



The Grassmannian of affine subspaces

Lek-Heng Lim¹ · Ken Sze-Wai Wong² · Ke Ye³

Received: 24 August 2018 / Revised: 9 March 2020 / Accepted: 19 March 2020 /
Published online: 27 May 2020
© SFoCM 2020

Abstract

The Grassmannian of affine subspaces is a natural generalization of both the Euclidean space, points being 0-dimensional affine subspaces, and the usual Grassmannian, linear subspaces being special cases of affine subspaces. We show that, like the Grassmannian, the affine Grassmannian has rich geometrical and topological properties: It has the structure of a homogeneous space, a differential manifold, an algebraic variety, a vector bundle, a classifying space, among many more structures; furthermore, it affords an analogue of Schubert calculus and its (co)homology and homotopy groups may be readily determined. On the other hand, like the Euclidean space, the affine Grassmannian serves as a concrete computational platform on which various distances, metrics, probability densities may be explicitly defined and computed via numerical linear algebra. Moreover, many standard problems in machine learning and statistics—linear regression, errors-in-variables regression, principal components analysis, support vector machines, or more generally any problem that seeks linear relations among variables that either best represent them or separate them into components—may be naturally formulated as problems on the affine Grassmannian.

Keywords Affine Grassmannian · Affine subspaces · Schubert calculus · homotopy and (co)homology · Probability densities · Distances and metrics · Multivariate data analysis

Mathematics Subject Classification 14M15 · 22F30 · 46T12 · 53C30 · 57R22 · 62H10

1 Introduction

The Grassmannian of affine subspaces, denoted $\text{Graff}(k, n)$, is an analogue of the usual Grassmannian $\text{Gr}(k, n)$. Just as $\text{Gr}(k, n)$ parameterizes k -dimensional *linear* subspaces in \mathbb{R}^n , $\text{Graff}(k, n)$ parameterizes k -dimensional *affine* subspaces in \mathbb{R}^n , i.e.,

Communicated by Alan Edelman.

Extended author information available on the last page of the article

$\mathbb{A} + b$ where the k -dimensional linear subspace $\mathbb{A} \subseteq \mathbb{R}^n$ is translated by a displacement vector $b \in \mathbb{R}^n$.

To the best of our knowledge, the Grassmannian of affine subspaces was first described in an elegant little volume [21] based on Gian-Carlo Rota's 1986 'Lezioni Lincee' lectures at the Scuola Normale Superiore. The treatment in [21, pp. 86–87] was somewhat cursory as $\text{Graff}(k, n)$ played only an auxiliary role in Rota's lectures (on geometric probability). Aside from another equally brief mention in [29, Section 9.1.3], we are unaware of any other discussion. Compared to its universally known cousin $\text{Gr}(k, n)$, it is fair to say that $\text{Graff}(k, n)$ has received next to no attention. The goal of our article is to fill this gap. We will show that the Grassmannian of affine subspaces has rich algebraic, geometric, and topological properties; moreover, it is an important object that could rival the usual Grassmannian in practical applicability, serving as a computational and modeling platform for problems in statistical estimation and pattern recognition. We start by showing that $\text{Graff}(k, n)$ may be viewed from several perspectives, and in more than a dozen ways:

ALGEBRA: as collections of (i) Minkowski sums of sets, (ii) cosets in an additive group, (iii) $n \times (k + 1)$ matrices;

DIFFERENTIAL GEOMETRY: as a (iv) smooth manifold, (v) homogeneous space, (vi) Riemannian manifold, (vii) base space of the compact and noncompact affine Stiefel manifolds regarded as principal bundles;

ALGEBRAIC GEOMETRY: as a (viii) irreducible nonsingular algebraic variety, (ix) Zariski open dense subset of the Grassmannian, (x) real affine variety of projection matrices;

ALGEBRAIC TOPOLOGY: as a (xi) vector bundle, (xii) classifying space.

$\text{Graff}(k, n)$ may also be regarded, in an appropriate sense, as the complement of $\text{Gr}(k + 1, n)$ in $\text{Gr}(k + 1, n + 1)$, or, in a different sense, as the moduli space of k -dimensional affine subspaces in \mathbb{R}^n . Moreover one may readily define, calculate, and compute various objects on $\text{Graff}(k, n)$ of either theoretical or practical interests:

SCHUBERT CALCULUS: affine (a) flags, (b) Schubert varieties, (c) Schubert cycles;

ALGEBRAIC TOPOLOGY: (d) homotopy, (e) homology, (f) cohomology;

METRIC GEOMETRY: (g) distances, (h) geodesic, (i) metrics;

PROBABILITY: (j) uniform, (k) von Mises–Fisher, (l) Langevin–Gaussian distributions.

The main reason for our optimism that $\text{Graff}(k, n)$ may be no less important than $\text{Gr}(k, n)$ in applications is the observation that common problems in multivariate data analysis and machine learning are naturally optimization problems over $\text{Graff}(k, n)$:

STATISTICS: (1) linear regression, (2) error-in-variables regression, (3) principal component analysis, (4) support vector machines.

In retrospect this is no surprise, many statistical estimation problems involve a search for linear relations among variables and are therefore ultimately a problem of finding one or more affine subspaces that either best represent a given data set (regression) or best separate it into two or more components (classification).

In a companion article [26], we showed that in practical terms, optimization problems over $\text{Graff}(k, n)$ are no different from optimization problems over \mathbb{R}^n , which is

of course just $\text{Graff}(0, n)$. More precisely, we showed that, like the Euclidean space \mathbb{R}^n , $\text{Graff}(k, n)$ serves the role of a concrete computational platform on which tangent spaces, Riemannian metric, exponential maps, parallel transports, gradients and Hessians of real-valued functions, optimization algorithms such as steepest descent, conjugate gradient, Newton methods, may all be efficiently computed using only standard numerical linear algebra.

For brevity, we will use the term *affine Grassmannian* when referring to the Grassmannian of affine subspaces from this point onwards. The term is now used far more commonly to refer to another very different object¹ [3,13,24], but in this article, it will always be used in the sense of Definition 1. To resolve the conflicting nomenclature, an alternative might be to christen the Grassmannian of affine subspaces the *Rota Grassmannian*.

Unless otherwise noted, the results in this article have not appeared before elsewhere to the best of our knowledge, although some of them are certainly routine for the experts. It is inevitable that there is some slight overlap with our companion article [26], which shares notations and terminologies; but aside from a small number of direct quotes that are clearly labeled as such, any related results are always stated in different light and given different proofs so that each provides its own value. We have written our article in the hope that it would be read by applied and computational mathematicians, statisticians, and engineers—in an effort to improve its accessibility, we have provided more basic details than is customary.

2 Basic terminologies

We remind the reader of some basic terminologies. We will always work over \mathbb{R} and our ambient vector space is \mathbb{R}^n unless specified otherwise. We adopt the convention that all vectors in \mathbb{R}^n are regarded as column vectors. Row vectors will be denoted with a transpose, i.e., as x^T where x is a column vector. When enclosed by parentheses, a vector denoted $(x_1, \dots, x_n) \in \mathbb{R}^n$ would mean a *column* vector with entries $x_1, \dots, x_n \in \mathbb{R}$. We let I_n denote the $n \times n$ identity matrix.

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, \dots, a_k are the basis vectors. A *flag* is a strictly increasing sequence of nested linear subspaces, $\mathbb{A}_0 \subseteq \mathbb{A}_1 \subseteq \mathbb{A}_2 \subseteq \dots$. A flag is said to be *complete* if $\dim \mathbb{A}_k = k$, *finite* if $k = 0, 1, \dots, n$, and *infinite* if $k \in \mathbb{N} \cup \{0\}$. Throughout this article, a blackboard bold letter \mathbb{A} will always denote a subspace and the corresponding normal letter A will then denote a matrix whose column vectors a_1, \dots, a_k form a basis (often but not necessarily orthonormal) of \mathbb{A} . We write $\text{im}(A)$ for the image of a matrix $A \in \mathbb{R}^{n \times k}$ and $\text{span } S$ for the linear span of a set $S \subseteq \mathbb{R}^n$. So $\mathbb{A} = \text{im}(A) = \text{span}\{a_1, \dots, a_k\}$.

We write $\text{Gr}(k, n)$ for the *Grassmannian* of k -planes in \mathbb{R}^n , $\text{V}(k, n)$ for the *Stiefel manifold* of orthonormal k -frames, and $\text{O}(n) := \text{V}(n, n)$ for the *orthogonal group*.

¹ In certain areas of algebraic geometry and representation theory, notably Langland’s program, the term ‘affine Grassmannian’ widely refers to a functor associated with an algebraic group, which is completely unrelated to the sense in which it is used in this article.

We may regard $V(k, n)$ as a homogeneous space,

$$V(k, n) \cong O(n)/O(n - k), \tag{2.1}$$

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 $Gr(k, n)$, i.e.,

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

So $\mathbb{A} \in Gr(k, n)$ may be identified with the equivalence class of its orthonormal k -frames $\{AQ \in V(k, n) : Q \in O(k)\}$. Note that $\text{im}(AQ) = \text{im}(A)$ for $Q \in O(k)$. Readers unfamiliar with these notions may refer to [1,12] for a very accessible introduction.

There is also a purely algebraic counterpart to the last paragraph, useful for generalizing to k -planes in a vector space that may not have an inner product (e.g., over fields of nonzero characteristics). We follow the terminologies and notations in [2, Section 2]. The *noncompact Stiefel manifold* of k -frames, denoted $St(k, n)$, may either be regarded as the manifold of $n \times k$ matrices with full rank or as the homogeneous space

$$St(k, n) \cong GL(n)/R(n - k), \tag{2.3}$$

where

$$R(n - k) := \left\{ \begin{bmatrix} I_k & A \\ 0 & B \end{bmatrix} \in GL(n) : A \in \mathbb{R}^{k \times (n-k)}, B \in GL(n - k) \right\}.$$

There is a right action of the general linear group $GL(k)$ on $St(k, n)$: For $X \in GL(k)$ and $A \in St(k, n)$, the action yields $AX \in St(k, n)$ and the resulting homogeneous space is $Gr(k, n)$, i.e.,

$$Gr(k, n) \cong St(k, n)/GL(k) \cong GL(n)/(R(n - k) \times GL(k)). \tag{2.4}$$

So $\mathbb{A} \in Gr(k, n)$ may be identified with the equivalence class of its k -frames $\{AX \in St(k, n) : X \in GL(k)\}$. Note that $\text{im}(AX) = \text{im}(A)$ for $X \in GL(k)$. The reader would see that orthogonality has been avoided in this paragraph.

For easy reference, we summarize the notations introduced in Sects. 2, 3, and 4 in Table 1.

3 Algebra of the affine Grassmannian

We will begin by discussing the set-theoretic and algebraic properties of the affine Grassmannian and introducing its two infinite-dimensional counterparts. We start with a formal definition of our main object of study, the (finite-dimensional) affine Grassmannian, using the same notations as in [26].

Table 1 List of Lie groups and homogeneous spaces that appear in this article

	<i>Object</i>	<i>Dimension</i>	<i>Quotient</i>
$O(n)$	Orthogonal group	$\binom{n}{2}$	
$E(n)$	Euclidean group	$\binom{n+1}{2}$	
$GL(n)$	General linear group	n^2	
$GA(n)$	General affine group	$n(n + 1)$	
$R(n - k)$	$\left\{ \begin{bmatrix} I_k & * \\ 0 & * \end{bmatrix} \in GL(n) \right\}$	$n(n - k)$	
$V(k, n)$	Stiefel manifold	$k(2n - k - 1)/2$	$O(n)/O(n - k)$
$St(k, n)$	Noncompact Stiefel manifold	kn	$GL(n)/R(n - k)$
$Vaff(k, n)$	Affine Stiefel manifold	$(k + 1)(2n - k)/2$	$E(n)/O(n - k)$
$Staff(k, n)$	Noncompact affine Stiefel manifold	$(k + 1)n$	$GA(n)/R(n - k)$
$Gr(k, n)$	Grassmannian	$k(n - k)$	$V(k, n)/O(k)$
$Graff(k, n)$	Affine Grassmannian	$(k + 1)(n - k)$	$Vaff(k, n)/E(k)$

Definition 1 (*Affine Grassmannian*) Let $k \leq n$ be positive integers. The *Grassmannian of k -dimensional affine subspaces* in \mathbb{R}^n or *Grassmannian of k -flats* in \mathbb{R}^n , denoted by $Graff(k, n)$, is the set of all k -dimensional affine subspaces of \mathbb{R}^n . For an abstract vector space \mathbb{V} , we write $Graff_k(\mathbb{V})$ for the set of k -flats in \mathbb{V} .

This set-theoretic definition hardly reveals anything about the rich algebra, geometry, and topology of the affine Grassmannian, which we will examine over this and the next few sections.

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

$$\mathbb{A} + b := \{ \lambda_1 a_1 + \dots + \lambda_k a_k + b \in \mathbb{R}^n : \lambda_1, \dots, \lambda_k \in \mathbb{R} \}. \tag{3.1}$$

The notation $\mathbb{A} + b$ may be taken to mean (i) the *Minkowski sum* of the sets \mathbb{A} and $\{b\}$ in the Euclidean space \mathbb{R}^n , (ii) a *coset* of the subgroup \mathbb{A} in the additive group \mathbb{R}^n , or (iii) a coset of the subspace \mathbb{A} in the vector space \mathbb{R}^n . The dimension of $\mathbb{A} + b$ is defined to be the dimension of the vector space \mathbb{A} . As one would expect of a coset representative, the displacement vector b is not unique: For any $a \in \mathbb{A}$, we have $\mathbb{A} + b = \mathbb{A} + (a + b)$. We introduce a simple map that will be important later: the *deaffine map*

$$\tau : Graff(k, n) \rightarrow Gr(k, n), \quad \mathbb{A} + b \mapsto \mathbb{A} \tag{3.2}$$

takes any affine subspace to its corresponding linear subspace.

Let $\mathbb{A} + b \in Graff(k, n)$. By our notational convention, $\text{im}(A) = \mathbb{A}$ and therefore the matrix $[A, b] \in \mathbb{R}^{n \times (k+1)}$ determines the affine subspace $\mathbb{A} + b$ and we will call this its *affine coordinates*. If in addition, we have $A \in V(k, n)$, i.e., an orthonormal basis for \mathbb{A} , and we choose $b_0 \in \mathbb{R}^n$ to be orthogonal to \mathbb{A} with $\text{im}(A) + b_0 = \mathbb{A} + b$, then we call $[A, b_0] \in V(k, n) \times \mathbb{R}^n$ an *orthogonal affine coordinates* of $\mathbb{A} + b$. Note that $A^T A = I_k$,

$A^\top b_0 = 0$, and two orthogonal affine coordinates $[A, b_0], [A', b'_0] \in \mathbb{R}^{n \times (k+1)}$ of the same affine subspace $\mathbb{A} + b$ must have that $A' = A Q$ for some $Q \in O(k)$ and $b'_0 = b_0$.

We will also need to discuss the cases where $k = \infty$ and $n = \infty$ as they will be important in Sects. 7 and 8. For each $k \in \mathbb{N}$, the infinite flag $\{0\} \subseteq \mathbb{R} \subseteq \mathbb{R}^2 \subseteq \dots$ induces a directed system

$$\dots \subseteq \text{Graff}(k, n) \subseteq \text{Graff}(k, n + 1) \subseteq \dots, \tag{3.3}$$

and taking direct limit gives

$$\text{Graff}(k, \infty) := \varinjlim \text{Graff}(k, n),$$

which we will call the *infinite Grassmannian of k -dimensional affine linear subspaces* or *infinite affine Grassmannian* for short. This parameterizes k -dimensional flats in \mathbb{R}^n for all $n \geq k$ and is the affine analogue of the infinite or Sato Grassmannian $\text{Gr}(k, \infty)$ [32].

