$Y_1^H \tilde{Y}_1 = \text{diag}(1, \dots, 1, \cos((1 - \lambda)\theta_{m-k+1}), \dots, \cos((1 - \lambda)\theta_m)).$$

Similar to the above, we have

$$\rho(\mathcal{Y}, \tilde{\mathcal{Y}}) = \Phi(0, \dots, 0, (1 - \lambda)\theta_{m-k+1}, \dots, (1 - \lambda)\theta_m) = \beta_1 \leq \beta.$$

Since

$$\tilde{Y}_1^H \tilde{X}_1 = \text{diag}(\cos \theta_1, \dots, \cos \theta_{m-k}, 1, \dots, 1),$$

it follows that

$$\theta(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) = (\theta_1, \dots, \theta_{m-k}, 0, \dots, 0);$$

which implies that $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) \geq k$. This proves the necessity of the condition. \square

We believe that carefully formulated similar statements to those in Theorem 3 for the case when $m \neq l$ are also valid, though a proof is not available yet at this moment. Nevertheless, the important special case when $l = n - m$ can be treated as in the proof of Theorem 3 by following Corollary 2 instead. We state this as a corollary.

Corollary 3 *Let ρ be the angular metric on $\mathcal{G}_{m,n}$ and that on $\mathcal{G}_{n-m,n}$ corresponding to a symmetric gauge function Φ on $\mathbb{R}^{\min\{m,n-m\}}$. Let $\mathcal{X} \in \mathcal{G}_{m,n}$ and $\mathcal{Y} \in \mathcal{G}_{n-m,n}$. Then for $\alpha \geq 0$, $\beta \geq 0$, and $1 \leq k \leq \min\{m, n - m\}$,*

1. $\text{nul}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$ for all $\tilde{\mathcal{X}} \in \mathcal{G}_{m,n}$ and $\tilde{\mathcal{Y}} \in \mathcal{G}_{n-m,n}$ satisfying $\rho(\mathcal{X}, \tilde{\mathcal{X}}) \leq \alpha$ and $\rho(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \beta$ if and only if

$$\alpha + \beta < \Phi(0, \dots, 0, \theta_{m-k+1}(\mathcal{X}, \mathcal{Y}), \dots, \theta_{\min\{m,n-m\}}(\mathcal{X}, \mathcal{Y}));$$

2. $\text{def}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) < k$ for all $\tilde{\mathcal{X}} \in \mathcal{G}_{m,n}$ and $\tilde{\mathcal{Y}} \in \mathcal{G}_{n-m,n}$ satisfying $\rho(\mathcal{X}, \tilde{\mathcal{X}}) \leq \alpha$ and $\rho(\mathcal{Y}, \tilde{\mathcal{Y}}) \leq \beta$ if and only if

$$\alpha + \beta < \Phi(0, \dots, 0, \theta_{m-k+1}(\mathcal{X}_\perp, \mathcal{Y}_\perp), \dots, \theta_{\min\{m,n-m\}}(\mathcal{X}_\perp, \mathcal{Y}_\perp)).$$

4 When the Triangle Inequalities Become Equalities

Whenever we have an inequality, the analysis of the cases when the inequality becomes an equality is of special importance. Such analysis often reveals extra insight and structures possessed by the inequality. It also exhibits how sharp the inequality is. In applications, these extreme cases often represent either the worst or the best cases where special attentions are needed. In the following, the equality cases for the triangle inequalities for the angular metrics and the gap metrics will be analyzed and compared. They will be characterized in terms of the direct rotations between the subspaces involved, which can be viewed as a natural generalization of the equality case relating three one dimensional subspaces in \mathbb{R}^n . The proofs in this section turn out to be quite technical. We suggest the readers to skip proofs in their first reading.

In this section, we first characterize those triples of subspaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m,n}$ such that

$$\rho(\mathcal{X}, \mathcal{Z}) = \rho(\mathcal{X}, \mathcal{Y}) + \rho(\mathcal{Y}, \mathcal{Z}), \quad (16)$$

i.e., the triangle inequality becomes an equality, either for a particular angular metric or for all angular metrics. An angular metric $\rho = \Phi \circ \theta$ is said to be strictly convex if Φ is a strictly convex symmetric gauge function, i.e., $\Phi(\xi + \zeta) = \Phi(\xi) + \Phi(\zeta)$ only if ξ and ζ are linearly dependent. The following theorem shows that if (16) is satisfied for a strictly convex angular metric, then \mathcal{Y} has to be an intermediate position on the path when \mathcal{X} is rotated to \mathcal{Z} by a direct rotation, which in turn implies that (16) is satisfied for all angular metrics.

Theorem 4 *For $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m,n}$, the following statements are equivalent:*

1. (16) holds for all angular metrics;
2. (16) holds for a strictly convex angular metric;
3. there exists a direct rotation from \mathcal{X} to \mathcal{Z}

$$X \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) X^H,$$

where $C = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Z}) & 0_{m, n-2m} \end{bmatrix}$, and $\lambda \in [0, 1]$ such that

$$\mathcal{Y} = X \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) X^H \mathcal{X}.$$

Proof: It is obvious that statement 1 implies statement 2. We only need to prove that statement 2 implies statement 3 and that 3 implies 1. We do the latter first. If

$$\mathcal{Y} = X \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) X^H \mathcal{X},$$

then a direct rotation from \mathcal{X} to \mathcal{Y} is

$$R = X \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) X^H$$

and one from \mathcal{Y} to \mathcal{Z} is

$$\begin{aligned} S &= \left[X \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) \right] \exp \left(\begin{bmatrix} 0 & -(1-\lambda)C \\ (1-\lambda)C^H & 0 \end{bmatrix} \right) \left[X \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) \right]^H \\ &= X \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) X^H X \exp \left(\begin{bmatrix} 0 & \lambda C \\ -\lambda C^H & 0 \end{bmatrix} \right) X^H. \end{aligned}$$

Then $\theta(\mathcal{X}, \mathcal{Y}) = \lambda\theta(\mathcal{X}, \mathcal{Z})$ and $\theta(\mathcal{Y}, \mathcal{Z}) = (1 - \lambda)\theta(\mathcal{X}, \mathcal{Z})$. This immediately implies (16) for all angular metrics.

