SCHUBERT VARIETIES AND DISTANCES BETWEEN SUBSPACES OF DIFFERENT DIMENSIONS

KE YE AND LEK-HENG LIM

Abstract. We resolve a basic problem regarding subspace distances that has arisen considerably often in applications: How could one define a notion of distance between two linear subspaces of different dimensions in a way that generalizes the usual Grassmann distance between equidimensional subspaces? We show that a natural solution is given by the distance of a point to a Schubert variety within the Grassmannian. Aside from reducing to the usual Grassmann distance when the subspaces are equidimensional, this distance is intrinsic and does not depend on any embedding into a larger ambient space. Furthermore, it can be written down as concrete expressions involving principal angles, and is efficiently computable in numerically stable ways. Our results are also largely independent of the Grassmann distance — if desired, it may be substituted by any other common distances between subspaces. Central to our approach to these problems is a concrete algebraic geometric view of the Grassmannian that parallels the differential geometric perspective that is now well-established in applied and computational mathematics. A secondary goal of this article is to demonstrate that the basic algebraic geometry of Grassmannian can be just as accessible and useful to practitioners.

1. Introduction

Modern data sets are often described or characterized by their principal subspaces. Whether one begins with biological data (e.g. gene expression levels, metabolomic profile), image data (e.g. MRI tractographs, movie clips), text data (e.g. blogs, tweets), etc, it is customary to represent measurements as a collection of feature vectors \(a_1, \ldots, a_m \in \mathbb{R}^d\) corresponding to each of the objects, thereby allowing us to represent raw data conveniently as 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). The matrix \(A\) is known as a design matrix in statistics parlance and 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.

However the raw data \(A\) is often less interesting and informative than the spaces it defines, e.g. its row and column spaces or its principal subspaces. Fortunately for us, it is also often the case that while the dimensions of the raw data are large, their intrinsic dimensions are small, i.e., \(A\) can be represented or well-approximated by a subspace \(A \in \text{Gr}(k, n)\) where \(k \ll m\) and \(n \ll d\).

The process of getting from \(A\) to \(A\) is well-studied, e.g. randomly sample a subset of representative landmarks or compute principal components.

More generally, the use of subspaces in mathematical modeling arises in diverse applications ranging from computer vision [37, 48], bioinformatics [22], machine learning [24, 33], communication [36, 52], coding theory [4, 6, 13, 16], statistical classification [27], and system identification [37]. In computational mathematics, subspaces arise in the form of Krylov subspaces [35] and their variants [12], as subspaces of structured matrices (e.g. Toeplitz, Hankel, banded), and in recent developments such as compressive sensing (e.g. Grassmannian dictionaries [46]) and matrix completion (e.g. online matrix completion [11]).

One of the most basic problems, before setting out to do anything else with subspaces, is to define a notion of separation between them. Given two subspaces of the same dimension, the solution is...
natural and well-known: Subspaces of dimension $k$ in $\mathbb{R}^n$ are points on the Grassmannian $\text{Gr}(k,n)$, and the geodesic distance between two points on $\text{Gr}(k,n)$, a Riemannian manifold, 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 two $k$-dimensional subspaces $A, B \in \text{Gr}(k,n)$, form matrices $A$ and $B \in \mathbb{R}^{n \times k}$ whose columns are their respective orthonormal bases, then their Grassmann distance is given by

$$d(A, B) = \left( \sum_{i=1}^{k} \theta_i^2 \right)^{1/2},$$

where $\theta_i = \cos^{-1} (\sigma_i (A^T B))$ is the $i$th principal angle between $A$ and $B$. This is the geodesic distance on the Grassmannian viewed as a Riemannian manifold. But using it as our choice of distance on the Grassmannian is merely for convenience. As we will see later, we could have instead picked any of the other common distances in Table 2 — Asimov, Binet–Cauchy, chordal, Fubini–Study, Martin, Procrustes, projection, spectral — and called it our Grassmann distance. All results in this article will hold true as long as we appropriately modify constants appearing in bounds and substitute (1) with the corresponding expression in Table 2.

However what if the subspaces are of different dimensions? In fact, if one examines the aforementioned applications, one invariably finds that the most general and natural settings for each of them would fall under one of these situations. The restriction to equidimensional linear subspaces thus somewhat limits the relevance and utility of these applications.

While applications that require measuring distances between subspaces of different dimensions are less common (in our opinion, largely because such distances have never been properly developed), one can still find many examples. They arise in numerical linear algebra [7], perturbation theory [45], information retrieval [53], facial recognition [50], motion segmentation [14], EEG signal analysis [19], mechanical engineering [26], economics [42], network analysis [43], blog spam detection [34], and decoding colored barcodes [5]. In the context of the first two paragraphs, 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 $B$.

1.1. Main Contributions. One of our key results, Theorem 4.2, from which a definition of distance between subspaces of different dimensions naturally arises, can be stated in simple linear algebraic terms: Given any two subspaces in $\mathbb{R}^n$, $A$ of dimension $k$ and $B$ of dimension $l$, assuming $k < l$ without loss of generality, the distance of $A$ to the nearest $k$-dimensional subspace contained in $B$ equals the distance of $B$ to the nearest $l$-dimensional subspace that contains $A$. Their common value gives the distance between $A$ and $B$.

To establish the properties of this distance and to extend it to other circumstances would require additional insights that come most easily from an algebraic geometric point-of-view:

*The distance between subspaces of different dimensions is the distance between a point and a special Schubert variety, a Zariski closed subset of the Grassmannian.*

This distance has the following properties:

(a) readily computable via SVD;
(b) restrict 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 on the Grassmannian.

Elaborating on the last point, what we meant is that in place of the usual Grassmann (i.e., geodesic) distance, our constructions allow us to start with any of the distances for equidimensional linear subspaces in Table 2 and obtain corresponding distances for subspaces of different dimensions. These properties are established in Sections 4.
The applications mentioned in the last paragraph of Section 1 are all based on two existing proposals for a distance between subspaces of different dimensions: The *containment gap* \[31, pp. 197–199\] and the *symmetric directional distance* \[47, 49\]. These are however somewhat ad hoc and bear little relation to the natural geometry of subspaces, i.e., the Grassmannian. 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. While our objective in this article is to show that there is an alternative definition that does generalize the Grassmann distance, our work will shed light on these two distances as well. This is discussed in Section 7.

Evidently, the word ‘distance’ in (i) is used in the sense of a distance of a point to a set (inside a Grassmannian). For example, if a subspace is strictly contained in another, then the distance between them is necessarily zero, even though they are distinct. In other words, the distance in (i) is not a metric\[1\]. Nonetheless we show in Section 6 that it is still possible to define a metric on the set of subspaces of all dimensions based on the distance in (i). This is achieved with an analogue of our aforementioned result: Given any two subspaces in \(\mathbb{R}^n\), \(A\) of dimension \(k\) and \(B\) of dimension \(l\) with \(k < l\), the distance of \(A\) to the furthest \(k\)-dimensional subspace contained in \(B\) equals the distance of \(B\) to the furthest \(l\)-dimensional subspace that contains \(A\). Their common value gives the metric between \(A\) and \(B\). The most interesting metrics for subspaces of different dimensions can be found in Table 3.

In Section 10, we obtain a volumetric analogue to our key results with respect to the uniform probability measure on Grassmannians. Given two arbitrary subspaces in \(\mathbb{R}^n\), \(A\) of dimension \(k\) and \(B\) of dimension \(l\) with \(k < l\), we show that the probability a random \(l\)-dimensional subspace contains \(A\) equals the probability a random \(k\)-dimensional subspace is contained in \(B\).

1.2. Secondary Goal. This article is written with an applied and computational mathematics readership in mind. We assume only one prerequisite: Knowledge of the singular value decomposition. The proof of every major result in this article essentially boils down to the \(\text{svd}\), possibly in the form of principal angles and principal vectors. Everything else is explained within the article and accessible to anyone willing to accept a small handful of unfamiliar terminologies and facts on faith.

Thanks largely to the far-reaching work in \[18\] that presented the basic *differential geometry* of Stiefel and Grassmannian manifolds concretely in terms of matrices, these objects are now standard knowledge in applied and computational mathematics. Subsequent works, notably \[1, 2, 3\], have further enriched and elucidated this concrete matrix-based approach. The line of work initiated in \[18\] has launch many new applications, too numerous to list here, including a whole subfield of optimization — *manifold optimization*. This article would have been much harder to write without this groundwork laid over the last 15 years.

A secondary objective of this article is to demonstrate that the basic *algebraic geometry* of Grassmannians can also be made very concrete and accessible, and more importantly, can often be a source of useful tools for applied and computational mathematicians. Why is the Grassmannian so useful? The reason, we think, is primarily because it provides a simple model with rich geometry for the set of all \(k\)-dimensional linear subspaces in an ambient space \(\mathbb{R}^n\). In the course of developing (i) and (ii), we will have the opportunity to introduce other similar models, listed in Table 1, that we think are as useful as the Grassmannian but have yet to find their ways into applied and computational mathematics literature.