To be more precise, the direct limit above is taken in the directed system given by the natural inclusions $\iota_n : \text{Graff}(k, n) \rightarrow \text{Graff}(k, n + 1)$ for $n \geq k$. If $\mathbb{A} + b \in \text{Graff}(k, n)$ has affine coordinates $[A, b] \in \mathbb{R}^{n \times (k+1)}$, then $\iota_n(\mathbb{A} + b) = \mathbb{A}' + b'$ where $\mathbb{A}' = \text{im} \begin{bmatrix} A \\ 0 \end{bmatrix}, b' = \begin{bmatrix} b \\ 0 \end{bmatrix}$, i.e., $\mathbb{A}' + b' \in \text{Graff}(k, n + 1)$ has affine coordinates $\begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+1) \times (k+1)}$. Readers unfamiliar with direct limits may simply identify $[A, b]$ with $\begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix}$ and thereby regard

$$\text{Graff}(k, n) \subseteq \text{Graff}(k, n + 1) \quad \text{and} \quad \text{Graff}(k, \infty) = \bigcup_{n=k}^\infty \text{Graff}(k, n).$$

It is straightforward to verify that the deaffine map $\tau : \text{Graff}(k, n) \rightarrow \text{Gr}(k, n)$ is compatible with the directed systems $\{\text{Graff}(k, n)\}_{n=k}^\infty$ and $\{\text{Gr}(k, n)\}_{n=k}^\infty$, i.e., the following diagram commutes:

$$\begin{array}{ccccccc} \dots & \hookrightarrow & \text{Graff}(k, n) & \xrightarrow{\iota_n} & \text{Graff}(k, n + 1) & \hookrightarrow & \dots \\ \downarrow \tau & & \downarrow \tau & & \downarrow \tau & & \downarrow \tau \\ \dots & \hookrightarrow & \text{Gr}(k, n) & \hookrightarrow & \text{Gr}(k, n + 1) & \hookrightarrow & \dots \end{array} \tag{3.4}$$

Note that one advantage afforded by $\text{Graff}(k, \infty)$ is that one may discuss a k -dimensional affine subspace without reference to an ambient space (although strictly speaking, points in $\text{Graff}(k, \infty)$ are k -flats in $\mathbb{R}^\infty := \varinjlim \mathbb{R}^n$). The *doubly infinite affine Grassmannian*, which parameterizes affine subspaces of all dimensions, may then be defined as the disjoint union

$$\text{Graff}(\infty, \infty) := \coprod_{k=1}^\infty \text{Graff}(k, \infty).$$

This is the affine analogue of $\text{Gr}(\infty, \infty)$, the doubly infinite Grassmannian of linear subspaces of all dimensions, defined in [41, Section 5].

For the affine Grassmannian, two groups will play the roles that $O(n)$ and $GL(n)$ play for the Grassmannian in Sect. 2. We defer the discussion to Sect. 4 but will introduce the relevant algebra here. The *group of orthogonal affine transformations* or *orthogonal affine group*, denoted $E(n)$, is the set $O(n) \times \mathbb{R}^n$ endowed with group operation

$$(Q_1, c_1)(Q_2, c_2) = (Q_1 Q_2, c_1 + Q_1 c_2).$$

In other words, it is a semidirect product: $E(n) = O(n) \ltimes_{\vartheta} \mathbb{R}^n$ where $\vartheta : O(n) \rightarrow \text{Aut}(\mathbb{R}^n) = GL(n)$ as inclusion. The *group of affine transformations* or *general affine group*, denoted $GA(n)$, is the set $GL(n) \times \mathbb{R}^n$ endowed with group operation

$$(X_1, c_1)(X_2, c_2) = (X_1 X_2, c_1 + X_1 c_2).$$

In other words, it is a semidirect product: $GA(n) = GL(n) \ltimes_{\iota} \mathbb{R}^n$ where $\iota : GL(n) \rightarrow \text{Aut}(\mathbb{R}^n) = GL(n)$ is the identity map. $GA(n)$ acts on \mathbb{R}^n naturally via

$$(X, c) \cdot v = Xv + c, \quad (X, c) \in GA(n), \quad v \in \mathbb{R}^n.$$

Clearly $E(n)$ is a subgroup of $GA(n)$ and therefore inherits this group action. We note that $E(n)$ has wide-ranging applications in engineering [11].

4 Differential geometry of the affine Grassmannian

The affine Grassmannian has rich geometric properties. We start by showing that it is a noncompact smooth manifold and then show that it is also homogeneous and Riemannian.

Before we begin, we note that while our description in Definition 1 is purely set-theoretic, $\text{Graff}(k, n)$ inherits a topology from $\text{Gr}(k, n)$ as follows: An open neighborhood of a point $\mathbb{X} + y \in \text{Graff}(k, n)$ is defined to be $U + y$ where $U \subseteq \text{Gr}(k, n)$ is an open neighborhood of the point $\mathbb{X} \in \text{Gr}(k, n)$; this system of open neighborhoods generates a topology on $\text{Graff}(k, n)$.

Proposition 1 *$\text{Graff}(k, n)$ is a noncompact smooth manifold with*

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

Proof Let $\mathbb{A} + b \in \text{Graff}(k, n)$ be represented by affine coordinates $[A, b_0] = [a_1, a_2, \dots, a_k, b_0] \in \mathbb{R}^{n \times (k+1)}$, where b_0 is chosen so that $b - b_0 \in \mathbb{A}$. We will show that there is a local chart around $\mathbb{A} + b$ with smooth transition functions. We may assume that $A \in \mathbb{R}^{n \times k}$, which has rank k , has a nonzero $k \times k$ leading principal minor. Indeed, any $T \in GL(n)$ determines an automorphism $\mathbb{X} + y \mapsto T\mathbb{X} + Ty$ on $\text{Graff}(k, n)$, and we may choose T such that the $k \times k$ leading principal minor of TA is nonzero. If U is a local chart for $T\mathbb{A} + Tb$, then $T^{-1}U$ is a local chart for $\mathbb{A} + b$. Let U be the set of all $\mathbb{X} + y \in \text{Graff}(k, n)$ 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 $\mathbb{A} + b$. Each $\mathbb{X} + y \in U$ 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 $\varphi : U \rightarrow \mathbb{R}^{(n-k)(k+1)}$, $\mathbb{X} + y \mapsto [\hat{X}, \hat{y}]$, is a homeomorphism and thus gives a local chart for U . We may likewise define other local charts by the nonvanishing of other $k \times k$ minors and verify that the transition functions $\varphi_1 \circ \varphi_2^{-1}$ are smooth for any two such local charts $\varphi_i : U_i \rightarrow \mathbb{R}^{(n-k)(k+1)}$, $i = 1, 2$. To see the noncompactness, take a sequence in $\text{Graff}(k, n)$ represented in orthogonal affine coordinates by $[A, mb]$ with $m \in \mathbb{N}$, $A = [a_1, \dots, a_k] \in V(k, n)$, and $0 \neq b \in \mathbb{R}^n$ such that $A^T b = 0$; observe that it has no convergent subsequence. \square

The manifold structure in Proposition 1 is identical to the one obtained in [26, Theorem 2.2]—the local chart (U, φ) around $\mathbb{X} + y \in \text{Graff}(k, n)$ constructed above is the preimage of a local chart (V, ψ) around $\mathbb{Y} = j(\mathbb{X} + y) \in \text{Gr}(k + 1, n + 1)$, where j is the embedding in (5.1).

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),

$$\text{Vaff}(k, n) \cong E(n) / O(n - k)$$

where $E(n)$ is the orthogonal affine group $E(n)$ introduced at the end of Sect. 3. We have the following characterizations of $\text{Graff}(k, n)$ as quotients of $E(n)$.

Proposition 2 *Graff(k, n) is a homogeneous Riemannian manifold. In fact, we have the following analogue of (2.2),*

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

Furthermore, $\text{Vaff}(k, n)$ is a principal $E(k)$ -bundle over $\text{Graff}(k, n)$.

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 $\text{Graff}(k, n)$. Equipped with this induced metric, $\text{Graff}(k, n)$ is a Riemannian manifold. The group $E(n)$ acts on $\text{Graff}(k, n)$ by $(Q, c) \cdot (\mathbb{A} + b) = Q \cdot \mathbb{A} + Qb + c$, where $(Q, c) \in E(n) = O(n) \times \mathbb{R}^n$, $\mathbb{A} + b \in \text{Graff}(k, n)$, and $Q \cdot \mathbb{A} := \text{im}(QA)$. It is easy to see that $E(n)$ acts on $\text{Graff}(k, n)$ transitively and so $\text{Graff}(k, n) \cong E(n) / \text{Stab}_{\mathbb{A}+b}(E(n))$, where $\text{Stab}_{\mathbb{A}+b}(E(n))$ is the stabilizer of any fixed affine linear subspace $\mathbb{A} + b \in \text{Graff}(k, n)$ in $E(n)$. Now $\text{Stab}_{\mathbb{A}+b}(E(n))$ consists of two types of actions. The first action is the affine action inside the plane \mathbb{A} , which is $E(k)$, while the second action is the rotation around the orthogonal complement of \mathbb{A} , which is $O(n - k)$. Hence

we obtain $\text{Stab}_{\mathbb{A}+b}(\mathbb{E}(n)) \cong \text{O}(n - k) \times \mathbb{E}(k)$, and the representation of $\text{Graff}(k, n)$ as a homogeneous Riemannian manifold follows. Clearly, $\text{Vaff}(k, n)$ is a principal $\mathbb{E}(k)$ -bundle over $\text{Vaff}(k, n)/\mathbb{E}(k) \cong \text{Graff}(k, n)$. \square

Let $\tau_v : \text{Vaff}(k, n) = \mathbb{V}(k, n) \times \mathbb{R}^n \rightarrow \mathbb{V}(k, n)$ be the projection. For any $k \leq n$, τ_v commutes with the deaffine map τ in (3.2):

$$\begin{CD} \text{Vaff}(k, n) @>\tau_v>> \mathbb{V}(k, n) \\ @V\pi_aVV @VV\pi V \\ \text{Graff}(k, n) @>\tau>> \text{Gr}(k, n) \end{CD} \tag{4.1}$$

where we view $\text{Graff}(k, n)$, $\text{Gr}(k, n)$, $\text{Vaff}(k, n)$, $\mathbb{V}(k, n)$ as homogeneous spaces. One may define $\mathbb{V}(k, \infty)$, the Stiefel manifold of orthogonal k -frames in \mathbb{R}^∞ , as the direct limit of the inclusions $\iota_n : \mathbb{V}(k, n) \rightarrow \mathbb{V}(k, n + 1)$, $Q \mapsto \begin{bmatrix} Q \\ 0 \end{bmatrix}$, and its affine counterpart as $\text{Vaff}(k, \infty) := \mathbb{V}(k, \infty) \times \mathbb{R}^\infty$, the *infinite affine Stiefel manifold*. Taking direct limit of (4.1), we obtain

$$\begin{CD} \text{Vaff}(k, \infty) @>\tau_v>> \mathbb{V}(k, \infty) \\ @V\pi_aVV @VV\pi V \\ \text{Graff}(k, \infty) @>\tau>> \text{Gr}(k, \infty) \end{CD} \tag{4.2}$$

The objects in (4.2) are all Hilbert manifolds although we will not use this fact.

From a computational perspective, one would prefer to work with orthogonal objects like $\mathbb{V}(k, n)$ and $\text{O}(k)$ rather than affine objects like $\text{Vaff}(k, n)$ and $\mathbb{E}(k)$. Roughly speaking, this is largely because orthogonal transformations preserve norm and do not magnify rounding errors during computations. With this in mind, we will seek to characterize the affine Grassmannian as an orbit space of the orthogonal group in a Stiefel manifold.

Let $\mathbb{A}+b \in \text{Graff}(k, n)$. Its orthogonal affine coordinates are $[A, b_0] \in \mathbb{V}(k, n) \times \mathbb{R}^n$ where $A^T b_0 = 0$, i.e., b_0 is orthogonal to the columns of A . However, as b_0 is in general not of unit norm, we may not regard $[A, b_0]$ as an element of $\mathbb{V}(k+1, n)$. The following variant² is a convenient system of coordinates for computations [26] and for defining various distances on $\text{Graff}(k, n)$ in Sect. 8.

Definition 2 Let $\mathbb{A}+b \in \text{Graff}(k, n)$ and $[A, b_0] \in \mathbb{R}^{n \times (k+1)}$ be its orthogonal affine coordinates, i.e., $A^T A = I_k$ and $A^T b_0 = 0$. The matrix of *Stiefel coordinates* for $\mathbb{A}+b$ is the $(n+1) \times (k+1)$ matrix with orthonormal columns,

$$Y_{\mathbb{A}+b} := \begin{bmatrix} A & b_0/\sqrt{1 + \|b_0\|^2} \\ 0 & 1/\sqrt{1 + \|b_0\|^2} \end{bmatrix} \in \mathbb{V}(k+1, n+1).$$

Two orthogonal affine coordinates $[A, b_0], [A', b'_0]$ of $\mathbb{A}+b$ give two corresponding matrices of Stiefel coordinates $Y_{\mathbb{A}+b}, Y'_{\mathbb{A}+b}$. By the remark after our definition of

² Definition 2 has appeared in [26, Definition 3.1]. We reproduce it here for the reader’s easy reference.

orthogonal affine coordinates, $A = A'Q'$ for some $Q' \in O(k)$ and $b_0 = b'_0$. Hence

$$Y_{\mathbb{A}+b} = \begin{bmatrix} A & b_0/\sqrt{1 + \|b_0\|^2} \\ 0 & 1/\sqrt{1 + \|b_0\|^2} \end{bmatrix} = \begin{bmatrix} A' & b'_0/\sqrt{1 + \|b'_0\|^2} \\ 0 & 1/\sqrt{1 + \|b'_0\|^2} \end{bmatrix} \begin{bmatrix} Q' & 0 \\ 0 & 1 \end{bmatrix} = Y'_{\mathbb{A}+b} Q \quad (4.3)$$

where $Q := \begin{bmatrix} Q' & 0 \\ 0 & 1 \end{bmatrix} \in O(k + 1)$. Hence two different matrices of Stiefel coordinates for the same affine subspace differ by an orthogonal transformation.

There is also an affine counterpart to the last paragraph of Sect. 2 that allows us to provide an analogue of Proposition 2 without reference to orthogonality, useful for studying the affine Grassmannian over a vector space without an inner product. The *noncompact affine Stiefel manifold* $\text{Staff}(k, n)$ may be defined in several ways:

$$\text{Staff}(k, n) = \text{GA}(n)/\text{R}(n - k) = (\text{GL}(n)/\text{R}(n - k)) \times \mathbb{R}^n = \text{St}(k, n) \times \mathbb{R}^n,$$

where $\text{GA}(n)$ is the general affine group in Sect. 3 and $\text{St}(k, n)$ the noncompact Stiefel manifold in Sect. 2.

Proposition 3

(i) *Dimensions of the compact and noncompact affine Stiefel manifolds are*

$$\dim \text{Vaff}(k, n) = \frac{1}{2}(2n - k)(k + 1), \quad \dim \text{Staff}(k, n) = n(k + 1).$$

(ii) *Whether as topological spaces, differential manifolds, or algebraic varieties, we have*

$$\text{Graff}(k, n) \cong \text{Staff}(k, n)/\text{GA}(k) \cong \text{GA}(n)/(\text{R}(n - k) \times \text{GA}(k)),$$

i.e., the isomorphism is a homeomorphism, diffeomorphism, and biregular map.

(iii) *Staff(k, n) is a principal GA(k)-bundle over Graff(k, n).*

Proof The inclusion $E(n) \hookrightarrow \text{GA}(n)$ as a subgroup naturally induces the commutative diagram:

$$\begin{array}{ccc} E(n) & \hookrightarrow & \text{GA}(n) \\ \downarrow & & \downarrow \\ \text{Vaff}(k, n) = E(n)/O(n - k) & \hookrightarrow & \text{Staff}(k, n) = \text{GA}(n)/\text{R}(n - k) \\ \downarrow \pi_a & & \downarrow \pi_s \\ \text{Graff}(k, n) = \text{Vaff}(k, n)/E(k) & \xrightarrow{j} & \text{Staff}(k, n)/\text{GA}(k) \\ \downarrow \tau & & \downarrow \tau_s \\ \text{Gr}(k, n) & \xrightarrow{\cong} & \text{GL}(n)/(\text{R}(n - k) \times \text{GL}(k)) \end{array} \quad (4.4)$$

where π_a and τ are as in (4.1), π_s is the quotient map, and τ_s is similarly defined as τ . The bottom isomorphism is (2.4), which is simultaneously an isomorphism of

topological spaces, differential manifolds, and algebraic varieties. (i) follows from the quotient space structures:

$$\begin{aligned} \dim \text{Vaff}(k, n) &= \dim E(n) - \dim O(n - k) = \left[\binom{n}{2} + n \right] - \binom{n - k}{2} \\ &= \frac{(k + 1)(2n - k)}{2}, \end{aligned}$$

$$\dim \text{Staff}(k, n) = \dim \text{GA}(n) - \dim \text{R}(n - k) = (n^2 + n) - n(n - k) = n(k + 1).$$

For (ii), it suffices to show that the restriction

$$j|_{\tau^{-1}(\mathbb{A})} : \tau^{-1}(\mathbb{A}) \rightarrow \tau_s^{-1}(\mathbb{A})$$

is an isomorphism for every $\mathbb{A} \in \text{Gr}(k, n)$, but this follows from the bottom isomorphism. (iii) follows from (ii) as $\text{Staff}(k, n)$ is a principal $\text{GA}(k)$ -bundle on $\text{Staff}(k, n)/\text{GA}(k) \cong \text{Graff}(k, n)$. □

5 Algebraic geometry of the affine Grassmannian

We now turn to the algebraic geometric aspects, characterizing $\text{Graff}(k, n)$ as (i) an irreducible nonsingular algebraic variety, (ii) a Zariski open dense subset of $\text{Gr}(k + 1, n + 1)$, and (iii) a real affine algebraic variety of projection matrices. In addition, just as $\text{Gr}(k, n)$ is a moduli space of k -dimensional linear subspaces in \mathbb{R}^n , $\text{Graff}(k, n)$ is a moduli space of k -dimensional affine subspaces in \mathbb{R}^n , although we have nothing to add beyond this observation. In Sect. 6, we will discuss affine Schubert varieties, an analogue of Schubert varieties, in $\text{Graff}(k, n)$.