To prove that 2 implies 3, let Φ be the strictly convex symmetric gauge function corresponding to the angular metric satisfying (16), and notice that equality (16) and the fact that $\theta(\mathcal{X}, \mathcal{Z})$ is weakly majorized by $\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})$ imply

$$\Phi(\theta(\mathcal{X}, \mathcal{Y})) + \Phi(\theta(\mathcal{Y}, \mathcal{Z})) = \Phi(\theta(\mathcal{X}, \mathcal{Z})) \leq \Phi(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})) \leq \Phi(\theta(\mathcal{X}, \mathcal{Y})) + \Phi(\theta(\mathcal{Y}, \mathcal{Z})).$$

This forces

$$\Phi(\theta(\mathcal{X}, \mathcal{Z})) = \Phi(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})) = \Phi(\theta(\mathcal{X}, \mathcal{Y})) + \Phi(\theta(\mathcal{Y}, \mathcal{Z})).$$

Since Φ is strictly convex, it follows from the second equality above that $\theta(\mathcal{X}, \mathcal{Y})$ and $\theta(\mathcal{Y}, \mathcal{Z})$ are linearly dependent. Also since $\theta(\mathcal{X}, \mathcal{Z})$ is weakly majorized by $\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})$, it follows from [30] that $\theta(\mathcal{X}, \mathcal{Z})$ is in the convex hull of

$$\{P(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})) : P \text{ is a signed permutation matrix}\}.$$

Again since Φ is strictly convex, $\Phi(\theta(\mathcal{X}, \mathcal{Z})) = \Phi(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z}))$ only if $\theta(\mathcal{X}, \mathcal{Z})$ is one of the vertices of the convex hull, i.e., $\theta(\mathcal{X}, \mathcal{Z}) = P(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z}))$ for a particular signed permutation matrix P . Since both of them are nonnegative and are decreasingly ordered, we have $\theta(\mathcal{X}, \mathcal{Z}) = \theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})$. In summary, we conclude that there exists $0 \leq \lambda \leq 1$ such that $\theta(\mathcal{X}, \mathcal{Y}) = \lambda\theta(\mathcal{X}, \mathcal{Z})$ and $\theta(\mathcal{Y}, \mathcal{Z}) = (1 - \lambda)\theta(\mathcal{X}, \mathcal{Z})$.

In the following, we only consider the nontrivial case when $0 < \lambda < 1$. By the discussion in Section 2, we know that there exists a direct rotation from \mathcal{X} to \mathcal{Y} given by

$$R = \hat{X} \exp \left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix} \right) \hat{X}^H$$

and a direct rotation from \mathcal{X} to \mathcal{Z} given by

$$T = \tilde{X} \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) \tilde{X}^H$$

where $\hat{X} = \begin{bmatrix} \hat{X}_1 & \hat{X}_2 \end{bmatrix} \in \mathcal{U}_n$, $\tilde{X} = \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 \end{bmatrix} \in \mathcal{U}_n$, and columns of \hat{X}_1 and those of \tilde{X}_1 both form (possibly different) orthonormal bases of \mathcal{X} . This shows that the columns of the matrices

$$X_1 = \tilde{X}_1, \quad Y_1 = \tilde{X} N \begin{bmatrix} \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix}, \quad Z_1 = \tilde{X} \begin{bmatrix} \cos(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) \\ \sin(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix}$$

form orthonormal bases of \mathcal{X} , \mathcal{Y} and \mathcal{Z} respectively, where $N = \tilde{X}^H \hat{X} = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}$, $N_1 \in \mathcal{U}_m$, $N_2 \in \mathcal{U}_{n-m}$.

By the definition of canonical angles, we know that there exist $E, F \in \mathcal{U}_m$ such that

$$\begin{aligned} & \cos(\text{diag } (1 - \lambda)\theta(\mathcal{X}, \mathcal{Z})) \\ &= EZ_1^H Y_1 F^H \\ &= E \cos(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) N_1 \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) F^H \\ &+ E \begin{bmatrix} \sin(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) & 0_{m, n-2m} \end{bmatrix} N_2 \begin{bmatrix} \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix} F^H. \end{aligned} \tag{17}$$

Let

$$\text{diag } \theta(\mathcal{X}, \mathcal{Z}) = \text{diag}(\phi_1 I_{m_1}, \phi_2 I_{m_2}, \dots, \phi_l I_{m_l}),$$

where $\pi/2 = \phi_1 > \phi_2 > \dots > \phi_l \geq 0$ and $m_1 + \dots + m_l = m$. By [27, Lemma II.3.4], we have

$$\text{Re tr}(E \cos(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) N_1 \cos(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z})) F^H) \leq \sum_{j=1}^l m_j \cos \phi_j \cos(\lambda \phi_j), \quad (18)$$

where $\text{Re tr}(\cdot)$ denotes the real part of the trace. Similarly, we have

$$\text{Re tr} \left(E \begin{bmatrix} \sin(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) & 0_{m, n-2m} \end{bmatrix} N_2 \begin{bmatrix} \sin(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix} F^H \right) \leq \sum_{j=1}^l m_j \sin \phi_j \sin(\lambda \phi_j). \quad (19)$$

From (17), (18), (19) and the fact that

$$\text{Re tr}(\cos(\text{diag}(1 - \lambda)\theta(\mathcal{X}, \mathcal{Z}))) = \sum_{j=1}^l m_j \cos \phi_j \cos(\lambda \phi_j) + \sum_{j=1}^l m_j \sin \phi_j \sin(\lambda \phi_j),$$

it follows that

$$\text{Re tr}(\cos(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) N_1 \cos(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z})) F^H E) = \sum_{j=1}^l m_j \cos \phi_j \cos(\lambda \phi_j) \quad (20)$$

and

$$\text{Re tr} \left(\begin{bmatrix} \sin(\text{diag } \theta(\mathcal{X}, \mathcal{Z})) & 0_{m, n-2m} \end{bmatrix} N_2 \begin{bmatrix} \sin(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix} F^H E \right) = \sum_{j=1}^l m_j \sin \phi_j \sin(\lambda \phi_j).$$

Suppose $D = N_1 \cos(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z})) F^H E$. It follows from the same lines as in the proof of [27, Lemma II.3.4] that $D = \text{diag}(D_1, \dots, D_l)$, where $D_j \in \mathcal{M}_{m_j}$. Then we have

$$\begin{aligned} \text{Re tr}(\cos(\text{diag}(\theta(\mathcal{X}, \mathcal{Z})D))) &= \text{Re tr}(\text{diag}(\cos \phi_1 D_1, \dots, \cos \phi_l D_l)) \\ &= \text{Re tr}(\cos \phi_2 D_2) + \dots + \text{Re tr}(\cos \phi_l D_l) \\ &\leq \sum_{i=1}^{m_2} \cos \phi_2 \sigma_i(D_2) + \dots + \sum_{i=1}^{m_l} \cos \phi_l \sigma_i(D_l) \\ &\leq \sum_{j=2}^l m_j \cos \phi_j \cos(\lambda \phi_j), \end{aligned} \quad (21)$$

where the second last inequality holds since

$$\text{Re tr}(\cos \phi_j D_j) \leq \sum_{i=1}^{m_j} \cos \phi_j \sigma_i(D_j)$$

for each $j = 2, \dots, l$ and the last inequality follows from the fact that the singular values of all D_j , $j = 1, \dots, l$, all together are the singular values of $\cos(\text{diag } \lambda \theta(\mathcal{X}, \mathcal{Z}))$.

