SCHUBERT VARIETIES AND DISTANCES BETWEEN SUBSPACES OF DIFFERENT DIMENSIONS

KE YE AND LEK-HENG LIM

Abstract. We resolve a basic problem on subspace distances that often arises in applications: How can the usual Grassmann distance between equidimensional subspaces be extended to subspaces of different dimensions? We show that a natural solution is given by the distance of a point to a Schubert variety within the Grassmannian. This distance reduces to the Grassmann distance when the subspaces are equidimensional and does not depend on any embedding into a larger ambient space. Furthermore, it has a concrete expression involving principal angles, and is efficiently computable in numerically stable ways. Our results are largely independent of the Grassmann distance — if desired, it may be substituted by any other common distances between subspaces. Our approach depends on a concrete algebraic geometric view of the Grassmannian that parallels the differential geometric perspective that is well-established in applied and computational mathematics.

1. Introduction

Biological data (e.g. gene expression levels, metabolomic profile), image data (e.g. MRI tractographs, movie clips), text data (e.g. blogs, tweets), etc., often come in the form of a set of feature vectors $a_1, \dots, a_m \in \mathbb{R}^d$ and can be conveniently represented by a matrix $A \in \mathbb{R}^{m\times d}$ (e.g. gene-microarray matrices of gene expression levels, frame-pixel matrices of grey scale values, term-document matrices of term frequencies-inverse document frequencies). In modern applications, it is often the case that one will encounter an exceedingly large sample size m (massive) or an exceedingly large number of variables d (high-dimensional) or both.

The raw data A is usually less interesting and informative than the spaces it defines, e.g. its row and column spaces or its principal subspaces. Moreoever, it often happens that A can be well-approximated by a subspace $\mathbf{A} \in \operatorname{Gr}(k, n)$ where $k \ll m$ and $n \ll d$. The process of getting from A to $\mathbf{A}$ is well-studied, e.g. randomly sample a subset of representative landmarks or compute principal components.

Subspace-valued data appears in a wide range of applications: computer vision [36, 45], bioinformatics [22], machine learning [23, 30], communication [34, 50], coding theory [4, 6, 12, 15], statistical classification [21], and system identification [36]. In computational mathematics, subspaces arise in the form of Krylov subspaces [32] and their variants [11], as subspaces of structured matrices (e.g. Toeplitz, Hankel, banded), and in recent developments such as compressive sensing (e.g. Grassmannian dictionaries [43], online matrix completion [10]).

One of the most basic problems with subspaces is to define a notion of separation between them. The solution is well-known for subspaces of the same dimension k in $\mathbb{R}^n$. These are points on the Grassmannian $\operatorname{Gr}(k, n)$, a Riemannian manifold, and the geodesic distance between them gives us an intrinsic distance. The *Grassmann distance* is independent of the choice of coordinates and can be readily related to principal angles and thus computed via the singular value decomposition (SVD): For subspaces $\mathbf{A}, \mathbf{B} \in \operatorname{Gr}(k, n)$, form matrices $A, B \in \mathbb{R}^{n\times k}$ whose columns are their respective

2010 *Mathematics Subject Classification.* 14M15, 15A18, 14N20, 51K99.

Key words and phrases. distances between inequidimensional subspaces, Grassmannian, Schubert variety, flag variety, probability densities on Grassmannian.

orthonormal bases, then

$$d(\mathbf{A}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^{k} \theta_i^2\Big)^{1/2}, \tag{1}$$

where $\theta_i = \cos^{-1}\big(\sigma_i(A^{\mathsf{T}}B)\big)$ is the ith principal angle between $\mathbf{A}$ and $\mathbf{B}$. This is the geodesic distance on the Grassmannian viewed as a Riemannian manifold. There are many other common distances defined on Grassmannians — Asimov, Binet–Cauchy, chordal, Fubini–Study, Martin, Procrustes, projection, spectral (see Table 2).

What if the subspaces are of different dimensions? In fact, if one examines the aforementioned applications, one invariably finds that the most general settings for each of them would fall under this situation. The restriction to equidimensional subspaces thus somewhat limits the utility of these applications. For example, the principal subspaces of two matrices A and B for a given noise level would typically be of different dimensions, since there is no reason to expect the number of singular values of A above a given threshold to be the same as that of B.

As such one may also find many applications that involve distances between subspaces of different dimensions: numerical linear algebra [8, 42], information retrieval [7, 51], facial recognition [16, 47], image classification [7, 16], motion segmentation [13, 35, 49], EEG signal analysis [18], mechanical engineering [24], economics [39], network analysis [40], blog spam detection [31], and decoding colored barcodes [5].

These applications are all based on two existing proposals for a distance between subspaces of different dimensions: The *containment gap* [28, pp. 197–199] and the *symmetric directional distance* [44, 46]. They are however somewhat ad hoc and bear little relation to the natural geometry of subspaces. Also, it is not clear what they are suppose to measure and neither restricts to the Grassmann distance when the subspaces are of the same dimension. Our main objective is to show that there is an alternative definition that does generalize the Grassmann distance but our work will also shed light on these two distances.

1.1. **Main Contributions.** Our main result (see Theorem 7) can be stated in simple linear algebraic terms: Given any two subspaces in $\mathbb{R}^n$, $\mathbf{A}$ of dimension k and $\mathbf{B}$ of dimension l, assuming $k < l$ without loss of generality, the distance from $\mathbf{A}$ to the nearest k-dimensional subspace contained in $\mathbf{B}$ *equals* the distance from $\mathbf{B}$ to the nearest l-dimensional subspace that contains $\mathbf{A}$. Their common value gives the distance between $\mathbf{A}$ and $\mathbf{B}$. Taking an algebraic geometric point-of-view:

(∗) *The distance between subspaces of different dimensions is the distance between a point and a certain Schubert variety within the Grassmannian.*

This distance has the following properties, established in Section 4:

(a) readily computable via SVD;
(b) restricts to the usual Grassmann distance (1) for subspaces of the same dimension;
(c) independent of the choice of local coordinates;
(d) independent of the dimension of the ambient space (i.e., n);
(e) may be defined in conjunction with other common distances in Table 2.

We will see in Section 7 that the two existing notions of distance between subspaces of different dimensions are special cases of (e).

Evidently, the word 'distance' in (∗) is used in the sense of a distance of a point to a set. For example, if a subspace is contained in another, then the distance between them is zero, even if they are distinct subspaces. Thus the distance in (∗) is not a metric[1]. In Section 6, we define a metric on the set of subspaces of all dimensions using an analogue of our main result: Given any two subspaces in $\mathbb{R}^n$, $\mathbf{A}$ of dimension k and $\mathbf{B}$ of dimension l with $k < l$, the distance from $\mathbf{A}$ to the furthest k-dimensional subspace contained in $\mathbf{B}$ *equals* the distance from $\mathbf{B}$ to the furthest

[1]We will see in Section 5 that this could be attributed to the fact that **Met**, the category of metric spaces and continuous contractions, does not admit coproduct.

l-dimensional subspace that contains $\mathbf{A}$. Their common value gives a metric between $\mathbf{A}$ and $\mathbf{B}$. The most interesting metrics for subspaces of different dimensions can be found in Table 3.

In Section 9, we obtain a volumetric analogue of our main result: Given two arbitrary subspaces in $\mathbb{R}^n$, $\mathbf{A}$ of dimension k and $\mathbf{B}$ of dimension l with $k < l$, we show that the probability a random l-dimensional subspace contains $\mathbf{A}$ *equals* the probability a random k-dimensional subspace is contained in $\mathbf{B}$.

The far-reaching work [17] popularized the basic *differential geometry* of Stiefel and Grassmannian manifolds by casting the discussions concretely in terms of matrices. Subsequent works, notably [1, 2, 3], have further enriched this concrete matrix-based approach. A secondary objective of our article is to do the same for the basic *algebraic geometry* of Grassmannians. In particular, we introduce some of the objects in Table 1 to an applied and computational mathematics readership. The proofs of our main results essentially use only the SVD. Everything else is explained within the article and accessible to anyone willing to accept a small handful of unfamiliar terminologies and facts on faith.

TABLE 1. Grassmannian and friends

Grassmannian	$\mathrm{Gr}(k, n)$	models k-dimensional subspaces in $\mathbb{R}^n$	§2
Infinite Grassmannian	$\mathrm{Gr}(k, \infty)$	models k-dimensional subspaces regardless of ambient space	§3
Doubly-infinite Grassmannian	$\mathrm{Gr}(\infty, \infty)$	models subspaces of all dimensions regardless of ambient space	§5
Flag variety	$\mathrm{Flag}(k_1, \dots, k_m, n)$	models nested sequences of subspaces in $\mathbb{R}^n$; $\mathrm{Flag}(k, n) = \mathrm{Gr}(k, n)$	§8
Schubert variety	$\Omega(\mathbf{X}_1, \dots, \mathbf{X}_m, n)$	'linearly constrained' subset of $\mathrm{Gr}(k, n)$	§8

2. Grassmannian of linear subspaces

We will selectively review some basic properties of the Grassmannian. The differential geometric perspectives are drawn from [27, 37], the more concrete matrix-theoretic view from [2, 17, 48], and the computational aspects from [20].

We fix the ambient space $\mathbb{R}^n$. A *k-plane* is a k-dimensional subspace of $\mathbb{R}^n$. A *k-frame* is an ordered orthonormal basis of a k-plane, regarded as an $n \times k$ matrix whose columns $a_1, \dots, a_k$ are the orthonormal basis vectors. A *flag* is a strictly increasing sequence of nested subspaces, $\mathbf{X}_0 \subset \mathbf{X}_1 \subset \cdots \subset \mathbf{X}_m \subset \mathbb{R}^n$; it is *complete* if $m = n$.

We write $\mathrm{Gr}(k, n)$ for the *Grassmannian* of k-planes in $\mathbb{R}^n$, $\mathrm{V}(k, n)$ for the *Stiefel manifold* of orthonormal k-frames, and $\mathrm{O}(n) \coloneqq \mathrm{V}(n, n)$ for the *orthogonal group* of $n \times n$ orthogonal matrices. $\mathrm{V}(k, n)$ may be regarded as a homogeneous space,

$$\mathrm{V}(k, n) \cong \mathrm{O}(n)/\mathrm{O}(n - k),$$

or more concretely as the set of $n \times k$ matrices with orthonormal columns.

There is a *right action* of $\mathrm{O}(k)$ on $\mathrm{V}(k, n)$: For $Q \in \mathrm{O}(k)$ and $A \in \mathrm{V}(k, n)$, the action yields $AQ \in \mathrm{V}(k, n)$ and the resulting homogeneous space is $\mathrm{Gr}(k, n)$, i.e.,

$$\mathrm{Gr}(k, n) \cong \mathrm{V}(k, n)/\mathrm{O}(k) \cong \mathrm{O}(n)/\big(\mathrm{O}(n - k) \times \mathrm{O}(k)\big). \tag{2}$$

In this picture, a subspace $\mathbf{A} \in \mathrm{Gr}(k, n)$ is identified with an equivalence class comprising all its k-frames $\{AQ \in \mathrm{V}(k, n) : Q \in \mathrm{O}(k)\}$. Note that $\operatorname{span}(AQ) = \operatorname{span}(A)$ for all $Q \in \mathrm{O}(k)$.

There is a *left action* of $\mathcal{O}(n)$ on $\mathrm{Gr}(k,n)$: For any $Q \in \mathcal{O}(n)$ and $\mathbf{A} = \mathrm{span}(A) \in \mathrm{Gr}(k,n)$ where A is a k-frame of $\mathbf{A}$, the action yields

$$Q \cdot \mathbf{A} \coloneqq \mathrm{span}(QA) \in \mathrm{Gr}(k,n). \tag{3}$$

This action is transitive as any k-plane can be rotated onto any other k-plane by some $Q \in \mathcal{O}(n)$. A k-plane $\mathbf{A} \in \mathrm{Gr}(k,n)$ will be denoted in boldface; the corresponding italized letter $A = [a_1, \dots, a_k] \in \mathrm{V}(k,n)$ will denote a k-frame of $\mathbf{A}$.

$\mathrm{Gr}(k,n)$ and $\mathrm{V}(k,n)$ are smooth manifolds of dimensions $k(n-k)$ and $nk-k(k+1)/2$ respectively. As a set of $n \times k$ matrices, $\mathrm{V}(k,n)$ is a submanifold of $\mathbb{R}^{n\times k}$ and inherits a Riemannian metric from the Euclidean metric on $\mathbb{R}^{n\times k}$, i.e., given $A = [a_1, \dots, a_k]$ and $B = [b_1, \dots, b_k]$ in $T_X \mathrm{V}(k,n)$, the tangent space at $X \in \mathrm{V}(k,n)$, the Riemannian metric g is defined by $g_X(A,B) = \sum_{i=1}^k a_i^{\mathsf{T}} b_i = \mathrm{tr}(A^{\mathsf{T}}B)$. As g is invariant under the action of $\mathcal{O}(k)$, it descends to a Riemannian metric on $\mathrm{Gr}(k,n)$ and in turn induces a geodesic distance on $\mathrm{Gr}(k,n)$ which we define below.