That $\text{Graff}(k, n)$ may be regarded as a Zariski dense subset of $\text{Gr}(k + 1, n + 1)$ is a noteworthy point. It is the key to our optimization algorithms in [26]. Also, it immediately implies that any probability densities [10] defined on the usual Grassmannian may be adapted to the affine Grassmannian, a fact that we will rely on in Sect. 9.

Theorem 1

- (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 $\text{Gr}(k + 1, n + 1)$,

$$j : \text{Graff}(k, n) \rightarrow \text{Gr}(k + 1, n + 1), \quad \mathbb{A} + b \mapsto \text{im}(\mathbb{A} \cup \{b + e_{n+1}\}), \quad (5.1)$$

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

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

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

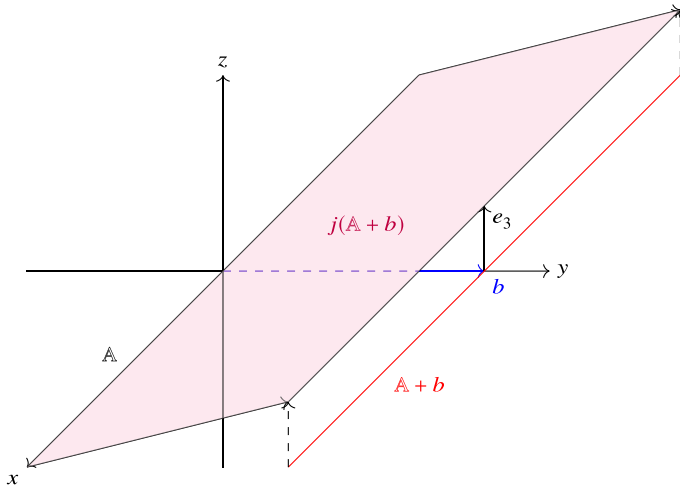


Fig. 1 Here our linear subspace \mathbb{A} is the x -axis. It is displaced by b along the y -axis to the affine subspace $\mathbb{A} + b$. The embedding $j : \text{Graff}(k, n) \rightarrow \text{Gr}(k + 1, n + 1)$ takes $\mathbb{A} + b$ to the smallest 2-plane containing \mathbb{A} and $b + e_3$, where e_3 is a unit vector along the z -axis

Proof Substituting ‘smooth’ with ‘regular’ and ‘differential manifold’ by ‘algebraic variety’ in the proof of Proposition 1, we see that $\text{Graff}(k, n)$ is a nonsingular algebraic variety. We use ‘algebraic variety’ 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} , i.e., $\mathbb{R}^n \supseteq \mathbb{A} + b \mapsto \text{span}(\mathbb{A} \cup \{b + e_{n+1}\}) \subseteq \mathbb{R}^{n+1}$. It maps \mathbb{R}^n onto $E_n := \text{span}\{e_1, \dots, e_n\} \subseteq \mathbb{R}^{n+1}$ where e_1, \dots, e_n, e_{n+1} are the standard basis vectors of \mathbb{R}^{n+1} . Linear subspaces $\mathbb{A} \subseteq \mathbb{R}^n$ are then mapped to $j(\mathbb{A}) \subseteq E_n$. Clearly j is an embedding, illustrated in Fig. 1 for the case $k = 1, n = 3$.

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 $\mathbb{B} \in \text{Gr}(k + 1, n + 1)$ is in X^c if and only if $\mathbb{B} \subseteq E_n$, which is to say that $X^c = \text{Gr}_{k+1}(E_n) \cong \text{Gr}(k + 1, n)$. We easily see that X is Zariski open because its complement X^c , comprising $(k + 1)$ -planes in E_n , is clearly Zariski closed. Since $\text{Gr}(k + 1, n + 1)$ is irreducible as an algebraic variety, its Zariski open subset X must also be irreducible, and thus so is $\text{Graff}(k, n)$. \square

Henceforth we will identify

$$\mathbb{R}^n \equiv \{(x_1, \dots, x_n, 0) \in \mathbb{R}^{n+1} : x_1, \dots, x_n \in \mathbb{R}\} \tag{5.2}$$

to obtain a complete flag

$$\{0\} \subseteq \mathbb{R}^1 \subseteq \mathbb{R}^2 \subseteq \dots \subseteq \mathbb{R}^n \subseteq \mathbb{R}^{n+1} \subseteq \dots,$$

which was essentially what we did in the proof of Theorem 1. With such an identification, our choice of e_{n+1} in the embedding j in (5.1) is the most natural one.

It is often desirable to uniquely represent elements of $\text{Graff}(k, n)$ as actual matrices instead of equivalence classes of matrices like the affine, orthogonal affine, and Stiefel coordinate representations in Sects. 3 and 4. For example, we will see that this is the case when we discuss probability distributions on $\text{Graff}(k, n)$ in Sect. 9. The Grassmannian has a well-known representation [29, Example 1.2.20] as the set of rank- k orthogonal projection³ matrices, or, equivalently, the set of trace- k idempotent symmetric matrices:

$$\text{Gr}(k, n) \cong \{P \in \mathbb{R}^{n \times n} : P^\top = P^2 = P, \text{tr}(P) = k\}. \tag{5.3}$$

Note that $\text{rank}(P) = \text{tr}(P)$ for an orthogonal projection matrix P . A straightforward affine analogue of (5.3) for $\text{Graff}(k, n)$ is the following.

Proposition 4 *Graff*(k, n) is a real affine algebraic variety given by

$$\text{Graff}(k, n) \cong \{[P, b] \in \mathbb{R}^{n \times (n+1)} : P^\top = P^2 = P, \text{tr}(P) = k, Pb = 0\}. \tag{5.4}$$

Proof Let $\mathbb{A} + b \in \text{Graff}(k, n)$ have orthogonal affine coordinates $[A, b_0] \in \mathbb{R}^{n \times (k+1)}$. Recall that if A is an orthonormal basis for the subspace \mathbb{A} , then AA^\top is the orthogonal projection onto \mathbb{A} . It is straightforward to check that the map $\mathbb{A} + b \mapsto [AA^\top, b_0]$ is independent of the choice of orthogonal affine coordinates and is bijective. \square

We will call the matrix $[P, b] \in \mathbb{R}^{n \times (n+1)}$ *projection affine coordinates* for $\mathbb{A} + b$. From a practical standpoint, we would like to represent points in $\text{Graff}(k, n)$ as orthogonal projection matrices; one reason is that such a coordinate system facilitates optimization algorithms on $\text{Graff}(k, n)$ (see [26]); another is that certain probability densities can be naturally expressed in such a coordinate system (see Sect. 9). Since $[P, b]$ is not an orthogonal projection matrix, we introduce the following variant.⁴

Definition 3 Let $\mathbb{A} + b \in \text{Graff}(k, n)$ and $[P, b] \in \mathbb{R}^{n \times (n+1)}$ be its projection affine coordinates. The matrix of *projection coordinates* for $\mathbb{A} + b$ is the orthogonal projection matrix

$$P_{\mathbb{A}+b} := \begin{bmatrix} P + bb^\top / (\|b\|^2 + 1) & b / (\|b\|^2 + 1) \\ b^\top / (\|b\|^2 + 1) & 1 / (\|b\|^2 + 1) \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}.$$

Alternatively, in terms of orthogonal affine coordinates $[A, b_0] \in \mathbb{R}^{n \times (k+1)}$,

$$P_{\mathbb{A}+b} = \begin{bmatrix} AA^\top + b_0 b_0^\top / (\|b_0\|^2 + 1) & b_0 / (\|b_0\|^2 + 1) \\ b_0^\top / (\|b_0\|^2 + 1) & 1 / (\|b_0\|^2 + 1) \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}.$$

It is easy to check that $P_{\mathbb{A}+b}$ is indeed an orthogonal projection matrix, i.e., $P_{\mathbb{A}+b}^2 = P_{\mathbb{A}+b} = P_{\mathbb{A}+b}^\top$. Unlike Stiefel coordinates, projection coordinates of a given affine subspace are unique.

³ A projection matrix satisfies $P^2 = P$ and an orthogonal projection matrix is in addition symmetric, i.e., $P^\top = P$. Despite its name, an orthogonal projection matrix P is not an orthogonal matrix unless P is an identity matrix.

⁴ Definition 3 has appeared in [26, Definition 3.4]. We reproduce it here for the reader’s easy reference.

6 Schubert calculus on the affine Grassmannian

We will show that basic aspects of Schubert calculus on the Grassmannian [22] could be readily extended to an “affine Schubert calculus” on the affine Grassmannian, with affine analogues of flags, Schubert varieties, Schubert cycles [33]. As is the case for (co)homology of the Grassmannian, the materials in this section will be important for our (co)homology calculations in Sect. 7.2; what is perhaps more surprising is that our study of distances between affine subspaces of different dimensions in Sect. 8.3 will also rely on affine Schubert varieties.

In this paragraph, we briefly review some basic terminologies and facts in Schubert calculus for the reader’s easy reference. The *Schubert variety* of a flag $\mathbb{A}_1 \subseteq \cdots \subseteq \mathbb{A}_k$ in \mathbb{R}^n is a subvariety of $\text{Gr}(k, n)$ defined by

$$\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k) := \{\mathbb{B} \in \text{Gr}(k, n) : \dim(\mathbb{B} \cap \mathbb{A}_j) \geq j, j = 1, \dots, k\}.$$

It is a standard fact [22, Proposition 4] that

$$\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k) \cong \Omega(\mathbb{B}_1, \dots, \mathbb{B}_k) \quad \text{if } \dim \mathbb{A}_j = \dim \mathbb{B}_j, j = 1, \dots, k. \quad (6.1)$$

So when the choice of the flag is unimportant, we may take it to be $\mathbb{R}^{d_1} \subseteq \cdots \subseteq \mathbb{R}^{d_k}$ and denote the corresponding Schubert variety by $\Omega(d_1, \dots, d_k)$. The ‘ \cong ’ in (6.1) may be taken to be either homeomorphism of topological spaces or biregular isomorphism of algebraic varieties, but it cannot in general be replaced by ‘=’ — two different flags of the same dimensions determine different varieties in $\text{Gr}(k, n)$. The following properties [9, 19, 28, 30] of Schubert varieties are also well known.

Facts 2

(i) *The dimension of a Schubert variety is given by*

$$\dim \Omega(d_1, \dots, d_k) = \sum_{j=1}^k d_j - \frac{1}{2}k(k + 1).$$

(ii) *The cycles determined by Schubert varieties of dimension i form a basis for the i th homology group $H_i(\text{Gr}(k, n), \mathbb{Z}_2)$ and cohomology group $H^i(\text{Gr}(k, n), \mathbb{Z}_2)$, which are isomorphic and*

$$H_i(\text{Gr}(k, n), \mathbb{Z}_2) \simeq H^i(\text{Gr}(k, n), \mathbb{Z}_2) \simeq \mathbb{Z}_2^{r_i},$$

where r_i is the number of Schubert varieties of dimension i , which equals the number of partitions of the integer i into at most k parts, each not exceeding $n - k$. Here $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$.

(iii) *The collection of Schubert varieties in $\text{Gr}(k, n)$ over all flags of length k in \mathbb{R}^n gives a cell decomposition⁵ for $\text{Gr}(k, n)$.*

We will now introduce an affine analogue of the Schubert variety in the affine Grassmannian using an *affine flag*, i.e., an increasing sequence of nested affine subspaces.

Definition 4 Let $\mathbb{A}_1 + b_1 \subseteq \dots \subseteq \mathbb{A}_k + b_k$ be an affine flag in \mathbb{R}^n . The corresponding *affine Schubert variety* is a subvariety of $\text{Graff}(k, n)$ defined by

$$\Psi(\mathbb{A}_1 + b_1, \dots, \mathbb{A}_k + b_k) := \{\mathbb{B} + c \in \text{Graff}(k, n) : \dim((\mathbb{B} + c) \cap (\mathbb{A}_j + b_j)) \geq j, j = 1, \dots, k\}.$$

We first show that the affine flag may always be chosen so that $b_1 = \dots = b_k$.

Lemma 1 *For any affine flag $\mathbb{A}_1 + b_1 \subseteq \dots \subseteq \mathbb{A}_k + b_k$, there exists a displacement vector $b \in \mathbb{R}^n$ such that $\mathbb{A}_j + b_j = \mathbb{A}_j + b, j = 1, \dots, k$. Thus every affine Schubert variety is of the form $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$.*

Proof Let $-b$ be any element in $\mathbb{A}_1 + b_1$. By definition, $-b \in \mathbb{A}_j + b_j, j = 1, \dots, k$. So $b_j \in \mathbb{A}_j + b$. Therefore $\mathbb{A}_j + b_j = \mathbb{A}_j + b, j = 1, \dots, k$. □

With Lemma 1, it is straightforward to derive an analogue of (6.1).

Proposition 5 *For any two affine flags $\mathbb{A}_1 + b \subseteq \dots \subseteq \mathbb{A}_k + b$ and $\mathbb{B}_1 + c \subseteq \dots \subseteq \mathbb{B}_k + c$ where $\dim \mathbb{A}_j = d_j = \dim \mathbb{B}_j, j = 1, \dots, k$, we have*

$$\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b) \cong \Psi(\mathbb{B}_1 + c, \dots, \mathbb{B}_k + c),$$

and so we may write $\Psi(d_1, \dots, d_k)$ when the specific affine flag is unimportant.

⁵ A cell decomposition of a topological space X is a partition of X into a disjoint union of open subsets $\{X_i\}_{i \in I}$ such that for each $i \in I$ there is a continuous map $f : B^{n_i} \rightarrow X$ from the unit closed ball B^{n_i} of dimension n_i to X satisfying (i) the restriction of f to the interior of B^{n_i} is a homeomorphism onto X_i ; and (ii) the image $f(\partial B^{n_i})$ is contained in the union of finitely many X_j 's with $\dim X_j < \dim X_i$.

Proof There is a general affine transformation $(X, y) \in \text{GA}(n)$ such that

$$(X, y) \cdot (\mathbb{A}_j + b) = X(\mathbb{A}_j) + Xb + y = \mathbb{B}_j + c, \quad j = 1, \dots, k,$$

i.e., $X(\mathbb{A}_j) = \mathbb{B}_j$ and $Xb + y = c$. The existence of $X \in \text{GL}(n)$ is guaranteed by the transitive action of $\text{GL}(n)$ on (linear) flags of fixed dimensions (d_1, \dots, d_k) . We then set $y := c - Xb$. □

We also provide an analogue of Fact 2(i), whose proof is somewhat more involved.

Theorem 3 *The dimension of an affine Schubert variety is*

$$\dim \Psi(d_1, \dots, d_k) = \sum_{j=1}^k d_j - \frac{k(k+1)}{2} + (d_1 - 1).$$

Proof Let j be the embedding in (5.1). The dimension of $j(\Psi(d_1, \dots, d_k))$ is clearly the same as that of $\Psi(d_1, \dots, d_k)$. We claim that

$$j(\Psi(d_1, \dots, d_k)) = \Omega(d_1, d_1 + 1, \dots, d_k + 1) \cap j(\text{Graff}(k, n)). \tag{6.2}$$

Since $\Omega(d_1, d_1 + 1, \dots, d_k + 1)$ is an irreducible subset of $\text{Gr}(k + 1, n + 1)$ and $j(\text{Graff}(k, n))$ is an affine open subset of $\text{Gr}(k + 1, n + 1)$, we obtain the required dimension via

$$\begin{aligned} \dim j(\Psi(d_1, \dots, d_k)) &= \dim \Omega(d_1, d_1 + 1, \dots, d_k + 1) \\ &= d_1 + \sum_{i=1}^k (d_i + 1) - \frac{(k + 1)(k + 2)}{2} \\ &= \sum_{i=1}^k d_i - \frac{k(k + 1)}{2} + (d_1 - 1), \end{aligned}$$

where we have used Fact 2(i) for the second equality.