Comparing (20) and (21), we see that inequalities in (21) have to be equalities. This is possible only if

$$\text{Re tr}(\cos(\phi_j) D_j) = \sum_{i=1}^{m_j} \cos(\phi_j) \sigma_i(D_j) = m_j \cos \phi_j \cos(\lambda \phi_j)$$

for $j = 2, \dots, l$. So there must exist $P_j \in \mathcal{U}_{m_j}$ such that $D_j = \cos(\lambda\phi_j)P_j$ with $\text{Re tr } P_j = m_j$. Hence $D_j = \cos(\lambda\phi_j)I_{m_j}$, $j = 2, \dots, l$. After determining D_2, \dots, D_l , the singular values D_1 have no choice but to equal to $\cos(\lambda\phi_1)$, i.e., $D_1 = \cos(\lambda\phi_1)W$ for some $W \in \mathcal{U}_{m_1}$. Hence we have proved that

$$\begin{aligned} N_1 \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z}))F^H E &= \text{diag}(\cos(\lambda\phi_1)W, \cos(\lambda\phi_2)I_{m_2}, \dots, \cos(\lambda\phi_l)I_{m_l}) \\ &= \begin{bmatrix} W & 0 \\ 0 & I_{m-m_1} \end{bmatrix} \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})). \end{aligned}$$

Similarly, we can prove that

$$N_2 \begin{bmatrix} \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0 \end{bmatrix} F^H E = \begin{bmatrix} \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0_{n-2m, m} \end{bmatrix}.$$

Define

$$X = \tilde{X} \begin{bmatrix} W & 0 \\ 0 & I_{n-m_1} \end{bmatrix}.$$

Then

$$X \exp\left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix}\right) X^H$$

is also a direct rotation from \mathcal{X} to \mathcal{Z} . We can see that

$$Y_1 F^H E = \tilde{X} \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix} \begin{bmatrix} \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0 \end{bmatrix} F^H E = X \begin{bmatrix} \cos(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ \sin(\text{diag } \lambda\theta(\mathcal{X}, \mathcal{Z})) \\ 0 \end{bmatrix}.$$

This shows

$$\mathcal{Y} = X \exp\left(\begin{bmatrix} 0 & -\lambda C \\ \lambda C^H & 0 \end{bmatrix}\right) X^H \mathcal{X}. \quad \square$$

The following example shows a case when direct rotations from \mathcal{X} to \mathcal{Y} are not unique and hence the paths from \mathcal{X} to \mathcal{Y} they define are not unique.

Example 1 Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m, 2m}$ be spanned by the columns of matrices

$$\begin{bmatrix} I \\ 0 \end{bmatrix}, \quad \begin{bmatrix} (\cos \alpha)U \\ (\sin \alpha)V \end{bmatrix}, \quad \begin{bmatrix} 0 \\ I \end{bmatrix}$$

respectively. Here $\alpha \in [0, \pi/2]$ and $U, V \in \mathcal{U}_m$. Then $\theta(\mathcal{X}, \mathcal{Y}) = (\alpha, \dots, \alpha)$, $\theta(\mathcal{Y}, \mathcal{Z}) = (\pi/2 - \alpha, \dots, \pi/2 - \alpha)$ and $\theta(\mathcal{X}, \mathcal{Z}) = (\pi/2, \dots, \pi/2)$. The triangle equality holds for all angular metrics $\rho = \Phi \circ \theta$ for all α, U, V .

Notice that for the gap metric,

$$\sin \theta_1(\mathcal{X}, \mathcal{Z}) = \sin \theta_1(\mathcal{X}, \mathcal{Y}) + \sin \theta_1(\mathcal{Y}, \mathcal{Z})$$

if and only if $\mathcal{Y} = \mathcal{X}$ or $\mathcal{Y} = \mathcal{Z}$. Also we can show that

$$\Phi(\sin \theta(\mathcal{X}, \mathcal{Z})) = \Phi(\sin \theta(\mathcal{X}, \mathcal{Y})) + \Phi(\sin \theta(\mathcal{Y}, \mathcal{Z}))$$

for a strictly convex symmetric gauge function Φ if and only if $\mathcal{Y} = \mathcal{X}$ or $\mathcal{Y} = \mathcal{Z}$. Hence the triangle equality holds for a gap metric $\Phi \circ \sin \theta$, where Φ is either Φ_1 or any strictly convex symmetric gauge function, only in the trivial cases. The same thing can be said for the Hausdorff metrics.

The Ky Fan k -functions are not strictly convex, so one can expect that $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m,n}$ may take more general positions when (16) is satisfied only for a Ky Fan function. This is indeed the case. It is easy to prove that

$$\Phi_m(\theta(\mathcal{X}, \mathcal{Z})) = \Phi_m(\theta(\mathcal{X}, \mathcal{Y})) + \Phi_m(\theta(\mathcal{Y}, \mathcal{Z}))$$

if there exists a direct rotation from \mathcal{X} to \mathcal{Z}

$$X \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) X^H,$$

where $C = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Z}) & 0_{m,n-2m} \end{bmatrix}$, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ where $\lambda_j \in [0, 1]$ such that

$$\mathcal{Y} = X \exp \left(\begin{bmatrix} 0 & -\Lambda C \\ C^H \Lambda & 0 \end{bmatrix} \right) X^H \mathcal{X}.$$

It would be nice to determine whether the condition is also necessary.

In contrast, for the gap metrics corresponding to the Ky Fan k -functions, we have the following results.

Proposition 2 *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \in \mathcal{G}_{m,n}$. The equality*

$$\Phi_k(\sin \theta(\mathcal{X}, \mathcal{Z})) = \Phi_k(\sin \theta(\mathcal{X}, \mathcal{Y})) + \Phi_k(\sin \theta(\mathcal{Y}, \mathcal{Z})) \quad (22)$$

holds for a particular k with $1 \leq k \leq m$ if and only if the collection

$$\{\theta_1(\mathcal{X}, \mathcal{Y}), \dots, \theta_k(\mathcal{X}, \mathcal{Y}), \theta_1(\mathcal{Y}, \mathcal{Z}), \dots, \theta_k(\mathcal{Y}, \mathcal{Z})\}$$

contains $\theta_1(\mathcal{X}, \mathcal{Z}), \dots, \theta_k(\mathcal{X}, \mathcal{Z})$ together with additional k zeros.