Let $\mathbf{A} \in \mathrm{Gr}(k,n)$ and $\mathbf{B} \in \mathrm{Gr}(l,n)$ respectively. Let $r \coloneqq \min(k,l)$. The *i*th *principal vectors* (p_i, q_i), $i = 1, \dots, r$, are defined recursively as solutions to the optimization problem

$$\begin{array}{ll} \text{maximize} & p^{\mathsf{T}} q \\ \text{subject to} & p \in \mathbf{A},\ p^{\mathsf{T}}p_1 = \cdots = p^{\mathsf{T}}p_{i-1} = 0,\ \|p\| = 1, \\ & q \in \mathbf{B},\ q^{\mathsf{T}}q_1 = \cdots = q^{\mathsf{T}}q_{i-1} = 0,\ \|q\| = 1, \end{array}$$

for $i = 1, \dots, r$. The *principal angles* are then defined by

$$\cos\theta_i = p_i^{\mathsf{T}} q_i, \quad i = 1, \dots, r.$$

Clearly $0 \le \theta_1 \le \cdots \le \theta_r \le \pi/2$. We will let $\theta_i(\mathbf{A}, \mathbf{B})$ denote the ith principal angle between $\mathbf{A} \in \mathrm{Gr}(k,n)$ and $\mathbf{B} \in \mathrm{Gr}(l,n)$.

Principal vectors and principal angles may be readily computed using QR and SVD [9, 20]. Let $A = [a_1, \dots, a_k]$ and $B = [b_1, \dots, b_l]$ be orthonormal bases and let

$$A^{\mathsf{T}}B = U\Sigma V^{\mathsf{T}} \tag{4}$$

be the full SVD of $A^{\mathsf{T}}B$, i.e., $U \in \mathcal{O}(k)$, $V \in \mathcal{O}(l)$, $\Sigma = \left[\begin{smallmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{smallmatrix}\right] \in \mathbb{R}^{k\times l}$ with $\Sigma_1 = \mathrm{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r\times r}$ where $\sigma_1 \ge \cdots \ge \sigma_r \ge 0$.

The principal angles $\theta_1 \le \cdots \le \theta_r$ are given by

$$\theta_i = \cos^{-1}\sigma_i, \quad i = 1, \dots, r. \tag{5}$$

It is customary to write $A^{\mathsf{T}}B = U(\cos\Theta)V^{\mathsf{T}}$, where $\Theta = \mathrm{diag}(\theta_1, \dots, \theta_r, 1, \dots, 1) \in \mathbb{R}^{k\times l}$ and $\Theta_1 = \mathrm{diag}(\theta_1, \dots, \theta_r) \in \mathbb{R}^{r\times r}$. Consider the column vectors,

$$AU = [p_1, \dots, p_k], \quad BV = [q_1, \dots, q_l].$$

The principal vectors are given by $(p_1, q_1), \dots, (p_r, q_r)$. Strictly speaking, principal vectors come in pairs but we will also call the vectors $p_{r+1}, \dots, p_k$ (if $r = l < k$) or $q_{r+1}, \dots, q_l$ (if $r = k < l$) principal vectors for lack of a better term.

We will use the following fact from [20, Theorem 6.4.2].

Proposition 1. *Let $r = \min(k,l)$ and $\theta_1, \dots, \theta_r$ and $(p_1, q_1), \dots, (p_r, q_r)$ be the principal angles and principal vectors between $\mathbf{A} \in \mathrm{Gr}(k,n)$ and $\mathbf{B} \in \mathrm{Gr}(l,n)$ respectively. If $m < r$ is such that $1 = \cos\theta_1 = \cdots = \cos\theta_m > \cos\theta_{m+1}$, then*

$$\mathbf{A} \cap \mathbf{B} = \mathrm{span}\{p_1, \dots, p_m\} = \mathrm{span}\{q_1, \dots, q_m\}.$$

If $k = l$, the geodesic distance between $\mathbf{A}$ and $\mathbf{B}$ in $\mathrm{Gr}(k,n)$ is called the *Grassmann distance* and is given by

$$d_{\mathrm{Gr}(k,n)}(\mathbf{A}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^k \theta_i^2\Big)^{1/2} = \|\cos^{-1}\Sigma\|_F. \tag{6}$$

An explicit expression for the geodesic [2] connecting $\mathbf{A}$ to $\mathbf{B}$ on $\operatorname{Gr}(k,n)$ that minimizes the Grassmann distance is given by $\gamma : [0,1] \to \operatorname{Gr}(k,n)$,

$$\gamma(t) = \operatorname{span}(AU \cos t\Theta + Q \sin t\Theta) \tag{7}$$

where $M = Q(\tan \Theta)U^{\mathsf{T}}$ is a condensed SVD of the matrix

$$M := (I - AA^{\mathsf{T}})B(A^{\mathsf{T}}B)^{-1} \in \mathbb{R}^{n\times k}$$

and where $U \in \mathcal{O}(k)$ and $\Theta = \operatorname{diag}(\theta_1, \dots, \theta_k) \in \mathbb{R}^{k\times k}$ are as in (4) and (5). Note that if $\cos \Theta = \Sigma$, then $\tan \Theta = (\Sigma^{-2} - I)^{1/2}$. Also, $\gamma(0) = \mathbf{A}$ and $\gamma(1) = \mathbf{B}$.

Aside from the Grassmann distance, there are many well-known distances between subspaces [6, 14, 15, 17, 21]. We present some of these in Table 2.

TABLE 2. Distances on $\operatorname{Gr}(k,n)$ in terms of principal angles and orthonormal bases.

	Principal angles	*Orthonormal bases*
Asimov	$d^{\alpha}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \theta_k$	$\cos^{-1}\|A^{\mathsf{T}}B\|_2$
Binet–Cauchy	$d^{\beta}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \left(1 - \prod_{i=1}^{k} \cos^2 \theta_i\right)^{1/2}$	$(1 - (\det A^{\mathsf{T}}B)^2)^{1/2}$
Chordal	$d^{\kappa}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \left(\sum_{i=1}^{k} \sin^2 \theta_i\right)^{1/2}$	$\frac{1}{\sqrt{2}}\|AA^{\mathsf{T}} - BB^{\mathsf{T}}\|_F$
Fubini–Study	$d^{\phi}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \cos^{-1}\left(\prod_{i=1}^{k} \cos \theta_i\right)$	$\cos^{-1}\lvert\det A^{\mathsf{T}}B\rvert$
Martin	$d^{\mu}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \left(\log \prod_{i=1}^{k} 1/\cos^2 \theta_i\right)^{1/2}$	$(-2 \log\det A^{\mathsf{T}}B)^{1/2}$
Procrustes	$d^{\rho}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = 2\left(\sum_{i=1}^{k} \sin^2(\theta_i/2)\right)^{1/2}$	$\|AU - BV\|_F$
Projection	$d^{\pi}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = \sin \theta_k$	$\|AA^{\mathsf{T}} - BB^{\mathsf{T}}\|_2$
Spectral	$d^{\sigma}_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = 2\sin(\theta_k/2)$	$\|AU - BV\|_2$

The value $\sin \theta_1$ is sometimes called the *max correlation distance* [21] or spectral distance [15] but it is not a distance in the sense of a metric (can be zero for a pair of distinct subspaces) and thus not listed. The spectral distance $d^{\sigma}_{\operatorname{Gr}(k,n)}$ is also called chordal 2-norm distance [6]. For each distance in Table 2 defined for equidimensional $\mathbf{A}$ and $\mathbf{B}$, Theorem 12 provides a corresponding version for when $\dim \mathbf{A} \neq \dim \mathbf{B}$.

The fact that all these distances in Table 2 depend on the principal angles is not a coincidence — the result [48, Theorem 3] implies the following.

Theorem 2. *Any notion of distance between k-dimensional subspaces in $\mathbb{R}^n$ that depends only on the relative positions of the subspaces, i.e., invariant under any rotation in $\mathcal{O}(n)$, must be a function of their principal angles. To be more specific, if a distance $d : \operatorname{Gr}(k,n) \times \operatorname{Gr}(k,n) \to [0,\infty)$ satisfies*

$$d(Q \cdot \mathbf{A}, Q \cdot \mathbf{B}) = d(\mathbf{A}, \mathbf{B}),$$

for all $\mathbf{A}, \mathbf{B} \in \operatorname{Gr}(k,n)$ and all $Q \in \mathcal{O}(n)$, where the action is as defined in (3)*, then d must be a function of $\theta_i(\mathbf{A},\mathbf{B})$, $i = 1, \dots, k$.*

We will next introduce the *infinite Grassmannian* $\operatorname{Gr}(k,\infty)$ to show that these distances between subspaces are independent of the dimension of their ambient space.

3. The Infinite Grassmannian

One way of defining a distance between $\mathbf{A} \in \operatorname{Gr}(k,n)$ and $\mathbf{B} \in \operatorname{Gr}(l,n)$ where $k \neq l$ is to first isometrically embed $\operatorname{Gr}(k,n)$ and $\operatorname{Gr}(l,n)$ into an ambient Riemannian manifold and then define the distance between $\mathbf{A}$ and $\mathbf{B}$ as their distance in the ambient space. This approach is taken in [12, 41], via an isometric embedding of $\operatorname{Gr}(0,n), \operatorname{Gr}(1,n), \dots, \operatorname{Gr}(n,n)$ into a sphere of dimension $(n-1)(n+2)/2$. Such a distance suffers from two shortcomings: It is not intrinsic to the Grassmannian and it depends on both the embedding and the ambient space.

The distance that we propose in Section 4 will depend only on the intrinsic distance of the Grassmannian and is independent of n, i.e., a k-plane $\mathbf{A}$ and an l-plane $\mathbf{B}$ in $\mathbb{R}^n$ will have the same distance if we regard them as subspaces in $\mathbb{R}^m$ for any $m \geq \min(k,l)$. We will first establish this for the special case $k = l$.

Consider the inclusion map $\iota_n : \mathbb{R}^n \to \mathbb{R}^{n+1}$, $\iota_n(x_1, \dots, x_n) = (x_1, \dots, x_n, 0)$. It is easy to see that ι_n induces a natural inclusion of $\operatorname{Gr}(k,n)$ into $\operatorname{Gr}(k,n+1)$ which we will also denote by ι_n. For any $m > n$, composition of successive natural inclusions gives the inclusion $\iota_{nm} : \operatorname{Gr}(k,n) \to \operatorname{Gr}(k,m)$, where $\iota_{nm} \coloneqq \iota_n \circ \iota_{n+1} \circ \cdots \circ \iota_{m-1}$. To be more concrete, if $A \in \mathbb{R}^{n\times k}$ has orthonormal columns, then

$$\iota_{nm} : \operatorname{Gr}(k,n) \to \operatorname{Gr}(k,m), \qquad \operatorname{span}(A) \mapsto \operatorname{span}\left(\begin{bmatrix} A \\ 0 \end{bmatrix}\right), \tag{8}$$

where the zero block matrix is $(m-n) \times k$ so that $\left[\begin{smallmatrix} A \\ 0 \end{smallmatrix}\right] \in \mathbb{R}^{m\times k}$.

For a fixed k, the family of Grassmannians $\{\operatorname{Gr}(k,n) : n \in \mathbb{N},\ n \geq k\}$ together with the inclusion maps $\iota_{nm} : \operatorname{Gr}(k,n) \to \operatorname{Gr}(k,m)$ for $m > n$ form a direct system. The *infinite Grassmannian* of k-planes is defined to be the direct limit of this system in the category of topological spaces and denoted by

$$\operatorname{Gr}(k,\infty) \coloneqq \varinjlim \operatorname{Gr}(k,n).$$

Those unfamiliar with the notion of direct limits may simply take

$$\operatorname{Gr}(k,\infty) = \bigcup\nolimits_{n=k}^{\infty} \operatorname{Gr}(k,n),$$

where we regard $\operatorname{Gr}(k,n) \subset \operatorname{Gr}(k,n+1)$ by identifying $\operatorname{Gr}(k,n)$ with $\iota_n\big(\operatorname{Gr}(k,n)\big)$. With this identification, we no longer need to distinguish between $\mathbf{A} \in \operatorname{Gr}(k,n)$ and its image $\iota_n(\mathbf{A}) \in \operatorname{Gr}(k,n+1)$ and may regard $\mathbf{A} \in \operatorname{Gr}(k,m)$ for all $m > n$.