It remains to establish (6.2). Let $\mathbb{A}_0 + b \subseteq \mathbb{A}_1 + b \subseteq \dots \subseteq \mathbb{A}_k + b$ be an affine flag with $\dim(\mathbb{A}_i + b) = d_i, i = 0, 1, \dots, k$, where we set $d_0 := d_1 - 1$.

Let $\mathbb{B} + c \in \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$. As $\dim(\mathbb{A}_1 + b) \cap (\mathbb{B} + c) \geq 1$, there is some $-x$ contained in both $\mathbb{A}_1 + b$ and $\mathbb{B} + c$. Since $\mathbb{A}_1 + b \subseteq \dots \subseteq \mathbb{A}_k + b$, we have

$$\mathbb{A}_i + b = \mathbb{A}_i + x, \quad \mathbb{B} + c = \mathbb{B} + x, \quad i = 1, \dots, k. \tag{6.3}$$

Therefore, for any $i = 1, \dots, k$,

$$\begin{aligned} \dim j(\mathbb{B} + c) \cap j(\mathbb{A}_i + b) &= \dim j(\mathbb{B} + x) \cap j(\mathbb{A}_i + x) \\ &= \dim(\mathbb{B} + c) \cap (\mathbb{A}_i + b) + 1 \geq i + 1. \end{aligned}$$

Since $j(\mathbb{A}_0 + b)$ is a codimension-one linear subspace of $j(\mathbb{A}_1 + b)$, we also have

$$\dim j(\mathbb{B} + c) \cap j(\mathbb{A}_0 + b) \geq \dim j(\mathbb{B} + c) \cap j(\mathbb{A}_1 + b) - 1 \geq 1.$$

Hence we must have

$$j(\mathbb{B} + c) \in \Omega(j(\mathbb{A}_0 + b), j(\mathbb{A}_1 + b), \dots, j(\mathbb{A}_k + b)) \cap j(\text{Graff}(k, n)).$$

This shows the “ \subseteq ” in (6.2).

Conversely, let $j(\mathbb{B} + c) \in \Omega(j(\mathbb{A}_0 + b), j(\mathbb{A}_1 + b), \dots, j(\mathbb{A}_k + b))$. On the one hand, we have

$$\dim j(\mathbb{B} + c) \cap j(\mathbb{A}_i + b) \geq i + 1;$$

and on the other hand, since j is an embedding, we have

$$j((\mathbb{B} + c) \cap (\mathbb{A}_i + b)) = j(\mathbb{B} + c) \cap j(\mathbb{A}_i + b),$$

for any $i = 0, 1, \dots, k$. Therefore, we have

$$\dim(\mathbb{B} + c) \cap (\mathbb{A}_i + b) + 1 = \dim j(\mathbb{B} + c) \cap j(\mathbb{A}_i + b) \geq i + 1, \quad i = 0, 1, \dots, k.$$

In other words, $\mathbb{B} + c \in \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$. This shows the “ \supseteq ” in (6.2). \square

In Sect. 7.2, we will give the affine analogues of Facts 2(ii) and (iii) as Theorem 6 and Proposition 9, respectively.

There are two affine Schubert varieties that deserve special mention because of their importance in our metric geometry discussions in Sect. 8.3 and, to a lesser extent, also the probability discussions in Sect. 9.

Definition 5 Let $\mathbb{A} + b \in \text{Graff}(k, n)$ and $\mathbb{B} + c \in \text{Graff}(l, n)$ where $k \leq l \leq n$. The *affine Schubert varieties* of l -flats containing $\mathbb{A} + b$ and k -flats contained in $\mathbb{B} + c$ are, respectively,

$$\begin{aligned} \Psi_+(\mathbb{A} + b) &:= \{\mathbb{X} + y \in \text{Graff}(l, n) : \mathbb{A} + b \subseteq \mathbb{X} + y\}, \\ \Psi_-(\mathbb{B} + c) &:= \{\mathbb{Y} + z \in \text{Graff}(k, n) : \mathbb{Y} + z \subseteq \mathbb{B} + c\}. \end{aligned} \tag{6.4}$$

The nomenclature in Definition 5 is justified as $\Psi_+(\mathbb{A} + b)$ is the affine Schubert variety of the affine flag

$$\{0\} =: \mathbb{A}_0 + b_0 \subseteq \mathbb{A}_1 + b_1 \subseteq \dots \subseteq \mathbb{A}_k + b_k := \mathbb{A} + b \subseteq \dots \subseteq \mathbb{A}_l + b_l, \tag{6.5}$$

where $\mathbb{A}_{k+i} + b_{k+i}$ is an affine subspace of dimension $n - l + (k + i)$, $i = 1, \dots, l - k$; and $\Psi_-(\mathbb{B} + c)$ is the affine Schubert variety of the affine flag

$$\{0\} =: \mathbb{B}_0 + c_0 \subseteq \mathbb{B}_1 + c_1 \subseteq \dots \subseteq \mathbb{B}_k + c_k := \mathbb{B} + c \tag{6.6}$$

where $\mathbb{B}_j + c_j$ is an affine subspace of dimension $l - k + j$, $j = 1, \dots, k$.

We next discuss the geometry of these sets, starting with the observation that $\Psi_+(\mathbb{A} + b)$ is isomorphic to a Grassmannian and $\Psi_-(\mathbb{B} + c)$ is isomorphic to an affine Grassmannian.

Proposition 6 *Let $\mathbb{A} + b \in \text{Graff}(k, n)$ and $\mathbb{B} + c \in \text{Graff}(l, n)$. Then*

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

as Riemannian manifolds and algebraic varieties. In particular, we have

$$\dim \Psi_+(\mathbb{A} + b) = (n - l)(l - k), \quad \dim \Psi_-(\mathbb{B} + c) = (k + 1)(l - k).$$

Proof We first observe that the map $\varphi : \Psi_+(\mathbb{A} + b) \rightarrow \Omega_+(\mathbb{A})$, $\mathbb{X} + y \mapsto \mathbb{X} + y - b$, is well-defined since $\mathbb{A} \subseteq \mathbb{X} + y - b$ by our choice of $\mathbb{X} + y$. Also, $\psi : \Omega_+(\mathbb{A}) \rightarrow \Psi_+(\mathbb{A} + b)$, $\mathbb{X} \mapsto \mathbb{X} + b$, is the inverse of φ and so it is an isomorphism. Together with [41, Proposition 21], we obtain the first isomorphism $\Psi_+(\mathbb{A} + b) \cong \Omega_+(\mathbb{A}) \cong \text{Gr}(n - l, n - k)$. For the second isomorphism, consider $\varphi' : \Psi_-(\mathbb{B} + c) \rightarrow \text{Graff}_k(\mathbb{B})$, $\mathbb{Y} + z \mapsto \mathbb{Y} + z - c$, which is well-defined since $\mathbb{Y} + z - c$ is an affine subspace of dimension k in \mathbb{B} . Its inverse is given by $\psi' : \text{Graff}_k(\mathbb{B}) \rightarrow \Psi_-(\mathbb{B} + c)$, $\mathbb{Y} + z \mapsto \mathbb{Y} + z + c$, and so it is an isomorphism. The required isomorphism then follows from $\Psi_-(\mathbb{B} + c) \cong \text{Graff}_k(\mathbb{B}) \cong \text{Graff}(k, l)$. □

The asymmetry in Proposition 6 is expected. $\Psi_+(\mathbb{A} + b)$ is a Grassmannian of linear subspaces since all affine subspaces containing $\mathbb{A} + b$ can be shifted back to the origin by the vector b . In the case of $\Psi_-(\mathbb{B} + c)$, shifting $\mathbb{B} + c$ back to the origin by c and then taking all affine subspaces contained in \mathbb{B} still gives a Grassmannian of affine subspaces. As a sanity check, note that the dimensions in Proposition 6 agree with their values given by Theorem 3 with respect to the affine flags (6.5) and (6.6).

We also have the following analogue of Proposition 4 that allows us to regard $\Psi_+(\mathbb{A} + b)$, $\Psi_-(\mathbb{B} + c)$ as subsets of $n \times (n + 1)$ matrices.

Proposition 7 *The affine Schubert varieties $\Psi_+(\mathbb{A} + b)$ and $\Psi_-(\mathbb{B} + c)$ are isomorphic to real affine algebraic varieties in $\mathbb{R}^{n \times (n+1)}$ given by*

$$\begin{aligned} \Psi_+(\mathbb{A} + b) &\cong \{[P, d] \in \mathbb{R}^{n \times (n+1)} : P^\top = P^2 = P, Pd = 0, \\ &\quad \text{tr}(P) = l, j(\mathbb{A} + b) \subseteq \text{im}([P, d])\}, \\ \Psi_-(\mathbb{B} + c) &\cong \{[P, d] \in \mathbb{R}^{n \times (n+1)} : P^\top = P^2 = P, Pd = 0, \\ &\quad \text{tr}(P) = k, \text{im}([P, d]) \subseteq j(\mathbb{B} + c)\}. \end{aligned}$$

7 Algebraic topology of the affine Grassmannian

We will determine the homotopy groups and (co)homology groups/rings of $\text{Graff}(k, n)$. With this in mind, we begin with yet another structure of $\text{Graff}(k, n)$, namely, as a vector bundle—in fact it is the universal quotient bundle of $\text{Gr}(k, n)$.

Recall that if S is a subbundle of a vector bundle E on a manifold M , then Q is called the *quotient bundle* on M of E by S if there is a short exact sequence of vector bundles

$$0 \rightarrow S \rightarrow E \rightarrow Q \rightarrow 0. \tag{7.1}$$

Recall also that the *tautological bundle* over $\text{Gr}(k, n)$ is the vector bundle whose fiber over $\mathbb{A} \in \text{Gr}(k, n)$ is simply \mathbb{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 (7.1), then the quotient bundle Q is called the *universal quotient bundle* of $\text{Gr}(k, n)$ [16,28].

Theorem 4

- (i) $\text{Graff}(k, n)$ is a rank- $(n - k)$ vector bundle over $\text{Gr}(k, n)$ with bundle projection $\tau : \text{Graff}(k, n) \rightarrow \text{Gr}(k, n)$, the deaffine map in (3.2).
- (ii) $\text{Graff}(k, n)$ is the universal quotient bundle of $\text{Gr}(k, n)$,

$$0 \rightarrow S \rightarrow \text{Gr}(k, n) \times \mathbb{R}^n \rightarrow \text{Graff}(k, n) \rightarrow 0, \tag{7.2}$$

where S is the tautological bundle.

Proof In affine coordinates, the deaffine map $\tau : \text{Graff}(k, n) \rightarrow \text{Gr}(k, n)$, $\mathbb{A} + b \mapsto \mathbb{A}$ takes the form $\tau([a_1, \dots, a_k, b_0]) = [a_1, \dots, a_k]$ where a_i 's and b_0 are chosen as in the proof of Proposition 1. Notice that the fiber $\tau^{-1}(\mathbb{A})$ for $\mathbb{A} \in \text{Gr}(k, n)$ is simply $\mathbb{R}^n / \mathbb{A}$, 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 $\text{Gr}(k, n)$. Let $q : \text{Gr}(k, n) \times \mathbb{R}^n \rightarrow \text{Graff}(k, n)$, $(\mathbb{A}, b) \mapsto \mathbb{A} + b$. 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 obtain the exact sequence in (7.2). This shows that $\text{Graff}(k, n)$ is the universal quotient bundle. □

7.1 Homotopy of Graff(k, n)

When $\text{Graff}(k, n)$ is regarded as a vector bundle on $\text{Gr}(k, n)$ as in Theorem 4(i), the base space $\text{Gr}(k, n)$ is homeomorphic to the zero section, which is a strong deformation retract of $\text{Graff}(k, n)$. Hence $\text{Gr}(k, n)$ and $\text{Graff}(k, n)$ have the same homotopy type and so

$$\pi_r(\text{Graff}(k, n)) \simeq \pi_r(\text{Gr}(k, n)), \quad r \in \mathbb{N}.$$

From the list of homotopy groups of $\text{Gr}(k, n)$ in [39, Section 10.8], we obtain those of $\text{Graff}(k, n)$.

Proposition 8 $\text{Graff}(k, n)$ is homotopy equivalent to $\text{Gr}(k, n)$. Therefore

- (i) for $n \geq k + 2$ and $0 < k < n/2$,

$$\pi_1(\text{Graff}(k, n)) \simeq \begin{cases} \mathbb{Z} & \text{if } k = 1, n = 2, \\ \mathbb{Z}_2 & \text{otherwise;} \end{cases}$$

(ii) for $0 \leq k < n/2$ and $2 \leq r < n - 2k$,

$$\pi_r(\text{Graff}(k, n)) \simeq \begin{cases} \mathbb{Z} & \text{if } r = 0, 4 \pmod 8, \\ \mathbb{Z}_2 & \text{if } r = 1, 2 \pmod 8, \\ 0 & \text{if } r = 3, 5, 6, 7 \pmod 8. \end{cases}$$

Since the deaffine map τ in (3.2) is a bundle projection by Theorem 4(i), it is straightforward to take direct limits in (3.4) and extend Proposition 8 to the infinite Grassmannian via the commutative diagram (4.2). This also shows that $\text{Graff}(k, \infty)$ is a classifying space [20].

Corollary 1 $\text{Graff}(k, \infty)$ is homotopy equivalent to $\text{Gr}(k, \infty)$. Therefore

$$\pi_1(\text{Graff}(k, \infty)) \simeq \mathbb{Z}_2;$$

and for $r \geq 2$,

$$\pi_r(\text{Graff}(k, \infty)) \simeq \begin{cases} \mathbb{Z} & \text{if } r = 0, 4 \pmod 8, \\ \mathbb{Z}_2 & \text{if } r = 1, 2 \pmod 8, \\ 0 & \text{if } r = 3, 5, 6, 7 \pmod 8. \end{cases}$$

Moreover, $\text{Graff}(k, \infty)$ is the classifying space of $E(n)$ and $GA(n)$ with total space $\text{Vaff}(k, \infty)$.

7.2 Homology and cohomology of $\text{Graff}(k, n)$

We show that the affine Schubert varieties in Sect. 6 play a role for the (co)homology of $\text{Graff}(k, n)$ similar to that of Schubert varieties for $\text{Gr}(k, n)$.

Let $\mathbb{A}_1 \subseteq \dots \subseteq \mathbb{A}_k$ be a flag in \mathbb{R}^n . For any $b \in \mathbb{R}^n$, the deaffine map $\tau : \text{Graff}(k, n) \rightarrow \text{Gr}(k, n)$ in (3.2), when restricted to $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$, defines a map

$$\tau_b : \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b) \rightarrow \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k), \quad \mathbb{A} + b \mapsto \mathbb{A}.$$

For any fixed $b \in \mathbb{R}^n$, it has a right inverse

$$s_b : \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k) \rightarrow \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b), \quad \mathbb{A} \mapsto \mathbb{A} + b.$$

Lemma 2 Let $\mathbb{A}_1 \subseteq \dots \subseteq \mathbb{A}_k$ be a flag in \mathbb{R}^n and $b \in \mathbb{R}^n$. Then the following diagram commutes:

$$\begin{array}{ccccc} \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b) & \hookrightarrow & \tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k)) & \hookrightarrow & \text{Graff}(k, n) \\ & \searrow \tau_b & \downarrow \tau & & \downarrow \tau \\ & \swarrow s_b & \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k) & \hookrightarrow & \text{Gr}(k, n) \end{array} \tag{7.3}$$

Proof The only point in (7.3) that needs verification is the inclusion $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b) \subseteq \tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$. Let $\mathbb{B} + c \in \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$. Then $\dim(\mathbb{A}_j + b) \cap (\mathbb{B} + c) \geq j, j = 1, \dots, k$. We need to show that $\dim \mathbb{A}_j \cap \mathbb{B} \geq j, j = 1, \dots, k$. By the same argument that led to (6.3), we may choose an $x \in \mathbb{R}^n$ so that

$$\mathbb{A}_j + b = \mathbb{A}_j + x, \quad \mathbb{B} + c = \mathbb{B} + x, \quad j = 1, \dots, k.$$

Therefore

$$\dim(\mathbb{A}_j + x) \cap (\mathbb{B} + x) = \dim(\mathbb{A}_j + b) \cap (\mathbb{B} + c) \geq j$$

and thus $\dim \mathbb{A}_j \cap \mathbb{B} \geq j, j = 1, \dots, k$. □

For a more precise relation between $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$ and $\tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$, we show that the fibers of τ_b are contractible.