Proof. The sufficiency is clear. To prove the necessity, denote $\alpha_j = \theta_j(\mathcal{X}, \mathcal{Y})$, $\beta_j = \theta_j(\mathcal{Y}, \mathcal{Z})$ and $\gamma_j = \theta_j(\mathcal{X}, \mathcal{Z})$, $j = 1, \dots, m$. Equality (22) becomes

$$\sum_{j=1}^k \sin \gamma_j = \sum_{j=1}^k (\sin \alpha_j + \sin \beta_j). \quad (23)$$

Let the columns of X_1, Y_1, Z_1 form orthonormal bases for subspaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. Denote

$$\begin{aligned} L &= X_1 X_1^H - Y_1 Y_1^H \\ M &= Y_1 Y_1^H - Z_1 Z_1^H \\ N &= X_1 X_1^H - Z_1 Z_1^H. \end{aligned}$$

Then the eigenvalues of L are $\sin \alpha_1, \dots, \sin \alpha_m, -\sin \alpha_m, \dots, -\sin \alpha_1, 0, \dots, 0$. Similar statements hold for M and N . Let $U \in \mathcal{U}_n$ satisfy $UNU^H = \text{diag}(N_1, N_2)$ where $N_1 \in \mathcal{M}_k$ has eigenvalues $\sin \gamma_1, \dots, \sin \gamma_k$. Let L_1 and M_1 be the leading $k \times k$ submatrices of ULU^H and UMU^H respectively. Since

$$ULU^H + UMU^H = UNU^H,$$

it follows that

$$\text{tr } L_1 + \text{tr } M_1 = \text{tr } N_1.$$

It is well-known that eigenvalues of a Hermitian matrix majorize its diagonal elements; see [20] for example. Then

$$\text{tr } L_1 \leq \sum_{j=1}^k \sin \alpha_j, \quad \text{tr } M_1 \leq \sum_{j=1}^k \sin \beta_j. \quad (24)$$

Equality (23) forces the equality signs in (24) to hold. By the extremal property in [19, Lemma 4.1], we have $ULL^H = \text{diag}(L_1, L_2)$ such that L_1 has eigenvalues $\sin \alpha_1, \dots, \sin \alpha_k$ and $UMU^H = \text{diag}(M_1, M_2)$ such that M_1 has eigenvalues $\sin \beta_1, \dots, \sin \beta_k$. Consequently, the unitary Hermitian matrices

$$2UX_1X_1^H U^H - I, \quad 2UY_1Y_1^H U^H - I, \quad 2UZ_1Z_1^H U^H - I,$$

whose differences are block diagonal, must share the same off-diagonal block, i.e., they take the respective forms

$$\begin{bmatrix} X_{11} & D \\ D^H & X_{22} \end{bmatrix}, \quad \begin{bmatrix} Y_{11} & D \\ D^H & Y_{22} \end{bmatrix}, \quad \begin{bmatrix} Z_{11} & D \\ D^H & Z_{22} \end{bmatrix},$$

where $D \in \mathcal{M}_{k, n-k}$. If we replace U by $\text{diag}(V_1, V_2)U$ and properly choose $V_1 \in \mathcal{U}_k$ and $V_2 \in \mathcal{U}_{n-k}$, we can diagonalize D by an SVD. Hence we can assume without loss of generality that $D = \begin{bmatrix} \text{diag}(d_1 I_{r_1}, \dots, d_l I_{r_l}) & 0_{k, n-2k} \end{bmatrix}$ where $0 = d_1 < \dots < d_l \leq 1$.

Since

$$X_{11}X_{11}^H + DD^H = X_{11}^2 + DD^H = I_k,$$

it follows that there exist certain unitary Hermitian matrices P_j and $0 \leq g_l < \dots < g_1 = 1$ satisfying $g_j^2 + d_j^2 = 1$, $j = 1, \dots, l$, such that $X_{11} = \text{diag}(g_1 P_1, \dots, g_l P_l)$. Similarly, $Y_{11} = \text{diag}(g_1 Q_1, \dots, g_l Q_l)$ and $Z_{11} = \text{diag}(g_1 R_1, \dots, g_l R_l)$, where Q_j, R_j are unitary Hermitian matrices.

Consider matrices P_1, Q_1, R_1 . Let $W_1 \in \mathcal{U}_{r_1}$ be chosen such that $W_1 Q_1 W_1^H = \text{diag}(I_{s_1}, -I_{t_1}) := J_{s_1, t_1}$. Since $W_1 P_1 W_1^H - W_1 Q_1 W_1^H \geq 0$, it follows that $W_1 P_1 W_1^H = \text{diag}(I_{s_1}, F_1)$ with $F_1 \in \mathcal{U}_{t_1}$. Similarly since $W_1 Q_1 W_1^H - W_1 R_1 W_1^H \geq 0$, we have $W_1 R_1 W_1^H = \text{diag}(G_1, -I_{t_1})$ with $G_1 \in \mathcal{U}_{s_1}$. Thus,

$$W_1(P_1 - Q_1)W_1^H = \text{diag}(0_{s_1}, F_1 + I_{t_1}), \quad W_1(Q_1 - R_1)W_1^H = \text{diag}(I_{s_1} - G_1, 0_{t_1}).$$

Applying the same arguments to P_j, Q_j, R_j , we have $W_j \in \mathcal{U}_{r_j}$ such that $W_j Q_j W_j^H = J_{s_j, t_j}$ and

$$W_j(P_j - Q_j)W_j^H = \text{diag}(0_{s_j}, F_j + I_{t_j}), \quad W_j(Q_j - R_j)W_j^H = \text{diag}(I_{s_j} - G_j, 0_{t_j})$$

for $j = 2, \dots, l$. Consequently, letting $W = \text{diag}(W_1, \dots, W_l)$ yields

$$\begin{aligned} WL_1 W^H &= \frac{1}{2} \text{diag}(0_{s_1}, F_1 + I_{t_1}, \dots, 0_{s_l}, F_l + I_{t_l}) \\ WM_1 W^H &= \frac{1}{2} \text{diag}(I_{s_1} - G_1, 0_{t_1}, \dots, I_{s_l} - G_l, 0_{t_l}) \end{aligned}$$

where $s = s_1 + \dots + s_l$, $t = t_1 + \dots + t_l$, and $t + s = k$. This gives

$$WN_1 W^H = WL_1 W^H + WM_1 W^H = \frac{1}{2} \text{diag}(I_{s_1} - G_1, F_1 + I_{t_1}, \dots, I_{s_l} - G_l, F_l + I_{t_l}).$$

Now it is easy to see that L_1 has at most t nonzero eigenvalues $\sin \alpha_1, \dots, \sin \alpha_t$ and M_1 has at most s nonzero eigenvalues $\sin \beta_1, \dots, \sin \beta_s$, and they constitute the largest k eigenvalues of N_1 , which are $\sin \gamma_1, \dots, \sin \gamma_k$. This completes the proof. \square

If equality (22) holds for $k = 1$, then either $\mathcal{Y} = \mathcal{X}$ or $\mathcal{Y} = \mathcal{Z}$. This implies that the equality (22) holds for all $1 \leq k \leq m$. If equality (22) holds for any other k but not for $k = 1$, then both $\theta_k(\mathcal{X}, \mathcal{Y}) = \theta_k(\mathcal{Y}, \mathcal{Z}) = 0$. In this case, $\theta_j(\mathcal{X}, \mathcal{Y}) = \theta_j(\mathcal{Y}, \mathcal{Z}) = 0$ for all $j \geq k$. This implies that $\theta_j(\mathcal{X}, \mathcal{Z}) = 0$ for all $j > k$ since otherwise we would have $\sum_{j=1}^{k+1} \theta_j(\mathcal{X}, \mathcal{Z}) > \sum_{j=1}^{k+1} \theta_j(\mathcal{X}, \mathcal{Y}) + \sum_{j=1}^{k+1} \theta_j(\mathcal{Y}, \mathcal{Z})$. This shows that (22) holds when k is replaced by $k+1, k+2, \dots, m$. Because of this, we can define the smallest k with (22) satisfied as k_{\min} . Then $k_{\min} = 1$ if and only if $\mathcal{Y} = \mathcal{X}$ or $\mathcal{Y} = \mathcal{Z}$. When $k_{\min} > 1$, we know that $\theta_{k_{\min}}(\mathcal{X}, \mathcal{Z}) > 0$ but $\theta_{k_{\min}+1}(\mathcal{X}, \mathcal{Z}) = 0$. This means that $\dim \mathcal{X} \cap \mathcal{Z} = m - k_{\min}$.