2. Grassmannian of linear subspaces

In this section we will selectively review some basic properties of the Grassmannian \(\text{Gr}(k, n)\) that will be useful to us later. The differential geometric perspectives of our discussions below are drawn
The action is transitive since any $k$-plane in a subspace $A$ is identified with an equivalence class comprising all its orthonormal $k$-frames, i.e.,

$$\text{Gr}(k,n) \cong V(k,n)/O(k) \cong O(n)/(O(n-k) \times O(k)).$$

(3)

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

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

$$Q \cdot A := \text{span}(QA) \in \text{Gr}(k,n).$$

(4)

This action is transitive since any $k$-plane can be rotated onto any other $k$-plane by some $Q \in O(n)$.

More generally, for an abstract vector space $V$, we write $\text{Gr}_k(V)$ and $V_k(V)$ for the sets of $k$-planes and $k$-frames in $V$ respectively. In this notation, $\text{Gr}(k,n) = \text{Gr}_k(\mathbb{R}^n)$ and $V(k,n) = V_k(\mathbb{R}^n)$. A $k$-plane $A \in \text{Gr}(k,n)$ will be denoted in boldfaced and the corresponding italicized letter $A = [a_1, \ldots, a_k] \in V(k,n)$ will denote an orthonormal $k$-frame of $A$.

$\text{Gr}(k,n)$ and $V(k,n)$ are smooth manifolds of dimensions $k(n-k)$ and $nk-k(k+1)/2$ respectively. Since we regard $V(k,n)$ as $n \times k$ matrices, it 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 $A = [a_1, \ldots, a_k]$ and $B = [b_1, \ldots, b_k]$ in $T_X V(k,n)$, the tangent space at $X \in V(k,n)$, the Riemannian metric $g$ is defined by $g_X(A,B) = \sum_{i=1}^k a_i^T b_i = \text{tr}(A^T B)$, the Frobenius inner product of $n \times k$ matrices. As the Riemannian metric $g$ is invariant under the action of $O(k)$, it descends to a Riemannian metric on $\text{Gr}(k,n)$ and in turn induces a geodesic distance on $\text{Gr}(k,n)$ which we define below.
Let $a_1, \ldots, a_k$ and $b_1, \ldots, b_l$ be bases (not necessarily orthonormal) for $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$ respectively. Let $r := \min(k, l)$. We define the $i$th principal vector $(p_i, q_i)$, $i = 1, \ldots, r$, recursively as solutions to the optimization problem

$$
\begin{align*}
\text{maximize} & \quad p^T q \\
\text{subject to} & \quad p \in A, \ q \in B, \\
& \quad p^T a_1 = \cdots = p^T a_{i-1} = 0, \ |p| = 1, \\
& \quad q^T b_1 = \cdots = q^T b_{l-1} = 0, \ |q| = 1,
\end{align*}
$$

for $i = 1, \ldots, r$ (for $i = 1$, the orthogonality conditions are vacuous). The principal angles are then defined by

$$
\cos \theta_i = p_i^T q_i, \quad i = 1, \ldots, r.
$$

Clearly $0 \leq \theta_1 \leq \cdots \leq \theta_k \leq \pi/2$. We will let $\theta_i(A, B)$ denote the $i$th principal angle between $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$.

Principal vectors and principal angles may be readily computed using QR and SVD [8, 21]. Let $A = [a_1, \ldots, a_k]$ and $B = [b_1, \ldots, b_l]$ be orthonormal bases and let

$$
A^T B = U \Sigma V^T
$$

be the full SVD of $A^T B$, i.e., $U \in O(k)$, $V \in O(l)$, $\Sigma = [\Sigma_1 0] \in \mathbb{R}^{k \times l}$ with $\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r \times r}$ where $\sigma_1 \geq \cdots \geq \sigma_r$ are the nonzero singular values.

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

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

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

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

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

We will be using the following fact from [21] Theorem 6.4.2.

**Proposition 2.1.** Let $r = \min(k, l)$ and $\theta_1, \ldots, \theta_r$ and $(p_1, q_1), \ldots, (p_r, q_r)$ be the principal angles and principal vectors between $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$ respectively. If $m < r$ is such that

$$
1 = \cos \theta_1 = \cdots = \cos \theta_m > \cos \theta_{m+1},
$$

then

$$
A \cap B = \text{span}\{p_1, \ldots, p_m\} = \text{span}\{q_1, \ldots, q_m\}.
$$

If $k = l$, one may show that [51] the geodesic distance between $A$ and $B$ on $\text{Gr}(k, n)$ is given by

$$
d_{\text{Gr}(k, n)}(A, B) = \left( \sum_{i=1}^k \theta_i^2 \right)^{1/2} = \|\cos^{-1} \Sigma\|_F. \tag{8}
$$

We will call this the Grassmann distance between subspaces.

Suppose $k = l$. To obtain an explicit expression for the geodesic [2] that connects $A$ to $B$ on $\text{Gr}(k, n)$ that minimizes the Grassmann distance, we consider the matrix

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

It is straightforward to verify that the condensed SVD of $M$ takes the form

$$
M = Q(\tan \Theta)U^T,
$$

where $U \in O(k)$ and $\Theta = \text{diag}(\theta_1, \ldots, \theta_k) \in \mathbb{R}^{k \times k}$ are as in [5] and [6]. The matrix $Q \in V(k, n)$ has no general simple expression in terms of earlier defined quantities. Note that if $\cos \Theta = \Sigma$, then
\[ \tan \Theta = (\Sigma^{-2} - I)^{1/2}. \]

With this, the shortest geodesic path from \( A \) to \( B \) on \( \text{Gr}(k, n) \) is given by
\[
\gamma : [0, 1] \rightarrow \text{Gr}(k, n),
\]
\[
\gamma(t) = \text{span}(AU \cos t\Theta + Q \sin t\Theta).
\]

One may show that \( \gamma(0) = A \) and \( \gamma(1) = B \).

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

<table>
<thead>
<tr>
<th>Principal angles</th>
<th>Orthonormal bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asimov ( d_{\text{Gr}(k,n)}^\phi(A,B) = \theta_k )</td>
<td>( \cos^{-1}|A^T B|_2 )</td>
</tr>
<tr>
<td>Binet–Cauchy ( d_{\text{Gr}(k,n)}^\beta(A,B) = \left( 1 - \prod_{i=1}^k \cos^2 \theta_i \right)^{1/2} )</td>
<td>( 1 - (\det A^T B)^2 )(^{1/2} )</td>
</tr>
<tr>
<td>Chordal ( d_{\text{Gr}(k,n)}^\mu(A,B) = \left( \sum_{i=1}^k \sin^2 \theta_i \right)^{1/2} )</td>
<td>( \frac{1}{\sqrt{2}} |AA^T - BB^T|_F )</td>
</tr>
<tr>
<td>Fubini–Study ( d_{\text{Gr}(k,n)}^\varphi(A,B) = \cos^{-1} \left( \prod_{i=1}^k \cos \theta_i \right)^{1/2} )</td>
<td>( \cos^{-1}</td>
</tr>
<tr>
<td>Martin ( d_{\text{Gr}(k,n)}^\delta(A,B) = \left( \log \prod_{i=1}^k 1/\cos^2 \theta_i \right)^{1/2} )</td>
<td>( -2 \log \det A^T B )(^{1/2} )</td>
</tr>
<tr>
<td>Procrustes ( d_{\text{Gr}(k,n)}^\alpha(A,B) = 2 \left( \sum_{i=1}^k \sin^2 (\theta_i/2) \right)^{1/2} )</td>
<td>( |AU - BV|_F )</td>
</tr>
<tr>
<td>Projection ( d_{\text{Gr}(k,n)}^\gamma(A,B) = \sin \theta_k )</td>
<td>( |AA^T - BB^T|_2 )</td>
</tr>
<tr>
<td>Spectral ( d_{\text{Gr}(k,n)}^\tau(A,B) = 2 \sin(\theta_k/2) )</td>
<td>( |AU - BV|_2 )</td>
</tr>
</tbody>
</table>

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 \( \text{O}(n) \), must be a function of their principal angles. To be more specific, if a distance \( d : \text{Gr}(k,n) \times \text{Gr}(k,n) \rightarrow [0, \infty) \) satisfies
\[
d(Q \cdot A, Q \cdot B) = d(A,B),
\]
for all \( A, B \in \text{Gr}(k,n) \) and all \( Q \in \text{O}(n) \), where the action is as defined in (4), then \( d \) must be a function of \( \theta_i(A,B) \), \( i = 1, \ldots, k \).

We will see later that our definition of a distance between subspaces of different dimension extends to all the above distances, i.e., for each of these distances, which of course defined between equidimensional \( A \) and \( B \), we have a corresponding version for when \( \text{dim } A \neq \text{dim } B \).