Lemma 3 *Let $\mathbb{A}_1 \subseteq \dots \subseteq \mathbb{A}_k$ be a flag in $\mathbb{R}^n, \mathbb{B} \in \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k)$, and $b \in \mathbb{R}^n$. Then*

$$\tau_b^{-1}(\mathbb{B}) = \{\mathbb{B} + c \in \text{Graff}(k, n) : \dim(\mathbb{B} + c) \cap (\mathbb{A}_j + b) \geq j, j = 1, \dots, k\}$$

is convex and therefore contractible.

Proof We first define an auxiliary set

$$\begin{aligned} C(\mathbb{B}, b) &:= \{c \in \mathbb{R}^n : \mathbb{B} + c \in \tau_b^{-1}(\mathbb{B})\} \\ &= \{c \in \mathbb{R}^n : \dim(\mathbb{B} + c) \cap (\mathbb{A}_j + b) \geq j, j = 1, \dots, k\}. \end{aligned} \tag{7.4}$$

If $c \in C(\mathbb{B}, b)$, then $c + b' \in C(\mathbb{B}, b)$ for any $b' \in \mathbb{B}$. Moreover, $\mathbb{B} + c = \mathbb{B} + c'$ if and only if $c' - c \in \mathbb{B}$. So we have a homeomorphism

$$C(\mathbb{B}, b)/\mathbb{B} \cong \tau_b^{-1}(\mathbb{B}), \tag{7.5}$$

where the left-hand side is regarded as a subset of the quotient vector space \mathbb{R}^n/\mathbb{B} . So the convexity of $\tau_b^{-1}(\mathbb{B})$ would follow from the convexity of $C(\mathbb{B}, b)$ in \mathbb{R}^n .

We remind the reader that if $A \in \mathbb{V}(k, n)$ is an orthonormal basis for $\mathbb{A} \in \text{Gr}(k, n)$, then $\mathbb{A} = \text{im}(A) = \ker(I_n - AA^T)$. Let $B \in \mathbb{V}(n, n - k)$ and $A_j \in \mathbb{V}(n - d_j, n)$ be orthonormal bases of \mathbb{B} and \mathbb{A}_j , respectively, $j = 1, \dots, k$. So

$$\mathbb{B} = \{y \in \mathbb{R}^n : (I_n - BB^T)y = 0\}, \quad \mathbb{A}_j = \{y \in \mathbb{R}^n : (I_n - A_j A_j^T)y = 0\},$$

and so

$$\begin{aligned} \mathbb{B} + c &= \{y \in \mathbb{R}^n : (I_n - BB^T)(y - c) = 0\}, \\ \mathbb{A}_j + b &= \{y \in \mathbb{R}^n : (I_n - A_j A_j^T)(y - b) = 0\}, \end{aligned} \tag{7.6}$$

for $j = 1, \dots, k$. Hence by (7.4) and (7.6),

$$C(\mathbb{B}, b) = \{c \in \mathbb{R}^n : \text{solution space of (7.6) has dimension } \geq j, j = 1, \dots, k\}.$$

With this characterization of $C(\mathbb{B}, b)$, convexity is straightforward: Let $c_1, c_2 \in C(\mathbb{B}, b)$ and $y_1, y_2 \in \mathbb{R}^n$ be such that

$$(I_n - BB^T)(y_i - c_i) = 0, \quad (I_n - A_j A_j^T)(y_i - b) = 0, \quad j = 1, \dots, k, \quad i = 1, 2. \tag{7.7}$$

For any $t \in [0, 1]$, $c_t = tc_1 + (1 - t)c_2$ and $y_t = ty_1 + (1 - t)y_2$ clearly also satisfy (7.7). □

This leads us to the following relation between $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$ and $\tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$.

Theorem 5 *The image $s_b(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$ is a strong deformation retract of $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$. In particular, $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$ is homotopy equivalent to $\tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$.*

Proof $s_b : \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k) \rightarrow \Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$ is a section of τ_b , i.e., $\tau_b \circ s_b = 1$. Hence it suffices to prove that the fiber $\tau_b^{-1}(\mathbb{B})$ deformation retracts to $s_b(\mathbb{B})$ for each $\mathbb{B} \in \Omega(\mathbb{A}_1, \dots, \mathbb{A}_k)$, but this is trivially true since $\tau_b^{-1}(\mathbb{B})$ is contractible by Lemma 3. □

By virtue of Theorem 5, we deduce next that the affine Schubert varieties form a natural basis for the (co)homology groups of $\text{Graff}(k, n)$. In this context, the (co)homology classes determined by affine Schubert varieties are called *affine Schubert cycles*.

Theorem 6 *Affine Schubert cycles form a basis for the (co)homology groups of an affine Grassmannian. In particular,*

$$H_i(\text{Graff}(k, n), \mathbb{Z}_2) \simeq H^i(\text{Graff}(k, n), \mathbb{Z}_2) \simeq \mathbb{Z}_2^{r_i},$$

where r_i is the number of partitions of the integer i into at most k parts, each not exceeding $n - k$. We also have a graded ring isomorphism

$$H^*(\text{Graff}(k, \infty), \mathbb{Z}_2) \simeq \mathbb{Z}_2[x_1, \dots, x_k]^{\mathfrak{S}_k}.$$

Proof By Fact 2(ii), as τ is a homotopy equivalence, the set of $\tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$ over all i -dimensional Schubert varieties $\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k)$ form a basis for the i th (co)homology group of $\text{Graff}(k, n)$. Therefore, by Theorems 3 and 5, the j -dimensional affine Schubert varieties $\Psi(\mathbb{A}_1 + b, \dots, \mathbb{A}_k + b)$ form a basis for the $(j - d_1 + 1)$ th (co)homology group of $\text{Graff}(k, n)$.

For the cohomology ring, the homotopy equivalence between $\text{Graff}(k, \infty)$ and $\text{Gr}(k, \infty)$ in Corollary 1 gives $H^*(\text{Graff}(k, \infty), \mathbb{Z}_2) \simeq H^*(\text{Gr}(k, \infty), \mathbb{Z}_2)$. The rest then follows from Fact 2(ii). □

We stated Theorem 6 with \mathbb{Z}_2 coefficients for simplicity, but in the same manner we may obtain $H^*(\text{Graff}(k, \infty), \mathbb{Z})$ and $H^*(\text{Graff}(k, \infty), \mathbb{Q})$ in terms of characteristic classes using the corresponding results for $\text{Gr}(k, \infty)$ in [7,8,35] and [37], respectively.

We conclude this section with a cell decomposition of $\text{Graff}(k, n)$.

Proposition 9 *The collection of preimages $\tau^{-1}(\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k))$ over all flags in \mathbb{R}^n of length k gives a cell decomposition of $\text{Graff}(k, n)$.*

Proof By Theorem 4, $\text{Graff}(k, n)$ is a vector bundle over $\text{Gr}(k, n)$ with bundle projection $\tau : \text{Graff}(k, n) \rightarrow \text{Gr}(k, n)$. By Fact 2(iii), the collection of $\Omega(\mathbb{A}_1, \dots, \mathbb{A}_k)$ over all flags of length k provides a cell-decomposition of $\text{Gr}(k, n)$. So the required result follows. □

8 Metric geometry of the affine Grassmannian

We have two goals in this section. The first is to extend various distances defined on Grassmannian to the affine Grassmannian; the results are summarized in Table 2—these are distances between affine subspaces of the *same* dimension. Following our earlier work in [41], our next goal is to further extend these distances in a natural way (using the affine Schubert varieties in Definition 5) to affine subspaces of *different* dimensions.

8.1 Issues in metricizing $\text{Graff}(k, n)$

A reason for the widespread applicability of the usual Grassmannian is that one has concrete, explicitly computable expressions for geodesics and distances on $\text{Gr}(k, n)$. In [2,12,40], these expressions were obtained from a purely differential geometric perspective. One might imagine that the differential geometric structures on $\text{Graff}(k, n)$ in Propositions 1, 2, or Theorem 4 would yield similar results. Surprisingly this is not the case.

A more careful examination of the arguments in [2,12,40] for obtaining explicit expressions for geodesics and geodesic distances on $\mathbb{V}(k, n)$ and $\text{Gr}(k, n)$ reveals that they rely on a somewhat obscure structure, namely, that of a *geodesic orbit space* [4,15]. 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 $\mathfrak{g}/\mathfrak{h}$ where \mathfrak{g} and \mathfrak{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 $\text{Gr}(k, n) = \text{O}(n)/(\text{O}(n - k) \times \text{O}(k))$ and $\mathbb{V}(k, n) = \text{O}(n)/\text{O}(n - k)$, as $\text{O}(n)$ is a compact semisimple Lie group, $\text{Gr}(k, n)$ and $\mathbb{V}(k, n)$ are geodesic orbit spaces. Furthermore, as $\text{O}(n)$ is a matrix Lie group, all its one-parameter subgroups are given by exponential maps, which in turn allows us to write down explicit expressions for the geodesics (and thus also the geodesic distances) on $\text{Gr}(k, n)$ and $\mathbb{V}(k, n)$. The difficulty in seeking similar expressions on $\text{Graff}(k, n) = \text{E}(n)/(\text{E}(n - k) \times \text{O}(k))$ is that it may not be a geodesic orbit space since $\text{E}(n)$ is not compact.

What about the vector bundle structure on $\text{Graff}(k, n)$ then? If E is a vector bundle over a Riemannian manifold M , then the *pullback* of the metric on M induces a metric on E . Nevertheless, this metric on E is uninteresting—by definition, it disregards the fibers of the bundle. In the context of Theorem 4, this is akin to defining the distance

between $\mathbb{A} + b$ and $\mathbb{B} + c \in \text{Graff}(k, n)$ as the usual Grassmann distance between \mathbb{A} and $\mathbb{B} \in \text{Gr}(k, n)$.

We will turn to the algebraic geometric properties of $\text{Graff}(k, n)$ in Theorem 1 to provide the framework for defining distances with explicitly computable expressions, first for equidimensional affine subspaces and next for inequidimensional affine subspaces.

8.2 Distances on $\text{Graff}(k, n)$

The Riemannian metric on $\text{Gr}(k, n)$ yields the following well-known *Grassmann distance* between two subspaces $\mathbb{A}, \mathbb{B} \in \text{Gr}(k, n)$,

$$d_{\text{Gr}(k,n)}(\mathbb{A}, \mathbb{B}) = \left(\sum_{i=1}^k \theta_i^2 \right)^{1/2}, \tag{8.1}$$

where $\theta_1, \dots, \theta_k$ are the principal angles between \mathbb{A} and \mathbb{B} . This distance is easily computable via SVD as $\theta_i = \cos^{-1} \sigma_i$, where σ_i is the i th singular value of the matrix $A^T B$ for any orthonormal bases A and B of \mathbb{A} and \mathbb{B} [14,41].

By Theorem 1(ii), we may identify $\text{Graff}(k, n)$ with its image $j(\text{Graff}(k, n))$ in $\text{Gr}(k + 1, n + 1)$. As a subset of $\text{Gr}(k + 1, n + 1)$, $\text{Graff}(k, n)$ inherits the Grassmann distance $d_{\text{Gr}(k+1,n+1)}$ on $\text{Gr}(k + 1, n + 1)$, giving us the distance in Theorem 7 that can also be readily computed using SVD. We will show in Proposition 11 that this distance is in fact *intrinsic*. In the following, Theorem 7 reproduces the statement (but not the proof) of [26, Theorem 4.2] for the reader’s easy reference. On the other hand, Lemma 4 and Corollary 2 collectively provide a full proof for [26, Corollary 4.3], wherein only a sketch was given.

Theorem 7 For any two affine k -flats $\mathbb{A} + b$ and $\mathbb{B} + c \in \text{Graff}(k, n)$,

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

where j is the embedding in (5.1), defines a notion of distance consistent with the Grassmann distance. If

$$Y_{\mathbb{A}+b} = \begin{bmatrix} A & b_0/\sqrt{1 + \|b_0\|^2} \\ 0 & 1/\sqrt{1 + \|b_0\|^2} \end{bmatrix}, \quad Y_{\mathbb{B}+c} = \begin{bmatrix} B & c_0/\sqrt{1 + \|c_0\|^2} \\ 0 & 1/\sqrt{1 + \|c_0\|^2} \end{bmatrix}$$

are the matrices of Stiefel coordinates for $\mathbb{A} + b$ and $\mathbb{B} + c$, respectively, then

$$d_{\text{Graff}(k,n)}(\mathbb{A} + b, \mathbb{B} + c) = \left(\sum_{i=1}^{k+1} \theta_i^2 \right)^{1/2}, \tag{8.2}$$

where $\theta_i = \cos^{-1} \sigma_i$ and σ_i is the i th singular value of $Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c} \in \mathbb{R}^{(k+1) \times (k+1)}$.

Proof Similar to [26, Theorem 4.2]. □

It is not difficult to see that the angles $\theta_1, \dots, \theta_{k+1}$ are independent of the choice of Stiefel coordinates. We define the following affine analogues of principal angles and principal vectors of linear subspaces [6,14,41] that will be useful later.

Definition 6 We will call θ_i the *i*th affine principal angles between the respective affine subspaces and denote it by $\theta_i(\mathbb{A} + b, \mathbb{B} + c)$. Consider the SVD,

$$Y_{\mathbb{A}+b}^\top Y_{\mathbb{B}+c} = U \Sigma V^\top \tag{8.3}$$

where $U, V \in O(k + 1)$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{k+1})$. Let

$$Y_{\mathbb{A}+b} U = [p_1, \dots, p_{k+1}], \quad Y_{\mathbb{B}+c} V = [q_1, \dots, q_{k+1}].$$

We will call the pair of column vectors (p_i, q_i) the *i*th affine principal vectors between $\mathbb{A} + b$ and $\mathbb{B} + c$.

We next show that the distance in Theorem 7 is the only possible distance on an affine Grassmannian compatible with the usual Grassmann distance on a Grassmannian. On any connected Riemannian manifold M with Riemannian metric g , there is an intrinsic distance function d_M on M with respect to g ,

$$d_M(x, y) := \inf\{L(\gamma) : \gamma \text{ is a piecewise smooth curve connecting } x \text{ and } y \text{ in } M\}.$$

Here $L(\gamma)$ is the length of the curve $\gamma : [0, 1] \rightarrow M$ defined by

$$L(\gamma) := \int_0^1 \|\gamma'(t)\| = \int_0^1 \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))}.$$

For a connected submanifold of $N \subseteq M$, there is a natural Riemannian metric g_N on N induced by g and therefore a corresponding intrinsic distance function,

$$d_N(x, y) := \inf\{L(\gamma) : \gamma \text{ is a piecewise smooth curve connecting } x \text{ and } y \text{ in } N\}.$$

On the other hand, we may also define a distance function $d_M|_N$ on N by simply restricting the distance function d_M to N —note that this is what we have done in Theorem 7 with $M = \text{Gr}(k + 1, n + 1)$ and $N = \text{Graff}(k, n)$. In general, $d_M|_N \neq d_N$. For example, for $N = \mathbb{S}^2$ embedded as the unit sphere in $M = \mathbb{R}^3$, the two distance functions on \mathbb{S}^2 are obviously different. However, for our embedding of $\text{Graff}(k, n)$ in $\text{Gr}(k + 1, n + 1)$, the two distances on $\text{Graff}(k, n)$ agree.

