DISTANCE BETWEEN SUBSPACES OF DIFFERENT DIMENSIONS

KE YE AND LEK-HENG LIM

Abstract. We resolve two problems regarding subspace distances that have arisen considerably often in applications: How could one define a notion of distance between (i) two linear subspaces of different dimensions, or (ii) two affine subspaces of the same dimension, in a way that generalizes the usual Grassmann distance between equidimensional linear subspaces? We show that (i) is the distance of a point to a Schubert variety, and (ii) is the distance in the Grassmannian of affine subspaces, both regarded as subvarieties in the Grassmannian. Combining (i) and (ii) yields a notion of distance between (iii) two affine subspaces of different dimensions. Aside from reducing to the usual Grassmann distance when the subspaces in (i) are equidimensional or when the affine subspaces in (ii) are linear subspaces, these distances are intrinsic and do not depend on any embedding. Furthermore, they may all be written down as concrete expressions involving principal angles and principal vectors, and are efficiently computable in numerical stable ways. We show that our results are largely independent of the Grassmann distance — if desired, it may be substituted by any other common distance between subspaces. Central to our approach to these problem 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

1.1. Motivations. 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. digital photos, MRI tractographs, movie clips), text data (blogs, emails, 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. One of the most fundamental problems, before setting out to do anything else with the data set, is to define a notion of distance between subspaces. The problem arises in diverse applications ranging from computer vision [41, 51], bioinformatics [25], machine learning [27], communication [40, 55], coding theory [5, 7, 14, 17], system identification [41], and statistical classification [30].

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 the 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 any orthonormal bases of \( A \) and \( B \) respectively, then the 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. 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 (1.1) with the corresponding expression in Table 2.

However what if the subspaces are of different dimensions and what if we have affine subspaces instead of linear subspaces? 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 \[8\], perturbation theory \[49\], information retrieval \[50\], facial recognition \[53\], motion segmentation \[15\], EEG signal analysis \[20\], mechanical engineering \[29\], economics \[46\], network analysis \[47\], blog spam detection \[39\], and decoding colored barcodes \[6\].

1.2. 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 \( l \), 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:

(i) 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.

(ii) The distance between affine subspaces is the distance within the Grassmannian of affine subspaces, a Zariski open subset of the Grassmannian.

Both distances in (i) and (ii) will have the following properties:

(a) readily computable via SVD;
(b) restrict to the usual Grassmann distance (1.1) for linear 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 linear and affine subspaces, possibly of different dimensions. These properties are established in Sections 4 and 11.

The applications mentioned in the last paragraph of Section 1.1 are all based on two existing proposals for a distance between subspaces of different dimensions: The containment gap \[36\], pp. 197–199] and the symmetric directional distance \[50\] \[52\]. 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 (in 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. 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 \( l \), 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 \( B \).

To demonstrate the utility of the distances defined, we will use them to study other problems. In Section 8, we give two examples: Defining a geometric mean of two subspaces and determining a furthest subspace from a given one where the subspaces may be of different dimensions.

The distance in (ii) is a metric on the Grassmannian of affine subspaces. Given that this object provides a natural geometric model for affine subspaces that is likely to be of independent interest, we develop some of its basic geometric properties in Section 10. We then discuss in Section 11 why the only way one could conceivably obtain a distance between affine subspaces is to regard the Grassmannian of affine subspaces as a Zariski open subset of the usual Grassmannian of linear subspaces. Given our consideration of the distance in (i), it is natural to ask if one may also define a notion of distance between affine subspaces of different dimensions. We will see the distances defined for (i) and (ii) are consistent and when combined yields a notion of distance for affine subspaces of different dimensions. This is also discussed in Section 11.

In Section 13 we obtain a volumetric analogue to our key result with respect to the natural probability measure on Grassmannians. Given two arbitrary subspaces in \( \mathbb{R}^n \), \( A \) of dimension \( k \) and \( B \) of dimension \( l \), we show that the probability a random \( l \)-dimensional subspace contains \( A \) equals the probability a random \( k \)-dimensional subspace is contained in \( B \).

1.3. 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 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 [19] 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 [19] 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.

\footnote{Not our fault: 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.}
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.

| Grassmannian | $\text{Gr}(k, n)$ | models $k$-dimensional linear subspaces in $\mathbb{R}^n$ \cite{2} |
| Infinite Grassmannian | $\text{Gr}(k, \infty)$ | models $k$-dimensional linear subspaces regardless of ambient space \cite{3} |
| Doubly-infinite Grassmannian | $\text{Gr}(\infty, \infty)$ | models linear subspaces of all dimensions regardless of ambient space \cite{5} |
| Flag variety | $\text{Flag}(k_1, \ldots, k_m, n)$ | models nested sequences of linear subspaces in $\mathbb{R}^n$; generalizes Grassmannian: $\text{Flag}(k, n) = \text{Gr}(k, n)$ \cite{8} |
| Schubert variety | $\Omega(X_1, \ldots, X_m, n)$ | a ‘linearly constrained’ subset of $\text{Gr}(k, n)$ much like a polytope is a linearly constrained subset of $\mathbb{R}^n$ \cite{8, 10, 12} |
| Grassmannian of affine subspaces | $\text{Graff}(k, n)$ | models $k$-dimensional affine subspaces in $\mathbb{R}^n$ \cite{10, 12} |

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 Grassmannian of affine subspaces $\text{Graff}(k, n)$ will be discussed separately in Section 10. The differential geometric perspectives of our discussions below are drawn from \cite{26, 31, 32}, the more concrete matrix-theoretic view from \cite{2, 19, 54}, and the computational aspects from \cite{22}.

We will sometimes use differential geometric terminologies. A $k$-plane is a $k$-dimensional linear subspace and a $k$-flat is a $k$-dimensional affine subspace. A $k$-frame is an ordered basis of a $k$-plane and we will regard it as an $n \times k$ matrix whose columns $a_1, \ldots, a_k$ are the basis vectors. A flag is a strictly increasing sequence of nested linear subspaces, $X_k \subseteq X_{k+1}$, also known as a filtration. A flag is said to be complete if $\dim X_k = k$, finite if $k = 0, 1, \ldots, n$, and infinite if $k \in \mathbb{N} \cup \{0\}$.

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

$$V(k, n) \cong O(n)/O(n-k), \quad (2.1)$$

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

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

$$\text{Gr}(k, n) \cong V(k, n)/O(k) \cong O(n)/O(n-k) \times O(k)). \quad (2.2)$$

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

$$Q \cdot A := \text{span}(QA) \in \text{Gr}(k, n). \quad (2.3)$$
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 $Gr_k(V)$ and $V_k(V)$ for the sets of $k$-planes and $k$-frames in $V$ respectively. In this notation, $Gr(k,n) = Gr_k(\mathbb{R}^n)$ and $V(k,n) = V_k(\mathbb{R}^n)$.

A $k$-plane $A \in Gr(k,n)$ 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 $A$.

$Gr(k,n)$ 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 $g_X(A,B) = \sum_{i=1}^k a_i^T b_i = tr(A^T B)$, 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 $Gr(k,n)$ and in turn induces a geodesic distance on $Gr(k,n)$ which we define below.

Let $a_1,\ldots,a_k$ and $b_1,\ldots,b_l$ be bases (not necessarily orthonormal) for $A \in Gr(k,n)$ and $B \in Gr(l,n)$ 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

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

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 $A \in Gr(k,n)$ and $B \in Gr(l,n)$.

In practice principal vectors and principal angles may be readily computed using QR and SVD [11, 22]. Let $A = [a_1,\ldots,a_k]$ and $B = [b_1,\ldots,b_l]$ be orthonormal bases and

\[A^T B = U\Sigma V^T\]  \hspace{1cm} (2.4)

are 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

\[\theta_i = \cos^{-1} \sigma_i, \quad i = 1,\ldots,r.\]  \hspace{1cm} (2.5)

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 of the following matrices,

\[AU = [p_1,\ldots,p_k], \quad BV = [q_1,\ldots,q_l].\]  \hspace{1cm} (2.6)

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 [22, 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 $A \in Gr(k,n)$ and $B \in Gr(l,n)$ 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 \([54]\) the geodesic distance between \( A \) and \( B \) on \( \text{Gr}(k, n) \) is given by
\[
d_{\text{Gr}(k,n)}(A, B) = \left( \sum_{i=1}^{k} \theta_i^2 \right)^{1/2} = \|\cos^{-1} \Sigma\|_F. \tag{2.7}
\]

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 start by considering the matrix
\[
M := (I - AA^T)B(A^TB)^{-1} \in \mathbb{R}^{n \times k}.
\]

It is straightforward to verify that the condensed SVD of \( M \) takes the form
\[
M = Q(\tan \Theta)U^T,
\]
where \( U \in \text{O}(k) \) and \( \Theta = \text{diag}(\theta_1, \ldots, \theta_k) \in \mathbb{R}^{k \times k} \) are as in \([23]\) and \([25]\). The matrix \( Q \in \text{V}(k, n) \) 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 \( \gamma : [0, 1] \to \text{Gr}(k, n) \),
\[
\gamma(t) = \text{span}(AU \cos t \Theta + Q \sin t \Theta). \tag{2.8}
\]

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

Aside from the Grassmann distance, there are many other well-known ways to define a distance between \( A \) and \( B \) that makes \( \text{Gr}(k, n) \) into a metric space \([7, 16, 17, 19, 30]\). We present some of them in Table 2. The quantities in the expressions for Procrustes and spectral distances on the right column of Table 2 are the respective principal angles, by virtue of \([26]\). The Martin distance \([16]\) is \(+\infty\) whenever \( \theta_k \), the largest principal angle of \( A \) and \( B \), equals \( \pi/2 \). The value \( \sin \theta_1 \), sometimes referred to as max correlation distance \([30]\) or spectral distance \([17]\), is not a distance in the sense of a metric since it can be zero for a pair of distinct subspaces. What we called spectral distance in Table 2 is called chordal 2-norm distance in \([7]\).

The fact that all these distances in Table 2 depend on the principal angles is not a coincidence — the result \([54]\) Theorem 3] implies the following.

**Theorem 2.2.** 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 \( \text{O}(n) \), must be a
function of their principal angles. To be more specific, if a distance $d : \text{Gr}(k, n) \times \text{Gr}(k, n) \to [0, \infty)$ satisfies

$$d(\mathbf{Q} \cdot \mathbf{A}, \mathbf{Q} \cdot \mathbf{B}) = d(\mathbf{A}, \mathbf{B}), \text{ for all } \mathbf{A}, \mathbf{B} \in \text{Gr}(k, n) \text{ and all } \mathbf{Q} \in O(n),$$

where the action is as defined in (2.3), then $d$ must be a function of $\theta_i(\mathbf{A}, \mathbf{B})$, $i = 1, \ldots, k$.

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 this distances, which of course defined between equidimensional $A$ and $B$, we have a corresponding version for when $\dim A \neq \dim B$.