We now define a distance $d_{\operatorname{Gr}(k,\infty)}$ on $\operatorname{Gr}(k,\infty)$ that is consistent with the Grassmann distance on $\operatorname{Gr}(k,n)$ for all n sufficiently large.

Lemma 3. *The natural inclusion $\iota_n : \operatorname{Gr}(k,n) \to \operatorname{Gr}(k,n+1)$ is isometric, i.e.,*

$$d_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = d_{\operatorname{Gr}(k,n+1)}\big(\iota_n(\mathbf{A}), \iota_n(\mathbf{B})\big). \tag{9}$$

Repeated applications of (9) *yields*

$$d_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = d_{\operatorname{Gr}(k,m)}\big(\iota_{nm}(\mathbf{A}), \iota_{nm}(\mathbf{B})\big) \tag{10}$$

for all $m > n$ and if we identify $\operatorname{Gr}(k,n)$ with $\iota_n\big(\operatorname{Gr}(k,n)\big)$, we may rewrite (10) *as*

$$d_{\operatorname{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = d_{\operatorname{Gr}(k,m)}(\mathbf{A},\mathbf{B})$$

for all $m > n$.

Proof. If $a \in \mathbb{R}^n$, we write $\hat{a} = \left[\begin{smallmatrix} a \\ 0 \end{smallmatrix}\right] \in \mathbb{R}^{n+1}$. Let $A = [a_1, \dots, a_k]$ and $B = [b_1, \dots, b_k]$ be any orthonormal bases of $\mathbf{A}$ and $\mathbf{B}$ respectively. By the definition of ι_n, $\iota_n(\mathbf{A})$ is the subspace in $\mathbb{R}^{n+1}$ spanned by an orthonormal basis that we will denote by $\iota_n(A) \coloneqq [\hat{a}_1, \dots, \hat{a}_k] \in \mathbb{R}^{(n+1)\times k}$. Hence we have

$$\iota_n(A)^{\mathsf{T}} \iota_n(B) = \begin{bmatrix} A^{\mathsf{T}} B \\ 0 \end{bmatrix}.$$

By the expression for Grassmann distance in (6), we see that (9) must hold. □

Since the inclusion of $\mathrm{Gr}(k,n)$ in $\mathrm{Gr}(k,n+1)$ is isometric, a geodesic in $\mathrm{Gr}(k,n)$ remains a geodesic in $\mathrm{Gr}(k,n+1)$. Given $\mathbf{A},\mathbf{B}\in\mathrm{Gr}(k,\infty)$, there must exist some n sufficiently large so that both $\mathbf{A},\mathbf{B}\in\mathrm{Gr}(k,n)$ and in which case we define the distance between $\mathbf{A}$ and $\mathbf{B}$ in $\mathrm{Gr}(k,\infty)$ to be

$$d_{\mathrm{Gr}(k,\infty)}(\mathbf{A},\mathbf{B}) \coloneqq d_{\mathrm{Gr}(k,n)}(\mathbf{A},\mathbf{B}).$$

By Lemma 3, this value is independent of our choice of n and is the same for all $m\geq n$. In particular, $d_{\mathrm{Gr}(k,\infty)}$ is well-defined and yields a distance on $\mathrm{Gr}(k,\infty)$. We summarize these observations below.

Corollary 4. *The Grassmann distance between two k-planes in $\mathrm{Gr}(k,n)$ is the geodesic distance in $\mathrm{Gr}(k,\infty)$ and is therefore independent of n. Also, the expression (7) for a distance minimizing geodesic in $\mathrm{Gr}(k,n)$ extends to $\mathrm{Gr}(k,\infty)$.*

Lemma 3 also holds for other distances on $\mathrm{Gr}(k,n)$ in Table 2, allowing us to define them on $\mathrm{Gr}(k,\infty)$.

Lemma 5. *For all $m>n$, the inclusion $\iota_{nm}:\mathrm{Gr}(k,n)\to\mathrm{Gr}(k,m)$ is isometric when $\mathrm{Gr}(k,n)$ and $\mathrm{Gr}(k,m)$ are both equipped with the same distance in Table 2, i.e.,*

$$d^*_{\mathrm{Gr}(k,n)}(\mathbf{A},\mathbf{B}) = d^*_{\mathrm{Gr}(k,m)}\big(\iota_{nm}(\mathbf{A}),\iota_{nm}(\mathbf{B})\big),$$

$* = \alpha,\beta,\kappa,\phi,\mu,\rho,\pi,\sigma$. Consequently $d^*_{\mathrm{Gr}(k,\infty)}$ is well-defined.

Proof. $d^*_{\mathrm{Gr}(k,n)}(\mathbf{A},\mathbf{B})$ and $d^*_{\mathrm{Gr}(k,n+1)}\big(\iota_n(\mathbf{A}),\iota_n(\mathbf{B})\big)$ depend only on the principal angles between $\mathbf{A}$ and $\mathbf{B}$, so the distance remains unchanged under ι_n. Repeated applications to $\iota_n\circ\iota_{n+1}\circ\cdots\circ\iota_{m-1}=\iota_{nm}$ yield the required isometry. □

4. Distances between subspaces of different dimensions

We now address our main problem. The proposed notion of distance will be that of a point $x\in X$ to a set $S\subset X$ in a metric space (X,d). Recall that this is defined by $d(x,S)\coloneqq\inf\{d(x,y):y\in S\}$. For us, X is a Grassmannian, therefore compact, and so $d(x,S)$ is finite. Also, S will be a closed subset and so we write min instead of inf. We will introduce two possible candidates for S.

Definition 6. Let $k,l,n\in\mathbb{N}$ be such that $k\leq l\leq n$. For any $\mathbf{A}\in\mathrm{Gr}(k,n)$ and $\mathbf{B}\in\mathrm{Gr}(l,n)$, we define the subsets

$$\Omega_+(\mathbf{A})\coloneqq\{\mathbf{X}\in\mathrm{Gr}(l,n):\mathbf{A}\subseteq\mathbf{X}\},\quad \Omega_-(\mathbf{B})\coloneqq\{\mathbf{Y}\in\mathrm{Gr}(k,n):\mathbf{Y}\subseteq\mathbf{B}\}.$$

We will call $\Omega_+(\mathbf{A})$ the *Schubert variety of l-planes containing* $\mathbf{A}$ and $\Omega_-(\mathbf{B})$ the *Schubert variety of k-planes contained in* $\mathbf{B}$.

As we will see in Section 8, $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are indeed Schubert varieties and therefore closed subsets of $\mathrm{Gr}(l,n)$ and $\mathrm{Gr}(k,n)$ respectively. Furthermore, $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are uniquely determined by $\mathbf{A}$ and $\mathbf{B}$ (see Proposition 20) and may be regarded as 'sub-Grassmannians' of $\mathrm{Gr}(l,n)$ and $\mathrm{Gr}(k,n)$ respectively (see Proposition 21).

How could one define the distance between a subspace $\mathbf{A}$ of dimension k and a subspace $\mathbf{B}$ of dimension l in $\mathbb{R}^n$ when $k\neq l$? We may assume $k<l\leq n$ without loss of generality. In which case a very natural solution is to define the required distance $\delta(\mathbf{A},\mathbf{B})$ as that between the k-plane $\mathbf{A}$ and the closest k-plane $\mathbf{Y}$ contained in $\mathbf{B}$, measured within $\mathrm{Gr}(k,n)$. In other words, we want the Grassmann distance from $\mathbf{A}$ to the closed subset $\Omega_-(\mathbf{B})$,

$$\delta(\mathbf{A},\mathbf{B})\coloneqq d_{\mathrm{Gr}(k,n)}\big(\mathbf{A},\Omega_-(\mathbf{B})\big)=\min\big\{d_{\mathrm{Gr}(k,n)}(\mathbf{A},\mathbf{Y}):\mathbf{Y}\in\Omega_-(\mathbf{B})\big\}. \tag{11}$$

This has the advantage of being intrinsic — the distance $\delta(\mathbf{A},\mathbf{B})$ is measured in $d_{\mathrm{Gr}(k,n)}$ and is defined wholly within $\mathrm{Gr}(k,n)$ without any embedding of $\mathrm{Gr}(k,n)$ into an arbitrary ambient space. Furthermore, by the property of $d_{\mathrm{Gr}(k,n)}$ in Corollary 4, $\delta(\mathbf{A},\mathbf{B})$ does not depend on n and takes the same value for any $m\geq n$. We illustrate this in Figure 1: The sphere is intended to be a depiction of $\mathrm{Gr}(1,3)$ though to be accurate antipodal points on the sphere should be identified.

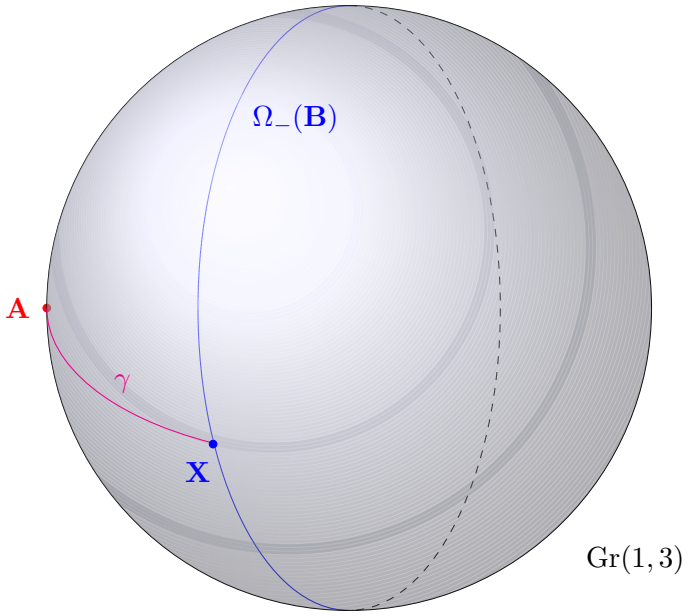

FIGURE 1. Distance between a line $\mathbf{A}$ and a plane $\mathbf{B}$ in $\mathbb{R}^3$. $\mathbf{X}$ is closest to $\mathbf{A}$ among all lines in $\mathbf{B}$. The length of the geodesic γ from $\mathbf{A}$ to $\mathbf{X}$ gives the distance.

However, it is equally natural to define $\delta(\mathbf{A}, \mathbf{B})$ as the distance between the l-plane $\mathbf{B}$ and the closest l-plane $\mathbf{Y}$ containing $\mathbf{A}$, measured within $\operatorname{Gr}(l, n)$. In other words, we could have instead defined it as the Grassmann distance from $\mathbf{B}$ to the closed subset $\Omega_+(\mathbf{A})$,

$$\delta(\mathbf{A}, \mathbf{B}) \coloneqq d_{\operatorname{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big) = \min\big\{d_{\operatorname{Gr}(l,n)}(\mathbf{B}, \mathbf{X}) : \mathbf{X} \in \Omega_+(\mathbf{A})\big\}. \tag{12}$$

It will have the same desirable features as the one in (11) except that the distance is now measured in $d_{\operatorname{Gr}(l,n)}$ and within $\operatorname{Gr}(l, n)$.

It turns out that the two values in (11) and (12) are equal, allowing us to define $\delta(\mathbf{A}, \mathbf{B})$ as their common value. We will establish this equality and the properties of $\delta(\mathbf{A}, \mathbf{B})$ in the remainder of this section. The results are summarized in Theorem 7. Our proof is constructive: In addition to showing the equality of (11) and (12), it shows how one may explicitly find the closest points on Schubert varieties $\mathbf{X} \in \Omega_-(\mathbf{B})$ and $\mathbf{Y} \in \Omega_+(\mathbf{A})$ to any given point in the respective Grassmannians.

Theorem 7. *Let $\mathbf{A}$ be a subspace of dimension k and $\mathbf{B}$ be a subspace of dimension l in $\mathbb{R}^n$. Suppose $k \leq l \leq n$. Then*

$$d_{\operatorname{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big) = d_{\operatorname{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big). \tag{13}$$

Their common value defines a distance $\delta(\mathbf{A}, \mathbf{B})$ between the two subspaces with the following properties:

(i) *$\delta(\mathbf{A}, \mathbf{B})$ is independent of the dimension of the ambient space n and is the same for all $n \geq l+1$;*
(ii) *$\delta(\mathbf{A}, \mathbf{B})$ reduces to the Grassmann distance between $\mathbf{A}$ and $\mathbf{B}$ when $k = l$;*
(iii) *$\delta(\mathbf{A}, \mathbf{B})$ may be computed explicitly as*

$$\delta(\mathbf{A}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^{\min\{k,l\}} \theta_i(\mathbf{A}, \mathbf{B})^2\Big)^{1/2} \tag{14}$$

where $\theta_i(\mathbf{A}, \mathbf{B})$ is the ith principal angle between $\mathbf{A}$ and $\mathbf{B}$, $i = 1, \dots, \min(k, l)$.