We will later discuss the independence of these distances between subspaces from the dimension of their ambient space and this discussion is most naturally formulated in terms of the infinite Grassmannian \( \text{Gr}(k, \infty) \). We will also discuss the construction of a metric on the set of subspaces of all dimensions and this discussion is most naturally formulated in terms of the doubly infinite Grassmannian \( \text{Gr}(\infty, \infty) \). These will be defined in Section 3 and Section 5 respectively.
3. The Infinite Grassmannian

A conceivable way of defining a distance between $\mathbf{A} \in \text{Gr}(k, n)$ and $\mathbf{B} \in \text{Gr}(l, n)$ where $k \neq l$ is to first isometrically embed $\text{Gr}(k, n)$ and $\text{Gr}(l, n)$ into an ambient Riemannian manifold and then define the distance between $\mathbf{A}$ and $\mathbf{B}$ to be their distance as measured in the ambient space. This is in fact the approach taken in [14], based on an isometric embedding of $\text{Gr}(0, n), \text{Gr}(1, n), \ldots, \text{Gr}(n, n)$ into a sphere of dimension $(n-1)(n+2)/2$ first proposed in [13]. 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.

Our proposed distance on the other hand depends only on the intrinsic distance in the Grassmannian and is furthermore 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 $m \geq \min(k, l)$. This depends on a simple property of the Grassmann distance stated in Corollary 3.2 that does not appear to be well-known and may be of independent interest.

Consider the inclusion map $\iota_n : \mathbb{R}^n \to \mathbb{R}^{n+1}$, $\iota_n(x_1, \ldots, x_n) = (x_1, \ldots, x_n, 0)$. It is easy to see that $\iota_n$ induces an inclusion of $\text{Gr}(k, n)$ into $\text{Gr}(k, n+1)$ which we will call natural inclusion and, with a slight abuse of notation, also denote by $\iota_n$. For any $m > n$, composition of successive inclusions gives the inclusion $\iota_{nm} : \text{Gr}(k, n) \to \text{Gr}(k, m)$ where

$$\iota_{nm} = \iota_n \circ \iota_{n+1} \circ \cdots \circ \iota_{m-1}.$$  

To be more concrete, if $\mathbf{A} \in \mathbb{R}^{n \times k}$ has orthonormal columns, then

$$\iota_{nm} : \text{Gr}(k, n) \to \text{Gr}(k, m), \quad \text{span}(\mathbf{A}) \mapsto \text{span} \begin{pmatrix} \mathbf{A} \\ 0 \end{pmatrix},$$

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

For a fixed $k$, the family of Grassmannians $\{\text{Gr}(k, n) : n \in \mathbb{N}, n \geq k\}$ together with the inclusion maps $\iota_{nm} : \text{Gr}(k, n) \to \text{Gr}(k, m)$ 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

$$\text{Gr}(k, \infty) := \lim \text{Gr}(k, n).$$

Those unfamiliar with the notion of direct limits may simply take

$$\text{Gr}(k, \infty) = \bigcup_{n=k}^{\infty} \text{Gr}(k, n),$$

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

We now show that one could define a distance function $d_{\text{Gr}(k, \infty)}$ on $\text{Gr}(k, \infty)$ that is consistent with the Grassmann distance on $\text{Gr}(k, n)$ for all $n$ sufficiently large. We also exhibit a simple way to calculate $d_{\text{Gr}(k, \infty)}$.

Lemma 3.1. The natural inclusion $\iota_n : \text{Gr}(k, n) \to \text{Gr}(k, n+1)$ is isometric, i.e.,

$$d_{\text{Gr}(k, n)}(\mathbf{A}, \mathbf{B}) = d_{\text{Gr}(k, n+1)}(\iota_n(\mathbf{A}), \iota_n(\mathbf{B})).$$  

Repeated applications of (11) yields

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

for all $m > n$ and if we identify $\text{Gr}(k, n)$ with $\iota_n(\text{Gr}(k, n))$, we may rewrite (12) as

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

for all $m > n$. 
Proof. If \( a \in \mathbb{R}^n \), we will write \( \hat{a} = [\hat{a}] \in \mathbb{R}^{n+1} \). Let \( A = [a_1, \ldots, a_k] \) and \( B = [b_1, \ldots, b_k] \) be any orthonormal bases of \( A \) and \( B \) respectively. By the definition of \( \iota_n \), \( \iota_n(A) \) is the subspace in \( \mathbb{R}^{n+1} \) spanned by an orthonormal basis that we will denote by \( \iota_n(A) := [\hat{a}_1, \ldots, \hat{a}_k, \epsilon_{n+1}] \in \mathbb{R}^{(n+1) \times (n+1)} \), with \( \epsilon_{n+1} \in \mathbb{R}^{n+1} \) a unit vector orthogonal to \( \mathbb{R}^n \). Hence we have that
\[
\iota_n(A)^T \iota_n(B) = \begin{bmatrix} A^T B & 0 \\ 0 & 1 \end{bmatrix}.
\]
By the expression for Grassmann distance in (8), one immediately sees that (11) must hold. \( \square \)

Since the inclusion of \( \text{Gr}(k, n) \) in \( \text{Gr}(k, n + 1) \) is isometric, a geodesic in \( \text{Gr}(k, n) \) remains a geodesic in \( \text{Gr}(k, n + 1) \). Given \( A, B \in \text{Gr}(k, \infty) \), there must exist some \( n \) sufficiently large so that both \( A, B \in \text{Gr}(k, n) \) and in which case we may define the distance between \( A \) and \( B \) in \( \text{Gr}(k, \infty) \) to be
\[
d_{\text{Gr}(k, \infty)}(A, B) := d_{\text{Gr}(k, n)}(A, B).
\]
By Lemma 3.1, this value is independent of our choice of \( n \) and is the same for all \( m \geq n \). In particular, \( d_{\text{Gr}(k, \infty)} \) is well-defined and yields a distance on \( \text{Gr}(k, \infty) \). We summarize these observations below.

Corollary 3.2. The Grassmann distance between two \( k \)-planes in \( \text{Gr}(k, n) \) may be regarded as the geodesic distance in \( \text{Gr}(k, \infty) \) and is therefore independent of \( n \). Also, the expression [1] for a distance minimizing geodesic connecting \( A \) and \( B \) in \( \text{Gr}(k, n) \) extends verbatim to \( \text{Gr}(k, \infty) \).

It is easy to see that Lemma 3.1 also holds for other notion of distances on \( \text{Gr}(k, n) \) described in Table 2, allowing us to define them on \( \text{Gr}(k, \infty) \).

Lemma 3.3. For all \( m > n \), the natural inclusion \( \iota_{nm} : \text{Gr}(k, n) \to \text{Gr}(k, m) \) is isometric when \( \text{Gr}(k, n) \) and \( \text{Gr}(k, m) \) are both equipped with one of the above distances, i.e.,
\[
d_{\text{Gr}(k, n)}^*(A, B) = d_{\text{Gr}(k, m)}^*(\iota_{nm}(A), \iota_{nm}(B)),
\]
where \( \iota_{nm} \) are uniquely determined by \( \text{Gr}(k, n) \) and \( \text{Gr}(k, m) \), respectively. Consequently \( d_{\text{Gr}(k, \infty)}(A, B) \) is well-defined for any \( A, B \in \text{Gr}(k, \infty) \).

Proof. \( d_{\text{Gr}(k, n)}^*(A, B) \) and \( d_{\text{Gr}(k, n+1)}^*(\iota_n(A), \iota_n(B)) \) depend only on the principal angles between \( A \) and \( B \), so the distance remains unchanged under \( \iota_n \). Repeated application then yields the required isometry. \( \square \)

4. Distances between linear subspaces of different dimensions

We now resolve 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 \( d(x, S) := \inf \{ d(x, y) : y \in S \} \). For us, \( X \) is a Grassmannian, therefore compact, and so \( d(x, S) \) will always be 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 4.1. Let \( k, l, n \in \mathbb{N} \) be such that \( k \leq l \leq n \). For any \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \), we define the subsets
\[
\Omega_+(A) := \{ X \in \text{Gr}(l, n) : A \subseteq X \}, \quad \Omega_-(B) := \{ Y \in \text{Gr}(k, n) : Y \subseteq B \}.
\]
We will call \( \Omega_+(A) \) the Schubert variety of \( l \)-planes containing \( A \) and \( \Omega_-(B) \) the Schubert variety of \( k \)-planes contained in \( B \).

As we will see in Section 8, \( \Omega_+(A) \) and \( \Omega_-(B) \) are indeed Schubert varieties and therefore closed subsets of \( \text{Gr}(l, n) \) and \( \text{Gr}(k, n) \) respectively. Furthermore, \( \Omega_+(A) \) and \( \Omega_-(B) \) are uniquely determined by \( A \) and \( B \) (cf. Proposition 8.3) and may be regraded as ‘sub-Grassmannians’ of \( \text{Gr}(l, n) \) and \( \text{Gr}(k, n) \) respectively (cf. Proposition 8.4).
How could one define the distance between a subspace \( A \) of dimension \( k \) and a subspace \( B \) 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 \( \delta(A, B) \) as that between the \( k \)-plane \( A \) and the closest \( k \)-plane \( Y \) contained in \( B \), measured within \( \text{Gr}(k, n) \). In other words, we want the Grassmann distance from \( A \) to the closed subset \( \Omega_-(B) \),

\[
\delta(A, B) := d_{\text{Gr}(k,n)}(A, \Omega_-(B)) = \min \{d_{\text{Gr}(k,n)}(A, Y) : Y \in \Omega_-(B)\}.
\] (15)

This has the advantage of being entirely intrinsic — the distance \( \delta(A, B) \) is measured in \( d_{\text{Gr}(k,n)} \) and is defined wholly within \( \text{Gr}(k, n) \) without any embedding of \( \text{Gr}(k, n) \) into an arbitrary ambient space. Furthermore, by the property of \( d_{\text{Gr}(k,n)} \) in Corollary 3.2, \( \delta(A, B) \) will 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 \( \text{Gr}(1, 3) \) though to be accurate antipodal points on the sphere should be identified.