Proposition 10 *Let K be a closed submanifold of codimension at least two in M and let N be the complement of K in M . Then $d_M|_N = d_N$.*

Proof We need to show that for any two distinct points $x, y \in N$, $d_M(x, y) = d_N(x, y)$. By the definition of d_M and d_N , it suffices to show that any piecewise

smooth curve γ in M connecting x, y can be approximated by a piecewise smooth curve in N connecting x, y . We remind the reader of the transversality theorem [18, Theorem 2.4]: Let $\varphi : L \rightarrow M$ be a smooth map between two manifolds and let $K \subseteq M$ be a submanifold. Then φ can be perturbed in an arbitrarily small neighborhood to some $\tilde{\varphi}$ such that

$$T_{\tilde{\varphi}(x)}(K) + \text{im}(d\tilde{\varphi}_x) = T_{\tilde{\varphi}(x)}(M)$$

for any $x \in \tilde{\varphi}^{-1}(K)$. In particular, if $\dim \text{im}(\varphi) + \dim K < \dim M$, then there exists a perturbation $\tilde{\varphi}$ in any small neighborhood of φ such that $\text{im}(\tilde{\varphi}) \cap K = \emptyset$. When applied to our case with $L = [0, 1]$, the transversality theorem permits us to perturb the curve $\gamma = \varphi(L) \subseteq M$ in any small neighborhood to avoid K as long as K has codimension at least two. It follows that γ can be approximated by curves in $N = M \setminus K$ connecting x and y . □

Proposition 11 *The distance $d_{\text{Graff}(k,n)}$ in Theorem 7 is intrinsic with respect to the Riemannian metric on $\text{Graff}(k, n)$ induced from that of $\text{Gr}(k + 1, n + 1)$.*

Proof By Theorem 1, the complement of $N = \text{Graff}(k, n)$ in M is $\text{Gr}(k + 1, n)$ and has codimension $k + 1 \geq 2$. Hence Proposition 10 applies. □

Proposition 11 justifies calling the distance in (8.2) the *Grassmann distance* on $\text{Graff}(k, n)$. We next determine an expression for the geodesic connecting two points on $\text{Graff}(k, n)$ that attains their minimum Grassmann distance. There is one caveat—this geodesic may contain a point lying outside $\text{Graff}(k, n)$ as it is not a geodesically complete manifold.

Lemma 4 *Let $\mathbb{A} + b, \mathbb{B} + c \in \text{Graff}(k, n)$ and let*

$$Y_{\mathbb{A}+b} = \begin{bmatrix} A & b_0/\sqrt{1 + \|b_0\|^2} \\ 0 & 1/\sqrt{1 + \|b_0\|^2} \end{bmatrix}, \quad Y_{\mathbb{B}+c} = \begin{bmatrix} B & c_0/\sqrt{1 + \|c_0\|^2} \\ 0 & 1/\sqrt{1 + \|c_0\|^2} \end{bmatrix}$$

be their Stiefel coordinates. If $Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c}$ is invertible, then there is at most one point on the distance-minimizing geodesic in $\text{Gr}(k + 1, n + 1)$ connecting $\mathbb{A} + b$ and $\mathbb{B} + c$ which lies outside $j(\text{Graff}(k, n))$. Here j is the embedding in (5.1).

Proof Let $U \in \text{O}(k + 1)$ and the diagonal matrix Σ be as in (8.3). Let $\Theta := \text{diag}(\theta_1, \dots, \theta_{k+1}) = \cos^{-1} \Sigma$ be the diagonal matrix of affine principal angles. By [2] the geodesic $\gamma : [0, 1] \rightarrow \text{Gr}(k + 1, n + 1)$ connecting $j(\mathbb{A} + b)$ and $j(\mathbb{B} + c)$ is given by $\gamma(t) = \text{span}(Y_{\mathbb{A}+b} U \cos(t\Theta) + Q \sin(t\Theta))$, where Q is an $(n + 1) \times (k + 1)$ orthonormal matrix such that the RHS of

$$(I_{n+1} - Y_{\mathbb{A}+b} Y_{\mathbb{A}+b}^T) Y_{\mathbb{B}+c} (Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c})^{-1} = Q(\tan(\Theta))U^T$$

gives an SVD of the matrix on the LHS. Let the last row of U be $[u_{k+1,1}, \dots, u_{k+1,k+1}]^T$ and that of Q be $[q_{k+1,1}, \dots, q_{k+1,k+1}]^T$. Then $\gamma(t) \in \text{Gr}(k + 1, n + 1) \setminus j(\text{Graff}(k, n))$

if and only if the entries on the last row of $\gamma(t)$ are all zero, i.e.,

$$\frac{u_{k+1,i} \cos(t\theta_i)}{\sqrt{\|b\|^2 + 1}} + q_{k+1,i} \sin(t\theta_i) = 0,$$

for all $i = 1, \dots, k + 1$. So at most one point on γ lies outside $\text{Graff}(k, n)$. □

Corollary 2 *Let $\mathbb{A} + b$ and $\mathbb{B} + c \in \text{Graff}(k, n)$. The distance-minimizing geodesic $\gamma : [0, 1] \rightarrow \text{Graff}(k, n)$ connecting $\mathbb{A} + b$ and $\mathbb{B} + c$ is given by*

$$\gamma(t) = j^{-1}(\text{span}(Y_{\mathbb{A}+b}U \cos t\Theta + Q \sin t\Theta)), \tag{8.4}$$

where $Q \in O(n + 1)$, $U \in O(k + 1)$ and the diagonal $(n + 1) \times (k + 1)$ matrix Θ are determined by the SVD

$$(I_{n+1} - Y_{\mathbb{A}+b}Y_{\mathbb{A}+b}^T)Y_{\mathbb{B}+c}(Y_{\mathbb{A}+b}^TY_{\mathbb{B}+c})^{-1} = Q(\tan \Theta)U^T.$$

The matrix U is the same as that in (8.3) and $\Theta = \text{diag}(\theta_1, \dots, \theta_{k+1})$ is the diagonal matrix of affine principal angles. The geodesic γ attains the distance in (8.2) and its derivative at $t = 0$ is given by

$$\gamma'(0) = j^{-1}(Q\Theta U^T). \tag{8.5}$$

We next give an example to illustrate Corollary 2 when $k = 1, n = 2$.

Example 1 Let $r, s \in \mathbb{R}$ with $r < s$. Consider the two affine lines in \mathbb{R}^2 ,

$$\begin{aligned} \mathbb{A} + b &= \{(x, y) \in \mathbb{R}^2 : x = r\} = \{r\} \times \mathbb{R}, \\ \mathbb{B} + c &= \{(x, y) \in \mathbb{R}^2 : x = s\} = \{s\} \times \mathbb{R}. \end{aligned}$$

By definition, these are two points on the affine Grassmannian $\text{Graff}(1, 2)$ and we may represent them in orthogonal affine coordinates as $[A, b_0], [B, c_0] \in \mathbb{R}^{2 \times 2}$ with

$$A = B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^{2 \times 1}, \quad b_0 = \begin{bmatrix} r \\ 0 \end{bmatrix} \in \mathbb{R}^2, \quad c_0 = \begin{bmatrix} s \\ 0 \end{bmatrix} \in \mathbb{R}^2.$$

Let $\theta, \phi \in (-\pi/2, \pi/2)$ be given by $\tan \theta = r$ and $\tan \phi = s$. By (8.4), the distance-minimizing geodesic curve $\gamma : [0, 1] \rightarrow \text{Graff}(1, 2)$ is given by

$$\gamma(t) = j^{-1}\left(\text{im} \begin{bmatrix} 0 & \sin((1-t)\theta+t\phi) \\ 1 & 0 \\ 0 & \cos((1-t)\theta+t\phi) \end{bmatrix}\right) = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x = \tan((1-t)\theta + t\phi) \right\}.$$

So the image of γ consists of affine lines of the form $\{(x, y) \in \mathbb{R}^2 : x = \tau\} = \{\tau\} \times \mathbb{R}$ where $r \leq \tau \leq s$. See Fig. 2 for an illustration. The geodesic distance is then

$$d_{\text{Graff}(1,2)}(\{r\} \times \mathbb{R}, \{s\} \times \mathbb{R}) = |\theta - \phi| = \tan^{-1}(s) - \tan^{-1}(r). \tag{8.6}$$

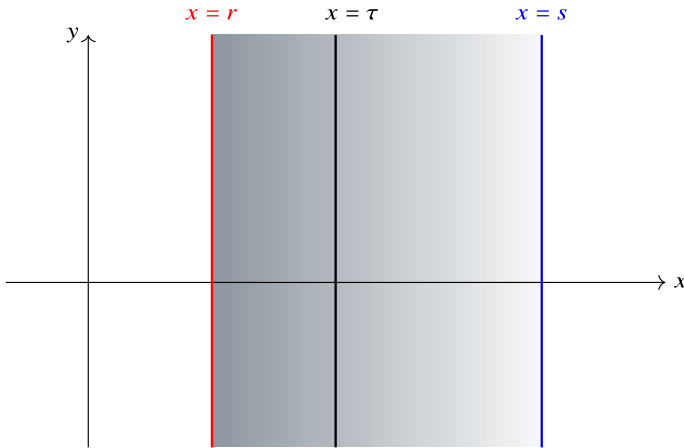


Fig. 2 The geodesic curve on $\text{Graff}(1, 2)$ in Example 1 is indicated by the shaded strip. Note that a point on $\text{Graff}(1, 2)$ is an affine line in \mathbb{R}^2 . The blue line $x = s$ is the initial point of this geodesic curve and the red line $x = r$ the terminal point. The black line $x = \tau$ is an arbitrary point on this geodesic curve. The lines are packed more closely together as we move to the left

It is worthwhile emphasizing that the geodesic distance between $\mathbb{A} + b$ (red line) and $\mathbb{B} + c$ (blue line) in Fig. 2 is not $r - s$ —this potential pitfall shows the fallacy of regarding $\text{Graff}(k, n)$ as a subset of $\text{Gr}(k, n) \times \mathbb{R}^n$; the value $r - s$ is the product distance on the latter, not the geodesic distance on $\text{Graff}(k, n)$ as a Riemannian manifold. Note that the Riemannian structure on $\text{Graff}(k, n)$ in this section comes from regarding it as a subset of $\text{Gr}(k + 1, n + 1)$ as in Sect. 5. In particular, the geodesic distance between any two points is finite since $\text{Gr}(k + 1, n + 1)$ is compact. For instance, by (8.6), if $r = -1$ and $s = 1$, then $d = \pi/2$; and if $r \rightarrow -\infty$ and $s \rightarrow \infty$, then $d \rightarrow \pi$.

The dark-to-light shading in Fig. 2 is intended to convey the idea that the affine lines get packed more closely together as we move to the left, i.e., for any $\delta > 0$,

$$d_{\text{Graff}(1,2)}(\{r\} \times \mathbb{R}, \{s\} \times \mathbb{R}) > d_{\text{Graff}(1,2)}(\{r + \delta\} \times \mathbb{R}, \{s + \delta\} \times \mathbb{R}).$$

For instance, $d_{\text{Graff}(1,2)}(\{0\} \times \mathbb{R}, \{1\} \times \mathbb{R}) \approx 0.785 > 0.322 \approx d_{\text{Graff}(1,2)}(\{1\} \times \mathbb{R}, \{2\} \times \mathbb{R})$. □

The Grassmann distance in (8.1) is the best known distance on the Grassmannian. But there are in fact several common distances on the Grassmannian [41, Table 2] and we may extend them to the affine Grassmannian by applying the embedding $j : \text{Graff}(k, n) \rightarrow \text{Gr}(k + 1, n + 1)$ and emulating our arguments in this section. We summarize these distances in Table 2.

8.3 Distances on $\text{Graff}(\infty, \infty)$

The problem of defining distances between *linear* subspaces of different dimensions has recently been resolved in [41]. We show here that the framework in [41] may be adapted for *affine* subspaces. This is expected to be important in modeling mixtures

Table 2 Distances on $\text{Graff}(k, n)$ in terms of affine principal angles and Stiefel coordinates. The matrices $U, V \in O(k + 1)$ in the right column of Table 2 are the ones in (8.3)

	Affine principal angles	Stiefel coordinates
Asimov	$d_{\text{Graff}(k,n)}^\alpha(\mathbb{A} + b, \mathbb{B} + c) = \theta_{k+1}$	$\cos^{-1} \ Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c}\ _2$
Binet–Cauchy	$d_{\text{Graff}(k,n)}^\beta(\mathbb{A} + b, \mathbb{B} + c) = \left(1 - \prod_{i=1}^{k+1} \cos^2 \theta_i\right)^{1/2}$	$(1 - (\det Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c})^2)^{1/2}$
Chordal	$d_{\text{Graff}(k,n)}^\kappa(\mathbb{A} + b, \mathbb{B} + c) = \left(\sum_{i=1}^{k+1} \sin^2 \theta_i\right)^{1/2}$	$\frac{1}{\sqrt{2}} \ Y_{\mathbb{A}+b} Y_{\mathbb{A}+b}^T - Y_{\mathbb{B}+c} Y_{\mathbb{B}+c}^T\ _F$
Fubini–Study	$d_{\text{Graff}(k,n)}^\phi(\mathbb{A} + b, \mathbb{B} + c) = \cos^{-1} \left(\prod_{i=1}^{k+1} \cos \theta_i\right)$	$\cos^{-1} \det Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c} $
Martin	$d_{\text{Graff}(k,n)}^\mu(\mathbb{A} + b, \mathbb{B} + c) = \left(\log \prod_{i=1}^{k+1} 1/\cos^2 \theta_i\right)^{1/2}$	$(-2 \log \det Y_{\mathbb{A}+b}^T Y_{\mathbb{B}+c})^{1/2}$
Procrustes	$d_{\text{Graff}(k,n)}^\rho(\mathbb{A} + b, \mathbb{B} + c) = 2 \left(\sum_{i=1}^{k+1} \sin^2(\theta_i/2)\right)^{1/2}$	$\ Y_{\mathbb{A}+b} U - Y_{\mathbb{B}+c} V\ _F$
Projection	$d_{\text{Graff}(k,n)}^\tau(\mathbb{A} + b, \mathbb{B} + c) = \sin \theta_{k+1}$	$\ Y_{\mathbb{A}+b} Y_{\mathbb{A}+b}^T - Y_{\mathbb{B}+c} Y_{\mathbb{B}+c}^T\ _2$
Spectral	$d_{\text{Graff}(k,n)}^\sigma(\mathbb{A} + b, \mathbb{B} + c) = 2 \sin(\theta_{k+1}/2)$	$\ Y_{\mathbb{A}+b} U - Y_{\mathbb{B}+c} V\ _2$

of affine subspaces of different dimensions [34]. The proofs of Lemma 5, Theorems 8 and 9 are similar to those of their linear counterparts [41, Lemma 3, Theorems 7 and 12] and are omitted.

Our first observation is that the Grassmann distance (8.2) on $\text{Graff}(k, n)$ does not depend on the ambient space \mathbb{R}^n and may thus be extended to $\text{Graff}(k, \infty)$.

Lemma 5 *The value $d_{\text{Graff}(k,n)}(\mathbb{A} + b, \mathbb{B} + c)$ of two k -flats $\mathbb{A} + b$ and $\mathbb{B} + c \in \text{Graff}(k, n)$ 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 $\text{Graff}(k, \infty)$.*

Our second observation is that for a k -dimensional affine subspace $\mathbb{A} + b$ and an l -dimensional affine subspace $\mathbb{B} + c$, assuming $k \leq l$ without loss of generality, (i) the distance from $\mathbb{A} + b$ to the set of k -dimensional affine subspaces contained in $\mathbb{B} + c$ equals (ii) the distance from $\mathbb{B} + c$ to the set of l -dimensional affine subspaces containing $\mathbb{A} + b$. Their common value then defines a natural distance between $\mathbb{A} + b$ and $\mathbb{B} + c$.

Note that (i) is a distance in $\text{Graff}(k, n)$, whereas (ii) is a distance in $\text{Graff}(l, n)$. Furthermore, the set in (i) is precisely $\Psi_+(\mathbb{A} + b)$ and the set in (ii) is precisely $\Psi_-(\mathbb{B} + c)$ —the affine Schubert varieties introduced in Definition 5.

Theorem 8 *Let $k \leq l \leq n$. For any $\mathbb{A} + b \in \text{Graff}(k, n)$ and $\mathbb{B} + c \in \text{Graff}(l, n)$, the following distances are equal,*

$$d_{\text{Graff}(k,n)}(\mathbb{A} + b, \Psi_-(\mathbb{B} + c)) = d_{\text{Graff}(l,n)}(\mathbb{B} + c, \Psi_+(\mathbb{A} + b)), \tag{8.7}$$

and their common value $\delta(\mathbb{A} + b, \mathbb{B} + c)$ may be computed explicitly as

$$\delta(\mathbb{A} + b, \mathbb{B} + c) = \left(\sum_{i=1}^{\min(k,l)+1} \theta_i(\mathbb{A} + b, \mathbb{B} + c)^2\right)^{1/2}. \tag{8.8}$$

The affine principal angles $\theta_1, \dots, \theta_{\min(h,l)+1}$ are as defined in Theorem 7 except that now they correspond to the singular values of the rectangular matrix

$$Y_{\mathbb{A}+b}^\top Y_{\mathbb{B}+c} = \begin{bmatrix} A & b_0/\sqrt{1 + \|b_0\|^2} \\ 0 & 1/\sqrt{1 + \|b_0\|^2} \end{bmatrix}^\top \begin{bmatrix} B & c_0/\sqrt{1 + \|c_0\|^2} \\ 0 & 1/\sqrt{1 + \|c_0\|^2} \end{bmatrix} \in \mathbb{R}^{(k+1) \times (l+1)}.$$

Like its counterpart for linear subspaces in [41, Theorem 7], δ defines a distance between the respective affine subspaces in the sense of a distance of a point to a set. It reduces to the Grassmann distance $d_{\text{Grass}}(k,n)$ in (8.2) when $\dim \mathbb{A} = \dim \mathbb{B} = k$.