Rewriting (13) as

$$\min_{\mathbf{X} \in \Omega_+(\mathbf{A})} d_{\operatorname{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) = \min_{\mathbf{Y} \in \Omega_-(\mathbf{B})} d_{\operatorname{Gr}(k,n)}(\mathbf{Y}, \mathbf{A}),$$

the equation says that the distance from $\mathbf{B}$ to the nearest l-dimensional subspace that contains $\mathbf{A}$ equals the distance from $\mathbf{A}$ to the nearest k-dimensional subspace contained in $\mathbf{B}$. This relation has several parallels. We will see that:

(a) the Grassmann distance may be replaced by any of the distances in Table 2 (see Theorem 12);
(b) 'nearest' may be replaced by 'furthest' and 'min' above replaced by 'max' when n is sufficiently large (see Proposition 16);
(c) 'distance' may be replaced by 'volume' with respect to the intrinsic uniform probability density on the Grassmannian (see Section 9).

We will prove Theorem 7 by way of the next two lemmas.

Lemma 8. *Let $k \leq l \leq n$ be positive integers. Let $\delta : \operatorname{Gr}(k,n) \times \operatorname{Gr}(l,n) \to [0,\infty)$ be the function defined by*

$$\delta(\mathbf{A},\mathbf{B}) = \Big(\sum\nolimits_{i=1}^{k} \theta_i^2\Big)^{1/2}$$

where $\theta_i \coloneqq \theta_i(\mathbf{A},\mathbf{B})$, $i = 1,\dots,k$. Then

$$\delta(\mathbf{A},\mathbf{B}) \geq d_{\operatorname{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big).$$

Proof. It suffices to find an $\mathbf{X} \in \Omega_+(\mathbf{A})$ such that $\delta(\mathbf{A},\mathbf{B}) = d_{\operatorname{Gr}(l,n)}(\mathbf{X},\mathbf{B})$. Let $(p_1,q_1),\dots,(p_k,q_k)$ be the principal vectors between $\mathbf{A}$ and $\mathbf{B}$. We will extend $q_1,\dots,q_k$ into an orthonormal basis of $\mathbf{B}$ by appending appropriate orthonormal vectors $q_{k+1},\dots,q_l$. The principal angles are given by $\theta_i = \cos^{-1} p_i^{\mathsf{T}} q_i$, $\|p_i\| = \|q_i\| = 1$. If we take $\mathbf{X} \in \operatorname{Gr}(l,n)$ to be the subspace spanned by $p_1,\dots,p_k,q_{k+1},\dots,q_l$, then

$$\begin{aligned} d_{\operatorname{Gr}(l,n)}(\mathbf{X},\mathbf{B}) &= [(\cos^{-1} p_1^{\mathsf{T}} q_1)^2 + \cdots + (\cos^{-1} p_k^{\mathsf{T}} q_k)^2 \\ &\qquad + (\cos^{-1} q_{k+1}^{\mathsf{T}} q_{k+1})^2 + \cdots + (\cos^{-1} q_l^{\mathsf{T}} q_l)^2]^{1/2} \\ &= [\theta_1^2 + \cdots + \theta_k^2 + 0^2 + \cdots + 0^2]^{1/2} = \delta(\mathbf{A},\mathbf{B}). \end{aligned} \tag{15}$$

□

We state the following well-known fact [26, Corollary 3.1.3] for easy reference and deduce a corollary that will be useful for Lemma 11.

Proposition 9. *Let $k \leq l \leq n$ be positive integers. Suppose $B \in \mathbb{R}^{n\times l}$ and $B_k \in \mathbb{R}^{n\times k}$ is a submatrix obtained by removing any $l-k$ columns from B. Then the ith singular values satisfy $\sigma_i(B_k) \leq \sigma_i(B)$ for $i = 1,\dots,k$.*

Corollary 10. *Let B and B_k be as in Proposition 9 and $\mathbf{B}$ and $\mathbf{B}_k$ be subspaces of $\mathbb{R}^n$ spanned by the column vectors of B and B_k respectively. Then for any subspace $\mathbf{A}$ of $\mathbb{R}^n$, the principal angles between the respective subspaces satisfy*

$$\theta_i(\mathbf{A},\mathbf{B}) \leq \theta_i(\mathbf{A},\mathbf{B}_k)$$

for $i = 1,\dots,\min(\dim \mathbf{A}, \dim \mathbf{B}_k)$.

Proof. By appropriate orthogonalization if necessary, we may assume that B and its submatrix B_k are orthonormal bases of $\mathbf{B}$ and $\mathbf{B}_k$. Let A be an orthonormal basis of $\mathbf{A}$. Then $\sigma_i(A^{\mathsf{T}}B)$ and $\sigma_i(A^{\mathsf{T}}B_k)$ take values in $[0,1]$. Since $\theta_i(\mathbf{A},\mathbf{B}) = \cos^{-1}(\sigma_i\big(A^{\mathsf{T}}B\big)\big)$ and $\cos^{-1}$ is monotone decreasing in $[0,1]$, the result follows from $\sigma_i(A^{\mathsf{T}}B) \geq \sigma_i(A^{\mathsf{T}}B_k)$, by Proposition 9 applied to the submatrix $A^{\mathsf{T}}B_k$ of $A^{\mathsf{T}}B$. □

Lemma 11. *Let $\mathbf{A}$, $\mathbf{B}$ be as in Lemma 8. Then $\delta(\mathbf{A},\mathbf{B}) \leq d_{\operatorname{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big)$.*

Proof. Let $\mathbf{Y} \in \Omega_-(\mathbf{B})$. Then $\mathbf{Y}$ is a k-dimensional subspace contained in $\mathbf{B}$ and in the notation of Corollary 10, we may write $\mathbf{Y} = \mathbf{B}_k$. By the same corollary we get $\theta_i(\mathbf{A}, \mathbf{B}) \leq \theta_i(\mathbf{A}, \mathbf{Y})$ for $i = 1, \dots, k$. Hence

$$\delta(\mathbf{A}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^{k} \theta_i(\mathbf{A}, \mathbf{B})^2\Big)^{1/2} \leq \Big(\sum\nolimits_{i=1}^{k} \theta_i(\mathbf{A}, \mathbf{Y})^2\Big)^{1/2} = d_{\mathrm{Gr}(k,n)}(\mathbf{A}, \mathbf{Y}). \tag{16}$$

The desired inequality follows since this holds for arbitrary $\mathbf{Y} \in \Omega_-(\mathbf{B})$. □

Proof of Theorem 7. Recall that Grassmannians satisfy an isomorphism

$$\mathrm{Gr}(k, n) \cong \mathrm{Gr}(n - k, n)$$

that takes a k-plane $\mathbf{Y}$ to the $(n-k)$-plane $\mathbf{Y}^\perp$ of linear forms vanishing on $\mathbf{Y}$. It is easy to see that this isomorphism is an isometry. Using this isometric isomorphism, together with Lemma 8 and Lemma 11, we can immediately deduce that

$$\delta(\mathbf{A}, \mathbf{B}) \leq d_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-\big(\mathbf{B}\big)\big) = d_{\mathrm{Gr}(n-k,n)}\big(\mathbf{A}^\perp, \Omega_+\big(\mathbf{B}^\perp\big)\big) \leq \delta(\mathbf{A}^\perp, \mathbf{B}^\perp).$$

On the other hand, by results in [29], we have $\delta(\mathbf{A}, \mathbf{B}) = \delta(\mathbf{A}^\perp, \mathbf{B}^\perp)$ and hence

$$\delta(\mathbf{A}, \mathbf{B}) = d_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big).$$

Similarly we can obtain

$$\delta(\mathbf{A}, \mathbf{B}) = d_{\mathrm{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big).$$

Hence we have the required equalities (13) and (14) in Theorem 7. Property (ii) is obvious from (14) and Property (i) follows from Lemma 3. □

The proof of Lemma 8 provides a simple way to find a point $\mathbf{X} \in \Omega_+(\mathbf{A})$ that realizes the distance $d_{\mathrm{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big) = \delta(\mathbf{A}, \mathbf{B})$. Similarly we may explicitly determine a point $\mathbf{Y} \in \Omega_-(\mathbf{B})$ that realizes the distance $d_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big) = \delta(\mathbf{A}, \mathbf{B})$.

One might wonder whether or not Theorem 7 still holds if we replace $d_{\mathrm{Gr}(k,n)}$ by other distance functions described in Table 2. The answer is yes.

Theorem 12. *Let $k \leq l \leq n$. Let $\mathbf{A} \in \mathrm{Gr}(k, n)$ and $\mathbf{B} \in \mathrm{Gr}(l, n)$. Then*

$$d^*_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big) = d^*_{\mathrm{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big),$$

for $* = \alpha, \beta, \kappa, \phi, \mu, \rho, \pi, \sigma$. Their common value $\delta^*(\mathbf{A}, \mathbf{B})$ is given by:

$$\begin{aligned}
&\delta^\alpha(\mathbf{A}, \mathbf{B}) = \theta_k, && \delta^\beta(\mathbf{A}, \mathbf{B}) = \Big(1 - \prod\nolimits_{i=1}^{k} \cos^2\theta_i\Big)^{1/2},\\
&\delta^\kappa(\mathbf{A}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^{k} \sin^2\theta_i\Big)^{1/2}, && \delta^\phi(\mathbf{A}, \mathbf{B}) = \cos^{-1}\big(\prod\nolimits_{i=1}^{k} \cos\theta_i\big),\\
&\delta^\mu(\mathbf{A}, \mathbf{B}) = \Big(\log \prod\nolimits_{i=1}^{k} \frac{1}{\cos^2\theta_i}\Big)^{1/2}, && \delta^\rho(\mathbf{A}, \mathbf{B}) = \Big(2\sum\nolimits_{i=1}^{k} \sin^2(\theta_i/2)\Big)^{1/2},\\
&\delta^\pi(\mathbf{A}, \mathbf{B}) = \sin\theta_k, && \delta^\sigma(\mathbf{A}, \mathbf{B}) = 2\sin(\theta_k/2),
\end{aligned}$$

or more generally with $\min(k, l)$ in place of the index k when we do not require $k \leq l$.

Proof. This follows from observing that our proof of Theorem 7 only involves principal angles between $\mathbf{A}$ and $\mathbf{B}$ and the diffeomorphism between $\mathrm{Gr}(k, n)$ and $\mathrm{Gr}(n - k, n)$ remains an isometry under these distances. In particular, both (15) and (16) still hold with any of these distances in place of the Grassmann distance. □

We will see in Section 7 that the projection distance δ^π in Theorem 12 is equivalent to the containment gap, a measure of distance between subspaces of different dimensions originally proposed in operator theory [28].

5. Grassmannian of subspaces of all dimensions

We view the equality of $d_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big)$ and $d_{\mathrm{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big)$ as the strongest evidence that their common value $\delta(\mathbf{A}, \mathbf{B})$ provides the most natural notion of distance between subspaces of different dimensions. As we pointed out earlier, δ is a distance in the sense of a distance from a point to a set, but not a distance in the sense of a metric on the set of all subspaces of all dimensions. For instance, δ does not satisfy the separation property: $\delta(\mathbf{A}, \mathbf{B}) = 0$ for any $\mathbf{A} \subsetneq \mathbf{B}$. In fact, it is easy to observe the following.

Lemma 13. *Let $\mathbf{A} \in \mathrm{Gr}(k, n)$ and $\mathbf{B} \in \mathrm{Gr}(l, n)$. Then $\delta(\mathbf{A}, \mathbf{B}) = 0$ iff $\mathbf{A} \subseteq \mathbf{B}$ or $\mathbf{B} \subseteq \mathbf{A}$.*

δ also does not satisfy the triangle inequality: For a line $\mathbf{L}$ not contained in a subspace $\mathbf{A}$, the triangle inequality, if true, would imply

$$\delta(\mathbf{L}, \mathbf{A}) = \delta(\mathbf{L}, \mathbf{A}) + \delta(\mathbf{A}, \mathbf{B}) \geq \delta(\mathbf{L}, \mathbf{B}),$$
$$\delta(\mathbf{L}, \mathbf{B}) = \delta(\mathbf{L}, \mathbf{B}) + \delta(\mathbf{A}, \mathbf{B}) \geq \delta(\mathbf{L}, \mathbf{A}),$$

giving $\delta(\mathbf{L}, \mathbf{A}) = \delta(\mathbf{L}, \mathbf{B})$ for any subspace $\mathbf{B}$, which is evidently false by Lemma 13 (e.g. take $\mathbf{B} = \mathbf{A} \oplus \mathbf{L}$).