Theorem 5 *The equality (22) holds for some k with $1 \leq k \leq m$ if and only if there exists a direct rotation from \mathcal{X} to \mathcal{Z}*

$$X \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) X^H,$$

where $C = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Z}) & 0_{m, n-2m} \end{bmatrix}$, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ where $\lambda_j \in \{0, 1\}$ such that

$$\mathcal{Y} = X \exp \left(\begin{bmatrix} 0 & -\Lambda C \\ C^H \Lambda & 0 \end{bmatrix} \right) X^H \mathcal{X}.$$

Proof. The sufficiency is clear. To simplify the proof of the necessity, we will prove the following fact first: if $k_{\min} > 1$, then $\mathcal{X} \cap \mathcal{Z} \subset \mathcal{Y} \subset \mathcal{X} + \mathcal{Z}$. Adopting the same notation as in the proof of Proposition 2, we choose unitary matrix $U^H = \begin{bmatrix} U_1^H & U_2^H & U_3^H \end{bmatrix}$ such that the columns of U_2^H span $\mathcal{X} \cap \mathcal{Z}$ and those of U_3^H span $(\mathcal{X} + \mathcal{Z})_{\perp}$. Then

$$UX_1X_1^H U^H = \text{diag}(\tilde{X}, I, 0), \quad UZ_1Z_1^H U^H = \text{diag}(\tilde{Z}, I, 0),$$

which gives

$$UNU^H = \text{diag}(\tilde{X} - \tilde{Z}, 0, 0)$$

where $\tilde{X} - \tilde{Z}$ has eigenvalues $\sin \gamma_1, \dots, \sin \gamma_{k_{\min}}, -\sin \gamma_{k_{\min}}, \dots, -\sin \gamma_1$. Furthermore U_1 can be chosen so that

$$UNU^H = \text{diag}(N_{11}, N_{22}, 0, 0),$$

$N_{11} > 0$ and $N_{22} < 0$. Because of

$$ULU^H + UMU^H = UNU^H,$$

using twice the majorization and extremal property argument as in the beginning of the proof of Proposition 2, we can see that

$$ULU^H = \text{diag}(L_{11}, L_{22}, 0, 0), \quad UMU^H = \text{diag}(M_{11}, M_{22}, 0, 0).$$

This shows

$$UY_1Y_1^H U^H = UMU^H + UZ_1Z_1^H U^H = \text{diag}(\text{diag}(M_{11}, M_{22}) + \tilde{Z}, I, 0),$$

i.e., the orthogonal projection onto \mathcal{Y} is identity when restricted to $\mathcal{X} \cap \mathcal{Z}$ and is zero when restricted to $(\mathcal{X} + \mathcal{Z})_{\perp}$. Therefore $\mathcal{X} \cap \mathcal{Z} \subset \mathcal{Y} \subset \mathcal{X} + \mathcal{Z}$.

Back to the proof of the necessity, we assume without loss of generality that $k_{\min} = m > 1$ since the case when $k_{\min} = 1$ is trivial and it follows from $\mathcal{X} \cap \mathcal{Z} \subset \mathcal{Y}$ that for the case when $1 < k_{\min} < m$, we can work on k_{\min} -dimensional subspaces $\mathcal{X}_0 = \mathcal{X} \ominus (\mathcal{X} \cap \mathcal{Z})$, $\mathcal{Y}_0 = \mathcal{Y} \ominus (\mathcal{X} \cap \mathcal{Z})$, $\mathcal{Z}_0 = \mathcal{Z} \ominus (\mathcal{X} \cap \mathcal{Z})$ instead. With this assumption, $\sin \theta_m(\mathcal{X}, \mathcal{Z}) > 0$. We also assume without loss of generality that $n = 2m$ since it follows from $\mathcal{Y} \subset \mathcal{X} + \mathcal{Z}$ that we can always confine our attention to a $2m$ -dimensional subspace of \mathbb{F}^n containing $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subset \mathcal{X} + \mathcal{Z}$.

We will carry further the idea in the proof of Proposition 2. We know that the eigenvalues of $X_{11} - Z_{11}$ are given by $2 \sin \gamma_1, \dots, 2 \sin \gamma_m$ which are all positive and $X_{11} - Z_{11} = \text{diag}(g_1(P_1 - R_1), \dots, g_l(P_l - R_l))$. Since $W_j(P_j - R_j)W_j^H = \text{diag}(I_{s_j} - G_j, F_j + I_{t_j})$, $j = 1, \dots, l$, where G_j and F_j are unitary Hermitian matrices, we can see that $g_j > 0$, $F_j = I_{t_j}$, and $G_j = -I_{s_j}$. Hence $P_j = I_{r_j}$ and $R_j = -I_{r_j}$.

Notice that matrices $2UX_1X_1^H U^H - I$, $2UY_1Y_1^H U^H - I$, $2UZ_1Z_1^H U^H - I$ are unitary Hermitian whose traces are all zero. Also notice that $g_l = 1$ and $d_l = 0$. Hence if we replace U by

$$\text{diag}(W_1, \dots, W_l, \tilde{W}_1, \dots, W_l)U$$

with some properly chosen \tilde{W}_1 , we will have

$$2UX_1X_1^H U^H - I = \begin{bmatrix} P & D \\ D & -P \end{bmatrix}, \quad 2UY_1Y_1^H U^H - I = \begin{bmatrix} Q & D \\ D & -Q \end{bmatrix} \quad \text{and} \quad 2UZ_1Z_1^H U^H - I = \begin{bmatrix} -P & D \\ D & P \end{bmatrix},$$

where

$$P = \text{diag}(g_1 I_{r_1}, \dots, g_l I_{r_l}) \quad \text{and} \quad Q = \text{diag}(g_1 J_{s_1, t_1}, \dots, g_l J_{s_l, t_l}).$$