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

\[
\delta(A, B) := d_{\text{Gr}(l,n)}(B, \Omega_+(A)) = \min \{d_{\text{Gr}(l,n)}(B, X) : X \in \Omega_+(A)\}.
\] (16)

It will have same desirable features as the one in (15) except that the distance is now measured in \( d_{\text{Gr}(l,n)} \) and within \( \text{Gr}(l, n) \). But why should we use one rather than the other?

Fortunately the two values in (15) and (16) turn out to be one and the same, allowing us to define \( \delta(A, B) \) as their common value. We will establish this equality and the properties of \( \delta(A, B) \) in the remainder of this section. The results are summarized in Theorem 4.2. Our proof is constructive: In addition to showing the equality of (15) and (16), it shows how one may explicitly find the closest points on Schubert varieties \( X \in \Omega_-(B) \) and \( Y \in \Omega_+(A) \) to any given point in the respective Grassmannians.

**Theorem 4.2.** Let \( A \) be a subspace of dimension \( k \) and \( B \) be a subspace of dimension \( l \) in \( \mathbb{R}^n \). Suppose \( k \leq l \leq n \). Then

\[
d_{\text{Gr}(k,n)}(A, \Omega_-(B)) = d_{\text{Gr}(l,n)}(B, \Omega_+(A)).
\] (17)
Their common value defines a distance $\delta(A, B)$ between the two subspaces with the following properties.

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

$$\delta(A, B) = \left(\sum_{i=1}^{\min(k, l)} \theta_i(A, B)^2\right)^{1/2} \tag{18}$$

where $\theta_i(A, B)$ is the $i$th principal angle between $A$ and $B$, $i = 1, \ldots, \min(k, l)$.

Rewriting (17) as

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

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

(a) the Grassmann distance may be replaced by any of the distances in Table 2 (cf. Theorem 4.7);
(b) ‘nearest’ may be replaced by ‘furthest’ and ‘min’ above replaced by ‘max’ when $n$ is sufficiently large (cf. Proposition 6.2);
(c) ‘distance’ may be replaced by ‘volume’ with respect to the intrinsic probability density on the Grassmannian (cf. Section 10).

We will prove Theorem 4.2 by way of the next two lemmas.

**Lemma 4.3.** Let $k \leq l \leq n$ be positive integers. Let $\delta : Gr(k, n) \times Gr(l, n) \to [0, \infty)$ be the function defined by

$$\delta(A, B) = \left(\sum_{i=1}^{k} \theta_i^2\right)^{1/2}$$

where $\theta_i := \theta_i(A, B)$, $i = 1, \ldots, k$. Then

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

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

$$d_{Gr(l,n)}(X, B) = \left[\cos^{-1} p_1^T q_1^2 + \cdots + \cos^{-1} p_k^T q_k^2 + \cdots + \cos^{-1} q_{k+1}^T q_{k+1}^2 + \cdots + \cos^{-1} q_l^T q_l^2\right]^{1/2} \tag{19}$$

$$= \left[\theta_1^2 + \cdots + \theta_k^2 + 0^2 + \cdots + 0^2\right]^{1/2} = \delta(A, B).$$

The following fact is well-known in numerical linear algebra [29 Corollary 3.1.3]. We state it here for easy reference and deduce a corollary that will be useful for Lemma 4.6.

**Proposition 4.4.** 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 respective $i$th singular value satisfy $\sigma_i(B_k) \leq \sigma_i(B)$ for $i = 1, \ldots, k$.

**Corollary 4.5.** Let $B$ and $B_k$ be as in Proposition 4.4 and $B$ and $B_k$ be subspaces of $\mathbb{R}^n$ spanned by the column vectors of $B$ and $B_k$ respectively. Then for any subspace $A$ of $\mathbb{R}^n$, the principal angles between the respective subspaces satisfy

$$\theta_i(A, B) \leq \theta_i(A, B_k)$$

for $i = 1, \ldots, \min(\dim A, \dim B_k)$. 
realizes the distance $d$ follows from Lemma 3.1. From which we have (17) and (18) in Theorem 4.2. Property (ii) is obvious from (18) and Property (i).

Hence we obtain the required equality

$$\sigma_i(A^TB) \geq \sigma_i(A^TB_k)$$

but this follows from Proposition 4.4 and the fact that $A^TB_k$ is a submatrix of $A^TB$.

Lemma 4.6. Let $A$ and $B$ be as in Lemma 4.3. Then

$$d_{Gr(k,n)}(A, \Omega_-(B)) \geq \delta(A, B).$$

Proof. Let $Y \in \Omega_-(B)$. Then $Y$ is a $k$-dimensional subspace contained in $B$ and in the notation of Corollary 4.5, we may write $Y = B_k$. By the same corollary we get

$$\theta_i(A, B) \leq \theta_i(A, Y)$$

for $i = 1, \ldots, k$. Hence

$$\delta(A, B) = \left(\sum_{i=1}^{k} \theta_i(A, B)^2\right)^{1/2} \leq \left(\sum_{i=1}^{k} \theta_i(A, Y)^2\right)^{1/2} = d_{Gr(k,n)}(A, Y). \tag{20}$$

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

Proof of Theorem 4.2. Recall that Grassmannians satisfy an isomorphism

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

that takes a $k$-plane $Y$ to the $(n - k)$-plane $Y^\perp$ of linear forms vanishing on $Y$. It is easy to see that this isomorphism is an isometry. Using this isometric isomorphism, together with Lemma 4.3 and Lemma 4.6, we can immediately deduce that

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

But on the other hand, by results in [32] we have

$$\delta(A, B) = \delta(A^\perp, B^\perp),$$

and hence

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

Similarly we can obtain

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

Hence we obtain the required equality

$$d_{Gr(l,n)}(B, \Omega_+(A)) = d_{Gr(k,n)}((A, \Omega_-(B))),$$

from which we have (17) and (18) in Theorem 4.2. Property (ii) is obvious from (18) and Property (i) follows from Lemma 4.1.

The proof of Lemma 4.3 gives a simple way to find a point $X \in \Omega_+(A)$ that realizes the distance $d_{Gr(l,n)}(B, \Omega_+(A)) = \delta(A, B)$. Similarly we may explicitly determine a point $Y \in \Omega_-(B)$ that realizes the distance $d_{Gr(k,n)}((A, \Omega_-(B)) = \delta(A, B)$.

One might wonder that whether or not Theorem 4.2 still holds if we replace $d_{Gr(k,n)}$ by other distance functions described in Table 2. The answer is yes.
**Theorem 4.7.** Let $k \leq l \leq n$. Let $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$. Then
\[
d_{\text{Gr}(k, n)}^*(A, \Omega_-(B)) = d_{\text{Gr}(l, n)}^*(B, \Omega_+(A)),
\]
for $* = \alpha, \beta, \kappa, \phi, \mu, \pi, \sigma$. Their common value $\delta^*(A, B)$ is given by:
\[
\begin{align*}
\delta^\alpha(A, B) &= \theta_k, \\
\delta^\epsilon(A, B) &= \left(\sum_{i=1}^{k} \sin^2 \theta_i\right)^{1/2}, \\
\delta^\mu(A, B) &= \left(\log \prod_{i=1}^{k} \frac{1}{\cos^2 \theta_i}\right)^{1/2}, \\
\delta^\pi(A, B) &= \sin \theta_k,
\end{align*}
\]
\[
\begin{align*}
\delta^\beta(A, B) &= \left(1 - \prod_{i=1}^{k} \cos^2 \theta_i\right)^{1/2}, \\
\delta^\phi(A, B) &= \cos^{-1}\left(\prod_{i=1}^{k} \cos \theta_i\right), \\
\delta^\rho(A, B) &= \left(2 \sum_{i=1}^{k} \sin^2(\theta_i/2)\right)^{1/2}, \\
\delta^\sigma(A, B) &= 2\sin(\theta_k/2),
\end{align*}
\]
or more generally with $\min(k, l)$ in place of the index $k$ when we do not require $k \leq l$.

**Proof.** This follows by observing that our proof of Theorem 4.2 only involves principal angles between $A$ and $B$ and the diffeomorphism between $\text{Gr}(k, n)$ and $\text{Gr}(n - k, n)$ remains an isometry under these distances. In particular, both \cite{19} and \cite{20} would still hold with any of these other distance functions in place of the Grassmannian distance. \hfill \Box

We will see in Section 7 that the projection distance $\delta^\pi$ in Theorem 4.7 turns out to be equivalent to the containment gap, a measure of distance between subspaces of different dimensions proposed originally in operator theory \cite{31}.