These observations also apply verbatim to all the other similarly-defined distances δ^* in Theorem 12, i.e., none of them are metrics.

The set of all subspaces of all dimensions is parameterized by $\mathrm{Gr}(\infty, \infty)$, the *doubly infinite Grassmannian* [19], which may be viewed informally as the disjoint union of all k-dimensional subspaces[2] over all $k \in \mathbb{N}$,

$$\mathrm{Gr}(\infty, \infty) = \coprod_{k=1}^{\infty} \mathrm{Gr}(k, \infty).$$

To define a metric between any pair of subspaces of arbitrary dimensions is to define one on $\mathrm{Gr}(\infty, \infty)$. It is easy to define metrics on $\mathrm{Gr}(\infty, \infty)$ that bear little relation to the geometry of Grassmannian but we will propose one in Section 6 that is consistent with δ and with $d_{\mathrm{Gr}(k,n)}$ for all $k \leq n$.

We will require the formal definition of $\mathrm{Gr}(\infty, \infty)$, namely, it is the direct limit of the direct system of Grassmannians $\{\mathrm{Gr}(k, n) : (k, n) \in \mathbb{N} \times \mathbb{N}\}$ with inclusion maps $\iota_{nm}^{kl} : \mathrm{Gr}(k, n) \to \mathrm{Gr}(l, m)$ for all $k \leq l$ and $n \leq m$ such that $l - k \leq m - n$. For $A \in \mathbb{R}^{n \times k}$ with orthonormal columns, the embedding is given by

$$\iota_{nm}^{kl} : \mathrm{Gr}(k, n) \to \mathrm{Gr}(l, m), \qquad \operatorname{span}(A) \mapsto \operatorname{span}\left(\begin{bmatrix} A & 0 \\ 0 & 0 \\ 0 & I_{l-k} \end{bmatrix}\right), \tag{17}$$

where $I_{l-k} \in \mathbb{R}^{(l-k)\times(l-k)}$ is an identity matrix and we have $(m - n) - (l - k)$ zero rows in the middle so that the 3×2 block matrix is in $\mathbb{R}^{m \times l}$. Note that for a fixed k, ι_{nm}^{kk} reduces to ι_{nm} in (8).

Since our distance $\delta(\mathbf{A}, \mathbf{B})$ is defined for subspaces $\mathbf{A}$ and $\mathbf{B}$ of all dimensions, it defines a function $\delta : \mathrm{Gr}(\infty, \infty) \times \mathrm{Gr}(\infty, \infty) \to \mathbb{R}$ that is a *premetric* on $\mathrm{Gr}(\infty, \infty)$, i.e., $\delta(\mathbf{A}, \mathbf{B}) \geq 0$ and $\delta(\mathbf{A}, \mathbf{A}) = 0$ for all $\mathbf{A}, \mathbf{B} \in \mathrm{Gr}(\infty, \infty)$. This in turn defines a topology τ on $\mathrm{Gr}(\infty, \infty)$ in a standard way: The ε-ball centered at $\mathbf{A}$ is

$$B_\varepsilon(\mathbf{A}) \coloneqq \{\mathbf{X} \in \mathrm{Gr}(\infty, \infty) : \delta(\mathbf{A}, \mathbf{X}) < \varepsilon\},$$

and $U \subseteq \mathrm{Gr}(\infty, \infty)$ is defined to be open if for any $\mathbf{A} \in U$, there is an ε-ball $B_\varepsilon(\mathbf{A}) \subseteq U$. The topology τ is consistent with the usual topology of Grassmannians (but it is not the disjoint union topology). If we restrict τ to $\mathrm{Gr}(k, \infty)$, then the subspace topology is the same as the topology induced by the metric $d_{\mathrm{Gr}(k,\infty)}$ on $\mathrm{Gr}(k, \infty)$ as defined in Section 3. Nevertheless this apparently natural topology on $\mathrm{Gr}(\infty, \infty)$ turns out to be a strange one.

[2]As discussed in Section 3, these are independent of the dimension of their ambient space and may be viewed as an element of the infinite Grassmannian $\mathrm{Gr}(k, \infty)$.

Proposition 14. *The topology τ on $\operatorname{Gr}(\infty,\infty)$ is non-Hausdorff and therefore non-metrizable.*

Proof. τ is not Hausdorff since it is not possible to separate $\mathbf{A} \subsetneq \mathbf{B}$ by open subsets, as we saw in Lemma 13. Metrizable spaces are necessarily Hausdorff. □

Even though τ restricts to the metric space topology on $\operatorname{Gr}(k,\infty)$ induced by the Grassmann distance $d_{\operatorname{Gr}(k,\infty)}$ for every $k \in \mathbb{N}$, it is not itself a metric space topology. We view this as a consequence of a more general phenomenon, namely, the category **Met** of metric spaces (objects) and continuous contractions (morphisms) has no coproduct, i.e., given a collection of metric spaces, there is in general no metric space that will behave like the disjoint union of the collection of metric spaces. To see this, take metric spaces (X_1, d_1) and (X_2, d_2) where $X_1 = \{x_1\}$, $X_2 = \{x_2\}$. Suppose a coproduct (X, d) of (X_1, d_1) and (X_2, d_2) exists. Let $Y = \{y_1, y_2\}$ and let d_Y be the metric on Y induced by $d_Y(y_1, y_2) = 2d(x_1, x_2) \neq 0$. Now define $\varphi_i : X_i \to Y$ by $\varphi_i(x_i) = y_i$, $i = 1, 2$. One sees that no morphism $\varphi : X \to Y$ in **Met** is compatible with φ_1 and φ_2, contradicting the assumption that X is the coproduct of X_1 and X_2.

If we instead look at the category of metric spaces with continuous or uniformly continuous maps as morphisms, then coproducts always exist [25]. In Section 6, we will relax our requirement and construct a metric $d_{\operatorname{Gr}(\infty,\infty)}$ on $\operatorname{Gr}(\infty,\infty)$ that restricts to $d_{\operatorname{Gr}(k,\infty)}$ for all $k \in \mathbb{N}$ but without requiring that it comes from a coproduct of $\{(\operatorname{Gr}(k,\infty), d_{\operatorname{Gr}(k,\infty)}) : k \in \mathbb{N}\}$ in **Met**.

6. Metrics for subspaces of all dimensions

We will describe a simple recipe for turning the distances δ^* in Theorem 12 into metrics on $\operatorname{Gr}(\infty,\infty)$. Suppose $k \leq l$ and we have $\mathbf{A} \in \operatorname{Gr}(k,n)$ and $\mathbf{B} \in \operatorname{Gr}(l,n)$. In this case there are k principal angles between $\mathbf{A}$ and $\mathbf{B}$, $\theta_1, \dots, \theta_k$, as defined in (5). First we will set $\theta_{k+1} = \cdots = \theta_l = \pi/2$. Then we take the Grassmann distance δ or any of the distances δ^* in Theorem 12, replace the index k by l, and call the resulting expressions $d_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B})$ (for Grassmann distance) and $d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B})$ (for other distances) respectively.

When n is sufficiently large, setting $\theta_{k+1}, \dots, \theta_l$ all equal to $\pi/2$ is equivalent to completing $\mathbf{A}$ to an l dimensional subspace of $\mathbb{R}^n$, by adding $l-k$ vectors orthonormal to the subspace $\mathbf{B}$. Hence the distance between $\mathbf{A}$ and $\mathbf{B}$ is defined by the distance function on the Grassmannian $\operatorname{Gr}(l,n)$. We show in Proposition 15 that these expressions will indeed define metrics on $\operatorname{Gr}(\infty,\infty)$.

Applying the above recipe to the Grassmann, chordal, and Procrustes distances yield the *Grassmann*, *chordal*, and *Procrustes metrics* on $\operatorname{Gr}(\infty,\infty)$ given in Table 3.

Table 3. Metrics on $\operatorname{Gr}(\infty,\infty)$ in terms of principal angles.

Grassmann metric	$d_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B}) = \Big(\lvert k-l\rvert\pi^2/4 + \sum_{i=1}^{\min(k,l)} \theta_i^2\Big)^{1/2}$
Chordal metric	$d^{\kappa}_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B}) = \Big(\lvert k-l\rvert + \sum_{i=1}^{\min(k,l)} \sin^2\theta_i\Big)^{1/2}$
Procrustes metric	$d^{\rho}_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B}) = \Big(\lvert k-l\rvert + 2\sum_{i=1}^{\min(k,l)} \sin^2(\theta_i/2)\Big)^{1/2}$

Evidently the metrics in Table 3 are all of the form

$$d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B}) = \sqrt{\delta^*(\mathbf{A},\mathbf{B})^2 + c_*^2\epsilon(\mathbf{A},\mathbf{B})^2}, \tag{18}$$

where $\epsilon(\mathbf{A},\mathbf{B}) \coloneqq |\dim \mathbf{A} - \dim \mathbf{B}|^{1/2}$. On the other hand, applying the above recipe to other distances in Table 2 yield the *Asimov*, *Binet–Cauchy*, *Fubini–Study*, *Martin*, *projection*, and *spectral metrics* on $\operatorname{Gr}(\infty,\infty)$ given by

$$d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B}) = \begin{cases} d^*_{\operatorname{Gr}(k,\infty)}(\mathbf{A},\mathbf{B}) & \text{if } \dim \mathbf{A} = \dim \mathbf{B} = k, \\ c_* & \text{if } \dim \mathbf{A} \neq \dim \mathbf{B}, \end{cases} \tag{19}$$

for $* = \alpha, \beta, \phi, \mu, \pi, \sigma$, respectively. The constants $c_* > 0$ can be seen to be

$$c = c_\alpha = \pi/2, \quad c_\sigma = \sqrt{2}, \quad c_\mu = \infty, \quad c_\beta = c_\phi = c_\pi = c_\kappa = c_\rho = 1.$$

In all cases, for subspaces $\mathbf{A}$ and $\mathbf{B}$ of equal dimension k, these metrics on $\operatorname{Gr}(\infty, \infty)$ restrict to the corresponding ones on $\operatorname{Gr}(k, \infty)$, i.e.,

$$d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}) = d^*_{\operatorname{Gr}(k,\infty)}(\mathbf{A}, \mathbf{B}),$$

where the latter is as described in Corollary 4 and Lemma 5. These metrics on $\operatorname{Gr}(\infty, \infty)$ are the amalgamation of two pieces of information, the distance $\delta^*(\mathbf{A}, \mathbf{B})$ and the difference in dimensions $|\dim \mathbf{A} - \dim \mathbf{B}|$, either via a root mean square or an indicator function.

The Grassmann metric has a natural interpretation (see Proposition 16):

> $d_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B})$ *is the distance from* $\mathbf{B}$ *to the furthest* l*-dimensional subspace that contains* $\mathbf{A}$*, which equals the distance from* $\mathbf{A}$ *to the furthest* k*-dimensional subspace contained in* $\mathbf{B}$*.*

The chordal metric in Table 3 is equivalent to the symmetric directional distance, a metric on subspaces of different dimensions [44, 46] popular in machine learning [5, 7, 13, 16, 18, 24, 31, 35, 39, 40, 47, 49, 51] (see Section 7).

Proposition 15. *The expressions in Table 3 and* (19) *are metrics on* $\operatorname{Gr}(\infty, \infty)$*.*

Proof. It is trivial to see that the expression defined in (19) yields a metric on $\operatorname{Gr}(\infty, \infty)$ for $* = \alpha, \beta, \mu, \pi, \sigma, \phi$, and so we just need to check the remaining three cases that take the form in (18). Of the four defining properties of a metric, only the triangle inequality is not immediately clear from (18).