Therefore we have

$$2UX_1X_1^H U^H = \begin{bmatrix} I+P & D \\ D & I-P \end{bmatrix}, \quad 2UY_1Y_1^H U^H = \begin{bmatrix} I+Q & D \\ D & I-Q \end{bmatrix}, \quad 2UZ_1Z_1^H U^H = \begin{bmatrix} I-P & D \\ D & I+P \end{bmatrix}.$$

Since the column space of $UX_1X_1^H U^H$ is the same as the column space of UX_1 , we have

$$U\mathcal{X} = \text{span} \begin{bmatrix} I+P \\ D \end{bmatrix}, \quad U\mathcal{Z} = \text{span} \begin{bmatrix} D \\ I+P \end{bmatrix}.$$

Normalizing the bases and using trigonometry, we get

$$U\mathcal{X} = \text{span} \begin{bmatrix} \cos \frac{1}{2}(\frac{\pi}{2}I - C) \\ \sin \frac{1}{2}(\frac{\pi}{2}I - C) \end{bmatrix}, \quad U\mathcal{Z} = \text{span} \begin{bmatrix} \cos \frac{1}{2}(\frac{\pi}{2}I + C) \\ \sin \frac{1}{2}(\frac{\pi}{2}I + C) \end{bmatrix}.$$

Let $\Theta_{\mathcal{X}} = \frac{1}{2}(\frac{\pi}{2}I - C)$ and $\Theta_{\mathcal{Z}} = \frac{1}{2}(\frac{\pi}{2}I + C)$. Then $\Theta_{\mathcal{Z}} = \Theta_{\mathcal{X}} + C$. Immediately,

$$\begin{bmatrix} \cos \Theta_{\mathcal{Z}} \\ \sin \Theta_{\mathcal{Z}} \end{bmatrix} = \exp \left(\begin{bmatrix} 0 & -\Theta_{\mathcal{X}} \\ \Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & \Theta_{\mathcal{X}} \\ -\Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right) \begin{bmatrix} \cos \Theta_{\mathcal{X}} \\ \sin \Theta_{\mathcal{X}} \end{bmatrix}.$$

Therefore,

$$\exp \left(\begin{bmatrix} 0 & -\Theta_{\mathcal{X}} \\ \Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & \Theta_{\mathcal{X}} \\ -\Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right)$$

is a direct rotation from $U\mathcal{X}$ to $U\mathcal{Z}$. Set

$$\Lambda = \frac{1}{2} [I_m - \text{diag}(J_{t_1, s_1}, \dots, J_{t_l, s_l})].$$

Then

$$\begin{aligned} & \begin{bmatrix} I+Q \\ D \end{bmatrix} (I - \Lambda) + \begin{bmatrix} D \\ I-Q \end{bmatrix} \Lambda \\ &= \exp \left(\begin{bmatrix} 0 & -\Theta_{\mathcal{X}} \\ \Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & -\Lambda C \\ C^H \Lambda & 0 \end{bmatrix} \right) \exp \left(\begin{bmatrix} 0 & \Theta_{\mathcal{X}} \\ -\Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right) \begin{bmatrix} I+P \\ D \end{bmatrix} \end{aligned}$$

which spans $U\mathcal{Y}$. Finally, if we let

$$X = U^H \exp \left(\begin{bmatrix} 0 & -\Theta_{\mathcal{X}} \\ \Theta_{\mathcal{X}}^H & 0 \end{bmatrix} \right),$$

then

$$X \exp \left(\begin{bmatrix} 0 & -C \\ C^H & 0 \end{bmatrix} \right) X^H$$

is a direct rotation from \mathcal{X} to \mathcal{Z} , and

$$\mathcal{Y} = X \exp \left(\begin{bmatrix} 0 & -\Lambda C \\ C^H \Lambda & 0 \end{bmatrix} \right) X^H \mathcal{X}. \quad \square$$

Summarizing the discussions in this section, we conclude that the triangle inequality for the angular metrics are tighter than those for the gap metrics in the sense that for a strictly convex symmetric gauge function or a Ky Fan k -function Φ , if

$$\Phi(\sin \theta(\mathcal{X}, \mathcal{Z})) = \Phi(\sin \theta(\mathcal{X}, \mathcal{Y})) + \Phi(\sin \theta(\mathcal{Y}, \mathcal{Z})),$$

then

$$\Phi(\theta(\mathcal{X}, \mathcal{Z})) = \Phi(\theta(\mathcal{X}, \mathcal{Y})) + \Phi(\theta(\mathcal{Y}, \mathcal{Z})).$$

5 Related Questions and Extensions

In this section, we will discuss three issues. The first is on the extremal properties of the direct rotation between two subspaces. The second is on possible generalization of the inequalities in the form of (12) and (13). The third is on the possibility to use functions more general than the symmetric gauge functions to define metrics on $\mathcal{G}_{m,n}$.

As we have seen in Section 2, for $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$, $m \leq n/2$, a direct rotation from \mathcal{X} to \mathcal{Y} is a map of the form

$$R = X \begin{bmatrix} \Gamma & -\Sigma & 0 \\ \Sigma & \Gamma & 0 \\ 0 & 0 & I_{n-2m} \end{bmatrix} X^H = \exp \left(X \begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} X^H \right)$$

where $A = \begin{bmatrix} \text{diag } \theta(\mathcal{X}, \mathcal{Y}) & 0_{m,n-2m} \end{bmatrix} \in \mathcal{M}_{m,n-m}$ and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ is a special unitary matrix such that the columns of X_1 span \mathcal{X} and those of X_2 span \mathcal{X}_\perp . Subspaces \mathcal{X}, \mathcal{Y} are said to be acute if the canonical angles in between are all less than $\pi/2$, or equivalently, $\mathcal{X} \cap \mathcal{Y}_\perp = \{0\}$. It was shown in [3] that the direct rotation between acute subspaces is unique.

Davis [3] studied some extremal properties of R among all unitary transformations from \mathcal{X} to \mathcal{Y} . He showed that R differs from the identity map in the least amount in some sense; in particular, for all unitarily invariant norm $\|\cdot\|$,

1. $\|(I - V)|_{\mathcal{X}}\|$ is minimized when $V = R$.
2. $\|(I - V)(I - V^H)\|$ is minimized when $V = R$.
3. $\|I - V\|$ is minimized when $V = R$ if $\theta_1(\mathcal{X}, \mathcal{Y}) \leq \pi/3$.

A by-product of the development in Section 2 results in another extremal property, which roughly says that the direct rotation rotates the vectors in \mathbb{F}^n with the smallest amount of angles. For a unitary matrix V , let $\log(V)$ be defined to be the unique skew-Hermitian matrix G with spectrum in $i(-\pi, \pi]$ such that $V = \exp(G)$.