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. DISTANCE BETWEEN LINEAR SUBSPACES OF THE SAME DIMENSION

A conceivable way of defining a distance between $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$ 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 $A$ and $B$ to be their distance as measured in the ambient space. This is in fact the approach taken in [18], 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 [14]. 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 $A$ and an $l$-plane $B$ 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 Grassmannian distance stated in Corollary 3.2 that does not appear to be well-known and may be of independent interest.

Consider the inclusion map $\iota_n : \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by $\iota_n(a_1, \ldots, a_n) = (a_1, \ldots, a_n, 0)$. 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 $A \in \mathbb{R}^{n \times k}$ has orthonormal columns, then

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

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

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 $m \geq n$ 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

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

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 $A \in \text{Gr}(k, n)$ and its image $\iota_n(A) \in \text{Gr}(k, n + 1)$ and may in fact regard $A \in \text{Gr}(k, m)$ 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.,

\[
d_{\text{Gr}(k,n)}(A, B) = d_{\text{Gr}(k,n+1)}(\iota_n(A), \iota_n(B)). \tag{3.2}
\]

Repeated applications of (3.2) yields

\[
d_{\text{Gr}(k,n)}(A, B) = d_{\text{Gr}(k,m)}(\iota_{nm}(A), \iota_{nm}(B)) \tag{3.3}
\]

for all \( m > n \) and if we identify \( \text{Gr}(k, n) \) with \( \text{Gr}(k, \infty) \), we may then rewrite (3.3) as

\[
d_{\text{Gr}(k,n)}(A, B) = d_{\text{Gr}(k,m)}(A, B) \tag{3.4}
\]

for all \( m > n \).

**Proof.** 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 \([a_1, \ldots, a_k, e_{n+1}]\) where \( e_{n+1} \) is a unit vector in \( \mathbb{R}^{n+1} \) that is 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 (2.7), one immediately sees that (3.2) 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 (2.8) 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 the infinite Grassmannian \( \text{Gr}(k, \infty) \).

**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.,

\[
d^*_{\text{Gr}(k,n)}(A, B) = d^*_{\text{Gr}(k,m)}(\iota_{nm}(A), \iota_{nm}(B)), \quad * = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi.
\]

Consequently \( d^*_{\text{Gr}(k,\infty)}(A, B) \) is well-defined for any \( A, B \in \text{Gr}(k, \infty) \).

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

\[
\Omega_+(A) := \{ X \in \text{Gr}(l,n) : A \subseteq X \} \quad \text{and} \quad \Omega_-(B) := \{ Y \in \text{Gr}(k,n) : Y \subseteq B \}. \quad (4.1)
\]

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, these are indeed Schubert varieties. At this point, it suffices to know that they are closed subsets of \( \text{Gr}(k,n) \) and \( \text{Gr}(l,n) \) respectively.

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 \( \Omega_-(B) \),

\[
\delta(A,B) := d_{\text{Gr}(k,n)}(A,\Omega_-(B)) = \min\{d_{\text{Gr}(k,n)}(A,Y) : Y \in \Omega_-(B)\}. \quad (4.2)
\]

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.

**Figure 1.** Distance between a line \( A \) and a plane \( B \) in \( \mathbb{R}^3 \). \( X \) is closest to \( A \) among all lines in \( B \). The length of the geodesic \( \gamma \) from \( A \) to \( X \) gives the distance.
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 \( \Omega_+(A) \),

\[
\delta(A, B) := d_{\text{Gr}(l, n)}(B, \Omega_+(A)) = \min \{ d_{\text{Gr}(l, n)}(B, X) : X \in \Omega_+(A) \}. 
\] (4.3)

It will have same desirable features as the one in (4.2) 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 (4.2) and (4.3) 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 (4.2) and (4.3), 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 (cf. Algorithm 9.1).

**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

\[
d_{\text{Gr}(k, n)}(A, \Omega_-(B)) = d_{\text{Gr}(l, n)}(B, \Omega_+(A)). 
\] (4.4)

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

\[
\delta(A, B) = \left( \sum_{i=1}^{\min\{k, l\}} \theta_i(A, B)^2 \right)^{1/2} 
\] (4.5)

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

Rewriting (4.4) as

\[
\min_{X \in \Omega_-(A)} d_{\text{Gr}(l, n)}(X, B) = \delta(A, B) = \min_{Y \in \Omega_+(B)} d_{\text{Gr}(k, n)}(Y, A), 
\] (4.6)

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’ in (4.6) replaced by ‘max’ when \( n \) is sufficiently large (cf. Proposition 9.4);
(c) ‘linear’ may be replaced by ‘affine’ (cf. Section 11);
(d) ‘distance’ may be replaced by ‘volume’ with respect to the intrinsic probability density on the Grassmannian (cf. Section 13).

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 \( \delta : \text{Gr}(k, n) \times \text{Gr}(l, n) \to [0, \infty) \) 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

\[
\delta(A, B) \geq d_{\text{Gr}(l, n)}(B, \Omega_+(A)). 
\]
Proof. It suffices to find an \( X \in \Omega_+(A) \) such that \( \delta(A, B) = d_{\text{Gr}(l,n)}(X, B) \). 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 \( \theta_i = \cos^{-1} p_i^T q_i, \| p_i \| = \| q_i \| = 1 \). If we take \( X \in \text{Gr}(l,n) \) to be the subspace spanned by \( p_1, \ldots, p_k, q_{k+1}, \ldots, q_l \), then

\[
d_{\text{Gr}(l,n)}(X, B) = [(\cos^{-1} p_1^T q_1)^2 + \cdots + (\cos^{-1} p_k^T q_k)^2 + (\cos^{-1} q_{k+1}^T q_{k+1})^2 + \cdots (\cos^{-1} q_l^T q_l)^2]^{1/2} = [\theta_1^2 + \cdots + \theta_k^2 + 0^2 + \cdots + 0^2]^{1/2} = \delta(A, B).
\]

(4.7)

\[ \square \]

The following fact is well-known in numerical linear algebra \[ 32, \text{Corollary 3.1.3} \]. 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 \( i = 1, \ldots, \min(\dim A, \dim B_k) \).

**Proof.** By appropriate orthogonalization if necessary, we may assume that \( B \) and its submatrix \( B_k \) are orthonormal bases of \( B \) and \( B_k \) respectively. Let \( A \) be an orthonormal basis of \( A \). Then \( \sigma_i(A^T B) \) and \( \sigma_i(A^T B_k) \) both take values in \([0, 1]\). Since \( \theta_i(A, B) = \cos^{-1}(\sigma_i(A^T B)) \) and \( \cos^{-1} \) is monotone decreasing in \([0, 1]\), it suffices to prove that

\[
\sigma_i(A^T B) \geq \sigma_i(A^T B_k)
\]

but this follows from Proposition 4.4 and the fact that \( A^T B_k \) is a submatrix of \( A^T B \). The other inequality is proved in the similar way. \[ \square \]

**Lemma 4.6.** Let \( A \) and \( B \) be as in Lemma 4.4. Then

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

**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

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

(4.8)

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

**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

\[
\delta(A, B) \leq d_{\text{Gr}(k,n)}(A, \Omega_-(B)) = d_{\text{Gr}(n-k,n)}(A^\perp, \Omega_+(B^\perp)) \leq \delta(A^\perp, B^\perp).
\]
But on the other hand, by results in [38] 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
\[ d_{Gr(l,n)}(B, \Omega^+(A)) = \delta(A, B) = d_{Gr(k,n)}(A, \Omega^-(B)) \]
from which we obtain (4.4) and (4.5) in Theorem 4.2. Property (ii) is obvious from (4.5) and
Property (i) follows from Lemma 3.1. □

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 \( d_{Gr(k,n)}(A, \Omega^-(B)) = \delta(A, B) \). We state these in Algorithm 9.1.

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 \( A \in Gr(k,n) \) and \( B \in Gr(l,n) \). Then
\[ d^*_{Gr(k,n)}(A, \Omega^-(B)) = d^*_{Gr(l,n)}(B, \Omega^+(A)), \quad * = \alpha, \beta, \kappa, \mu, \rho, \sigma, \phi. \]
Their common value \( \delta^*(A, B) \) is given by:
\[
\begin{align*}
\delta^\alpha(A, B) &= \theta_k, \\
\delta^\beta(A, B) &= \left(1 - \prod_{i=1}^k \cos^2 \theta_i \right)^{1/2}, \\
\delta^\kappa(A, B) &= \left(\sum_{i=1}^k \sin^2 \theta_i \right)^{1/2}, \\
\delta^\mu(A, B) &= \sin \theta_k, \\
\delta^\rho(A, B) &= \left(\log \prod_{i=1}^k \frac{1}{\cos^2 \theta_i} \right)^{1/2}, \\
\delta^\sigma(A, B) &= \cos^{-1}\left(\prod_{i=1}^k \cos \theta_i \right), \\
\delta^\phi(A, B) &= 2 \sin(\theta_k/2), \\
\delta^\phi(A, B) &= \left(2 \sum_{i=1}^k \sin^2(\theta_i/2) \right)^{1/2},
\end{align*}
\]
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 \( Gr(k,n) \) and \( Gr(n-k,n) \) is an isometry under these distances. In particular, both (4.7) and (4.8) would still hold with any of these other
distance functions in place of the Grassmann distance. □

We will see in Section 7 that the projection distance \( \delta^p \) 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 [36].

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 \( Gr(k,n) \). For
the special case of the Euclidean space \( \mathbb{R}^n = Gr(0,n) \), the problem often arise in applications [18, 45]. 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 [31]; for
general varieties, it is at least as hard as deciding Hilbert Nullstellensatz, which is NP-complete
in the Blum–Shub–Smale sense [12, 13]. 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

The equality of \( d_{\text{Gr}(k,n)}(A, \Omega_-(B)) \) and \( d_{\text{Gr}(l,n)}(B, \Omega_+(A)) \) is, in our opinion, 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

\[
\delta(L, A) = \delta(L, A) + \delta(A, B) \geq \delta(L, B)
\]

and

\[
\delta(L, B) = \delta(L, B) + \delta(A, B) \geq \delta(L, A),
\]

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** \([21]\), which may be viewed informally as the disjoint union of all \( k \)-dimensional subspaces\(^2\) over all \( k \in \mathbb{N} \),

\[
\text{Gr}(\infty, \infty) = \bigsqcup_{k=1}^{\infty} \text{Gr}(k, \infty).
\]

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 \( i_{nm}^{kl} : \text{Gr}(k,n) \to \text{Gr}(l,m) \) 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

\[
i_{nm}^{kl} : \text{Gr}(k,n) \to \text{Gr}(l,m), \quad \text{span}(A) \mapsto \text{span} \left( \begin{bmatrix} A & 0 \\ 0 & I_{l-k} \end{bmatrix} \right), \tag{5.1}\]

where \( I_{l-k} \in \mathbb{R}^{(l-k) \times (l-k)} \) is an identity matrix and we append \( (m-n)-(l-k) \) zero rows at the bottom so that the \( 3 \times 2 \) block matrix is in \( \mathbb{R}^{m \times l} \). Note that for a fixed \( k \), \( i_{nm}^{kk} \) reduces to \( i_{nm} \) in (3.1).