Let $k = \dim \mathbf{A}$, $l = \dim \mathbf{B}$, and $m = \dim \mathbf{C}$. We may assume WLOG that $k \leq l \leq m \leq n$ where n is chosen sufficiently large so that $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are subspaces in $\mathbb{R}^n$. Let $A \in \mathbb{R}^{n\times k}$, $B \in \mathbb{R}^{n\times l}$, $C \in \mathbb{R}^{n\times m}$ be matrices whose columns are orthonormal bases of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ respectively. Consider the following $(n + m - k) \times m$ matrices:

$$A' = \begin{bmatrix} A & 0 \\ 0 & I_{m-k} \end{bmatrix}, \quad B' = \begin{bmatrix} B & 0 \\ 0 & 0 \\ 0 & I_{m-l} \end{bmatrix}, \quad C' = \begin{bmatrix} C \\ 0 \end{bmatrix}.$$

and set $\mathbf{A}' = \operatorname{span}(A')$, $\mathbf{B}' = \operatorname{span}(B')$, $\mathbf{C}' = \operatorname{span}(C')$; note that these are just $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ embedded in $\operatorname{Graff}(m, n + m - k)$ via (17). The expressions in Table 3 satisfy

$$\begin{aligned} d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}) &= d^*_{\operatorname{Gr}(m,n+m-k)}(\mathbf{A}', \mathbf{B}'), \\ d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{B}, \mathbf{C}) &= d^*_{\operatorname{Gr}(m,n+m-k)}(\mathbf{B}', \mathbf{C}'), \\ d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{C}) &= d^*_{\operatorname{Gr}(m,n+m-k)}(\mathbf{A}', \mathbf{C}'). \end{aligned}$$

Since $\mathbf{A}', \mathbf{B}', \mathbf{C}' \in \operatorname{Gr}(m, n + m - k)$, the triangle inequality for $d^*_{\operatorname{Gr}(m,n+m-k)}$ immediately yields the triangle inequality for $d^*_{\operatorname{Gr}(\infty,\infty)}$. □

The proof shows that for any $\mathbf{A} \in \operatorname{Gr}(k, n)$ and $\mathbf{B} \in \operatorname{Gr}(l, n)$ where $k \leq l \leq n$,

$$d^*_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}) = d^*_{\operatorname{Gr}(l,n+l-k)}\big(\iota^{k,l}_{n,n+l-k}(\mathbf{A}), \iota^{l,l}_{n,n+l-k}(\mathbf{B})\big).$$

The embeddings $\iota^{k,l}_{n,n+l-k} : \operatorname{Gr}(k, n) \to \operatorname{Gr}(l, n + l - k)$ and $\iota^{l,l}_{n,n+l-k} : \operatorname{Gr}(l, n) \to \operatorname{Gr}(l, n + l - k)$ are as defined in (17) and are isometric for all $k \leq l \leq n$.

Proposition 16. *Let* $k \leq l \leq n/2$ *and* $\mathbf{A} \in \operatorname{Gr}(k, n)$*,* $\mathbf{B} \in \operatorname{Gr}(l, n)$*. Then*

$$\max_{\mathbf{X}\in\Omega_+(\mathbf{A})} d_{\operatorname{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) = \max_{\mathbf{Y}\in\Omega_-(\mathbf{B})} d_{\operatorname{Gr}(k,n)}(\mathbf{Y}, \mathbf{A}) = d_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}), \tag{20}$$

i.e., $d_{\operatorname{Gr}(\infty,\infty)}$ *is the distance between* furthest *subspaces.*

Proof. We assume WLOG that $\mathbf{A} \cap \mathbf{B} = \{0\}$ by Proposition 1. Since

$$d_{\mathrm{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) = \delta(\mathbf{X}, \mathbf{B}) = \Big(\sum\nolimits_{i=1}^{l} \theta_i(\mathbf{X}, \mathbf{B})^2\Big)^{1/2},$$

and by Corollary 10, $\theta_i(\mathbf{X}, \mathbf{B}) \leq \theta_i(\mathbf{A}, \mathbf{B})$, $i = 1, \dots, k$, we obtain

$$d_{\mathrm{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) \leq \Big(\delta(\mathbf{A}, \mathbf{B})^2 + \sum\nolimits_{i=k+1}^{l} \theta_i(\mathbf{X}, \mathbf{B})^2\Big)^{1/2}.$$

Let $(a_1, b_1), \dots, (a_k, b_k)$ be the principal vectors between $\mathbf{A}$ and $\mathbf{B}$. We extend $b_1, \dots, b_k$ to obtain an orthonormal basis $b_1, \dots, b_k, b_{k+1}, \dots, b_l$ of $\mathbf{B}$. Let $\mathbf{X} \cap \mathbf{A}^\perp$ be the orthogonal complement of $\mathbf{A}$ in $\mathbf{X}$ and let $\mathbf{B}_0 \coloneqq \operatorname{span}\{b_{k+1}, \dots, b_l\}$. Then

$$\Big(\sum\nolimits_{i=k+1}^{l} \theta_i(\mathbf{X}, \mathbf{B})^2\Big)^{1/2} = \delta(\mathbf{X} \cap \mathbf{A}^\perp, \mathbf{B}_0),$$

and the last inequality becomes

$$d_{\mathrm{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) \leq \sqrt{\delta(\mathbf{A}, \mathbf{B})^2 + \delta(\mathbf{X} \cap \mathbf{A}^\perp, \mathbf{B}_0)^2}.$$

If $n \geq 2l$, then there exist $l-k$ vectors $c_1, \dots, c_{l-k}$ orthogonal to $\mathbf{A}$ and $\mathbf{B}$ simultaneously. Choosing $\mathbf{X} = \operatorname{span}\{a_1, \dots, a_k, c_1, \dots, c_{l-k}\}$, we attain the required maximum:

$$d_{\mathrm{Gr}(l,n)}(\mathbf{X}, \mathbf{B}) = \sqrt{\delta(\mathbf{A}, \mathbf{B})^2 + (l-k)\pi^2/4} = d_{\mathrm{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}).$$

The second equality in (20) follows from $d_{\mathrm{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}) = d_{\mathrm{Gr}(\infty,\infty)}(\mathbf{B}, \mathbf{A})$, given that $d_{\mathrm{Gr}(\infty,\infty)}$ is a metric by Proposition 15. □

The existence of the metrics $d^*_{\mathrm{Gr}(\infty,\infty)}$ as defined in (18) and (19) does not contradict our earlier discussion about the general nonexistence of coproduct in **Met** as these metrics do not respect continuous *contractions*. Take the Grassmann metric on $\mathrm{Gr}(\infty, \infty)$ for instance. $(\mathrm{Gr}(\infty, \infty), d_{\mathrm{Gr}(\infty,\infty)})$ is an object of the category **Met** but it is *not* the coproduct of $\{(\mathrm{Gr}(k, \infty), d_{\mathrm{Gr}(k,\infty)}) : k \in \mathbb{N}\}$. Indeed, let $Y = \{y_1, y_2\}$ with metric defined by $d_Y(y_1, y_2) = 1$. Consider a family of maps $f_k : \mathrm{Gr}(k, \infty) \to Y$,

$$f_k(\mathbf{A}) = \begin{cases} y_1 & \text{if } k = 2, \\ y_2 & \text{otherwise.} \end{cases}$$

Then f_k is a continuous contraction between $\mathrm{Gr}(k, \infty)$ and Y. So $\{f_k : k \in \mathbb{N}\}$ is a family of morphisms in **Met** compatible with $\{(\mathrm{Gr}(k, \infty), d_{\mathrm{Gr}(k,\infty)}) : k \in \mathbb{N}\}$. If $(\mathrm{Gr}(\infty, \infty), d_{\mathrm{Gr}(\infty,\infty)})$ is the coproduct of this family, then there must be a continuous contraction $f : \mathrm{Gr}(\infty, \infty) \to Y$ such that $f \circ \iota_k = f_k$ with ι_k being the natural inclusion of $\mathrm{Gr}(k, \infty)$ into $\mathrm{Gr}(\infty, \infty)$. But taking $\mathbf{A} \in \mathrm{Gr}(2, \infty)$ and $\mathbf{B} \in \mathrm{Gr}(3, \infty)$, we see that

$$d_{\mathrm{Gr}(\infty,\infty)}(\mathbf{A}, \mathbf{B}) \geq \frac{\pi}{2} > 1 = d_Y\big(f(\mathbf{A}), f(\mathbf{B})\big),$$

contradicting the surmise that f is a contraction. Similarly, one may show that $(\mathrm{Gr}(\infty, \infty), d^*_{\mathrm{Gr}(\infty,\infty)})$ is not a coproduct in **Met** for any $* = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi$.

$(\mathrm{Gr}(\infty, \infty), d_{\mathrm{Gr}(\infty,\infty)})$ is also not the coproduct of $\{(\mathrm{Gr}(k, \infty), d_{\mathrm{Gr}(k,\infty)}) : k \in \mathbb{N}\}$ in the category of metric spaces with continuous (or uniformly continuous) maps as morphisms. The coproduct in this category is simply $\mathrm{Gr}(\infty, \infty)$ with the metric induced by the disjoint union topology, which is too fine (in the sense of topology) to be interesting. In particular, such a metric is unrelated to the distance δ.

7. Comparison with existing works

There are two existing proposals for a distance between subspaces of different dimensions — the containment gap and the symmetric directional distance. These turn out to be special cases of our distance in Section 4 and our metric in Section 6.

Let $\mathbf{A} \in \operatorname{Gr}(k,n)$ and $\mathbf{B} \in \operatorname{Gr}(l,n)$. The *containment gap* is defined as

$$\gamma(\mathbf{A},\mathbf{B}) \coloneqq \max_{a\in\mathbf{A}} \min_{b\in\mathbf{B}} \frac{\|a-b\|}{\|a\|}.$$

This was proposed in [28, pp. 197–199] and used in numerical linear algebra [42] for measuring separation between Krylov subspaces [8]. It is equivalent to our *projection distance* δ^π in Theorem 12. It was observed in [8, p. 495] that

$$\gamma(\mathbf{A},\mathbf{B}) = \sin\big(\theta_k(\mathbf{A},\mathbf{Y})\big)$$

where $\mathbf{Y} \in \Omega_-(\mathbf{B})$ is nearest to $\mathbf{A}$ in the projection distance $d^\pi_{\operatorname{Gr}(k,n)}$. By Theorem 12, we deduce that it can also be realized as

$$\gamma(\mathbf{A},\mathbf{B}) = \sin\big(\theta_l(\mathbf{B},\mathbf{X})\big)$$

where $\mathbf{X} \in \Omega_+(\mathbf{A})$ is nearest to $\mathbf{B}$ in the projection distance $d^\pi_{\operatorname{Gr}(l,n)}$, a fact about the containment gap that had not been observed before. Indeed, by Theorem 12, we get

$$\gamma(\mathbf{A},\mathbf{B}) = \delta^\pi(\mathbf{A},\mathbf{B})$$

for all $\mathbf{A} \in \operatorname{Gr}(k,n)$ and $\mathbf{B} \in \operatorname{Gr}(l,n)$.

The *symmetric directional distance* is defined as

$$d_\Delta(\mathbf{A},\mathbf{B}) \coloneqq \Big(\max(k,l) - \sum\nolimits_{i,j=1}^{k,l} (a_i^\mathsf{T} b_j)^2\Big)^{1/2} \tag{21}$$

where $A = [a_1,\dots,a_k]$ and $B = [b_1,\dots,b_l]$ are the respective orthonormal bases. This was proposed in [44, 46], and is widely used [5, 7, 13, 16, 18, 24, 31, 35, 39, 40, 47, 49, 51]. The definition (21) is equivalent to our *chordal metric* $d^\kappa_{\operatorname{Gr}(\infty,\infty)}$ in Table 3,

$$d^\kappa_{\operatorname{Gr}(\infty,\infty)}(\mathbf{A},\mathbf{B})^2 = |k-l| + \sum_{i=1}^{\min(k,l)} \sin^2\theta_i = \max(k,l) - \sum_{i,j=1}^{k,l} (a_i^\mathsf{T} b_j)^2 = d_\Delta(\mathbf{A},\mathbf{B})^2,$$

since $|k-l| = \max(k,l) - \min(k,l)$, and

$$\sum\nolimits_{i,j=1}^{k,l} (a_i^\mathsf{T} b_j)^2 = \|A^\mathsf{T} B\|_F^2 = \sum\nolimits_{i=1}^{\min(k,l)} \cos^2\theta_i = \min(k,l) - \sum\nolimits_{i=1}^{\min(k,l)} \sin^2\theta_i.$$

8. Geometry of $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$

Up to this point, $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$, as defined in Definition 6, are treated as mere subsets of $\operatorname{Gr}(l,n)$ and $\operatorname{Gr}(k,n)$ respectively. We will see that $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ have rich geometric properties. Firstly, we will show that they are Schubert varieties, justifying their names.