Theorem 6 *Let $\|\cdot\|$ be a unitarily invariant norm. Assume V is a unitary matrix satisfying $V\mathcal{X} = \mathcal{Y}$. Then $\|\log(V)\|$ is minimized when $V = R$.*

Proof: The proof will be sketchy since it follows from the same idea as in the development of Section 2. We know that V only differs from R locally in \mathcal{X} and \mathcal{X}_\perp , i.e., there exist unitary matrix $Q = \text{diag}(Q_1, Q_2)$ with $Q_1 \in \mathcal{U}_m$ and $Q_2 \in \mathcal{U}_{n-m}$ such that

$$V = X \exp \left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} \right) Q X^H.$$

Hence

$$X^H V X = \exp \left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} \right) Q. \quad (25)$$

Taking inverse, we get

$$X^H V^H X = Q^H \exp \left(\begin{bmatrix} 0 & A \\ -A^H & 0 \end{bmatrix} \right).$$

Multiplying $J = \text{diag}(I_m, -I_{n-m})$ from both sides and noticing that J commutes with Q , we get

$$J X^H V^H X J = Q^H \exp \left(\begin{bmatrix} 0 & -A \\ A^H & 0 \end{bmatrix} \right). \quad (26)$$

Multiplying (25) and (26), we obtain

$$\exp \left(\begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} \right) = X^H V X J X^H V^H X J.$$

Let $G = \log(V)$, we have

$$\exp \left(\begin{bmatrix} 0 & -2A \\ 2A^H & 0 \end{bmatrix} \right) = \exp(U_1 G U_1^H + U_2 G^H U_2^H)$$

for some $U_1, U_2 \in \mathcal{U}_n$. The rest of the proof follows from that of Theorem 1. \square

The extremal property of the direct rotation for $\|\log(V)\|$ is free of the unpleasant condition required in that for $\|I - V\|$.

Example 2 Compare the reflection $V = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$ and direct rotation $R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ mapping $\text{span} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ to $\text{span} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$. Clearly $\pi = \|\log(V)\|_2 \geq \|\log(R)\|_2 = 2\theta$, whereas $\|I - V\|_2 = 2$ and $\|I - R\|_2 = 4 \sin(\theta/2)$, which can be ordered either way depending on θ , as pointed out in [3].

The second issue concerns the possible extension of inequalities in the form of (12-13) on canonical angles between three subspaces, which resemble the corresponding inequalities (6-7) on singular values of matrix sums. In the case when $\theta(\mathcal{X}, \mathcal{Y})$ and $\theta(\mathcal{Y}, \mathcal{Z})$ are sufficiently small so that $\theta_1(\mathcal{X}, \mathcal{Y}) + \theta_1(\mathcal{Y}, \mathcal{Z}) < \pi/2$, inequalities (12-13) immediately follow inequalities (6-7) after taking the logarithm from both sides of (5). There are some recent results on developing more inequalities of the form

$$\sum_{t \in T} \sigma_t(C) \leq \sum_{r \in R} \sigma_r(A) + \sum_{s \in S} \sigma_s(B) \quad (27)$$

where $R = \{r_1 < \dots < r_k\}$, $S = \{s_1 < \dots < s_k\}$, $T = \{t_1 < \dots < t_k\}$ are increasingly ordered subsets of $\{1, \dots, \min\{m, n\}\}$ with the same cardinality $k \leq \min\{m, n\}$ [8]. An example of a family of such inequalities beyond those in (6-7) is given by R, S, T satisfying $r_k + s_k \leq \min\{m, n\} + k$ and $t_j = r_j + s_j - j$ for $1 \leq j \leq k$. It is natural to conjecture that

$$\sum_{t \in T} \theta_t(\mathcal{X}, \mathcal{Z}) \leq \sum_{r \in R} \theta_r(\mathcal{X}, \mathcal{Y}) + \sum_{s \in S} \theta_s(\mathcal{Y}, \mathcal{Z}) \quad (28)$$

whenever the sets R, S, T make inequality (27) satisfied (with $\min\{m, n\}$ replaced by m). However this is not immediately clear, even in the case when $\theta(\mathcal{X}, \mathcal{Y})$ and $\theta(\mathcal{Y}, \mathcal{Z})$ are sufficiently small, due to the unpleasant fact that the right hand side of (5) involves the sum of three terms, two of which are related to $\theta(\mathcal{Y}, \mathcal{Z})$.

The last issue is whether we can use more general functions of the canonical angles to define unitarily metrics on $\mathcal{G}_{m,n}$. One may consider a function $\Psi : [0, \pi/2]^m \rightarrow \mathbb{R}$ and define $\delta(\mathcal{X}, \mathcal{Y}) = \Psi(\theta(\mathcal{X}, \mathcal{Y}))$ for $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$. It would be interesting to determine the necessary and sufficient conditions on Ψ for δ to be a metric on $\mathcal{G}_{m,n}$. Evidently, such a function δ always satisfies $\delta(\mathcal{X}, \mathcal{Y}) = \delta(\mathcal{Y}, \mathcal{X})$ and $\delta(U\mathcal{X}, U\mathcal{Y}) = \delta(\mathcal{X}, \mathcal{Y})$ for all $U \in \mathcal{U}_n$. To ensure that $\delta(\mathcal{X}, \mathcal{Y}) \geq 0$ so that equality holds if and only if $\mathcal{X} = \mathcal{Y}$, the function Ψ must satisfy the following condition.

(P1) Ψ is positive, i.e., $\Psi(\xi) \geq 0$ for all $\xi \in [0, \pi/2]^m$, and the equality holds if and only if $\xi = 0$.

To ensure triangle inequality, the following is necessary.

(P2) Ψ is subadditive, i.e., $\Psi(\xi + \zeta) \leq \Psi(\xi) + \Psi(\zeta)$ for any $\xi, \zeta \in [0, \pi/2]^m$ satisfying $\xi + \zeta \in [0, \pi/2]^m$.

Suppose Ψ also satisfies the following.

(P3) Ψ is strongly Schur convex (or strongly isotone), i.e., $\Psi(\xi) \leq \Psi(\zeta)$ whenever $\xi, \zeta \in [0, \pi/2]^m$ and ξ is weakly majorized by ζ .

Then we have

$$\begin{aligned} \delta(\mathcal{X}, \mathcal{Z}) &= \Psi(\theta(\mathcal{X}, \mathcal{Z})) \leq \Psi(\theta(\mathcal{X}, \mathcal{Y}) + \theta(\mathcal{Y}, \mathcal{Z})) \\ &\leq \Psi(\theta(\mathcal{X}, \mathcal{Y})) + \Psi(\theta(\mathcal{Y}, \mathcal{Z})) = \delta(\mathcal{X}, \mathcal{Y}) + \delta(\mathcal{Y}, \mathcal{Z}), \end{aligned}$$

i.e., function δ satisfies the triangle inequality and hence is a unitarily invariant metric. This may allow us to go a small step beyond the set of symmetric gauge functions. However, it is not clear at this stage what are functions satisfying (P1), (P2) and (P3) other than symmetric gauge functions and those of the form $f \circ \Phi_1$ where f is an increasing subadditive scalar function. Also, (P3) is far from a necessary condition for δ to be a metric on $\mathcal{G}_{m,n}$. It is not satisfied by the gap metrics defined in (8), where $\Psi(\xi) = \Phi(\sin \xi)$, nor the Hausdorff metric defined in (9), where $\Psi(\xi) = \Phi(2 \sin(\xi/2))$. It would be interesting to weaken the condition (P3) so that it is satisfied by all known metrics so far, or even becomes necessary.