We end this section with a remark about the complexity of computing $\delta^*$, which falls under the general problem of computing distance of a point to a subvariety in a Grassmannian $\text{Gr}(k, n)$. For the special case of the Euclidean space $\mathbb{R}^n = \text{Gr}(0, n)$, the problem often arises in applications \cite{17, 41}. Nonetheless there are abundant examples of simple varieties where the problem is intractable: E.g., for a 3-factor Segre variety, the problem is NP-hard in the Cook–Karp–Levin sense \cite{28}; for general varieties, it is at least as hard as deciding Hilbert Nullstellensatz, which is NP-complete in the Blum–Shub–Smale sense \cite{9, 10}. Having a Grassmannian instead of a Euclidean space as the ambient space further complicates the problem since distances in Grassmannians require more effort to compute than Euclidean distance (i.e., $l^2$-norm). It is therefore somewhat surprising that all the distances in Theorems 4.2 and 4.7 can be readily computed in polynomial time to any fixed accuracy via the SVD.

5. **Grassmannian of linear subspaces of all dimensions**

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

**Lemma 5.1.** Let $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$. Then $\delta(A, B) = 0$ iff $A \subseteq B$ or $B \subseteq A$.

Note that $\delta$ also does not satisfy the triangle inequality: For a line $L$ not contained in a subspace $A$, the triangle inequality, if true, would imply
\[
\delta(L, A) = \delta(L, A) + \delta(A, B) \geq \delta(L, B),
\]
\[
\delta(L, B) = \delta(L, B) + \delta(A, B) \geq \delta(L, A),
\]
giving $\delta(L, A) = \delta(L, B)$ for any subspace $B$, which is evidently false by Lemma 5.1 (e.g. take $B = A \oplus L$).
These observations also apply verbatim to all the other similarly-defined distances $\delta^*$ in Theorem 4.7, i.e., none of them are metrics.

The set of all linear subspaces of all dimensions is parameterized by $\text{Gr}(\infty, \infty)$, the doubly infinite Grassmannian $^[20]$, which may be viewed informally as the disjoint union of all $k$-dimensional subspaces over all $k \in \mathbb{N}$,

$$\text{Gr}(\infty, \infty) = \prod_{k=1}^{\infty} \text{Gr}(k, \infty).$$

To define a metric on the set of subspaces of all dimensions is to define one on $\text{Gr}(\infty, \infty)$. It is of course trivial to define arbitrary metrics that bear little relation to the geometry of Grassmannian. What we would like is a metric that is consistent with $\delta$ and with $d_{\text{Gr}(k,n)}$ for all $k \leq n$. We will discuss this below.

We will say a few more words about the doubly infinite Grassmannian given that it is not widely known. More formally, we may define $\text{Gr}(\infty, \infty)$ as the direct limit of the direct system of Grassmannians $\{\text{Gr}(k,n) : (k,n) \in \mathbb{N} \times \mathbb{N}\}$ with inclusion maps $i^{kl}_{nm} : \text{Gr}(k,n) \to \text{Gr}(l,m)$ 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

$$i^{kl}_{nm} : \text{Gr}(k,n) \to \text{Gr}(l,m), \quad \text{span}(A) \mapsto \text{span}\left(\begin{bmatrix} A & 0 \\ 0 & 0 \\ 0 & I_{l-k} \end{bmatrix}\right),$$

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 $3 \times 2$ block matrix is in $\mathbb{R}^{m \times l}$. Note that for a fixed $k$, $i^{kk}_{nm}$ reduces to $\iota_{nm}$ in $\textbf{10}$.

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

$$B_\varepsilon(A) := \{X \in \text{Gr}(\infty, \infty) : \delta(A,X) < \varepsilon\},$$

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

**Proposition 5.2.** The topology $\tau$ on $\text{Gr}(\infty, \infty)$ is non-Hausdorff and therefore non-metrizable.

**Proof.** $\tau$ is not Hausdorff since it is not possible to separate $A \subsetneq B$ by open subsets, as we have seen. Metrizable spaces are necessarily Hausdorff.

In other words, even though $\tau$ restricts to the metric space topology on $\text{Gr}(k,\infty)$ induced by the Grassmann distance $d_{\text{Gr}(k,\infty)}$ 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 $\textbf{Met}$ of metric spaces (objects) and continuous contractions (morphisms) has no coproduct. 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. For illustration, take $(X_1,d_1)$ and $(X_2,d_2)$ to be one point metric spaces. Suppose a coproduct $(X,d)$ of $(X_1,d_1)$ and $(X_2,d_2)$ exist. Let $Y = \{y_1,y_2\}$ be a space with two points 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 $\varphi_i(x_i) = y_i$ for $x_i \in X_i$, $i = 1,2$. It is easy to verify that there is no morphism $\varphi : X \to Y$ in $\textbf{Met}$

$^[2]$As discussed in Section $\textbf{3}$ these are independent of the dimension of their ambient space and may be viewed as an element of the infinite Grassmannian $\text{Gr}(k,\infty)$. 
that will be compatible with \( \varphi_1 \) and \( \varphi_2 \). This contradicts the assumption that \( X \) is the coproduct of \( X_1 \) and \( X_2 \).

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

6. Metrics for Linear Subspaces of All Dimensions

We will describe a simple recipe for turning the distances \( \delta^* \) in Theorem 4.7 into metrics on \( Gr(\infty, \infty) \). We explain later why their existence do not contradict our comment about coproducts in Section 5.

Suppose \( k \leq l \) and we have \( A \in Gr(k, n) \) and \( B \in Gr(l, n) \). In this case there are \( k \) principal angles between \( A \) and \( B \), \( \theta_1, \ldots, \theta_k \), as defined in \([6]\). First we will set

\[
\theta_{k+1} = \cdots = \theta_l = \pi/2.
\]

Then we take the Grassmann distance \( \delta \) or any of the distances \( \delta^* \) in Theorem 4.7, replace the index \( k \) by \( l \), and call the resulting expressions \( d_{Gr(\infty, \infty)}(A, B) \) (for Grassmann distance) and \( d_{Gr(k, \infty)}^*(A, B) \) (for other distances) respectively. We want to point out that when \( n \) is sufficiently large, setting \( \theta_{k+1}, \ldots, \theta_l \) all equal to \( \pi/2 \) is equivalent to completing \( A \) to an \( l \) dimensional subspace of \( \mathbb{R}^n \), by adding \((l-k)\) vectors orthonormal to the subspace \( B \). Hence the distance between \( A \) and \( B \) is defined by the distance function on the Grassmannian \( Gr(l, n) \).

We will show in Proposition 6.1 that these expressions will indeed define metrics on \( Gr(\infty, \infty) \).

Applying the recipe in the previous paragraph to the Grassmann, chordal, and Procrustes distances yield the Grassmann, chordal, and Procrustes metrics on \( Gr(\infty, \infty) \) given in Table 3. It is evident that the metrics in Table 3 are all of the form

\[
d_{Gr(\infty, \infty)}^*(A, B) = \sqrt{\delta^*(A, B)^2 + \epsilon(A, B)^2},
\]

where \( \epsilon(A, B) := |\dim A - \dim B|^{1/2} \).

On the other hand, applying the aforementioned recipe to other distances in Table 2 yield the Asimov, Binet–Cauchy, Fubini–Study, Martin, projection, and spectral metrics on \( Gr(\infty, \infty) \) given by

\[
d_{Gr(\infty, \infty)}^*(A, B) = \begin{cases} 
d_{Gr(k, \infty)}^*(A, B) & \text{if } \dim A = \dim B = k, \\
c_* & \text{if } \dim A \neq \dim B,
\end{cases}
\]

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

\[
c_c = \pi/2, \quad c_\phi = \sqrt{2}, \quad c_\mu = \infty, \quad c_\beta = c_\phi = c_\pi = c_\sigma = c_\rho = 1.
\]

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

\[
d_{Gr(\infty, \infty)}^*(A, B) = d_{Gr(k, \infty)}^*(A, B),
\]

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

<table>
<thead>
<tr>
<th>Metric Type</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grassmann</td>
<td>( d_{Gr(\infty, \infty)}(A, B) = \left(</td>
</tr>
<tr>
<td>Chordal</td>
<td>( d_{Gr(\infty, \infty)}^*(A, B) = \left(</td>
</tr>
<tr>
<td>Procrustes</td>
<td>( d_{Gr(\infty, \infty)}^*(A, B) = \left(</td>
</tr>
</tbody>
</table>
where the latter is as described in Corollary 3.2 and Lemma 3.3. Clearly, these metrics on \( \text{Gr}(\infty, \infty) \)
are really nothing more than the amalgamation of two pieces of information, the distance \( \delta^*(A, B) \)
and the difference in dimensions \(|\text{dim} A - \text{dim} B|\), either via a root mean square or an indicator
function.

We will see in Section 7 that the chordal metric in Table 3 turns out to be equivalent to the
symmetric directional distance, a metric on subspaces of different dimensions \([47, 49]\) popular in
machine learning. We will also see in Proposition 6.2 that the Grassmann metric in Table 3 has
the following interpretation: \( d_{\text{Gr}(\infty, \infty)}(A, B) \) is the distance of the furthest \( l \)-dimensional subspace
from \( B \) that contains \( A \), which equals the distance of the furthest \( k \)-dimensional subspace from \( A \)
contained in \( B \) for sufficiently large \( n \).