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

\(^2\) As discussed in Section 4 these are independent of the dimension of their ambient space and may be viewed as an element of the infinite Grassmannian \( \text{Gr}(k,\infty) \).
the topology induced by the metric $d_{\text{Gr}(k,\infty)}$ on $\text{Gr}(k, \infty)$ as defined in Section 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 $A \subseteq B$ by open subsets, as we have seen. Metrizable spaces are necessarily Hausdorff. \hfill $\square$

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 $\text{Met}$ 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 $\text{Met}$ 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 [28]. In the following, we will relax our requirement and construct a metric $d_{\text{Gr}(\infty, \infty)}$ on $\text{Gr}(\infty, \infty)$ that restricts to $d_{\text{Gr}(k,\infty)}$ for all $k \in \mathbb{N}$ but without requiring that it comes from a coproduct of $\{(\text{Gr}(k, \infty), d_{\text{Gr}(k,\infty)) : k \in \mathbb{N}\}$. In $\text{Met}$. \hfill \footnote{Without assuming $k \leq l$, we would have $|k - l|$ in place of $l - k$ and $\min(k,l)$ in place of $k$ in the sums.}

### 6. Metric for subspaces of all dimensions

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

Suppose $k \leq l$ and we have $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$. In this case there are $k$ principal angles between $A$ and $B$, $\theta_1, \ldots, \theta_k$, as defined in [25]. 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 $d_{\text{Gr}(\infty, \infty)}(A, B)$ (for Grassmann distance) and $d_{\text{Gr}(\infty, \infty)}^*(A, B)$ (for other distances) respectively. We will show in Proposition 6.1 that these expressions will indeed define metrics on $\text{Gr}(\infty, \infty)$. For now, we observe that the Grassmann, chordal, and Procrustes metrics on $\text{Gr}(\infty, \infty)$ are given respectively by the following expressions:\footnote{Without assuming $k \leq l$, we would have $|k - l|$ in place of $l - k$ and $\min(k,l)$ in place of $k$ in the sums.}

\[
d_{\text{Gr}(\infty, \infty)}(A, B) = \left(\sum_{i=1}^l \theta_i^2\right)^{1/2} = \left(\frac{l - k}{2} \pi^2 + \sum_{i=1}^k \theta_i^2\right)^{1/2}, \quad (6.1)
\]

\[
d_{\text{Gr}(\infty, \infty)}^*(A, B) = \left(\sum_{i=1}^l \sin^2 \theta_i\right)^{1/2} = \left(l - k + \sum_{i=1}^k \sin^2 \theta_i\right)^{1/2}, \quad (6.2)
\]

\[
d_{\text{Gr}(\infty, \infty)}^0(A, B) = \left(2 \sum_{i=1}^l \sin^2 (\theta_i/2)\right)^{1/2} = \left(l - k + 2 \sum_{i=1}^k \sin^2 (\theta_i/2)\right)^{1/2}. \quad (6.3)
\]

In short, they are all of the form

\[
d_{\text{Gr}(\infty, \infty)}^*(A, B) = \sqrt{\delta^*(A, B)^2 + c^2 e(A, B)^2}, \quad (6.4)
\]
where $\epsilon(A, B) := |\dim A - \dim B|^{1/2}$. On the other hand, the Asimov, Binet–Cauchy, Fubini–Study, Martin, projection, and spectral metrics on $Gr(\infty, \infty)$ are given by

$$d_{Gr(\infty, \infty)}^*(A, B) = \begin{cases} d_{Gr(k, \infty)}(A, B) & \text{if } \dim A = \dim B = k, \\ c_\ast & \text{if } \dim A \neq \dim B, \end{cases} \quad (6.5)$$

for $\ast = \alpha, \beta, \phi, \mu, \pi, \sigma$, respectively. The constants $c_\ast > 0$ are given by

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

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

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

where the latter are as described in Corollary 3.2 and Lemma 3.3. Clearly, these metrics on $Gr(\infty, \infty)$ are really nothing more than the combination of two pieces of information, the distance $\delta^*(A, B)$ and the difference in dimensions $|\dim A - \dim B|$, either via a root mean square or an indicator function.

We will see in Section 7 that the chordal metric in (6.2) turns out to be equivalent to the symmetric directional distance, a metric on subspaces of different dimensions $[50, 52]$ popular in machine learning. We will see in Proposition 9.4 that the Grassmann metric in (6.1) has following interpretation: $d_{Gr(\infty, \infty)}^*(A, B)$ 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 (6.4) and (6.5) are metrics on $Gr(\infty, \infty)$.

**Proof.** It is trivial to see that the expression defined in (6.5) yields a metric on $Gr(\infty, \infty)$ for $\ast = \alpha, \beta, \mu, \pi, \sigma, \phi$, and so we just need to check the remaining three cases that take the form in (6.4). Moreover, of the four defining properties of a metric, only the triangle inequality is not immediately clear from (6.4).

Let $k = \dim A$, $l = \dim B$, and $m = \dim C$. 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. We construct the following $n \times (m + n - k)$ matrices

$$A' = \begin{bmatrix} A & 0 \\ 0 & I_{n-k} \end{bmatrix}, \quad B' = \begin{bmatrix} B & 0 \\ 0 & I_{m-l} \end{bmatrix}, \quad C' = \begin{bmatrix} C \\ 0 \end{bmatrix} \in \mathbb{R}^{(n+m-k) \times m},$$

and set $A' = \text{span}(A')$, $B' = \text{span}(B')$ and $C' = \text{span}(C')$. It is straightforward to check that the expressions in (6.1), (6.2), (6.3) satisfy

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

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

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

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

$$d_{Gr(m, n+m-k)}^*(A', B') + d_{Gr(m, n+m-k)}^*(B', C') \geq d_{Gr(m, n+m-k)}^*(A', C'),$$

implies the triangle inequality for $d_{Gr(\infty, \infty)}^*$. \hfill \Box

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

$$d_{Gr(\infty, \infty)}^*(A, B) = d_{Gr(l, n+l-k)}^*(\ell_{n,n+l-k}^k(A), \ell_{n,n+l-k}^l(B)).$$
The embeddings $t_{n+l-k}^{k,l} : \text{Gr}(k, n) \to \text{Gr}(l, n + l - k)$ and $\iota_{n,n+l-k}^{l} : \text{Gr}(l, n) \to \text{Gr}(l, n + l - k)$ are as defined in (5.1). In other words, these are isometric embeddings for all integers $k \leq l \leq n$.

In case the reader is curious why the existence of the metrics $d^*_{\text{Gr}(\infty, \infty)}$ as defined in (6.4) and (6.5) 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. Even though $(\text{Gr}(\infty, \infty), d_{\text{Gr}(\infty, \infty)})$ is an object of the category $\text{Met}$, 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 $* = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi$.

We 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

$$\gamma(A, B) := \max_{a \in A} \min_{b \in B} \frac{\|a - b\|}{\|a\|}. \quad (7.1)$$

This was proposed in [36] pp. 197–199 and used in numerical linear algebra [49], particularly for measuring separation between Krylov subspaces [8]. We see here that it is in fact equivalent to our projection distance $\delta^\pi$ in Theorem 4.7. It was observed in [8] 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 $d^\pi_{\text{Gr}(l, n)}$, a fact about the containment gap that had not been observed before. Indeed, by Theorem 4.7 we see that

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

for all $A \in \text{Gr}(k, n)$ and $B \in \text{Gr}(l, n)$. 
The symmetric directional distance is defined as

\[ d_\Delta(A, B) := \left( \max(k, l) - \sum_{i,j=1}^{k,l} (a_i^T b_j)^2 \right)^{1/2} \]  (7.2)

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 [50, 52], and has been widely used [6, 15, 20, 29, 39, 46, 47, 53, 56]. The definition (7.2) turns out to be identical to our chordal metric \( d_{\text{Gr}(\infty, \infty)}^\kappa \) in (6.2),

\[ d_{\text{Gr}(\infty, \infty)}^\kappa(A, B)^2 = |k - l| + \sum_{i=1}^{\min(k, l)} \sin^2 \theta_i = \max(k, l) - \sum_{i,j=1}^{k,l} (a_i^T b_j)^2 = d_\Delta(A, B)^2, \]

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

\[ \sum_{i,j=1}^{k,l} (a_i^T b_j)^2 = \|A^T B\|^2_F = \sum_{i=1}^{\min(k, l)} \cos^2 \theta_i = \min(k, l) - \sum_{i=1}^{\min(k, l)} \sin^2 \theta_i. \]

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.** A Schubert variety in \( \text{Gr}(k, n) \) 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 \( \mathbb{R}^n \). We denote it by

\[ \Omega(X_1, \ldots, X_k, n) = \{ Y \in \text{Gr}(k, n) : \dim(Y \cap X_i) \geq i, \quad i = 1, \ldots, k \}. \]

Let \( 0 =: k_0 < k_1 < \cdots < k_m < k_{m+1} := n \) 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

\[ \text{Flag}(k_1, \ldots, k_m, n) = \{(X_1, \ldots, X_m) \in \text{Gr}(k_1, n) \times \cdots \times \text{Gr}(k_m, n) : \]

\[ \dim X_i = k_i, \quad X_i \subset X_{i+1}, \quad i = 1, \ldots, m \}. \]

Observe that a Schubert variety depends on a specific increasing sequence of subspaces whereas a flag variety depends only an increasing sequence of dimensions (of subspaces). Flag varieties may be viewed as a generalization of Grassmannians since if \( m = 1 \), then \( \text{Flag}(k, n) = \text{Gr}(k, n) \). In fact, like Grassmannians, \( \text{Flag}(k_1, \ldots, k_m, n) \) may also be viewed as a smooth manifold and in this context is often called a flag manifold. The parallel goes further, \( \text{Flag}(k_1, \ldots, k_m, n) \) is a homogeneous space,

\[ \text{Flag}(k_1, \ldots, k_m, n) \cong O(n)/(O(d_1) \times \cdots \times O(d_{m+1})) \]  (8.1)

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,

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

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

\[ \Omega_+(A) = \Omega(A_1, \ldots, A_k, n) \quad \text{and} \quad \Omega_-(B) = \Omega(B_0, \ldots, B_{n-k}, n). \]
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 $\text{Gr}(n-k, n)$. More importantly, this observation implies 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

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

In fact the resemblance between $\Omega_+(A)$ and $\Omega_-(B)$ goes further — they are both isomorphic to Grassmannians. In particular, $\Omega_+(A)$ and $\Omega_-(B)$ are also flag varieties.