Definition 17. Let $\mathbf{X}_1 \subset \mathbf{X}_2 \subset \cdots \subset \mathbf{X}_k$ be a fixed flag in $\mathbb{R}^n$. The *Schubert variety* $\Omega(\mathbf{X}_1,\dots,\mathbf{X}_k,n)$ is the set of k-planes $\mathbf{Y}$ satisfying the Schubert conditions $\dim(\mathbf{Y}\cap\mathbf{X}_i) \geq i$, $i = 1,\dots,k$, i.e.,

$$\Omega(\mathbf{X}_1,\dots,\mathbf{X}_k,n) = \{\mathbf{Y} \in \operatorname{Gr}(k,n) : \dim(\mathbf{Y}\cap\mathbf{X}_i) \geq i,\ i = 1,\dots,k\}.$$

Definition 18. Let $0 =: k_0 < k_1 < \cdots < k_{m+1} \coloneqq n$ be a sequence of increasing nonnegative integers. The associated *flag variety* is the set of flags satisfying the condition $\dim \mathbf{X}_i = k_i$, $i = 0,1,\dots,m+1$. We denote it by $\operatorname{Flag}(k_1,\dots,k_m,n)$, i.e.,

$$\{(\mathbf{X}_1,\dots,\mathbf{X}_m) \in \operatorname{Gr}(k_1,n)\times\cdots\times\operatorname{Gr}(k_m,n) : \mathbf{X}_i \subset \mathbf{X}_{i+1},\ i = 1,\dots,m\}.$$

Observe that a Schubert variety depends on a specific increasing sequence of subspaces whereas a flag variety depends only on an increasing sequence of dimensions (of subspaces). Flag varieties may be viewed as a generalization of Grassmannians since if $m = 1$, then $\operatorname{Flag}(k, n) = \operatorname{Gr}(k, n)$. Like Grassmannians, $\operatorname{Flag}(k_1, \dots, k_m, n)$ is a smooth manifold and sometimes called a *flag manifold*. The parallel goes further, $\operatorname{Flag}(k_1, \dots, k_m, n)$ is a homogeneous space,

$$\operatorname{Flag}(k_1, \dots, k_m, n) \cong \mathcal{O}(n)/\big(\mathcal{O}(d_1) \times \cdots \times \mathcal{O}(d_{m+1})\big) \tag{22}$$

where $d_i = k_i - k_{i-1}$ for $i = 1, \dots, m+1$, generalizing (2).

Let $\mathbf{A} \in \operatorname{Gr}(k, n)$ and $\mathbf{B} \in \operatorname{Gr}(l, n)$ with $k \leq l$. Then

$$\Omega_+(\mathbf{A}) = \Omega(\mathbf{A}_1, \dots, \mathbf{A}_l, n), \qquad \Omega_-(\mathbf{B}) = \Omega(\mathbf{B}_1, \dots, \mathbf{B}_k, n),$$

are Schubert varieties in $\operatorname{Gr}(l, n)$ and $\operatorname{Gr}(k, n)$ respectively with the flags

$$\{0\} =: \mathbf{A}_0 \subset \mathbf{A}_1 \subset \cdots \subset \mathbf{A}_k \coloneqq \mathbf{A} \subset \mathbf{A}_{k+1} \cdots \subset \mathbf{A}_l,$$
$$\{0\} =: \mathbf{B}_0 \subset \mathbf{B}_1 \subset \cdots \subset \mathbf{B}_k \coloneqq \mathbf{B}.$$

where $\mathbf{A}_{k+i}$ is a subspace of $\mathbb{R}^n$ containing $\mathbf{A}$ of dimension $n - l + (k + i)$ for $1 \leq i \leq l - k$.

The isomorphism $\operatorname{Gr}(l, n) \cong \operatorname{Gr}(n-l, n)$ (resp. $\operatorname{Gr}(k, n) \cong \operatorname{Gr}(n-k, n)$) that sends $\mathbf{X}$ to $\mathbf{X}^\perp$ takes $\Omega_+(\mathbf{A})$ to $\Omega_-(\mathbf{A}^\perp)$ (resp. $\Omega_-(\mathbf{B})$ to $\Omega_+(\mathbf{B}^\perp)$). Thus $\Omega_+(\mathbf{A})$ (resp. $\Omega_-(\mathbf{B})$) may also be viewed as Schubert varieties in $\operatorname{Gr}(n - l, n)$ (resp. $\operatorname{Gr}(n - k, n)$). More importantly, this observation implies that $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$, despite superficial difference in their definitions, are essentially the same type of objects.

Proposition 19. *For any $\mathbf{A} \in \operatorname{Gr}(k, n)$ and $\mathbf{B} \in \operatorname{Gr}(l, n)$, we have*

$$\Omega_+(\mathbf{A}) \cong \Omega_-(\mathbf{A}^\perp) \quad \text{and} \quad \Omega_-(\mathbf{B}) \cong \Omega_+(\mathbf{B}^\perp).$$

Also, $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are uniquely determined by $\mathbf{A}$ and $\mathbf{B}$ respectively.

Proposition 20. *Let $\mathbf{A}, \mathbf{A}' \in \operatorname{Gr}(k, n)$ and $\mathbf{B}, \mathbf{B}' \in \operatorname{Gr}(l, n)$. Then*

$$\Omega_+(\mathbf{A}) = \Omega_+(\mathbf{A}') \quad \text{if and only if} \quad \mathbf{A} = \mathbf{A}',$$
$$\Omega_-(\mathbf{B}) = \Omega_-(\mathbf{B}') \quad \text{if and only if} \quad \mathbf{B} = \mathbf{B}'.$$

Proof. Suppose $\Omega_+(\mathbf{A}) = \Omega_+(\mathbf{A}')$. Observe that the intersection of all l-planes containing $\mathbf{A}$ is exactly $\mathbf{A}$ and ditto for $\mathbf{A}'$. So

$$\mathbf{A} = \bigcap\nolimits_{\mathbf{X} \in \Omega_+(\mathbf{A})} \mathbf{X} = \bigcap\nolimits_{\mathbf{X} \in \Omega_+(\mathbf{A}')} \mathbf{X} = \mathbf{A}'.$$

The converse is obvious. The statement for Ω_- then follows from Proposition 19. □

This observation allows us to treat subspaces of different dimensions on the same footing by regarding them as *subsets* in the same Grassmannian. If we have a collection of subspaces of dimensions $k \leq k_1 < k_2 < \cdots < k_m \leq l$, the injective map $\mathbf{A} \mapsto \Omega_+(\mathbf{A})$ takes all of them into distinct subsets of $\operatorname{Gr}(l, n)$. Alternatively, the injective map $\mathbf{B} \mapsto \Omega_-(\mathbf{B})$ takes all of them into distinct subsets of $\operatorname{Gr}(k, n)$.

The resemblance between $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ in Proposition 19 goes further — we may view them as 'sub-Grassmannians'.

Proposition 21. *Let $k \leq l \leq n$ and $\mathbf{A} \in \operatorname{Gr}(k, n)$, $\mathbf{B} \in \operatorname{Gr}(l, n)$. Then*

$$\Omega_+(\mathbf{A}) \cong \operatorname{Gr}(l - k, n - k), \quad \Omega_-(\mathbf{B}) \cong \operatorname{Gr}(k, l),$$

isomorphic as algebraic varieties and diffeomorphic as smooth manifolds. Thus

$$\dim \Omega_+(\mathbf{A}) = (n - l)(l - k), \quad \dim \Omega_-(\mathbf{B}) = k(l - k).$$

Proof. The first isomorphism is the quotient map $\varphi : \Omega_+(\mathbf{A}) \to \mathrm{Gr}_{l-k}(\mathbb{R}^n/\mathbf{A})$, $\mathbf{X} \mapsto \mathbf{X}/\mathbf{A} \subseteq \mathbb{R}^n/\mathbf{A}$, composed with the isomorphism $\mathrm{Gr}_{l-k}(\mathbb{R}^n/\mathbf{A}) \cong \mathrm{Gr}(l-k, n-k)$. The second isomorphism is obtained by regarding a k-dimensional subspace $\mathbf{Y}$ of $\mathbb{R}^n$ in $\Omega_-(\mathbf{B})$ as a k-dimensional subspace of $\mathbf{B}$, i.e., $\Omega_-(\mathbf{B}) = \mathrm{Gr}_k(\mathbf{B}) \cong \mathrm{Gr}(k, l)$. □

That $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are Grassmannians allows us to infer the following:

(i) as topological spaces, they are compact and path-connected;
(ii) as algebraic varieties, they are irreducible and nonsingular;
(iii) as differential manifolds, they are smooth and any two points on them can be connected by a length-minimizing geodesic.

The topology in (i) refers to the metric space topology, not Zariski topology. A consequence of compactness is that the distance $d_{\mathrm{Gr}(k,n)}\big(\mathbf{A}, \Omega_-(\mathbf{B})\big) = d_{\mathrm{Gr}(l,n)}\big(\mathbf{B}, \Omega_+(\mathbf{A})\big)$ can be attained by points in $\Omega_-(\mathbf{B})$ and $\Omega_+(\mathbf{A})$ respectively. We constructed these closest points explicitly when we proved Theorem 7.

Many more topological and geometric properties of $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ follow from Proposition 21 as they inherit everything that we know about Grassmannians (e.g. coordinate ring, cohomology ring, Plücker relations, etc.); in particular, $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are also flag varieties.

The last property in (iii) requires a proof. The length-minimizing geodesic is not unique and so $\Omega_+(\mathbf{A})$ and $\Omega_-(\mathbf{B})$ are not *geodesically convex* [38, Definition 4.1.35].

Proposition 22. *Any two points in $\Omega_-(\mathbf{B})$ (resp. $\Omega_+(\mathbf{A})$) can be connected by a length-minimizing geodesic in $\mathrm{Gr}(k, n)$ (resp. $\mathrm{Gr}(l, n)$).*

Proof. By Proposition 19, it suffices to show that any two points in $\Omega_-(\mathbf{B})$ can be connected by a geodesic curve in $\Omega_-(\mathbf{B})$. By Proposition 21, $\Omega_-(\mathbf{B})$ is the image of $\mathrm{Gr}(k, l)$ embedded isometrically in $\mathrm{Gr}(k, n)$. So by Lemma 3, for any $\mathbf{X}_1, \mathbf{X}_2 \in \mathrm{Gr}(k, l)$, $d_{\mathrm{Gr}(k,n)}(\mathbf{X}_1, \mathbf{X}_2) = d_{\mathrm{Gr}(k,l)}(\mathbf{X}_1, \mathbf{X}_2) = d_{\Omega_-(\mathbf{B})}(\mathbf{X}_1, \mathbf{X}_2)$, where the last is the geodesic distance in $\Omega_-(\mathbf{B})$. Hence if $d_{\Omega_-(\mathbf{B})}(\mathbf{X}_1, \mathbf{X}_2)$ is realized by a geodesic curve γ in $\Omega_-(\mathbf{B})$, then γ must also be a geodesic curve in $\mathrm{Gr}(k, n)$. □

We have represented $\mathrm{Gr}(k, n)$ as a set of *equivalence classes* of matrices but it may also be represented as a set of *actual matrices* [38, Example 1.2.20], namely, idempotent symmetric matrices of trace k:

$$\mathrm{Gr}(k, n) \cong \{P \in \mathbb{R}^{n\times n} : P^{\mathsf{T}} = P^2 = P,\ \mathrm{tr}(P) = k\}.$$

The isomorphism maps each subspace $\mathbf{A} \in \mathrm{Gr}(k, n)$ to $P_{\mathbf{A}} \in \mathbb{R}^{n\times n}$, the unique orthogonal projection onto $\mathbf{A}$, and its inverse takes an orthogonal projection P to the subspace $\mathrm{im}(P) \in \mathrm{Gr}(k, n)$. P is an orthogonal projection iff it is symmetric and idempotent, i.e., $P^{\mathsf{T}} = P^2 = P$. The eigenvalues of an orthogonal projection onto a subspace of dimension k are 1's and 0's with multiplicities k and $n-k$, so $\mathrm{tr}(P) = k$ is equivalent to $\mathrm{rank}(P) = k$, ensuring $\mathrm{im}(P)$ has dimension k. In this representation,

$$\Omega_+(\mathbf{A}) \cong \{P \in \mathbb{R}^{n\times n} : P^{\mathsf{T}} = P^2 = P,\ \mathrm{tr}(P) = l,\ \mathrm{im}(A) \subseteq \mathrm{im}(P)\},$$
$$\Omega_-(\mathbf{B}) \cong \{P \in \mathbb{R}^{n\times n} : P^{\mathsf{T}} = P^2 = P,\ \mathrm{tr}(P) = k,\ \mathrm{im}(P) \subseteq \mathrm{im}(B)\},$$

allowing us to treat $\mathrm{Gr}(k, n)$, $\mathrm{Gr}(l, n)$, $\Omega_+(\mathbf{A})$, $\Omega_-(\mathbf{B})$ all as subvarieties of $\mathbb{R}^{n\times n}$.