**Proposition 6.1.** The expressions in Table 6 and (24) are metrics on \( \text{Gr}(\infty, \infty) \).

Proof. It is trivial to see that the expression defined in (24) yields a metric on \( \text{Gr}(\infty, \infty) \) for
\( * = \alpha, \beta, \mu, \pi, \sigma, \phi \), and so we just need to check the remaining three cases that take the form
in (23). Moreover, of the four defining properties of a metric, only the triangle inequality is not
immediately clear from (23).

Let \( k = \text{dim} A \), \( l = \text{dim} B \), and \( m = \text{dim} C \). We may assume WLOG that \( k \leq l \leq m \leq n \)
where \( n \) is chosen sufficiently large so that \( A, B, C \) 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 \( A, B, C \) respectively. Consider
the following \( (n + m - k) \times m \) matrices:

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

and set \( A' = \text{span}(A'), \quad B' = \text{span}(B'), \quad C' = \text{span}(C'); \) note that these are just \( A, B, C \) embeded
in \( \text{Graff}(m, n + m - k) \) via (22). It is straightforward to check that the expressions in Table 3 satisfy

\[
d_{\text{Gr}(\infty, \infty)}^*(A, B) = d_{\text{Gr}(m, n + m - k)}^*(A', B'),
\]

\[
d_{\text{Gr}(\infty, \infty)}^*(B, C) = d_{\text{Gr}(m, n + m - k)}^*(B', C'),
\]

\[
d_{\text{Gr}(\infty, \infty)}^*(A, C) = d_{\text{Gr}(m, n + m - k)}^*(A', C').
\]

Since \( A', B', C' \in \text{Gr}(m, n + m - k) \), the triangle inequality for \( d_{\text{Gr}(m, n + m - k)}^* \) immediately yields
the triangle inequality for \( d_{\text{Gr}(\infty, \infty)}^* \).

An alternative way to view the above proof is that for any \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \) where
\( k \leq l \leq n \),

\[
d_{\text{Gr}(\infty, \infty)}^*(A, B) = d_{\text{Gr}(l, n + l - k)}^*(i_{n, n+l-k}^l(A), i_{n, n+l-k}^l(B)).
\]

The embeddings \( i_{n, n+l-k}^l : \text{Gr}(k, n) \to \text{Gr}(l, n + l - k) \) and \( i_{n, n+l-k}^l : \text{Gr}(l, n) \to \text{Gr}(l, n + l - k) \) are
as defined in (22). In other words, these are isometric embeddings for all integers \( k \leq l \leq n \).

Our Grassmann metric \( d_{\text{Gr}(\infty, \infty)} \) in Table 3 has a nice interpretation as the distance between
furthest subspaces.

**Proposition 6.2.** Let \( k \leq l \leq n/2 \). Given \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \), we have

\[
\max_{X \in \Omega_+^l(A)} d_{\text{Gr}(l, n)}(X, B) = \max_{Y \in \Omega_+^l(B)} d_{\text{Gr}(k, n)}(Y, A) = d_{\text{Gr}(\infty, \infty)}(A, B).
\] (25)

Proof. Without loss of generality, we may assume that \( A \cap B = \{0\} \) by Proposition 2.1. Since

\[
d_{\text{Gr}(l, n)}(X, B) = \delta(X, B) = \left( \sum_{i=1}^l \theta_i(X, B)^2 \right)^{1/2},
\]

and by Corollary 4.5

\[
\theta_i(X, B) \leq \theta_i(A, B), \quad i = 1, \ldots, k,
\]
we obtain
\[ d_{Gr(l,n)}(X, B) \leq (\delta(A, B)^2 + \sum_{i=k+1}^l \theta_i(X, B)^2)^{1/2}. \]

Let \((a_1, b_1), \ldots, (a_k, b_k)\) be the principal vectors between \(A\) and \(B\). We extend \(b_1, \ldots, b_k\) to obtain an orthonormal basis \(b_1, \ldots, b_k, b_{k+1}, \ldots, b_l\) of \(B\). Let \(X \cap A^\perp\) be the orthogonal complement of \(A\) in \(X\) and let \(B_0 := \text{span}\{b_{k+1}, \ldots, b_l\}\). Then we have
\[ (\sum_{i=k+1}^l \theta_i(X, B)^2)^{1/2} = \delta(X \cap A^\perp, B_0), \]
and the last inequality becomes
\[ d_{Gr(l,n)}(X, B) \leq \sqrt{\delta(A, B)^2 + \delta(X \cap A^\perp, B_0)^2}. \]

If \(n \geq 2l\), then there exist \(l-k\) vectors \(c_1, \ldots, c_{l-k}\) orthogonal to \(A\) and \(B\) simultaneously. Choosing \(X = \text{span}\{a_1, \ldots, a_k, c_1, \ldots, c_{l-k}\}\), we attain the required maximum:
\[ d_{Gr(l,n)}(X, B) = \sqrt{\delta(A, B)^2 + (l-k)\pi^2/4} = d_{Gr(\infty, \infty)}(A, B). \]

The second equality in (25) follows from \(d_{Gr(\infty, \infty)}(A, B) = d_{Gr(\infty, \infty)}(B, A)\), given that \(d_{Gr(\infty, \infty)}\) is a metric by Proposition 6.1.

In case the reader is curious why the existence of the metrics \(d_{Gr(\infty, \infty)}\) as defined in (23) and (24) does not contradict our earlier discussion about the general nonexistence of coproduct in \(\text{Met}\), the reason is that these metrics do not respect continuous contractions.

Take the Grassmann metric on \(\text{Gr}(\infty, \infty)\) for instance. \((\text{Gr}(\infty, \infty), d_{Gr(\infty, \infty)})\) is an object of the category \(\text{Met}\) but it is not the coproduct of \(\{(\text{Gr}(k, \infty), d_{\text{Gr}(k, \infty)}) : k \in \mathbb{N}\}\). Indeed, let \(Y = \{y_1, y_2\}\) be a two-point set with a metric defined by \(d_Y(y_1, y_2) = 1\). We consider a family of maps \(f_k : \text{Gr}(k, \infty) \to Y\) defined by
\[ f_k(A) = \begin{cases} y_1 & \text{if } k = 2, \\ y_2 & \text{otherwise.} \end{cases} \]

Then \(f_k\) is a continuous contraction between \(\text{Gr}(k, \infty)\) and \(Y\). So \(\{f_k : k \in \mathbb{N}\}\) is a family of morphisms in \(\text{Met}\) compatible with \(\{(\text{Gr}(k, \infty), d_{\text{Gr}(k, \infty)}) : k \in \mathbb{N}\}\). If \(\text{Gr}(\infty, \infty), d_{\text{Gr}(\infty, \infty)})\) is the coproduct of this family, then there must exist a continuous contraction \(f : \text{Gr}(\infty, \infty) \to Y\) such that \(f \circ \iota_k = f_k\) with \(\iota_k\) being the natural inclusion of \(\text{Gr}(k, \infty)\) into \(\text{Gr}(\infty, \infty)\). But taking \(A \in \text{Gr}(2, \infty)\) and \(B \in \text{Gr}(3, \infty)\), we see that
\[ d_{\text{Gr}(\infty, \infty)}(A, B) \geq \frac{\pi}{2} > 1 = d_Y(f(A), f(B)), \]
contradicting the surmise that \(f\) is a contraction. Similarly, one may show that \((\text{Gr}(\infty, \infty), d_{\text{Gr}(\infty, \infty)})\) is not a coproduct in \(\text{Met}\) for any \(* = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi\).

Note that \((\text{Gr}(\infty, \infty), d_{\text{Gr}(\infty, \infty)})\) is also not the coproduct of \(\{(\text{Gr}(k, \infty), d_{\text{Gr}(k, \infty)}) : k \in \mathbb{N}\}\) in the category of metric spaces with continuous (or uniformly continuous) maps as morphisms. The coproduct in this category is simply \(\text{Gr}(\infty, \infty)\) with the metric induced by the disjoint union topology, which is much too fine (in the sense of topological spaces) to be interesting. For example, such a metric will not be related to the distance \(\delta\) in any way.

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. They turn out to be special cases of our distance in Section 4 and our metric in Section 6. In the following, let \(A \in \text{Gr}(k, n)\) and \(B \in \text{Gr}(l, n)\) be arbitrary subspaces.
The containment gap is defined as
\[
\gamma(A, B) := \max_{a \in A} \min_{b \in B} \frac{\|a - b\|}{\|a\|}.
\] (26)
This was proposed in [31, pp. 197–199] and used in numerical linear algebra [45], particularly for measuring separation between Krylov subspaces [7]. We see here that it is in fact equivalent to our projection distance \(\delta^\pi\) in Theorem 4.7. It was observed in [7, p. 495] that
\[
\gamma(A, B) = \sin(\theta_k(A, Y))
\]
where \(Y \in \Omega_-(B)\) is nearest to \(A\) in the projection distance \(d^\pi_{\text{Gr}(k,n)}\). By Theorem 4.7, we deduce that it can also be realized as
\[
\gamma(A, B) = \sin(\theta_l(B, X))
\]
where \(X \in \Omega_+(A)\) is nearest to \(B\) in the projection distance \(d^\pi_{\text{Gr}(k,n)}\), a fact about the containment gap that had not been observed before. Indeed, by Theorem 4.7 we get
\[
\gamma(A, B) = \delta^\pi(A, B)
\]
for all \(A \in \text{Gr}(k,n)\) and \(B \in \text{Gr}(l,n)\).