**Proposition 8.3.** 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 the following:

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

which are both isomorphisms of algebraic varieties and diffeomorphisms of smooth manifolds. Consequently we have

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

**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.,

$$\Omega_-(B) = \text{Gr}_k(B) \cong \text{Gr}(k, l). \quad \square$$

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 8.3 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. One property that does not quite follow from Proposition 8.3 is the geodesic convexity claimed in (iii) and we provide a short proof below.

**Proposition 8.4.** $\Omega_+(A)$ 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 8.3 $\Omega_-(B)$ 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

$$d_{\text{Gr}(k,n)}(X_1, X_2) = d_{\text{Gr}(k,l)}(X_1, X_2) = d_{\Omega_-(B)}(X_1, X_2), \quad (8.2)$$

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 (8.2). \qed

\footnote{Henceforth the term isomorphism will mean a map that is both an isomorphism of algebraic varieties and a diffeomorphism of smooth manifolds.}
9. Nearest subspaces, geometric mean of subspaces, and furthest subspaces

As we mentioned in Section 4, the nearest subspaces attaining $\delta(\A, \B)$ can be constructed explicitly. Note that there are two such subspaces when the dimensions of $\A$ and $\B$ are different: For $\A \in \Gr(k, n)$ and $\B \in \Gr(l, n)$, there is a nearest $l$-dimensional subspace to $\B$ on the Schubert variety $\Omega_+(\A)$ and a nearest $k$-dimensional subspace to $\A$ on $\Omega_-(\B)$, both attaining $\delta(\A, \B)$. We state this construction formally in Algorithm 9.1. Computationally, the algorithm depends only

\begin{algorithm}
\caption{Nearest subspaces}
\begin{algorithmic}[1]
\Procedure{Nearest}{$\A, \B$} \Comment{bases of $\A$ and $\B$}
\State $[A, R] \leftarrow QR(A)$, $[B, R] \leftarrow QR(B)$ \Comment{orthonormalize with condensed QR}
\State $[U, \Sigma, V] \leftarrow \text{svd}(A^T B)$ \Comment{compute full SVD}
\State $[p_1, \ldots, p_k] \leftarrow AU$, $[q_1, \ldots, q_l] \leftarrow BV$ \Comment{compute principal vectors}
\State $X \leftarrow [p_1, \ldots, p_k, q_{k+1}, \ldots, q_l]$ \Comment{nearest $l$-plane to $\B$ containing $\A$}
\State $Y \leftarrow [q_1, \ldots, q_k]$ \Comment{nearest $k$-plane to $\A$ contained in $\B$}
\State \Return $X$, $Y$
\EndProcedure
\end{algorithmic}
\end{algorithm}

on two QR factorizations and an SVD and can therefore be done efficiently with a multitude of standard, numerically stable algorithms [11, 22].

**Proposition 9.1.** Let $X$ and $Y$ be the output of Algorithm 9.1. Then $\text{span}(X) \in \Gr(l, n)$ and $\text{span}(Y) \in \Gr(k, n)$ attain

\[
\min_{X \in \Omega_+(\A)} d_{\Gr(l,n)}(X, \B) \quad \text{and} \quad \min_{Y \in \Omega_-(\B)} d_{\Gr(k,n)}(Y, \A)
\]

(9.1) respectively. This remains true for all other distances in Table 2.

**Proof.** This follows from our constructive proofs of Lemmas 4.3, 4.6 and Theorem 4.2. In general, both the distance of a point to a set and the nearest point attaining it will depend on the choice of distance function used. It is no different for a Grassmannian — the choice of distance will matter if we replace the Schubert varieties $\Omega_+(\B)$ and $\Omega_-(\A)$ in (9.1) by arbitrary subsets of $\Gr(k, n)$ and $\Gr(l, n)$. The reason it holds for $\Omega_+(\B)$ and $\Omega_-(\A)$ is because of Theorem 4.7. \hfill $\square$

Algorithm 9.1 enables one to define \textit{geometric mean of two subspaces} of any dimensions. For equidimensional subspaces, the geodesic midpoint $\M$ of $\A, \B$ in $\Gr(k, n)$, i.e., the unique point satisfying

\[
d_{\Gr(k,n)}(\A, \M) = \frac{1}{2} d_{\Gr(k,n)}(\A, \B) = d_{\Gr(k,n)}(\B, \M),
\]
gives a notion of geometric mean for subspaces. This has been proposed for subspaces in [33] although the idea also applies to other Riemannian manifolds, notably the set of positive definite matrices equipped with its natural hyperbolic metric [9, 10].

Recall that the shortest geodesic $\gamma$ connecting $\A$ to $\B$ takes the form in (2.8). Since geodesics are of constant velocity, $\M$ is given by (2.8) with $t = 1/2$ and we have Algorithm 9.2.

Obtaining a geometric mean for subspaces $\A \in \Gr(k, n)$ and $\B \in \Gr(l, n)$ of different dimensions then reduces to using Algorithm 9.1 in conjunction with Algorithm 9.2. Evidently this involves a choice of whether we want our geometric mean to be of the same dimension as $\A$ or as $\B$. The algorithm below gives both: The midpoint $\M \in \Gr(k, n)$ on the shortest geodesic from $\A$ to $\Omega_-(\B)$ and the midpoint $\N \in \Gr(l, n)$ on the shortest geodesic from $\B$ to $\Omega_+(\A)$. 

...
Problem 9.2. Let \( k \leq l \leq n \).

(i) Find the largest separation between \( k \)- and \( l \)-dimensional subspaces

\[
\max \left\{ \delta(Y, X) : Y \in \text{Gr}(k, n), \ X \in \text{Gr}(l, n) \right\}.
\]

(ii) Given \( A \in \text{Gr}(k, n) \), find \( X \in \text{Gr}(l, n) \) furthest from \( A \),

\[
\max \left\{ \delta(A, X) : X \in \text{Gr}(l, n) \right\}.
\]

(iii) Given \( A \in \text{Gr}(k, n) \), find \( X \) in \( \text{Gr}(l, n) \) furthest from all subspaces containing \( A \),

\[
\max \left\{ d_{\text{Gr}(l,n)}(X, \Omega_+(A)) : X \in \text{Gr}(l, n) \right\}.
\]

(iv) Given \( A \in \text{Gr}(k, n) \) and \( B \in \text{Gr}(l, n) \), find \( X \in \text{Gr}(l, n) \) containing \( A \) and is furthest from \( B \),

\[
\max \left\{ d_{\text{Gr}(l,n)}(X, B) : X \in \Omega_+(A) \right\}.
\]

As another illustration of the use of Theorem 4.2 in answering various questions involving separation of subspaces, we examine the furthest subspace problem, partly motivated by considerations of optimal subspace packings [5, 7, 14, 17]. There are several natural formulations:

**Proposition 9.3.** The solutions to (i), (ii), (iii) are all equal,

\[
\max_{Y \in \text{Gr}(k,n), \ X \in \text{Gr}(l,n)} \delta(Y, X) = \max_{X \in \text{Gr}(l,n)} \delta(A, X) = \max_{X \in \text{Gr}(l,n)} d_{\text{Gr}(l,n)}(X, \Omega_+(A)),
\]
and their common value $\delta_{kln}$ depends only on $k, l, n$, but not on $A$,

$$\delta_{kln} := \begin{cases} \frac{\sqrt{n-k}}{\sqrt{2}} & n \geq k + l, \\ \frac{\sqrt{n}}{\sqrt{2}} & \text{otherwise}. \end{cases}$$

**Proof.** We start by explicitly constructing $X \in \text{Gr}(l,n)$ and $Y \in \text{Gr}(k,n)$ that attain the maximum value in (i), which we denote by $\delta$.

By the expression for $\delta$ in (15), we must have $d_1 \leq \frac{\sqrt{n}}{\sqrt{2}}$.

**Case 1:** $n \geq k + l$. Take any $l$-dimensional subspace $X$ and any $k$-dimensional subspace $Y$ such that $X$ is perpendicular to $Y$ in $\mathbb{R}^n$. Then the upper bound $\frac{\sqrt{n}}{\sqrt{2}}$ is attained by $X$ and $Y$.

**Case 2:** $n < k + l$. Here we must have $\dim(X \cap Y) \geq k + l - n > 0$. By Proposition 2.1

$$\delta(Y, X) = \delta(Y/(X \cap Y), X/(X \cap Y)),$$

and so $\delta(Y, X)$ is bounded above by $\frac{\sqrt{n}}{\sqrt{2}}$. Since $X$ and $Y$ are arbitrary, $d_1 \leq \frac{\sqrt{n}}{\sqrt{2}}$ in this case.

To attain this upper bound, notice that $X/(X \cap Y)$ and $Y/(X \cap Y)$ may be regarded as an $(n-k)$-dimensional subspace and an $(n-l)$-dimensional subspace of $\mathbb{R}^n/(X \cap Y) \cong \mathbb{R}^{2n-k-l}$. Since $2n - k - l = (n - l) + (n - k)$, we may apply Case 1 with $n - l, n - k, 2n - k - l$ playing the roles of $k, l, n$ respectively, and conclude that the upper bound $\frac{\sqrt{n}}{\sqrt{2}}$ can be attained.

The maximum values in (i) and (ii) are equal because of Theorem 4.2. For any given $A \in \text{Gr}(k,n)$ and $X \in \text{Gr}(l,n)$,

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

The maximum values in (i) and (iii) are also equal. Note that the function $\delta$ is invariant under left multiplication by orthogonal matrices and this action of $O(n)$ is transitive on $\text{Gr}(k,n)$, i.e.,

$$\delta(Y, X) = \delta(QY, QX) \quad \text{for all } Q \in O(n),$$

and for any given $A \in \text{Gr}(k,n)$ we can always choose a $Q_0 \in O(n)$ such that $Q_0Y = A$ and thus $\delta(Y, X) = \delta(A, Q_0X)$. Hence we see that

$$\max_{Y \in \text{Gr}(k,n), X \in \text{Gr}(l,n)} \delta(Y, X) = \max_{X \in \text{Gr}(l,n)} \delta(A, Q_0X) = \max_{X \in \text{Gr}(l,n)} \delta(A, X).$$

Roughly speaking, the reason for the equality in Proposition 3.3 is that the value $\delta(X, Y)$ depends only on the relative positions of $X$ and $Y$ in $\mathbb{R}^n$. Problem 4.2.4 is however dependent on $A$ and $B$ and different in nature from (i), (ii), and (iii). It has a neat solution when the dimension of the ambient space is sufficiently large.