9. Probability density on the Grassmannian

We determine the relative volumes of $\Omega_+(\mathbf{A})$, $\Omega_-(\mathbf{B})$ and prove a volumetric analogue of (13) in Theorem 7:

> *Given k-dimensional subspace $\mathbf{A}$ and l-dimensional subspace $\mathbf{B}$ in $\mathbb{R}^n$, the probability that a randomly chosen l-dimensional subspace in $\mathbb{R}^n$ contains $\mathbf{A}$ equals the probability that a randomly chosen k-dimensional subspace in $\mathbb{R}^n$ is contained in $\mathbf{B}$.*

Every Riemannian metric on a Riemannian manifold yields a *volume density* that is consistent with the metric [38, Example 3.4.2]. The Riemannian metric[3] on $\operatorname{Gr}(k,n)$ that gives us the Grassmann distance in (6) and the geodesic in (7) also gives a density $d\gamma_{k,n}$ on $\operatorname{Gr}(k,n)$. The volume of $\operatorname{Gr}(k,n)$ is then

$$\operatorname{Vol}\big(\operatorname{Gr}(k,n)\big) = \int_{\operatorname{Gr}(k,n)} |d\gamma_{k,n}| = \binom{n}{k} \frac{\prod_{j=1}^{n} \omega_j}{\big(\prod_{j=1}^{k} \omega_j\big)\big(\prod_{j=1}^{n-k} \omega_j\big)}, \tag{23}$$

where $\omega_m \coloneqq \pi^{m/2}/\Gamma(1+m/2)$, volume of the unit ball in $\mathbb{R}^m$ [38, Proposition 9.1.12].

The normalized density $d\mu_{k,n} \coloneqq \operatorname{Vol}\big(\operatorname{Gr}(k,n)\big)^{-1} |d\gamma_{k,n}|$ defines a natural *uniform probability density* on $\operatorname{Gr}(k,n)$. With respect to this, the probability of landing on $\Omega_+(\mathbf{A})$ in $\operatorname{Gr}(l,n)$ equals the probability of landing on $\Omega_-(\mathbf{B})$ in $\operatorname{Gr}(k,n)$.

Corollary 23. *Let $k \leq l \leq n$ and $\mathbf{A} \in \operatorname{Gr}(k,n)$, $\mathbf{B} \in \operatorname{Gr}(l,n)$. The relative volumes of $\Omega_+(\mathbf{A})$ in $\operatorname{Gr}(l,n)$ and $\Omega_-(\mathbf{B})$ in $\operatorname{Gr}(k,n)$ are equal and their common value depends only on k, l, n,*

$$\mu_{l,n}\big(\Omega_+(\mathbf{A})\big) = \mu_{k,n}\big(\Omega_-(\mathbf{B})\big) = \frac{l!(n-k)! \prod_{j=l-k+1}^{l} \omega_j}{n!(l-k)! \prod_{j=n-k+1}^{n} \omega_j}.$$

Proof. By Proposition 21, $\Omega_+(\mathbf{A})$ is isometric to $\operatorname{Gr}(n-l, n-k)$ and $\Omega_-(\mathbf{B})$ is isometric to $\operatorname{Gr}(k,l)$, so by (23) their volumes are

$$\binom{n-k}{n-l} \frac{\prod_{j=1}^{n-k} \omega_j}{\big(\prod_{j=1}^{n-l} \omega_j\big)\big(\prod_{j=1}^{l-k} \omega_j\big)}, \qquad \binom{l}{k} \frac{\prod_{j=1}^{l} \omega_j}{\big(\prod_{j=1}^{k} \omega_j\big)\big(\prod_{j=1}^{l-k} \omega_j\big)}$$

respectively. Now divide by the volumes of $\operatorname{Gr}(l,n)$ and $\operatorname{Gr}(k,n)$ respectively. □

By definition, relative volume depends on the volume of ambient space and the dependence on n is expected, a slight departure from Theorem 7(i).

10. Conclusions

We provided what we hope is a thorough study of subspace distances, a topic of wide-ranging interest. We investigated the topic from different angles and filled in the most glaring gap in our existing knowledge — defining distances and metrics for inequidimensional subspaces. We also developed simple geometric models for subspaces of all dimensions and enriched the existing differential geometric view of Grassmannians with algebraic geometric perspectives. We expect these to be of independent interest to applied and computational mathematicians. Most of the topics discussed in this article have been extended to affine subspaces in [33].

Acknowledgment. We are very grateful to the two anonymous referees for their invaluable suggestions, both mathematical and stylistic. We thank Sayan Mukherjee for telling us about the importance of measuring distances between inequidimensional subspaces, Frank Sottile for invaluable discussions that led to the Schubert variety approach, and Lizhen Lin, Tom Luo, Giorgio Ottaviani for helpful comments.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation," *Acta Appl. Math.*, **80** (2004), no. 2, pp. 199–220.

[3] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, "A Grassmann–Rayleigh quotient iteration for computing invariant subspaces," *SIAM Rev.*, **44** (2002), no. 1, pp. 57–73.

[3] Discussed at length in [2, 17]; we did not specify it since we have no use for it except implicitly.

[4] A. Ashikhmin and A. R. Calderbank, "Grassmannian packings from operator Reed–Muller codes," *IEEE Trans. Inform. Theory*, **56** (2003), no. 11, pp. 5689–5714.
[5] H. Bagherinia and R. Manduchi, "A theory of color barcodes," *Proc. IEEE Int. Conf. Comput. Vis.* (ICCV), **14** (2011), pp. 806–813.
[6] A. Barg and D. Yu. Nogin, "Bounds on Packings of Spheres in the Grassmannian Manifold," *IEEE Trans. Inform. Theory*, **48** (2002), no. 9, pp. 2450–2454.
[7] R. Basri, T. Hassner, and L. Zelnik-Manor, "Approximate nearest subspace search," *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (2011), no. 2, pp. 266–278.
[8] C. A. Beattie, M. Embree, and D. C. Sorensen, "Convergence of polynomial restart Krylov methods for eigenvalue computations," *SIAM Rev.*, **47** (2005), no. 3, pp. 492–515.
[9] Å. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Math. Comp.*, **27** (1973), no. 123, pp. 579–594.
[10] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," *Annual Allerton Conf. Commun. Control Comput.*, **48** (2010), pp. 704–711.
[11] S.-C. T. Choi, "Minimal residual methods for complex symmetric, skew symmetric, and skew Hermitian systems," *preprint*, Report ANL/MCS-P3028-0812, Computation Institute, University of Chicago, IL, 2013.
[12] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: packings in Grassmannian spaces," *Experiment. Math.*, **5** (1996), no. 2, pp. 83–159.
[13] N. P. Da Silva and J. P. Costeira, "The normalized subspace inclusion: Robust clustering of motion subspaces," *Proc. IEEE Int. Conf. Comput. Vis.* (ICCV), **12** (2009), pp. 1444–1450.
[14] M. M. Deza and E. Deza, *Encyclopedia of Distances*, 2nd Ed., Springer, Heidelberg, 2013.
[15] I. S. Dhillon, R. W. Heath, Jr., T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Experiment. Math.*, **17** (2008), no. 1, pp. 9–35.
[16] B. Draper, M. Kirby, J. Marks, T. Marrinan, and C. Peterson, "A flag representation for finite collections of subspaces of mixed dimensions," *Linear Algebra Appl.*, **451** (2014), pp. 15–32.
[17] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, **20** (1999), no. 2, pp. 303–353.
[18] N. Figueiredo, P. Georgieva, E. W. Lang, I. M. Santos, A. R. Teixeira, and A. M. Tomé, "SSA of biomedical signals: A linear invariant systems approach," *Stat. Interface*, **3** (2010), no. 3, pp. 345–355.
[19] R. Fioresi and C. Hacon, "On infinite-dimensional Grassmannians and their quantum deformations," *Rend. Sem. Mat. Univ. Padova*, **111** (2004), pp. 1–24.
[20] G. Golub and C. Van Loan, *Matrix Computations*, 4th Ed., John Hopkins University Press, Baltimore, MD, 2013.
[21] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: A unifying view on subspace-based learning," *Proc. Internat. Conf. Mach. Learn.* (ICML), **25** (2008), pp. 376–383.
[22] T. F. Hansen and D. Houle, "Measuring and comparing evolvability and constraint in multivariate characters," *J. Evolution. Biol.*, **21** (2008), no. 5, pp. 1201–1219.
[23] G. Haro, G. Randall, and G. Sapiro, "Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds," *Proc. Adv. Neural Inform. Process. Syst.* (NIPS), **26** (2006), pp. 553–560.
[24] Q. He, F. Kong, and R. Yan, "Subspace-based gearbox condition monitoring by kernel principal component analysis, *Mech. Systems Signal Process.*, **21** (2007), no. 4, pp. 1755–1772.
[25] A. Ya. Helemskii, *Lectures and Exercises on Functional Analysis*, Translations of Mathematical Monographs, **233**, AMS, Providence, RI, 2006.
[26] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[27] D. Husemoller, *Fibre Bundles*, 3rd Ed., Graduate Texts in Mathematics, **20**, Springer, New York, NY, 1994.
[28] T. Kato, *Perturbation Theory for Linear Operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995.
[29] A. V. Knyazev and M. E. Argentati, "Majorization for changes in angles between subspaces, Ritz values, and graph Laplacian spectra," *SIAM J. Matrix Anal. Appl*, **29** (2006), no. 1, pp. 15–32.
[30] G. Lerman and T. Zhang, "Robust recovery of multiple subspaces by geometric l_p minimization," *Ann. Statist.*, **39** (2011), no. 5, pp. 2686–2715.
[31] H. Li and A. Li, "Utilizing improved Bayesian algorithm to identify blog comment spam," *Proc. IEEE Symp. Robot. Appl.* (ISRA), **1** (2012), pp. 423–426.
[32] J. Liesen and Z. Strakoš, *Krylov Subspace Methods*, Oxford University Press, Oxford, 2013.
[33] L.-H. Lim, K. S.-W. Wong, and K. Ye, "Statistical estimation and the grassmannian of affine subspaces," *prepirint*, (2016), `http://www.stat.uchicago.edu/~lekheng/work/affine.pdf`.
[34] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, **49** (2003), no. 10, pp. 2735–2747.
[35] D. Luo and H. Huang, "Video motion segmentation using new adaptive manifold denoising model," pp. 65–72, *Proc. IEEE Conf. Computer Vis. Pattern Recognit.* (CVPR), Columbus, OH, 2014.

[36] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, **50** (2008), no. 3, pp. 413–458.
[37] J. W. Milnor and J. D. Stasheff, *Characteristic Classes*, Annals of Mathematics Studies, **76**, Princeton University Press, Princeton, NJ, 1974.
[38] L. I. Nicolaescu, *Lectures on the Geometry of Manifolds*, 2nd Ed., World Scientific, Hackensack, NJ, 2007.
[39] M. Peng, D. Bu, and Y. Wang, "The measure of income mobility in vector space," *Physics Procedia*, **3** (2010), no. 5, pp. 1725–1732.
[40] E. Sharafuddin, N. Jiang, Y. Jin, and Z.-L. Zhang, "Know your enemy, know yourself: Block-level network behavior profiling and tracking," *Proc. IEEE Global Telecomm. Conf.* (GLOBECOM), **53** (2010), pp. 1–6.
[41] B. St. Thomas, L. Lin, L.-H. Lim, and S. Mukherjee, "Learning subspaces of different dimensions," *preprint*, (2014), `http://arxiv.org/abs/1404.6841`.
[42] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, MA, 1990.
[43] T. Strohmer and R. W. Heath, Jr., "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, **14** (2003), no. 3, pp. 257–275.
[44] X. Sun, L. Wang, and J. Feng, "Further results on the subspace distance," *Pattern Recognition*, **40** (2007), no. 1, pp. 328–329.
[45] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, **27** (2005), no. 12, pp. 1945–1959.
[46] L. Wang, X. Wang, and J. Feng, "Subspace distance analysis with application to adaptive Bayesian algorithm for face recognition," *Pattern Recognition*, **39** (2006), no. 3, pp. 456–464.
[47] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (CVPR), **26** (2008), pp. 1–8.
[48] Y.-C. Wong, "Differential geometry of Grassmann manifolds," *Proc. Nat. Acad. Sci.*, **57** (1967), no. 3, pp. 589–594.
[49] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," pp. 94–106, *Proc. European Conf. Computer Vis.* (ECCV), Graz, Austria, 2006.
[50] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, **48** (2002), no. 2, pp. 359–383.
[51] G. Zuccon, L. A. Azzopardi, C. J. van Rijsbergen, "Semantic spaces: Measuring the distance between different subspaces," pp. 225–236, P. Bruza, D. Sofge, W. Lawless, K. van Rijsbergen, M. Klusch (Eds), *Quantum Interaction*, Lecture Notes in Artificial Intelligence, **5494**, Springer-Verlag, Berlin, 2009.

Department of Mathematics, University of Chicago, Chicago, IL 60637.
E-mail address: `kye@math.uchicago.edu`

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