The symmetric directional distance is defined as
\[
d_{\Delta}(A, B) := \left(\max(k, l) - \sum_{i,j=1}^{k,l} \sin^2 \theta_i = \max(k, l) - \sum_{i,j=1}^{k,l} (a_i^T b_j)^2 = \delta_{\Delta}(A, B)^2,\right)
\] (27)
where \(A = [a_1, \ldots, a_k]\) and \(B = [b_1, \ldots, b_l]\) are, as usual, the respective orthonormal bases. This was proposed in [17, 49], and has been widely used [5, 14, 19, 26, 34, 42, 43, 50, 53]. The definition (27) turns out to be identical to our chordal metric \(d^c_{\text{Gr}(\infty, \infty)}\) in Table 3,
\[
d^c_{\text{Gr}(\infty, \infty)}(A, B)^2 = |k - l| + \sum_{i=1}^{\min(k,l)} \sin^2 \theta_i = \max(k, l) - \sum_{i,j=1}^{k,l} (a_i^T b_j)^2 = \delta_{\Delta}(A, B)^2,
\]
since \(|k - l| = \max(k, l) - \min(k, l)\), and
\[
\sum_{i,j=1}^{k,l} (a_i^T b_j)^2 = \|A^T B\|_F^2 = \sum_{i=1}^{\min(k,l)} \cos^2 \theta_i = \min(k, l) - \sum_{i=1}^{\min(k,l)} \sin^2 \theta_i.
\]

8. Geometry of \(\Omega_+(A)\) and \(\Omega_-(B)\)

Up to this point, \(\Omega_+(A)\) and \(\Omega_-(B)\), as defined in Definition 4.1, have been treated as mere subsets of \(\text{Gr}(l,n)\) and \(\text{Gr}(k,n)\) respectively. We will see in this section that \(\Omega_+(A)\) and \(\Omega_-(B)\) have rich geometric properties. First and foremost, we will see that they are Schubert varieties, thereby justifying their names. We give a concrete definition of Schubert variety and the closely related notion of flag variety below.

Definition 8.1. The Schubert variety \(\Omega(X_1, \ldots, X_k, n)\) is the set of \(k\)-planes \(Y\) satisfying the Schubert conditions
\[
\dim(Y \cap X_i) \geq i, \quad i = 1, \ldots, k,
\]
where \(X_1 \subset X_2 \subset \cdots \subset X_k\) is a fixed flag of linear subspaces of \(\mathbb{R}^n\), i.e.,
\[
\{Y \in \text{Gr}(k,n) : \dim(Y \cap X_i) \geq i, \quad i = 1, \ldots, k\}.
\]
Let \(0 =: k_0 < k_1 < \cdots < k_{m+1} := n\) be a sequence of increasing nonnegative integers. The associated flag variety is the set of flags satisfying the condition
\[
\dim X_i = k_i, \quad i = 0, 1, \ldots, m + 1.
\]
We denote it by \(\text{Flag}(k_1, \ldots, k_m, n)\), i.e.,
\[
\{(X_1, \ldots, X_m) \in \text{Gr}(k_1, n) \times \cdots \times \text{Gr}(k_m, n) : X_i \subset X_{i+1}, \quad i = 1, \ldots, m\}.
\]
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 Flag(\( k, n \)) = Gr(k, n). In fact, like Grassmannians, Flag(\( k_1, \ldots, k_m, n \)) may also be viewed as a smooth manifold and in this context is often called a flag manifold. The parallel goes further, Flag(\( k_1, \ldots, k_m, n \)) is a homogeneous space,

\[
\text{Flag}(k_1, \ldots, k_m, n) \cong O(n)/(O(d_1) \times \cdots \times O(d_{m+1}))
\]

where \( d_i = k_i - k_{i-1} \) for \( i = 1, \ldots, m+1 \), generalizing

\[
\text{Gr}(k, n) \cong O(n)/(O(k) \times O(n-k)).
\]

Let \( A, B \in \text{Gr}(k, n) \). By Definition 8.1, we see that \( \Omega_+(A) \) and \( \Omega_-(B) \) are Schubert varieties in \( \text{Gr}(k, n) \): Choose the following flags,

\[
\{0\} = A_0 \subset A_1 \subset \cdots \subset A_k := A, \quad B_0 \subset B_1 \subset \cdots \subset B_{n-k} := \mathbb{R}^n,
\]

(note that these are complete flags in \( A \) and \( \mathbb{R}^n/B \) respectively). Then

\[
\Omega_+(A) = \Omega(A_1, \ldots, A_k), \quad \Omega_-(B) = \Omega(B_0, \ldots, B_{n-k}).
\]

The isomorphism \( \text{Gr}(k, n) \cong \text{Gr}(n-k, n) \) that sends \( X \) to \( X^\perp \) takes \( \Omega_+(A) \) to \( \Omega_-(A^\perp) \) and \( \Omega_-(B) \) to \( \Omega_+(B^\perp) \). Thus \( \Omega_+(A) \) and \( \Omega_-(B) \) may also be viewed as Schubert varieties in \( \text{Gr}(n-k, n) \).

An important observation is that \( \Omega_+(A) \) and \( \Omega_-(B) \), despite superficial difference in their definitions, are essentially identical type of objects.

**Proposition 8.2.** For any \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \), we have

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

**Proposition 8.3.** Let \( A, A' \in \text{Gr}(k, n) \) and \( B, B' \in \text{Gr}(l, n) \). Then

\[
\Omega_+(A) = \Omega_+(A') \quad \text{if and only if} \quad A = A',
\]

\[
\Omega_-(B) = \Omega_-(B') \quad \text{if and only if} \quad B = B'.
\]

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

\[
A = \bigcap_{X \in \Omega_+(A)} X = \bigcap_{X \in \Omega_+(A')} X = A'.
\]

The converse is obvious. The statement for \( \Omega_- \) then follows from Proposition 8.2 \( \square \)

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

The resemblance between \( \Omega_+(A) \) and \( \Omega_-(B) \) in Proposition 8.2 goes further — we may view them as ‘sub-Grassmannians’.

**Proposition 8.4.** Let \( k \leq l \leq n \) be positive integers. For any \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \), we have:

\[
\Omega_+(A) \cong \text{Gr}(l-k, n-k), \quad \Omega_-(B) \cong \text{Gr}(k, l),
\]
which are both isomorphisms of algebraic varieties and diffeomorphisms of smooth manifolds. Consequently,
\[ \dim \Omega_+(A) = (n-l)(l-k), \quad \dim \Omega_-(B) = k(l-k). \]

**Proof.** The first isomorphism is given by the projection sending \( X \in \Omega_+(A) \) to its quotient by \( A \),
\[ \varphi : \Omega_+(A) \to \text{Gr}_{l-k}(\mathbb{R}^n/A), \quad X \mapsto X/A \subseteq \mathbb{R}^n/A, \]
and observing that \( \text{Gr}_{l-k}(\mathbb{R}^n/A) \cong \text{Gr}(l-k, n-k) \). The second isomorphism can be seen by regarding a \( k \)-dimensional subspace \( Y \) of \( \mathbb{R}^n \) in \( \Omega_-(B) \) as a \( k \)-dimensional subspace of \( B \), i.e.,
\[ \Omega_-(B) = \text{Gr}_k(B) \cong \text{Gr}(k,l). \]

The observation that \( \Omega_+(A) \) and \( \Omega_-(B) \) are essentially Grassmannians allows us to immediately infer the following basic properties:

(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 geodesically convex.

The topology in (i) refers to the metric space topology, not Zariski topology. A consequence of compactness is that the distance \( d_{\text{Gr}(k,n)}(A, \Omega_-(B)) = d_{\text{Gr}(l,n)}(B, \Omega_+(A)) \) can be attained by points in \( \Omega_-(B) \) and \( \Omega_+(A) \) respectively. We constructed these closest points explicitly when we proved Theorem 4.2.

We could of course deduce much more about topological and geometric properties of \( \Omega_+(A) \) and \( \Omega_-(B) \) since Proposition 8.4 implies that they inherit everything that we know about Grassmannians (coordinate ring, cohomology ring, Plücker relations, etc) but we see no point in such an exercise. In particular, \( \Omega_+(A) \) and \( \Omega_-(B) \) are also flag varieties.

One property that does not quite follow from Proposition 8.4 is the geodesic convexity claimed in (iii) and we provide a short proof below.