**Proposition 9.4.** If $n \geq 2l$, then the solution to (iv) is given by the distance between $A$ and $B$, measured via the Grassmann metric on $\text{Gr}(\infty, \infty)$ in (6.1). In fact we have

$$\max_{X \in \Omega_+(A)} d_{\text{Gr}(l,n)}(X, B) = d_{\text{Gr}(\infty, \infty)}(A, B) = \max_{Y \in \Omega_-(B)} d_{\text{Gr}(k,n)}(Y, A). \quad (9.2)$$

**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

$$d_{\text{Gr}(l,n)}(X, B) \leq \left(\delta(A, B)^2 + \sum_{i=k+1}^{l} \theta_i(X, B)^2\right)^{1/2}. \quad (9.3)$$

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

$$\left(\sum_{i=k+1}^{l} \theta_i(X, B)^2\right)^{1/2} = \delta(X \cap A^\perp, B_0),$$

and by Proposition 9.4.
and (9.3) 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 value:
\[ d_{Gr(l,n)}(X, B) = \sqrt{\delta(A, B)^2 + (l-k)\pi^2/4} = d_{Gr(\infty, \infty)}(A, B). \]
The second equality in (9.2) 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.

If \( n < 2l \), the value \( \sqrt{\delta(A, B)^2 + (l-k)\pi^2/4} \) is clearly still an upper bound for \( d_{Gr(l,n)}(X, B) \). However the upper bound cannot be attained in this case: Otherwise we would be able to find \( l-k \) vectors orthogonal to both \( A \) and \( B \) as in the proof above, which is impossible by a dimension count. We do not think that there is a simple expression for the maximum value when \( n < 2l \).

10. Grassmannian of affine subspaces

The Grassmannian of affine subspaces\(^5\) was first described in [37] but has received relatively little attention compared to the usual Grassmannian \( \text{Gr}(k, n) \). Aside from a brief discussion in [44, Section 9.1.3], we are unaware of any other systematic treatments. Nevertheless, given that it naturally parameterizes all affine subspaces in \( \mathbb{R}^n \), we think that the Grassmannian of affine subspaces would be a singularly important object that could rival the usual Grassmannian of linear subspaces in practical applicability, once its properties are properly established.

In this section we will establish some basic properties of the Grassmannian of affine subspaces that elucidates its structure. We view this section as a public service — the results here are neither difficult nor surprising, certainly routine to the experts, but to the best of our knowledge they have not appeared before elsewhere.

**Definition 10.1.** Let \( k < n \) be positive integers. The **Grassmannian of \( k \)-dimensional affine subspaces** in \( \mathbb{R}^n \) or Grassmannian of \( k \)-flats in \( \mathbb{R}^n \), denoted \( \text{Graff}(k, n) \), is the set of all \( k \)-dimensional affine subspaces of \( \mathbb{R}^n \). More generally, given an abstract vector space \( V \), we write \( \text{Graff}_k(V) \) for the set of \( k \)-flats in \( V \).

This set-theoretic definition does not reveal much about the rich geometry behind \( \text{Graff}(k, n) \). We will examine it below as (i) a differential manifold, (ii) a vector bundle, (iii) a homogeneous space, and (iv) an algebraic variety.

We denote a \( k \)-dimensional affine subspace as \( A + b \in \text{Graff}(k, n) \) where \( A \in \text{Gr}(k, n) \) is a \( k \)-dimensional linear subspace and \( b \in \mathbb{R}^n \) is the displacement of \( A \) from the origin. If \( A = [a_1, \ldots, a_k] \in \mathbb{R}^{n \times k} \) is a basis of \( A \), then
\[ A + b := \{\lambda_1a_1 + \cdots + \lambda_ka_k + b \in \mathbb{R}^n : \lambda_1, \ldots, \lambda_k \in \mathbb{R}\}. \]

The notation \( A + b \) may be taken to mean a coset of the subgroup \( A \) in the additive group \( \mathbb{R}^n \) or the Minkowski sum of the sets \( A \) and \( \{b\} \) in the Euclidean space \( \mathbb{R}^n \). The dimension of \( A + b \) is defined to be the dimension of the vector space \( A \). As one would expect of a coset representative, the displacement vector \( b \) is not unique: For any \( a \in A \), we have \( A + b = A + (a + b) \).

We may always choose an orthonormal basis for \( A \) so that \( A \in V(k, n) \) and we may always choose \( b \) to be orthogonal to \( A \) so that \( A^Tb = 0 \). Hence we may always represent \( A + b \in \text{Graff}(k, n) \) by a matrix \([A, b_0] \in \mathbb{R}^{n \times (k+1)} \) where \( A^TA = I \) and \( A^Tb_0 = 0 \); in this case we call \([A, b_0] \) **orthogonal affine coordinates**. Our distance between two \( k \)-flats in Section 11 will be expressed in terms of orthogonal affine coordinates for simplicity. However, a departure from the case of Grassmannian (of linear subspaces) is that we will not insist on using orthogonal affine coordinates all the time as

\(^5\)We were tempted to call this the affine Grassmannian as in [37] [44] but noticed that this term is now used far more commonly to refer to another very different object.
they can be unnecessarily restrictive (especially in proofs). Without these orthogonality conditions, a matrix \([A,b_0] \in \mathbb{R}^{n \times (k+1)}\) that represents an affine subspace \(A + b\) in the sense of (10.1) is called its affine coordinates.

**Proposition 10.2.** \(\text{Graff}(k,n)\) is a smooth manifold.

**Proof.** The proof is similar to showing that \(\text{Gr}(k,n)\) is a smooth manifold. Let \(A + b \in \text{Graff}(k,n)\) be represented by affine coordinates \([A,b_0] = [a_1,a_2,\ldots,a_k,b_0] \in \mathbb{R}^{n \times (k+1)}\), where \(b_0\) is chosen so that \(b - b_0 \in A\). We will show that there is a local chart around \(A + b\) with smooth transition functions. Let \(U\) be the set of all \(X + y \in \text{Graff}(k,n)\) whose affine coordinates \([X,y_0]\) have nonzero \(k \times k\) leading principal minors. Then \(U\) is an open subset of \(\text{Graff}(k,n)\) containing \(A + b\). Each \(X + y \in U\) has unique affine coordinates \([\hat{X},\hat{y}] \in \mathbb{R}^{n \times (k+1)}\) of the form

\[
[\hat{X},\hat{y}] = \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
\hat{x}_{k+1,1} & \hat{x}_{k+1,2} & \cdots & \hat{x}_{k+1,k} & \hat{y}_{k+1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\hat{x}_{n,1} & \hat{x}_{n,2} & \cdots & \hat{x}_{n,k} & \hat{y}_n
\end{bmatrix}.
\]

It is routine to verify that the map

\[
\varphi : U \to \mathbb{R}^{(n-k)(k+1)}, \quad X + y \mapsto [\hat{X},\hat{y}],
\]

is a homeomorphism and thus gives a local chart for \(U\). One may likewise define local charts for open subsets defined by the nonvanishing of other \(k \times k\) minors. It is also routine to verify that the transition functions \(\varphi_1 \circ \varphi_2^{-1}\) are smooth for any two such local charts \(\varphi_i : U_i \to \mathbb{R}^{(n-k)(k+1)}\), \(i = 1,2\).

It turns out that \(\text{Graff}(k,n)\) may be viewed as a vector bundle over \(\text{Gr}(k,n)\). Recall that if \(S\) is a subbundle of the vector bundle \(E\), then \(Q\) is called the *quotient bundle* of \(E\) by \(S\) if there is a short exact sequence of vector bundles

\[
0 \to S \to E \to Q \to 0. \tag{10.2}
\]

In the context of Grassmannians, there is a special vector bundle over \(\text{Gr}(k,n)\), called the *tautological bundle*, whose fiber over \(A \in \text{Gr}(k,n)\) is simply \(A\) itself. One may view this as a subbundle of the *trivial vector bundle* \(\text{Gr}(k,n) \times \mathbb{R}^n\). If \(S\) is the tautological bundle and \(E\) is the trivial bundle in (10.2), then the quotient bundle \(Q\) is called the *universal quotient bundle* of \(\text{Gr}(k,n)\) \([24,43]\).

**Proposition 10.3.** \(\text{Graff}(k,n)\) is the universal quotient bundle on \(\text{Gr}(k,n)\).

**Proof.** There is a natural map

\[
p : \text{Graff}(k,n) \to \text{Gr}(k,n), \quad A + b \mapsto A,
\]

that translates an affine space back to the origin. In terms of affine coordinates,

\[
p([a_1,\ldots,a_k,b_0]) = [a_1,\ldots,a_k]
\]

where \(a_i\)'s and \(b_0\) are chosen as in the proof of Proposition 10.2. Notice that the fiber \(p^{-1}(A)\) for \(A \in \text{Gr}(k,n)\) is simply \(\mathbb{R}^n/A\), a linear subspace of dimension \(n - k\). Local trivializations of \(\text{Graff}(k,n)\) are obtained from local charts of \(\text{Gr}(k,n)\) by construction. Hence \(\text{Graff}(k,n)\) is a vector bundle over \(\text{Gr}(k,n)\). Moreover we have

\[
q : \text{Gr}(k,n) \times \mathbb{R}^n \to \text{Graff}(k,n), \quad (A,b) \mapsto A + b.
\]
It is straightforward to check that \( q \) is a surjective bundle map and the kernel of \( q \) is the tautological vector bundle \( S \) over \( \text{Gr}(k, n) \), i.e., we have an exact sequence

\[
0 \to S \to \text{Gr}(k, n) \times \mathbb{R}^n \to \text{Graff}(k, n) \to 0.
\]

This shows that \( \text{Graff}(k, n) \) is the universal quotient bundle. \( \square \)

By either Proposition \( \text{[10.2]} \) or Proposition \( \text{[10.3]} \) we can see that

\[
\dim \text{Graff}(k, n) = (n - k)(k + 1).
\]

Unlike \( \text{Gr}(k, n) \), \( \text{Graff}(k, n) \) is a non-compact manifold. To see this, simply take a sequence of points in \( \text{Graff}(k, n) \) represented in coordinates by \( [A, mb] \) with \( m \in \mathbb{N}, A = [a_1, \ldots, a_k] \in V(k, n), \) and \( 0 \neq b \in \mathbb{R}^n \) such that \( A^T b = 0 \); observe that it has no convergent subsequence.