Example 3 Consider $\Psi(\xi_1, \xi_2) = \sin \xi_1 + \sin \xi_2$. Then the corresponding function δ is the gap metric on $\mathcal{G}_{2,n}$ corresponding to the Ky Fan 2-function. However, for $\xi = (\pi/4, \pi/4)$ and $\zeta = (\pi/2, 0)$, it follows that ξ is majorized by ζ , but $\Psi(\xi) = \sqrt{2} > 1 = \Psi(\zeta)$.

Acknowledgement

We thank Professor Chandler Davis for some helpful discussions. In particular, he suggested a different approach (using integration theory) which may lead to a proof of Theorem 1. We also would like to thank Nam-Kiu Tsing, Yimin Wei, Roy Mathias, and Pierre-Antoine Absil for helpful discussions and Wasin So for bringing the reference [31] to the authors' attention.

The research of L. Qiu and Y. Zhang was supported by a HK RGC grant. The research of C. K. Li was support by a USA NSF grant and a HK RGC grant.

References

- [1] S. Agnihotri and C. Woodward, "Eigenvalues of products of unitary matrices and quantum Schubert calculus," *Math. Res. Lett.*, Vol. 5, pp. 817–836, 1998.

- [2] S. N. Afriat, "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces," *Proceedings of the Cambridge Philosophical Society*, Vol. 53, pp. 800-816, 1957.
- [3] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a perturbation. III," *SIAM Journal on Numerical Analysis*, Vol. 7, pp. 1-46, 1970.
- [4] K. De Cock and B. De Moor, "Subspace angles between ARMA models," *Systems & Control Letters*, Vol. 46, pp. 265-270, 2002.
- [5] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, Vol. 1, pp. 211-218, 1936.
- [6] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, Vol. 20, pp. 303-353, 1998.
- [7] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *Proceedings of the National Academy of Sciences*, Vol.37, pp. 760-766, 1951.
- [8] W. Fulton, "Eigenvalues, invariant factors, highest weights, and Schubert calculus," *Bull. Amer. Math. Soc.*, Vol. 37, pp. 209-249, 2000.
- [9] T. T. Georgiou and M. C. Smith, "Optimal robustness in the gap metric," *IEEE Transactions on Automatic Control*, vol. 35, pp. 673-685, 1990.
- [10] I. C. Gohberg and M. G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*, American Mathematical Society, Providence, Rhode Island, 1969.
- [11] G. H. Golub and H. Zha, "Perturbation analysis of the canonical correlations of matrix pairs," *Linear Algebra and Its Applications*, Vol. 210, pp. 3-28, 1994.
- [12] H. Hotelling, "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, Vol. 24, pp. 417-441 and 498-520, 1933.
- [13] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28, pp. 321-377, 1936.
- [14] C. Jordan, "Mémoire sur les formes bilinéaires," *Journal de Mathématiques Pures et Appliquées, Deuxième Série*, Vol. 19, pp. 35-54, 1874.
- [15] C. Jordan, "Essai sur la géométrie à n dimensions," *Bulletin de la Société Mathématique*, Vol. 3, pp. 103-174, 1875.
- [16] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin; New York, 1995.
- [17] V. C. Klema and A. J. Laub, "The singular value decomposition: its computation and some applications," *IEEE Transactions on Automatic Control*, Vol. AC-25, pp. 164-176, 1980.
- [18] A. V. Knyazev and M. E. Argentati, "Principal angles between subspaces in an A -based scalar product: algorithms and perturbation estimates," *SIAM Journal of Scientific Computing*, Vol. 23, pp. 2008-2040, 2002.
- [19] C. K. Li, "Matrices with some extremal properties," *Linear Algebra and Its Applications*, Vol. 101, pp. 255-267, 1988.
- [20] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic, 1979.

- [21] R. J. Martin, "A metric for ARMA processes," *IEEE Transactions on Signal Processing*, Vol. 48, pp. 1164-1169, 2000.
- [22] L. Mirsky, "Results and problems in the theory of doubly stochastic matrices," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* Vol. 1, pp. 319-334, 1963.
- [23] J. Miao and A. Ben-Israel, "On principal angles between subspaces in \mathbb{R}^n ," *Linear Algebra and Its Applications*, Vol. 171, pp. 81-98, 1992.
- [24] L. Qiu and E. J. Davison, "Pointwise gap metrics on transfer matrices," *IEEE Transactions on Automatic Control*, Vol. 37, pp. 741-758, 1992.
- [25] L. Qiu and E. J. Davison, "Feedback stability under simultaneous gap metric uncertainties in plant and controller," *Systems & Control Letters*, Vol. 18, pp. 9-22, 1992.
- [26] E. Schmidt, "Zur theorie der linearen und nichtlinearen integralgleichungen. I Tiel. entwicklung willkürlichen funktionen nach system vorgeschriebener," *Mathematische Annalen*, Vol. 63, pp. 433-476, 1907.
- [27] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
- [28] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Review*, Vol. 15, pp. 727-764, 1973.
- [29] J.-G. Sun, "Perturbation of angles between linear subspaces," *Journal of Computational Mathematics*, Vol. 5 pp. 58-61, 1987.
- [30] R. C. Thompson, "Singular values, diagonal elements, and convexity," *SIAM Journal on Applied Mathematics*, Vol. 32, pp. 39-63, 1977.
- [31] R. C. Thompson, "Proof of a conjectured exponential formula," *Linear and Multilinear Algebra*, Vol. 9, pp. 187-197, 1986.
- [32] G. Vinnicombe, "Frequency domain uncertainty and the graph topology," *IEEE Transactions on Automatic Control*, Vol. 38, pp. 1371-1383, 1993.
- [33] P. van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory-Implementation-Applications*, Kluwer Academic Publisher, Boston, 1996.
- [34] J. von Neumann, "Some matrix-inequalities and metrization of matrix-space," *Tomsk. Univ. Rev.*, Vol. 1, pp. 286-300. 1937.
- [35] P.-Å. Wedin, "On angles between subspaces," *Matrix Pencils*, pp. 263-285. Springer, New York, 1983.
- [36] A. Weinstein, "Almost invariant submanifolds for compact group actions", *Journal of European Mathematical Society*, Vol. 2, pp. 53-86, 2000.
- [37] H. K. Wimmer, "Canonical angles of unitary spaces and perturbations of direct complements," *Linear Algebra and its Applications*, Vol. 287, pp. 373-379, 1999.
- [38] Y.-C. Wong, "Differential geometry of Grassmann manifold," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 57, pp. 589-594, 1967.