**Proposition 8.5.** \( \Omega_+(A) \) and \( \Omega_-(B) \) are geodesically convex.

**Proof.** By Proposition 8.2 it suffices to show that \( \Omega_-(B) \) is geodesically convex, i.e., any two points in \( \Omega_-(B) \) can be connected by a geodesic curve in \( \Omega_-(B) \). By Proposition 8.4 \( \Omega_-(B) \) is the image of \( \text{Gr}(k,l) \) embedded isometrically in \( \text{Gr}(k,n) \). So for any \( X_1, X_2 \in \text{Gr}(k,l) \), we have by Lemma 3.1
\[ d_{\text{Gr}(k,n)}(X_1, X_2) = d_{\text{Gr}(k,l)}(X_1, X_2) = d_{\Omega_-(B)}(X_1, X_2), \quad (29) \]
where \( d_{\Omega_-(B)} \) denotes the geodesic distance in \( \Omega_-(B) \). If \( d_{\Omega_-(B)}(X_1, X_2) \) is realized by a geodesic curve \( \gamma \) in \( \Omega_-(B) \), then \( \gamma \) must also be a geodesic curve in \( \text{Gr}(k,n) \) by \([29]\). \( \square \)

9. Grassmannians as matrix varieties

While we have thus far been thinking of \( \text{Gr}(k,n) \) as a set of equivalence classes of matrices, there is another well-known representation \( [10] \) Example 1.2.20 of \( \text{Gr}(k,n) \) as a set of actual matrices, namely, the set of idempotent symmetric matrices of trace \( k \):
\[ \text{Gr}(k,n) \cong \{ P \in \mathbb{R}^{n \times n} : P^T = P^2 = P, \ \text{tr}(P) = k \}. \quad (30) \]
Such a representation allows one to regard \( \text{Gr}(k,n) \) as a subvariety of \( \mathbb{R}^{n \times n} \). The purpose of this short section is to present the Schubert varieties \( \Omega_+(A) \) and \( \Omega_-(B) \) in this alternate form.

The isomorphism in \([30]\) maps each subspace \( A \in \text{Gr}(k,n) \) to \( P_A \in \mathbb{R}^{n \times n} \), the unique orthogonal projection onto \( A \), and its inverse takes an orthogonal projection \( P \) to the subspace \( \text{im}(P) \in \text{Gr}(k,n) \). Note that a matrix is an orthogonal projection if and only if it is symmetric and idempotent; the first two equalities in \([30]\) ensure that \( P \) is as such. Note also that the eigenvalues of an orthogonal\(^3\) isomorphism will mean both an isomorphism of algebraic varieties and a diffeomorphism of smooth manifolds.
projection onto a subspace of dimension \( k \) are 1’s with multiplicity \( k \) and 0’s with multiplicity \( n-k \); so the condition \( \text{tr}(P) = k \) is equivalent to \( \text{rank}(P) = k \), ensuring that \( P \) projects onto a subspace of dimension \( k \).

\( \Omega_+(A) \) and \( \Omega_-(B) \) may be represented in a manner consistent with (30),

\[
\Omega_+(A) \cong \{ P \in \mathbb{R}^{n \times n} : P^T P^2 = P, \text{tr}(P) = l, \text{im}(A) \subseteq \text{im}(P) \}, \\
\Omega_-(B) \cong \{ P \in \mathbb{R}^{n \times n} : P^T P^2 = P, \text{tr}(P) = k, \text{im}(P) \subseteq \text{im}(B) \}.
\]

Observe that such a matrix representation allows us to simultaneously embed \( \text{Gr}(k,n) \), \( \text{Gr}(l,n) \), \( \Omega_+(A) \), \( \Omega_-(B) \) into \( \mathbb{R}^{n \times n} \).

10. Probability density on the Grassmannian

In this section, we determine the relative volumes of the Schubert varieties \( \Omega_+(A) \), \( \Omega_-(B) \). We prove a volumetric analogue of (17) in Theorem 4.2. Given any \( k \)-dimensional subspace \( A \) and \( l \)-dimensional subspace \( B \) in \( \mathbb{R}^n \), the probability that a randomly chosen \( l \)-dimensional subspace in \( \mathbb{R}^n \) contains \( A \) equals the probability that a randomly chosen \( k \)-dimensional subspace in \( \mathbb{R}^s \) is contained in \( B \). This probability value is independent of our choices of \( A \) and \( B \) and only depends on \( k, l, n \).

Every Riemannian metric on a Riemannian manifold yields a volume density that in turn allows one to define a notion of volume on the manifold that is consistent with the metric [40, Example 3.4.2]. In the case of the Grassmannian, the Riemannian metric \( [40] \) on \( \text{Gr}(k,n) \) that gave rise to the Grassmann distance in [8] and the geodesic in [9] yields a density \( d\gamma_{k,n} \) on \( \text{Gr}(k,n) \). The volume of \( \text{Gr}(k,n) \) is then

\[
\text{Vol}(\text{Gr}(k,n)) = \int_{\text{Gr}(k,n)} |d\gamma_{k,n}|,
\]

and this can be evaluated explicitly. We reproduce [40 Proposition 9.1.12] (see also [38, pp. 48–53]) below for easy reference.

**Proposition 10.1.** The volume of \( \text{Gr}(k,n) \) is

\[
\text{Vol}(\text{Gr}(k,n)) = \binom{n}{k} \frac{\prod_{j=1}^{k} \omega_j}{(\prod_{j=1}^{n-k} \omega_j)},
\]

where \( \omega_m := \frac{\pi^{m/2}}{\Gamma(1 + m/2)} \) is the volume of the unit 2-norm ball in \( \mathbb{R}^m \).

The normalized density \( d\mu_{k,n} := \text{Vol}(\text{Gr}(k,n))^{-1} |d\gamma_{k,n}| \) then defines a natural uniform probability density on \( \text{Gr}(k,n) \). We will show that with respect to the uniform density, the probability of landing on \( \Omega_+(A) \) in \( \text{Gr}(l,n) \) and the probability of landing on \( \Omega_-(B) \) in \( \text{Gr}(k,n) \) are the same. This observation will be important in Bayesian inference of subspaces [4] — the implication being that it does not matter whether we work with \( A \) and \( \Omega_-(B) \) within \( \text{Gr}(k,n) \) or with \( B \) and \( \Omega_+(A) \) within \( \text{Gr}(l,n) \).

**Corollary 10.2.** Let \( k \leq l \leq n \). Let \( A \in \text{Gr}(k,n) \) and \( B \in \text{Gr}(l,n) \). The relative volumes of \( \Omega_+(A) \) in \( \text{Gr}(l,n) \) and \( \Omega_-(B) \) in \( \text{Gr}(k,n) \) are identical,

\[
\mu_{l,n}(\Omega_+(A)) = \mu_{k,n}(\Omega_-(B)).
\]

Their common value does not depend on the choices of \( A \) and \( B \) but only on \( k, l, n \) and is given by

\[
\frac{l!(n-k)!}{n!(l-k)!} \frac{\prod_{j=l-k+1}^{l} \omega_j}{\prod_{j=n-k+1}^{n} \omega_j}.
\]

\( ^4 \)The Riemannian metric on \( \text{Gr}(k,n) \) has been discussed at length in [2, 18]. Since we have no use for it except implicitly, we did not specify it in this article.
Proof. By Proposition 8.4, $\Omega_+(A)$ is isometric to $\text{Gr}(n-l,n-k)$ and so we have

$$\text{Vol}(\Omega_+(A)) = \binom{n-k}{n-l} \frac{\prod_{j=1}^{n-k} \omega_j}{(\prod_{j=1}^{n-l} \omega_j)(\prod_{j=1}^{l-k} \omega_j)}$$

by Proposition 10.1. Likewise, $\Omega_-(B)$ is isometric to $\text{Gr}(k,l)$ and so

$$\text{Vol}(\Omega_-(B)) = \binom{l}{k} \frac{\prod_{j=1}^{l} \omega_j}{(\prod_{j=1}^{k} \omega_j)(\prod_{j=1}^{l-k} \omega_j)}.$$}

Dividing by the volumes of $\text{Gr}(l,n)$ and $\text{Gr}(k,n)$ respectively yields the required results. 

By its definition, relative volume must be dependent on the volume of the ambient space and so the dependence on $n$ is expected. This is a slight departure the independence of $n$ in Theorem 4.2(i).

11. Conclusions

In this article, we provide what we hope is a thorough study of subspace distances, a topic of wide-ranging interests in applied mathematics, computations, and statistics. We investigated the topic from many different angles and filled in the most glaring gaps in our existing knowledge — defining distances and metrics for inequidimensional subspaces.

In the course of our investigations, we developed simple geometric models for linear subspaces of all dimensions, the infinite and doubly-infinite Grassmannians. We also enriched the existing differential geometric view of Grassmannians in applied and computational mathematics with algebraic geometric perspectives. We expect these to be of independent interests and our exposition takes a view towards making these models useful to applied and computational mathematicians.

Acknowledgment. We would like to thank Sayan Mukherjee for educating us about the importance of measuring distances between subspaces of different dimensions and Frank Sottile for invaluable discussions that led to the Schubert variety approach. We also gratefully acknowledge Lizhen Lin, Tom Luo, and Giorgio Ottaviani for their helpful suggestions. LHL and KY are partially supported by AFOSR FA9550-13-1-0133, DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, and DMS 1057064. In addition, KY’s work is also partially supported by NSF CCF 1017760.

References


Department of Mathematics, University of Chicago, Chicago, IL 60637-1514.
E-mail address: kye@math.uchicago.edu

Corresponding author. Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, Chicago, IL 60637-1514.
E-mail address, corresponding author: lekhen@galton.uchicago.edu