The group of orthogonal affine transformations is denoted \( E(n) \) and is the set \( O(n) \times \mathbb{R}^n \) equipped with the group operation \((Q_1, c_1)(Q_2, c_2) = (Q_1 Q_2, c_1 + Q_1 c_2)\). The affine Stiefel manifold is defined to be the product manifold \( \text{Vaff}(k, n) := V(k, n) \times \mathbb{R}^n \). It is a homogeneous space because of the following analogue of (2.1),

\[
\text{Vaff}(k, n) \cong E(n)/O(k).
\]

**Proposition 10.4.** \( \text{Graff}(k, n) \) is a reductive homogeneous Riemannian manifold and is geodesically complete. In fact, we have the following analogue of (2.2),

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

**Proof.** Since \( \text{Graff}(k, n) \) can be identified with an open subset of \( \text{Gr}(k + 1, n + 1) \), the Riemannian metric \( g_e \) on \( \text{Gr}(k + 1, n + 1) \) induces a metric on \( \text{Graff}(k, n) \). With this induced metric equipped, \( \text{Graff}(k, n) \) is a Riemannian manifold. The group \( E(n) \) acts on \( \text{Graff}(k, n) \) by

\[
(Q, c) \cdot (A + b) = Q \cdot A + Qb + c,
\]

where \( (Q, c) \in E(n) = O(n) \times \mathbb{R}^n, A + b \in \text{Graff}(k, n), \) and \( Q \cdot A := \text{span}(QA) \). It is easy to see that \( E(n) \) acts on \( \text{Graff}(k, n) \) transitively hence

\[
\text{Graff}(k, n) \cong E(n)/\text{Stab}_{A+b}(E(n)),
\]

where \( \text{Stab}_{A+b}(E(n)) \) is the stabilizer of any fixed affine linear subspace \( A + b \in \text{Graff}(k, n) \) in \( E(n) \). Now \( \text{Stab}_{A+b}(E(n)) \) consists of two types of actions. The first action is the affine action inside the plane \( A \), which is \( E(k) \), while the second action is the rotation around the orthogonal complement of \( A \), which is \( O(n - k) \). Hence we obtain

\[
\text{Stab}_{A+b}(E(n)) \cong O(n - k) \times E(k),
\]

and the representation of \( \text{Graff}(k, n) \) as a homogeneous Riemannian manifold follows. A reductive homogeneous Riemannian manifold is always geodesically complete \( \text{[35]} \).

We now turn to the algebraic geometric aspects of \( \text{Graff}(k, n) \). At this point it helps to remind ourselves that one of our main goals in studying the Grassmannian of affine subspaces is to define a notion of distance between two affine subspaces. We will see in Section \( \text{[11]} \) that it is the view of \( \text{Graff}(k, n) \) as an Zariski open dense subset of \( \text{Gr}(k + 1, n + 1) \) that will prove most useful in this regard. Our construction of this embedding is illustrated in Figure \( \text{[2]} \).

**Proposition 10.5.**

(i) \( \text{Graff}(k, n) \) is an algebraic variety that is irreducible and nonsingular.

(ii) \( \text{Graff}(k, n) \) may be embedded as a Zariski open subset of \( \text{Gr}(k + 1, n + 1) \),

\[
j : \text{Graff}(k, n) \to \text{Gr}(k + 1, n + 1), \quad A + b \mapsto \text{span}(A \cup \{b + e_{n+1}\}),
\]

where \( e_{n+1} = (0, \ldots, 0, 1)^T \in \mathbb{R}^{n+1} \). The image is open and dense in both the Zariski and manifold topologies.

---

\(^6\)In terms of semidirect product, \( E(n) = O(n) \ltimes_{\vartheta} \mathbb{R}^n \) where \( \vartheta : O(n) \to \text{Aut}(\mathbb{R}^n) = \text{GL}(n) \) is the inclusion.
Here our linear subspace $A$ is the $x$-axis. It is displaced by $b$ along the $y$-axis to the affine subspace $A + b$. The embedding $j : \text{Graff}(k, n) \to \text{Gr}(k+1, n+1)$ takes $A + b$ to the smallest 2-plane containing $A$ and $b + e_3$, where $e_3$ is a unit vector along the $z$-axis.

(iii) $\text{Gr}(k + 1, n + 1)$ may be regarded as the disjoint union of $\text{Gr}(k + 1, n)$ and $\text{Graff}(k, n)$,

$$\text{Gr}(k + 1, n + 1) = X \cup X^c, \quad X \cong \text{Graff}(k, n), \quad X^c \cong \text{Gr}(k + 1, n).$$

Proof. Substituting ‘smooth’ with ‘regular’ and ‘differential manifold’ by ‘algebraic variety’ in the proof of Proposition 10.2, we see that $\text{Graff}(k, n)$ is a nonsingular algebraic variety. Its irreducibility follows from Proposition 10.3 since $\text{Gr}(k, n)$ is irreducible and all fibers of $\text{Graff}(k, n) \to \text{Gr}(k, n)$ are irreducible and of the same dimension. Note that the term ‘algebraic variety’ is used here in the sense of an abstract algebraic variety, i.e., $\text{Graff}(k, n)$ is obtained by gluing together affine open subsets.

The embedding $j$ takes $k$-flats in $\mathbb{R}^n$ to $(k + 1)$-planes in $\mathbb{R}^{n+1}$,

$$\mathbb{R}^n \supseteq A + b \mapsto \text{span}(A \cup \{b + e_{n+1}\}) \subseteq \mathbb{R}^{n+1}.$$ 

In particular, $\mathbb{R}^n$ is mapped onto the subspace $E_n := \text{span}\{e_1, \ldots, e_n\} \subseteq \mathbb{R}^{n+1}$ where $e_1, \ldots, e_n, e_{n+1}$ are the standard basis vectors of $\mathbb{R}^{n+1}$. Linear subspaces $A \subseteq \mathbb{R}^n$ are then mapped to $j(A) \subseteq E_n$. Clearly $j$ is an embedding.

We set $X := j(\text{Graff}(k, n)) \subseteq \text{Gr}(k + 1, n + 1)$ and set $X^c$ to be the set-theoretic complement of $X$ in $\text{Gr}(k + 1, n + 1)$. By (ii), $X \cong \text{Graff}(k, n)$. By the definition of $X^c$, a $(k + 1)$-plane $B \in \text{Gr}(k + 1, n + 1)$ is in $X^c$ if and only if $B \subseteq E_n$, which is to say that

$$X^c = \text{Gr}_{k+1}(E_n) \cong \text{Gr}(k + 1, n).$$

Lastly we see that $X$ is Zariski open because its complement $X^c$, comprising $(k + 1)$-planes in $E_n$, is clearly Zariski closed.

11. Distance between affine subspaces

The differential geometric perspective taken in [2, 19, 54] leads one to an explicit expression for the geodesic distance on $\text{Gr}(k, n)$, and thus a natural notion of distance between linear subspaces. One might imagine that a notion of distance between affine subspaces could be similarly obtained.
from the differential geometric structures on Graff($k, n$) established in Propositions 10.2, 10.3 and 10.4. Surprisingly this is not the case.

By Proposition 10.4 Graff($k, n$) is geodesically complete; by the Hopf–Rinow Theorem [45], this implies that any two points on Graff($k, n$) can be connected by a distance minimizing geodesic. Once we have determined this geodesic, we may in principle compute the geodesic distance. In practice, however, it will be difficult to obtain an explicit expression for the geodesic without having additional structures on the manifold. Geodesic completeness only guarantees existence of such a geodesic but does not offer any clues in finding one.

How does one obtain an explicit expression for the geodesic and therefore the geodesic distance on the Stiefel manifold and Grassmannian? A careful examination of the arguments in [2, 19, 53] would reveal that the answer depends on a somewhat obscure structure on V($k, n$) and Gr($k, n$), namely, that of a geodesic orbit space [4, 23]. In general, if $G$ is a compact semisimple Lie group and $G/H$ is a reductive homogeneous space then there is a standard metric induced by the restriction of the Killing form on $g/h$ where $g$ and $h$ are the Lie algebras of $G$ and $H$ respectively. With this standard metric, $G/H$ is a geodesic orbit space, i.e., all geodesics are orbits of one parameter subgroups of $G$. In the case of $Gr(k, n) = O(n)/(O(n−k) × O(k))$ and Stiefel manifold $V(k, n) = O(n)/O(n−k)$, $O(n)$ is a compact semisimple Lie group and the Riemannian metrics we use on $Gr(k, n)$ and $V(k, n)$ are indeed the standard metrics. Hence they are geodesic orbit spaces. Moreover, for matrix Lie groups like $O(n)$, we know that all their one parameter subgroups are given by the exponential maps. These observations allow us to write down geodesics on $Gr(k, n)$ and $V(k, n)$ explicitly.

In seeking an expression for the geodesic distance between affine subspaces, it might appear that we could just apply the same arguments to Graff($k, n$), given that Proposition 10.4 guarantees the existence of a distance minimizing geodesic between any two $k$-dimensional affine subspaces in $\mathbb{R}^n$. The difficulty in this situation is that Graff($k, n$) = $E(n)/(E(n−k) × O(k))$ might not be a geodesic orbit space since $E(n) = O(n) × \mathbb{R}^n$ is not compact and therefore does not have a standard metric on Graff($k, n$) as in the case of Gr($k, n$) and V($k, n$).

What about the vector bundle structure on Graff($k, n$)? If $E$ is a vector bundle over a Riemannian manifold $M$, then there is always a metric induced on $E$ by the metric on $M$, namely, the pullback of the metric on $M$. Nevertheless, this metric on $E$ is evidently not very interesting — by definition, it disregards the fibers of the bundle. In the context of Proposition 10.3, this is akin to defining the distance between $A + b$ and $C + d$ ∈ Graff($k, n$) as the usual Grassmann distance between $A$ and $B$ ∈ Gr($k, n$), which ignores $b$ and $d$ totally.

In summary, the differential geometric structures on Graff($k, n$) established in Propositions 10.2, 10.3 and 10.4 do not really help us define a distance between two affine subspaces. We will instead turn to the algebraic geometric properties of Graff($k, n$) in Proposition 10.5 to provide the framework for defining such a distance. We will first describe the distance between two equidimensional affine subspaces and then extend it to affine subspaces of different dimensions along the same lines as in Section 9. The proofs of the results in this section are similar to those of their linear counterparts and are thus omitted.

By Proposition 10.5(ii), we may identify Graff($k, n$) with its image $j(Graff(k, n))$ in Gr($k, n$). As a Zariski open subset of Gr($k+1, n+1$), Graff($k, n$) inherits the Grassmann distance $d_{Gr(k+1, n+1)}$ on Gr($k+1, n+1$) and we obtain the following.

**Theorem 11.1.** For any two affine $k$-flats $A + b$ and $B + c$ ∈ Graff($k, n$), defining

\[
d_{\text{Graff}(k,n)}(A + b, B + c) := d_{\text{Gr}(k+1, n+1)}(j(A + b), j(B + c)),
\]

yields a notion of distance consistent with the Grassmann distance. If $[A, b_0]$ and $[B, c_0] ∈ \mathbb{R}^{n×(k+1)}$ are the orthogonal affine coordinates of the respective $k$-flats, then the analogue of (2.7) is given by

\[
d_{\text{Graff}(k,n)}(A + b, B + c) = \left(\sum_{i=1}^{k+1} \phi_i^2\right)^{1/2},
\]
where \( \phi_i = \cos^{-1} \tau_i \) and \( \tau_1 \geq \cdots \geq \tau_{k+1} \) are the singular values of the matrix
\[
\begin{bmatrix}
A & b_0 / \sqrt{1 + \|b_0\|^2} \\
0 & 1 / \sqrt{1 + \|b_0\|^2}
\end{bmatrix}
\begin{bmatrix}
B & c_0 / \sqrt{1 + \|c_0\|^2} \\
0 & 1 / \sqrt{1 + \|c_0\|^2}
\end{bmatrix}
\in \mathbb{R}^{(k+1) \times (k+1)},
\]
where \( \| \cdot \| \) denotes the 2-norm on \( \mathbb{R}^n \).