Our third observation is that, like $d_{\text{Grass}}(k,n)$, the distances in Table 2 may be extended in the same manner to affine subspaces of different dimensions.

Theorem 9 *Let $k \leq l \leq n$. Let $\mathbb{A} + b \in \text{Grass}(k, n)$, $\mathbb{B} + c \in \text{Grass}(l, n)$. Then*

$$d_{\text{Grass}(k,n)}^*(\mathbb{A} + b, \Psi_-(\mathbb{B} + c)) = d_{\text{Grass}(l,n)}^*(\mathbb{B} + c, \Psi_+(\mathbb{A} + b))$$

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

$$\begin{aligned} \delta^\alpha(\mathbb{A} + b, \mathbb{B} + c) &= \theta_{k+1}, & \delta^\beta(\mathbb{A} + b, \mathbb{B} + c) &= \left(1 - \prod_{i=1}^{k+1} \cos^2 \theta_i\right)^{1/2}, \\ \delta^\pi(\mathbb{A} + b, \mathbb{B} + c) &= \sin \theta_{k+1}, & \delta^\mu(\mathbb{A} + b, \mathbb{B} + c) &= \left(\log \prod_{i=1}^{k+1} \frac{1}{\cos^2 \theta_i}\right)^{1/2}, \\ \delta^\sigma(\mathbb{A} + b, \mathbb{B} + c) &= 2 \sin(\theta_{k+1}/2), & \delta^\phi(\mathbb{A} + b, \mathbb{B} + c) &= \cos^{-1}\left(\prod_{i=1}^{k+1} \cos \theta_i\right), \\ \delta^\kappa(\mathbb{A} + b, \mathbb{B} + c) &= \left(\sum_{i=1}^{k+1} \sin^2 \theta_i\right)^{1/2}, & \delta^\rho(\mathbb{A} + b, \mathbb{B} + c) &= \left(2 \sum_{i=1}^{k+1} \sin^2(\theta_i/2)\right)^{1/2}, \end{aligned}$$

where $\theta_1, \dots, \theta_{k+1}$ are as defined above.

Like the δ in Theorem 8, the δ^* 's in Theorem 9 are distances in the sense of distances from a point to a set, but they are not *metrics*. The doubly infinite Grassmannian of linear subspaces of all dimensions $\text{Gr}(\infty, \infty)$ has been shown to be metrizable [41, Section 5] with respect to any of the common distances between linear subspaces. Our last observation is that $\text{Grass}(\infty, \infty)$ can likewise be metricized, i.e., a metric can be defined between any pair of affine subspaces of arbitrary dimensions. The embedding $j : \text{Grass}(k, n) \rightarrow \text{Gr}(k + 1, n + 1)$ induces an embedding of sets $j_\infty : \text{Grass}(\infty, \infty) \rightarrow \text{Gr}(\infty, \infty)$. So $\text{Grass}(\infty, \infty)$ may be identified with $j_\infty(\text{Grass}(\infty, \infty))$ and regarded as a subset of $\text{Gr}(\infty, \infty)$. It inherits any metric on $\text{Gr}(\infty, \infty)$: If $\mathbb{A} + b$ and $\mathbb{B} + c$ are affine subspaces of possibly different dimensions, we may define

$$d_{\text{Grass}(\infty,\infty)}^*(\mathbb{A} + b, \mathbb{B} + c) := d_{\text{Gr}(\infty,\infty)}^*(j_\infty(\mathbb{A} + b), j_\infty(\mathbb{B} + c)),$$

for any choice of metric $d_{\text{Gr}(\infty,\infty)}^*$ on $\text{Gr}(\infty, \infty)$. For example, the metrics in Table 3 correspond to Grassmann, chordal, and Procrustes distances.

Table 3 Metrics on $\text{Graff}(\infty, \infty)$ in terms of affine principal angles and $k = \dim \mathbb{A}, l = \dim \mathbb{B}$

Grassmann metric	$d_{\text{Graff}(\infty, \infty)}(\mathbb{A} + b, \mathbb{B} + c) = \left(k - l \pi^2 / 4 + \sum_{i=1}^{\min(k+1, l+1)} \theta_i^2 \right)^{1/2}$
Chordal metric	$d_{\text{Graff}(\infty, \infty)}^k(\mathbb{A} + b, \mathbb{B} + c) = \left(k - l + \sum_{i=1}^{\min(k+1, l+1)} \sin^2 \theta_i \right)^{1/2}$
Procrustes metric	$d_{\text{Graff}(\infty, \infty)}^l(\mathbb{A} + b, \mathbb{B} + c) = \left(k - l + 2 \sum_{i=1}^{\min(k+1, l+1)} \sin^2(\theta_i/2) \right)^{1/2}$

9 Probability on the affine Grassmannian

To do statistical estimation and inference with affine subspace-valued data, i.e., with $\text{Graff}(k, n)$ in place of $\mathbb{R}^n = \text{Graff}(0, n)$, we will need reasonable notions of probability densities on $\text{Graff}(k, n)$. We introduce three here: uniform, Langevin (or von Mises–Fisher), and Langevin–Gaussian.

The Riemannian metric on $\text{Gr}(k, n)$ that induces the Grassmann distance in (8.1) also induces a volume density $d\gamma_{k,n}$ on $\text{Gr}(k, n)$ [29, Proposition 9.1.12] with

$$\text{Vol}(\text{Gr}(k, n)) = \int_{\text{Gr}(k, n)} |d\gamma_{k,n}| = \binom{n}{k} \frac{\prod_{j=1}^n \omega_j}{(\prod_{j=1}^k \omega_j)(\prod_{j=1}^{n-k} \omega_j)}, \tag{9.1}$$

where $\omega_m := \pi^{m/2} / \Gamma(1 + m/2)$, volume of the unit ball in \mathbb{R}^m . A natural uniform probability density on $\text{Gr}(k, n)$ is given by $d\mu_{k,n} := \text{Vol}(\text{Gr}(k, n))^{-1} |d\gamma_{k,n}|$.

By Theorem 1(ii), $\text{Graff}(k, n)$ is a Zariski open dense subset in $\text{Gr}(k + 1, n + 1)$ and we must have $\mu_{k+1, n+1}(\text{Graff}(k, n)) = 1$. Therefore the restriction of $\mu_{k+1, n+1}$ to $\text{Graff}(k, n)$ gives us a *uniform probability measure* on $\text{Graff}(k, n)$. It has an interesting property—a volumetric analogue of Theorem 8: The probability that a randomly chosen l -dimensional affine subspace contains $\mathbb{A} + b$ equals the probability that a randomly chosen k -dimensional affine subspace is contained in $\mathbb{B} + c$.

Theorem 10 *Let $k \leq l \leq n$ be such that $k + l \geq n$. Let $\mathbb{A} + b \in \text{Graff}(k, n)$ and $\mathbb{B} + c \in \text{Graff}(l, n)$. The relative volume of $\Psi_+(\mathbb{A} + b)$ in $\text{Graff}(l, n)$ and $\Psi_-(\mathbb{B} + c)$ in $\text{Graff}(k, n)$ are identical. Furthermore, their common value does not depend on the choices of $\mathbb{A} + b$ and $\mathbb{B} + c$ but only on k, l, n and is given by*

$$\mu_{l+1, n+1}(\Psi_+(\mathbb{A} + b)) = \mu_{k+1, n+1}(\Psi_-(\mathbb{B} + c)) = \frac{(l + 1)!(n - k)! \prod_{j=l-k+1}^{l+1} \omega_j}{(n + 1)!(l - k)! \prod_{j=n-k+1}^{n+1} \omega_j}.$$

Proof By Theorem 1(ii), we have

$$\text{Vol}(\text{Graff}(k, n)) = \text{Vol}(\text{Gr}(k + 1, n + 1)) = \binom{n + 1}{k + 1} \frac{\prod_{j=1}^{n+1} \omega_j}{(\prod_{j=1}^{k+1} \omega_j)(\prod_{j=1}^{n-k} \omega_j)}.$$

By Proposition 6, we have

$$\text{Vol}(\Psi_+(\mathbb{A} + b)) = \text{Vol}(\text{Gr}(n - l, n - k)) = \binom{n - k}{n - l} \frac{\prod_{j=1}^{n-k} \omega_j}{(\prod_{j=1}^{n-l} \omega_j)(\prod_{j=1}^{l-k} \omega_j)},$$

$$\text{Vol}(\Psi_-(\mathbb{B} + c)) = \text{Vol}(\text{Graff}(k, l)) = \binom{l + 1}{k + 1} \frac{\prod_{j=1}^{l+1} \omega_j}{(\prod_{j=1}^{k+1} \omega_j)(\prod_{j=1}^{l-k} \omega_j)}.$$

Dividing $\text{Vol}(\Psi_+(\mathbb{A}+b))$ and $\text{Vol}(\Psi_-(\mathbb{B}+c))$ by $\text{Vol}(\text{Graff}(l, n))$ and $\text{Vol}(\text{Graff}(k, n))$, respectively, completes the proof. □

In the following, we will use the projection coordinates in Definition 3. By embedding $\text{Graff}(k, n)$ as a subset $X = j(\text{Graff}(k, n)) \subseteq \text{Gr}(k + 1, n + 1)$ as in Theorem 1(ii) and noting that X is an open dense subset, we have $\mu(X) = 1$ for any Borel probability measure μ on $\text{Gr}(k + 1, n + 1)$ (and that $\mu(X^c) = 0$). Hence $\text{Graff}(k, n)$ inherits any continuous probability distribution on $\text{Gr}(k + 1, n + 1)$, in particular the Langevin distribution [10].

Definition 7 The *Langevin distribution*, also known as the *von Mises–Fisher distribution*, on $\text{Graff}(k, n)$ is given by the probability density function

$$f_L(P_{\mathbb{A}+b} | S) := \frac{1}{{}_1F_1(\frac{1}{2}(k + 1); \frac{1}{2}(n + 1); S)} \exp(\text{tr}(SP_{\mathbb{A}+b}))$$

for any $\mathbb{A} + b \in \text{Graff}(k, n)$. Here $S \in \mathbb{R}^{(n+1) \times (n+1)}$ is symmetric and ${}_1F_1$ is the confluent hypergeometric function of the first kind of a matrix argument [23].

The function ${}_1F_1(a; b; S)$ has well-known expressions as series and integrals and may be characterized via functional equations and recurrence relations. However, its explicit expression is unimportant for us—the only thing to note is that it can be efficiently evaluated [23] for any $a, b \in \mathbb{C}$ and symmetric $S \in \mathbb{C}^{(n+1) \times (n+1)}$.

Roughly speaking, the parameter $S \in \mathbb{R}^{(n+1) \times (n+1)}$ may be interpreted as a ‘mean direction’ and its eigendecomposition $S = V \Lambda V^T$ gives an ‘orientation’ $V \in O(n + 1)$ with ‘concentrations’ $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{n+1})$. In some sense, the Langevin distribution measures the first-order ‘spread’ on $\text{Graff}(k, n)$. If $S = 0$, then the distribution reduces to the uniform distribution, but if S is ‘large’ (i.e., $|\lambda_i|$ ’s are large), then the distribution concentrates about the orientation V . It may appear that a ‘Bingham distribution’ that measures second-order ‘spread’ can be defined on $\text{Graff}(k, n)$ by

$$f_B(P_{\mathbb{A}+b} | S) := \frac{1}{{}_1F_1(\frac{1}{2}(k + 1); \frac{1}{2}(n + 1); S)} \exp(\text{tr}(P_{\mathbb{A}+b}SP_{\mathbb{A}+b})),$$

but this is identical to the Langevin distribution as $\text{tr}(PSP) = \text{tr}(SP^2) = \text{tr}(SP)$ for any projection matrix P .

The Langevin distribution treats an affine subspace $\mathbb{A} + b \in \text{Graff}(k, n)$ as a single object, but there are occasions where it is desirable to distinguish between the

linear subspace $\mathbb{A} \in \text{Gr}(k, n)$ and the displacement vector $b \in \mathbb{R}^n$. We will show how a probability distribution on $\text{Graff}(k, n)$ may be constructed by amalgamating probability distributions on $\text{Gr}(k, n)$ and \mathbb{R}^n (or rather, \mathbb{R}^{n-k} , as we will see). First, we will identify $\text{Gr}(k, n)$ and $\text{Graff}(k, n)$ with their projection affine coordinates, i.e., imposing equality in (5.3) and (5.4),

$$\begin{aligned} \text{Gr}(k, n) &= \{P \in \mathbb{R}^{n \times n} : P^\top = P^2 = P, \text{tr}(P) = k\}, \\ \text{Graff}(k, n) &= \{[P, b] \in \mathbb{R}^{n \times (n+1)} : P \in \text{Gr}(k, n), Pb = 0\}. \end{aligned}$$

We will define a marginal density on the linear subspaces, and then impose a conditional density on the displacement vectors in the orthogonal complement of the respective linear subspaces.

For concreteness, we use the Langevin distribution $f_L(P | S)$ on the linear spaces $P \in \text{Gr}(k, n)$. Conditioning on P , we know there exists $Q \in O(n)$ such that $\ker(P) = \{b \in \mathbb{R}^n : Pb = 0\} = QE_{n-k} \cong \mathbb{R}^{n-k}$, where $E_{n-k} := \text{span}\{e_1, \dots, e_{n-k}\} \subseteq \mathbb{R}^{n+1}$. We may use any probability distribution on $\ker(P) \cong \mathbb{R}^{n-k}$ but again for concreteness, a natural choice is the *spherical Gaussian distribution* with probability density $f_G(x | \sigma^2) := (2\pi\sigma^2)^{-(n-k)/2} \exp(-\|x\|^2/2\sigma^2)$. The conditional density on $\ker(P)$ is then

$$f_G(b | P, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^{n-k}}} \exp\left(-\frac{\|b\|^2}{2\sigma^2}\right) \tag{9.2}$$

for any $b \in \ker(P)$. Note that $Q^\top b = \begin{bmatrix} b' \\ 0 \end{bmatrix}$ where $b' \in \mathbb{R}^{n-k}$ and since $\|b\| = \|Q^\top b\| = \|b'\|$, it is fine to have b instead of b' appearing on the RHS of (9.2). The construction gives us the following distribution.

Definition 8 The probability density function of the *Langevin–Gaussian distribution* on $\text{Graff}(k, n)$ is $f_{LG}([P, b] | S, \sigma^2) := f_L(P | S)f_G(b | P, \sigma^2)$, i.e.,

$$f_{LG}([P, b] | S, \sigma^2) = \frac{1}{{}_1F_1\left(\frac{1}{2}k; \frac{1}{2}n; S\right)\sqrt{(2\pi\sigma^2)^{n-k}}} \exp\left(\text{tr}(SP) - \frac{\|b\|^2}{2\sigma^2}\right),$$

where $S \in \mathbb{R}^{n \times n}$ is symmetric and $\sigma^2 > 0$.

10 Statistics on the affine Grassmannian

Even without going into the statistical analysis of affine subspace-valued data, we will see that the affine Grassmannian is hidden in plain sight in many standard problems of old-fashioned statistics and machine learning. Statistical estimation problems in multivariate data analysis and machine learning often seek linear relations among variables. This translates to finding an affine subspace from the sample data set that, in an appropriate sense, either best represents the data set or best separates it into components. In other words, statistical estimation problems are often optimization problems on the affine Grassmannian. We present four examples to illustrate this,

following conventional statistical notations $(n, p, X, \beta, \text{etc.})$. For the rest of this section, $\mathbb{1} = (1, \dots, 1)$ will denote a vector of all ones of appropriate dimension.

In the next two examples, we look at two common regression problems from this perspective, one that agrees with the intuitive idea of regression—finding a best-fitting affine hyperplane through a collection of n scattered data points $(x_i, y_i) \in \mathbb{R}^p \times \mathbb{R} = \mathbb{R}^{p+1}$, $i = 1, \dots, n$. The sole difference between Examples 2 and 3 is in the interpretation of “best-fitting.”