It is easy to see that \( \phi_1, \ldots, \phi_{k+1} \) are independent of our choice of orthogonal affine coordinates and we shall call \( \phi_i(A + b, B + c) \) the \( i \)th affine principal angles between the respective affine \( k \)-flats.

Following the arguments in Section 3 we may define the infinite Grassmannian of affine subspaces,
\[
\text{Graff}(k, \infty) := \lim_{n \to \infty} \text{Graff}(k, n) = \bigcup_{n=k}^{\infty} \text{Graff}(k, n),
\]
by taking direct limit of the directed system given by the natural inclusions \( \iota_n : \text{Graff}(k, n) \to \text{Graff}(k, n+1) \) for \( n \geq k \). For the same reasons given in the proof of Lemma 3.1 we obtain its analogue.

**Lemma 11.2.** The value \( d_{\text{Graff}(k, n)}(A + b, B + c) \) of two \( k \)-flats \( A + b \) and \( B + c \in \text{Graff}(k, n) \) is independent of \( n \), the dimension of their ambient space. Consequently, \( d_{\text{Graff}(k, n)} \) induces a distance \( d'_{\text{Graff}(k, \infty)} \) on \( \text{Graff}(k, \infty) \).

Following the arguments in Section 4 we obtain a notion of distance for affine subspaces of different dimensions. Let \( A + b \in \text{Graff}(k, n) \) and \( B + c \in \text{Graff}(l, n) \) where \( k \leq l \leq n \). We define the affine Schubert varieties of \( l \)-flats containing \( A + b \) and \( k \)-flats contained in \( B + c \) respectively as
\[
\Omega_+ (A + b) := \{ X + y \in \text{Graff}(l, n) : A + b \subseteq X + y \},
\]
\[
\Omega_- (B + c) := \{ Y + z \in \text{Graff}(k, n) : Y + z \subseteq B + c \}.
\]
We have the following analogue of Theorem 4.2.

**Theorem 11.3.** Let \( k \leq l \leq n \). For any \( A + b \in \text{Graff}(k, n) \) and \( B + c \in \text{Graff}(l, n) \), the following distances are equal,
\[
d_{\text{Graff}(k, n)}(A + b, \Omega_-(B + c)) = d_{\text{Graff}(l, n)}(B + c, \Omega_+(A + b)), \tag{11.1}
\]
and their common value \( \delta(A + b, B + c) \) may be computed explicitly as
\[
\delta(A + b, B + c) = \left( \sum_{i=1}^{\min(k, l)+1} \phi_i(A + b, B + c)^2 \right)^{1/2}. \tag{11.2}
\]

The affine principal angles \( \phi_1, \ldots, \phi_{\min(k, l)+1} \) are as defined in Theorem 11.1 except that this time they correspond to the singular values of a rectangular matrix
\[
\begin{bmatrix}
A & b_0 / \sqrt{1 + \|b_0\|^2} \\
0 & 1 / \sqrt{1 + \|b_0\|^2}
\end{bmatrix}
\begin{bmatrix}
B & c_0 / \sqrt{1 + \|c_0\|^2} \\
0 & 1 / \sqrt{1 + \|c_0\|^2}
\end{bmatrix}
\in \mathbb{R}^{(k+1) \times (l+1)}.
\]
Like its counterpart for linear subspaces, our \( \delta \) here defines a notion of distance between the respective affine subspaces in the sense of a distance of a point to a set. It is clearly reduces to the usual Grassmann distance \( d_{\text{Graff}(k, n)} \) when \( b = c = 0 \) and \( \dim A = \dim B = k \).

Another advantage of relying on an embedding of \( \text{Graff}(k, n) \) into \( \text{Gr}(k+1, n+1) \) for our definition of distance between affine subspaces is that \( \text{Graff}(k, n) \) automatically inherits the other distances on \( \text{Gr}(k+1, n+1) \) and we have an analogue Theorem 4.7.

**Theorem 11.4.** Let \( k \leq l \leq n \). Let \( A + b \in \text{Graff}(k, n) \) and \( B + c \in \text{Graff}(l, n) \). Then
\[
d_{\text{Graff}(k, n)}(A + b, \Omega_-(B + c)) = d_{\text{Graff}(l, n)}^*(B + c, \Omega_+(A + b)), \quad * = \alpha, \beta, \kappa, \mu, \pi, \rho, \sigma, \phi.
\]
Their common value $\delta^*(A + b, B + c)$ is given by:

$$
\begin{align*}
\delta^\alpha(A + b, B + c) &= \phi_{k+1}, \\
\delta^\sigma(A + b, B + c) &= \sin \phi_{k+1}, \\
\delta^\varphi(A + b, B + c) &= 2\sin(\phi_{k+1}/2), \\
\delta^\rho(A + b, B + c) &= \left(\frac{1}{1 - \prod_{i=1}^{k+1} \cos^2 \phi_i}\right)^{1/2}, \\
\delta^\delta(A + b, B + c) &= \left(\prod_{i=1}^{k+1} \frac{1}{\cos^2 \phi_i}\right)^{1/2}, \\
\delta^\theta(A + b, B + c) &= \left(\sum_{i=1}^{k+1} \sin^2 \phi_i\right)^{1/2}, \\
\delta^\mu(A + b, B + c) &= \left(\sum_{i=1}^{k+1} \sin^2(\phi_i/2)\right)^{1/2}, \\
\end{align*}
$$

where $\phi_1, \ldots, \phi_{k+1}$ are as defined above.

We have said relatively little about the properties of the two affine Schubert varieties $\Omega_+(A + b)$ and $\Omega_-(B + c)$. But this is intentional because the affine analogue of Proposition 8.3 holds: $\Omega_+(A + b)$ may be viewed a Grassmannian of linear subspaces while $\Omega_-(B + c)$ may be viewed as a Grassmannian of affine subspaces. In particular they inherit the rich geometric properties of $\text{Gr}(k, n)$ and $\text{Graff}(k, n)$ discussed earlier.

**Proposition 11.5.** Let $A + b \in \text{Graff}(k, n)$ and $B + c \in \text{Graff}(l, n)$. Then

$$
\Omega_+(A + b) \cong \text{Gr}(n-l, n-k) \quad\text{and}\quad \Omega_-(B + c) \cong \text{Graff}(k, l)
$$

as Riemannian manifolds or algebraic varieties.

**Proof.** We first observe that the map

$$
\varphi : \Omega_+(A + b) \to \Omega_+(A),
$$

$$
X + y \mapsto X + y - b,
$$

is well-defined since $A \subset X + y - b$ by our choice of $X + y$. Also,

$$
\psi : \Omega_+(A) \to \Omega_+(A + b),
$$

$$
X \mapsto X + b,
$$

is the inverse of $\varphi$ and so it is an isomorphism. Together with Proposition 8.3 we obtain the first isomorphism

$$
\Omega_+(A + b) \cong \Omega_+(A) \cong \text{Gr}(n-l, n-k).
$$

For the second isomorphism, consider

$$
\varphi' : \Omega_-(B + c) \to \text{Graff}_k(B),
$$

$$
Y + z \mapsto Y + z - c,
$$

which is well-defined since $Y + z - c$ is an affine subspace of dimension $k$ in $B$. Its inverse is given by

$$
\psi' : \text{Graff}_k(B) \to \Omega_-(B + c)
$$

$$
Y + z \mapsto Y + z + c,
$$

and so it is an isomorphism. The required isomorphism then follows from

$$
\Omega_-(B + c) \cong \text{Graff}_k(B) \cong \text{Graff}(k, l). \quad \square
$$

The reason for the asymmetry in Proposition 11.5 is as follows. $\Omega_+(A + b)$ is a Grassmannian of linear subspaces since all affine subspaces containing $A + b$ can be shifted back to the origin by the vector $b$. In the case of $\Omega_-(B + c)$, shifting $B + c$ back to the origin by $c$ and then taking all affine subspaces contained in $B$ still gives a Grassmannian of affine subspaces.
12. 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 \cite[Example 1.2.20]{44} of $\text{Gr}(k, n)$ as a set of actual matrices, namely, the set of idempotent symmetric matrices of trace $k$:

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

(12.1)

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 and affine Grassmannian discussed in earlier sections in this alternate form.

The isomorphism in (12.1) 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 iff it is symmetric and idempotent; the first two equalities in (12.1) ensure that $P$ is as such. Note also that the eigenvalues of an orthogonal projection onto a subspace of dimension $k$ are 1’s with multiplicity $k$ and 0’s with multiplicity $n-k$; so the condition tr$(P) = k$ is equivalent to 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 (12.1),

$$\Omega_+(A) \cong \{ P \in \mathbb{R}^{n \times n} : P^T = P^2 = P, \ \text{tr}(P) = l, \ \text{im}(A) \subseteq \text{im}(P) \},$$

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

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

There is also an affine analogue of (12.1) for $\text{Graff}(k, n)$. Given $A + b \in \text{Graff}(k, n)$ with orthogonal affine coordinates $[A, b_0] \in \mathbb{R}^{n \times (k+1)}$, we associate the unique matrix $[P_A \ b_0] \in \mathbb{R}^{(n+1) \times (n+1)}$. We obtain

$$\text{Graff}(k, n) \cong \{ [P \ b] : Pb = 0, \ P^T = P^2 = P, \ \text{tr}(P) = k \}.$$

Corresponding to such a representation, we get

$$\Omega_+(A + b) \cong \{ [P \ b] \in \mathbb{R}^{(n+1) \times (n+1)} : Pb = 0, \ P^T = P^2 = P, \ \text{tr}(P) = l, \ \text{im}(A) \subseteq \text{im}(P) \},$$

$$\Omega_-(B + c) \cong \{ [P \ c] \in \mathbb{R}^{(n+1) \times (n+1)} : Pc = 0, \ P^T = P^2 = P, \ \text{tr}(P) = k, \ \text{im}(P) \subseteq \text{im}(B) \}.$$

Observe that this allows us to simultaneously embed $\text{Graff}(k, n)$, $\text{Graff}(l, n)$, $\Omega_+(A + b)$, $\Omega_-(B + c)$ into $\mathbb{R}^{(n+1) \times (n+1)}$.

13. Probability density on the Grassmannian

In this section, we determine the relative volumes of the Schubert varieties $\Omega_+(A)$, $\Omega_-(B)$ and their affine analogues $\Omega_+(A+b)$, $\Omega_-(B+c)$. We prove volumetric analogues of (4.4) in Theorem 12.2 and (11.11) in Theorem 11.3. We will see that 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 $\mathbb{R}^n$ 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 \cite[Example 3.4.2]{44}. In the case of the Grassmannian, the Riemannian metric\footnote{The Riemannian metric on $\text{Gr}(k, n)$ has been discussed at length in \cite{2, 19}. Since we have no use for it except implicitly, we did not specify it in this article.} on $\text{Gr}(k, n)$ that gave rise to
the Grassmann distance in \((2.7)\) and the geodesic in \((2.8)\) 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 [44, Proposition 9.1.12] (see also [42, pp. 48–53]) below for easy reference.

**Proposition 13.1.** The volume of \(\text{Gr}(k, n)\) is
\[
\text{Vol}(\text{Gr}(k, n)) = \left(\frac{n}{k}\right) \frac{\prod_{j=1}^{n} \omega_j}{\left(\prod_{j=1}^{k} \omega_j\right) \left(\prod_{j=1}^{n-k} \omega_j\right)},
\]
where \(\omega_m := \pi^{m/2}/\Gamma(1 + m/2)\) is the volume of the unit 2-norm ball in \(\mathbb{R}^m\).

Let \(d\mu_{k,n} := \text{Vol}(\text{Gr}(k, n))^{-1}|d\gamma_{k,n}|\) be the corresponding probability density on \(\text{Gr}(k, n)\). We show that 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 particularly important in Bayesian inference of subspaces [48] — the implication being that it does not matter whether we work with \(A\) and \(\Omega_+(A)\) in \(\text{Gr}(k, n)\) or with \(B\) and \(\Omega_-(B)\) within \(\text{Gr}(l, n)\).

**Corollary 13.2.** 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
\[
\frac{l!(n-k)! \prod_{j=l-k+1}^{l} \omega_j}{n!(l-k)! \prod_{j=n-k+1}^{n} \omega_j}.
\]

**Proof.** By Proposition [3.3] \(\Omega_+(A)\) is isometric to \(\text{Gr}(n-l, n-k)\) and so we have
\[
\text{Vol}(\Omega_+(A)) = \left(\frac{n-k}{n-l}\right) \frac{\prod_{j=1}^{n-k} \omega_j}{\left(\prod_{j=1}^{l} \omega_j\right) \left(\prod_{j=1}^{n-l} \omega_j\right)}
\]
by Proposition [13.1] Likewise, \(\Omega_-(B)\) is isometric to \(\text{Gr}(k, l)\) and so
\[
\text{Vol}(\Omega_-(B)) = \left(\frac{l}{k}\right) \frac{\prod_{j=1}^{l} \omega_j}{\left(\prod_{j=1}^{k} \omega_j\right) \left(\prod_{j=1}^{l-k} \omega_j\right)}.
\]
Dividing by the volumes of \(\text{Gr}(l, n)\) and \(\text{Gr}(k, n)\) respectively yields the required results. \(\square\)

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 [12](i).

By Proposition [10.5](ii), since \(\text{Graff}(k, n)\) is open and dense in \(\text{Gr}(k+1, n+1)\), we must have \(\mu_{k+1,n+1}(\text{Graff}(k, n)) = 1\) and therefore the restriction of \(\mu_{k+1,n+1}\) to \(\text{Graff}(k, n)\) defines a probability measure on \(\text{Graff}(k, n)\).

**Corollary 13.3.** Let \(k \leq l \leq n\). Let \(A + b \in \text{Graff}(k, n)\) and \(B + c \in \text{Graff}(l, n)\). The relative volume of \(\Omega_+(A + b)\) in \(\text{Graff}(l, n)\) and \(\Omega_-(B + c)\) in \(\text{Graff}(k, n)\) are identical,
\[
\mu_{l+1,n+1}(\Omega_+(A + b)) = \mu_{k+1,n+1}(\Omega_-(B + c)).
\]
Their common value does not depend on the choices of \(A + b\) and \(B + c\) but only on \(k, l, n\) and is given by
\[
\frac{(l+1)!(n-k)! \prod_{j=l-k+1}^{l+1} \omega_j}{(n+1)!(l-k)! \prod_{j=n-k+1}^{n+1} \omega_j}.
\]
Proof. By Proposition 10.5(ii), we have
\[
\text{Vol}(\text{Graff}(k,n)) = \text{Vol}(\text{Gr}(k+1,n+1)) = \left(\frac{n+1}{k+1}\right) \frac{\prod_{j=1}^{n+1} \omega_j}{\prod_{j=1}^{k+1} \omega_j \prod_{j=1}^{n-k} \omega_j}.
\]
By Proposition 11.5, we have
\[
\text{Vol}(\Omega_+(A+b)) = \text{Vol}(\text{Gr}(n-l,n-k)) = \left(\frac{n-k}{n-l}\right) \frac{\prod_{j=1}^{n-k} \omega_j}{\prod_{j=1}^{n-l} \omega_j \prod_{j=1}^{n-k} \omega_j},
\]
\[
\text{Vol}(\Omega_-(B+c)) = \text{Vol}(\text{Graff}(k,l)) = \left(\frac{l+1}{k+1}\right) \frac{\prod_{j=1}^{l+1} \omega_j}{\prod_{j=1}^{k+1} \omega_j \prod_{j=1}^{l-k} \omega_j}.
\]
The rest follows as in the proof of Corollary 13.2. □

14. 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 non-equidimensional and affine subspaces.

In the course of our investigations, we developed simple geometric models for linear subspaces of all dimensions and for affine subspaces. 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, Giorgio Ottaviani for suggesting that we look at the geometric mean problem, and Frank Sottile for invaluable discussions that led to the Schubert variety approach. This work is partially supported by AFOSR FA9550-13-1-0133, NSF DMS 1209136, and NSF DMS 1057064.

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.
[4] D. Alekseevsky and A. Arvanitoyeorgos, “Riemannian flag manifolds with homogeneous geodesics,” Trans. Amer. Math. Soc., 359 (2007), no. 8, pp. 5689–5714.
[5] A. Ashikhmin and A. R. Calderbank, “Grassmannian packings from operator Reed–Muller codes,” IEEE Trans. Inform. Theory, 56 (2003), no. 11, pp. 5689–5714.
[6] H. Bagherinia and R. Manduchi, “A theory of color barcodes,” Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 14 (2011), pp. 806–813.
[7] 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.
[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] R. Bhatia, “On the exponential metric increasing property,” Linear Algebra Appl., 375 (2003), pp. 211–220.
[10] R. Bhatia and J. Holbrook, “Riemannian geometry and matrix geometric means,” Linear Algebra Appl., 413 (2006), no. 2–3, pp. 594–618.
[11] Å. Björck and G. H. Golub, “Numerical methods for computing angles between linear subspaces,” Math. Comp., 27 (1973), no. 123, pp. 579–594.
[12] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, NY, 1998.
[13] L. Blum, M. Shub, and S. Smale, “On a theory of computation and complexity over the real numbers,” *Bull. Amer. Math. Soc.*, 21 (1989), no. 1, pp. 1–46.
[14] 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.
[15] 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.
[16] M. M. Deza and E. Deza, *Encyclopedia of Distances*, 2nd Ed., Springer, Heidelberg, 2013.
[17] 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.
[18] J. Draisma, E. Horobet, G. Ottaviani, B. Sturmfels, and R. Thomas, “The Euclidean distance degree of an algebraic variety,” *Found. Comput. Math.*, to appear.
[19] 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.
[20] 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.
[21] R. Fioresi and C. Hacon, “On infinite-dimensional Grassmannians and their quantum deformations,” *Rend. Sem. Mat. Univ. Padova*, 111 (2004), pp. 1–24.
[22] G. Golub and C. Van Loan, *Matrix Computations*, 4th Ed., John Hopkins University Press, Baltimore, MD, 2013.
[23] C. S. Gordon, “Homogeneous Riemannian manifolds whose geodesics are orbits,” pp. 155–174, S. Gindikin, Ed., *Topics in Geometry, Progress in Nonlinear Differential Equations and their Applications*, 20, Birkhäuser, Boston, MA, 1996.
[24] P. Griffiths and J. Harris, *Principles of Algebraic Geometry*, John Wiley, New York, NY, 1994.
[25] 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.
[26] A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2002.
[27] 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.
[28] A. Ya. Helemskii, *Lectures and Exercises on Functional Analysis*, Translations of Mathematical Monographs, 233, AMS, Providence, RI, 2006.
[29] 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.
[30] 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. 370–383.
[31] C. J. Hillar and L.-H. Lim, “Most tensor problems are NP-hard,” *J. ACM*, 60 (2013), no. 6, Art. 45, 39 pp.
[32] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[33] K. Hupé, U. Helmke and S. Herzberg, “On the computation of means on Grassmann manifolds,” *Proc. Int. Symp. Math. Theory Networks Syst. (MTNS)*, 19 (2010), pp. 2439–2441.
[34] D. Husemoller, *Fibre Bundles*, 3rd Ed., Graduate Texts in Mathematics, 20, Springer, New York, NY, 1994.
[35] J. Jost, *Riemannian Geometry and Geometric Analysis*, 6th Ed., Springer, Heidelberg, 2011.
[36] T. Kato, *Perturbation Theory for Linear Operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995.
[37] D. A. Klain and G.-C. Rota, *Introduction to Geometric Probability*, Lezioni Lincee, Cambridge University Press, Cambridge, 1997.
[38] 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.
[39] 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.
[40] 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.
[41] 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.
[42] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces*, Cambridge Studies in Advanced Mathematics, 44, Cambridge University Press, Cambridge, 1995.
[43] J. W. Milnor and J. D. Stasheff, *Characteristic Classes*, Annals of Mathematics Studies, 76, Princeton University Press, Princeton, NJ, 1974.
[44] L. I. Nicolaescu, *Lectures on the Geometry of Manifolds*, 2nd Ed., World Scientific, Hackensack, NJ, 2007.
[45] G. Ottaviani, P.-J. Spaenlehauer, and B. Sturmfels, “Exact solutions in structured low-rank approximation,” preprint, (2013), http://arxiv.org/abs/1311.2376
[46] M. Peng, D. Bu, and Y. Wang, “The measure of income mobility in vector space,” Physics Procedia, 3 (2010), no. 5, pp. 1725–1732.
[47] 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.
[48] B. St. Thomas, L. Lin, L.-H. Lim, and S. Mukherjee, “Learning subspaces of different dimensions,” preprint, (2014), http://arxiv.org/abs/1404.6841
[49] G. W. Stewart and J. Sun, Matrix Perturbation Theory, Academic Press, Boston, MA, 1990.
[50] X. Sun, L. Wang, and J. Feng, “Further results on the subspace distance,” Pattern Recognition, 40 (2007), no. 1, pp. 328–329.
[51] R. Vidal, Y. Ma, and S. Sastry, “Generalized principal component analysis,” IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), no. 12, pp. 1945–1959.
[52] 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.
[53] 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.
[54] Y.-C. Wong, “Differential geometry of Grassmann manifolds,” Proc. Nat. Acad. Sci., 57 (1967), no. 3, pp. 589–594.
[55] 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.
[56] 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 60637-1514.
E-mail address: kye@math.uchicago.edu

COMPUTATIONAL AND APPLIED MATHEMATICS INITIATIVE, DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.
E-mail address, corresponding author: lekheng@galton.uchicago.edu