Example 2 (Linear Regression) Consider a linear regression problem with $X \in \mathbb{R}^{n \times p}$, a design matrix of explanatory variables, and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, a vector of response variables. In linear regression, we seek parameters $\beta = (\beta_0, \beta_1, \dots, \beta_p) \in \mathbb{R}^{p+1}$ such that

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n.$$

Note that unlike the next example, the assumption here is that the observational errors ε_i ’s occur only in the response variables y_i ’s but not in the explanatory variables x_{ij} ’s. The parameters $\beta_0, \beta_1, \dots, \beta_p$ are then estimated by minimizing the *ordinary least squares* error $\varepsilon_1^2 + \dots + \varepsilon_n^2$, or, equivalently, $\|[\mathbb{1}, X]\beta - y\|^2$. Geometrically, this yields an affine hyperplane in \mathbb{R}^{p+1} , i.e., $\mathbb{A} + b \in \text{Graff}(p, p + 1)$, given by

$$\begin{aligned} \mathbb{A} + b &= \{z \in \mathbb{R}^{p+1} : \beta_0 + \beta_1 z_1 + \dots + \beta_p z_p - z_{p+1} = 0\} \\ &= \{(w, \beta_0 + \beta_1 w_1 + \dots + \beta_p w_p) \in \mathbb{R}^{p+1} : w \in \mathbb{R}^p\}, \end{aligned} \tag{10.1}$$

where $\beta \in \mathbb{R}^{p+1}$ is the least squares estimator. The affine hyperplane $\mathbb{A} + b$ best fits the given data in the sense of ordinary least squares error, as illustrated on the left of Fig. 3. □

Example 3 (Errors-in-Variables Regression) We follow the same notations as in the above example. In errors-in-variables regression, we do not distinguish between response and explanatory variables and assume that observational errors occur in all variables. In this case, $\beta = (\beta_0, \beta_1, \dots, \beta_p) \in \mathbb{R}^{p+1}$ is obtained by minimizing over all $\beta \in \mathbb{R}^{p+1}$, $y' \in \mathbb{R}^n$, and $X' \in \mathbb{R}^{n \times p}$, the *total least squares error* $\|X' - X\|^2 + \|y' - y\|^2$ subjected to the condition $y' = [\mathbb{1}, X']\beta$. One can show that the total least squares estimator $\beta \in \mathbb{R}^{p+1}$ may also be obtained by minimizing

$$\frac{\|[\mathbb{1}, X]\beta - y\|^2}{1 + \|\beta\|^2 - \beta_0^2},$$

and the corresponding affine hyperplane $\mathbb{B} + c \in \text{Graff}(p, p + 1)$ is given by

$$\begin{aligned} \mathbb{B} + c &= \{z \in \mathbb{R}^{p+1} : \beta_0 + \beta_1 z_1 + \dots + \beta_p z_p - z_{p+1} = 0\} \\ &= \{(w, \beta_0 + \beta_1 w_1 + \dots + \beta_p w_p) \in \mathbb{R}^{p+1} : w \in \mathbb{R}^p\}. \end{aligned}$$

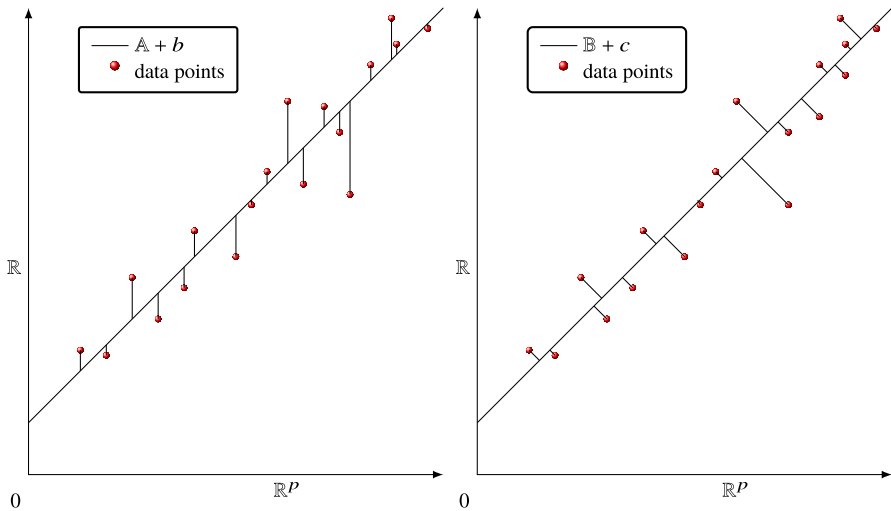


Fig. 3 Illustrations of linear regression (left) and errors-in-variables regression (right)

This appears to be identical to (10.1), but the difference here is that $\beta \in \mathbb{R}^{p+1}$ is the total least squares estimator, i.e., the affine hyperplane $\mathbb{B} + c$ best fits the given data in the sense of total least squares error, as illustrated on the right of Fig. 3. \square

Example 4 (Principal Component Analysis) Let $\bar{x} = \frac{1}{n} X^T \mathbb{1} \in \mathbb{R}^p$ be the sample mean of a data matrix $X \in \mathbb{R}^{n \times p}$. Write $\bar{X} = X - \mathbb{1}\bar{x}^T$ for the mean-centered data. Find $Z_k \in \mathbb{R}^{p \times k}$ that maximizes $\text{tr}(Z_k^T \bar{X}^T \bar{X} Z_k)$ subjected to $Z_k^T Z_k = I_k$. For $k \leq p$, the k -dimensional linear subspace $\text{im}(Z_k) \subseteq \mathbb{R}^p$ is the k th principal subspace of X and an orthonormal basis of $\text{im}(Z_k)$, defined successively for $k = 1, \dots, p$, gives the k largest principal components of X . The affine subspace

$$\text{im}(Z_k) + \bar{x} \in \text{Graff}(k, p)$$

then captures the greatest k -dimensional variability in the data X . \square

Example 5 (Support Vector Machine) Let $\{(x_i, y_i) \in \mathbb{R}^{p+1} : x_i \in \mathbb{R}^p, y_i = \pm 1, i = 1, \dots, n\}$ be a training set for binary classification. The maximum-margin hyperplane is given by $w^T x - \beta = 0$, where $(w, \beta) \in \mathbb{R}^p \times \mathbb{R}$ is found by minimizing $\|w\|$ subjected to $y_i(w^T x_i - \beta) \geq 1$ for all $i = 1, \dots, n$. Let $\hat{x} \in \mathbb{R}^p$ be any vector satisfying $w^T \hat{x} = \beta$. Then

$$\ker(w^T) + \hat{x} \in \text{Graff}(p, p + 1)$$

is the best separating hyperplane in the sense of support vector machines. \square

These four examples represent a sampling of the most rudimentary classical examples. It is straightforward to extend them to incorporate more modern considerations such

as sparsity or robustness by changing the objective function used; or have matrix variables in place of vector variables by considering affine subspaces within \mathbb{S}^n or $\mathbb{R}^{m \times n}$ instead of \mathbb{R}^n .

These simple examples may be solved in the usual manners with techniques in numerical linear algebra: least squares for linear regression, singular value decomposition for errors-in-variables regression, eigenvalue decomposition for principal component analysis, linear programming for support vector machines. Nevertheless, viewing them in their full generality as optimization problems on the affine Grassmannian allows us to treat them on equal footings and facilitates development of new multivariate statistics/machine learning techniques. More importantly, we argue that the prevailing approaches may be suboptimal. For instance, in Example 4 one circumvents the problem of finding a best-fitting affine subspace with a two-step heuristic: First find the empirical mean of the data set \bar{x} and then mean center to reduce the problem to one of finding a best-fitting linear subspace $\text{im}(Z)$. But there is no reason to expect $\text{im}(Z) + \bar{x}$ to be the best-fitting affine subspace. In [26], we developed various optimization algorithms—steepest decent, conjugate gradient, Newton method—that allow us to directly optimize real-valued functions on $\text{Graff}(k, n)$.

We would like to highlight another reason we expect the affine Grassmannian to be useful in data analytic problems. Over the past two decades, parameterizing a data set by geometric structures has become a popular alternative to probabilistic modeling, particularly when the intrinsic dimension of the data set is low or when it satisfies obvious geometric constraints. In this case, statistical estimation takes into account the intrinsic geometry of the data, and the deviation from the underlying geometric structures is used as a measure of accuracy of the statistical model. The two most common geometric structures employed are (a) a mixture of affine spaces [17,25,27] and (b) a manifold, which often reduces to (a) when it is treated as a collection of tangent spaces [38]—in fact, the first manifold learning techniques ISOMAP [36], LLE [31], and Laplacian Eigenmap [5] are essentially different ways to approximate a manifold by a collection of its tangent spaces. This provides another impetus for studying $\text{Graff}(k, n)$, which parameterizes all affine spaces of a fixed dimension in an ambient space; $\text{Graff}(k, \infty)$, which parameterizes all affine spaces of a fixed dimension; and $\text{Graff}(\infty, \infty)$, which parameterizes all affine spaces of all dimensions.

Acknowledgements The authors thank the two referees for their very helpful comments and suggestions. In particular, Example 1 was suggested by one of them. The work in this article is supported by DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, NSFC Grant no. 11801548, NSFC Grant no. 11688101 and National Key R&D Program of China Grant no. 2018YFA0306702. In addition, LHL's work is supported by a DARPA Director's Fellowship and the Eckhardt Faculty Fund; KY's work is supported by the Hundred Talents Program of the Chinese Academy of Sciences and the Recruitment Program of Global Experts of China.

References

1. ABSIL, P.-A., MAHONY, R., & SEPULCHRE, R. (2008) *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ.
2. ABSIL, P.-A., MAHONY, R., & SEPULCHRE, R. (2004) Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Appl. Math.*, **80**, no. 2, pp. 199–220.

3. ACHAR, P. N. & RIDER, L. (2015) Parity sheaves on the affine Grassmannian and the Mirković–Vilonen conjecture. *Acta Math.*, **215**, no. 2, pp. 183–216.
4. ALEKSEEVSKY, D. & ARVANITOYEORGOS, A. (2007) Riemannian flag manifolds with homogeneous geodesics. *Trans. Am. Math. Soc.*, **359**, no. 8, pp. 3769–3789.
5. BELKIN, M. & NIYOGE, P. (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. *Proc. Adv. Neural Inform. Process. Syst.*, **14**, pp. 586–691.
6. BJÖRCK, Å. & GOLUB, G. H. (1973) Numerical methods for computing angles between linear subspaces. *Math. Comput.*, **27** (1973), no. 123, pp. 579–594.
7. BOREL, A. (1953) Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts. *Ann. Math.*, **57**, pp. 115–207.
8. BROWN, E. H., JR (1982) The cohomology of BSO_n and BO_n with integer coefficients. *Proc. Am. Math. Soc.*, **85** (1982), no. 2, pp. 283–288.
9. CHERN, S.-S. (1948) On the multiplication in the characteristic ring of a sphere bundle. *Ann. Math.*, **49**, pp. 362–372.
10. CHIKUSE, Y. (2012) *Statistics on Special Manifolds*, Lecture Notes in Statistics, **174**, Springer, New York, NY.
11. CHIRIKJIAN, G. S. & KYATKIN, A. B. (2001) *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL.
12. EDELMAN, A., ARIAS, T., & SMITH, S. T. (1999) The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, **20**, no. 2, pp. 303–353.
13. FRENKEL, E. & GAITSGORY, D. (2009) Localization of \mathfrak{g} -modules on the affine Grassmannian. *Ann. Math.*, **170**, no. 3, pp. 1339–1381.
14. GOLUB, G. & VAN LOAN, C. (2013) *Matrix Computations*, 4th Ed., John Hopkins University Press, Baltimore, MD.
15. GORDON, C. S. (1996) 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.
16. GRIFFITHS, P. & HARRIS, J. (1994) *Principles of Algebraic Geometry*, Wiley, New York, NY.
17. HARO, G., RANDALL, G., & SAPIRO, G. (2006) Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds. *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, **26**, pp. 553–560.
18. HIRSCH, M. (1976) *Differential Topology*, Springer, New York, NY.
19. HODGE, W. V. D. & DANIEL, P. (1994) *Methods of Algebraic Geometry*, **2**, Cambridge University Press, Cambridge, UK.
20. HUSEMOLLER, D. (1994) *Fibre bundles*. Third edition. Graduate Texts in Mathematics, **20**. Springer, New York, 1994.
21. KLAIN, D. A. & ROTA, G.-C. (1997) *Introduction to Geometric Probability*, Cambridge University Press.
22. KLEIMAN, S. L., & LAKSOV, D. (1972) Schubert calculus. *Am. Math. Monthly*, **79** (1972), pp. 1061–1082.
23. KOEV, P. & EDELMAN, A. (2006) The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comput.*, **75**, no. 254, pp. 833–846.
24. LAM, T. (2008) Schubert polynomials for the affine Grassmannian. *J. Am. Math. Soc.*, **21**, no. 1, pp. 259–281.
25. LERMAN, G. & ZHANG, T. (2011) Robust recovery of multiple subspaces by geometric l_p minimization. *Ann. Statist.*, **39**, no. 5, pp. 2686–2715.
26. LIM, L.-H., WONG, K. S.-W., & YE, K. (2019) Numerical algorithms on the affine Grassmannian, *SIAM J. Matrix Anal. Appl.*, **40**, no. 2, pp. 371–393.
27. MA, Y., YANG, A., DERKSEN, H., & FOSSUM, R. (2008) Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Rev.*, **50**, no. 3, pp. 413–458.
28. MILNOR, J. W. & STASHEFF, J. D. (1974) *Characteristic Classes*, Annals of Mathematics Studies, **76**, Princeton University Press, Princeton, NJ.
29. NICOLAESCU, L. I. (2007) *Lectures on the Geometry of Manifolds*, 2nd Ed., World Scientific, Hackensack, NJ.
30. PRASOLOV, V. V. (2007) *Elements of Homology Theory*, Graduate Studies in Mathematics, **81**, AMS, Providence, RI.

31. ROWEIS, S. T. & SAUL, L. K. (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, no. 5500, pp. 2323–2326.
32. SATO, M. & SATO, Y. (1983) Soliton equations as dynamical systems on infinite-dimensional Grassmann manifold. *Nonlinear partial differential equations in applied science* (Tokyo, 1982), pp. 259–271, North-Holland Math. Stud., **81**, Lecture Notes Numer. Appl. Anal., **5**, North-Holland, Amsterdam.
33. SOTTILE, F. (2001) Schubert calculus, Schubert cell, Schubert cycle, and Schubert polynomials. pp. 343–346 in HAZEWINDEL, M. (Ed.) *Encyclopaedia of Mathematics*, Supplement III, Kluwer Academic Publishers, Dordrecht, Netherlands.
34. ST. THOMAS, B., LIN, L., LIM, L.-H., & MUKHERJEE, S. (2014) Learning subspaces of different dimensions. Preprint [arXiv:1404.6841](https://arxiv.org/abs/1404.6841).
35. TAKEUCHI, M. (1962) On Pontrjagin classes of compact symmetric spaces. *J. Fac. Sci. Univ. Tokyo Sect.*, Iss. **9**, pp. 313–328.
36. TENENBAUM, J., DE SILVA, V., & LANGFORD, J. (2000) A global geometric framework for nonlinear dimensionality reduction, *Science*, **290**, no. 5500, pp. 2319–2323.
37. THOMAS, E. (1960) On the cohomology of the real Grassmann complexes and the characteristic classes of n -plane bundles. *Trans. Am. Math. Soc.*, **96**, pp. 67–89.
38. TYAGI, H., VURAL, E., & FROSSARD, P. (2013) Tangent space estimation for smooth embeddings of Riemannian manifolds. *Inf. Inference*, **2**, no. 1, pp. 69–114
39. VIRO, O. Y. & FUCHS, D. B. *Homology and Cohomology*, Encyclopaedia Math. Sci., **24**, Topology II, pp. 95–196, Springer, Berlin, 2004.
40. WONG, Y.-C. (1967) Differential geometry of Grassmann manifolds. *Proc. Nat. Acad. Sci.*, **57**, no. 3, pp. 589–594.
41. YE, K. & LIM, L.-H. (2016) Schubert varieties and distances between linear spaces of different dimensions. *SIAM J. Matrix Anal. Appl.*, **37**, no. 3, pp. 1176–1197.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Lek-Heng Lim¹ · Ken Sze-Wai Wong² · Ke Ye³

✉ Lek-Heng Lim
lekheng@galton.uchicago.edu

Ken Sze-Wai Wong
kenwong@uchicago.edu

Ke Ye
keyk@amss.ac.cn

¹ Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, Chicago, IL 60637, USA

² Department of Statistics, University of Chicago, Chicago, IL 60637, USA

³ KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China