Fitting Flats to Flats

Gabriel Dogadov   Ugo Finnendahl   Marc Alexa

TU Berlin, Computer Graphics Group

`cg.tu-berlin.de`

Abstract

Affine subspaces of Euclidean spaces are also referred to as flats. A standard task in computer vision, or more generally in engineering and applied sciences, is fitting a flat to a set of points, which is commonly solved using the PCA. We generalize this technique to enable fitting a flat to a set of other flats, possibly of varying dimensions, based on representing the flats as squared distance fields. Compared to previous approaches such as Riemannian centers of mass in the manifold of affine Grassmannians, our approach is conceptually much simpler and computationally more efficient, yet offers desirable properties such as respecting symmetries and being equivariant to rigid transformations, leading to more intuitive and useful results in practice. We demonstrate these claims in a number of synthetic experiments and a multi-view reconstruction task of line-like objects.

1. Introduction

Affine subspaces of Euclidean spaces, also called *flats*, are the basic entities in affine geometry, and are foundational building blocks for computations in computer vision, computer graphics, and generally all engineering disciplines. To motivate this work with a concrete example, notice that a point observed in a calibrated camera gives rise to a line in 3-space. Observing the same point in several cameras allows recovering its position by fitting a point to the collection of lines. Minimizing the squared distances to the lines only requires solving a linear system. Now imagine the features observed in the images are lines, giving rise to planes in each camera view. We would like to fit a line to the set of observed planes. Surprisingly, despite the problem appearing almost identical, there is no established 'standard' solution to this problem.

Conversely, we often have to work with samples (observations) in some Euclidean space. The PCA [29] is arguably the standard way to fit a line or plane to the observation if they are points in space – but what do we do if the observations are lines or planes?

Approaching these problems in a systematic way commonly leads to *Grassmannians*, the manifold of *linear* subspaces. Similar to how homogeneous coordinates of Euclidean space lead to a representation of projective space, Grassmannians can be projectivized and then represent flats. The first and likely best-known example are Plücker coordinates for affine lines in $\mathbb{R}^3$ [5, 16, 27]. One may argue that the resulting spaces are not strictly representing flats, as they also contain ideal elements, i.e., flats 'at infinity'. This is rectified in the explicit construction of a Grassmannian of affine subspaces [23]. While the manifold of the former construction is represented by the Klein quadric (and its generalizations), the latter is described by orthogonality conditions, a so-called Stiefel manifold (we review several representations of flats in Sec. 2). It has been argued [22] that computation on Stiefel manifolds is more convenient and better established [1, 2, 8]. Still, the mathematical sophistication and complexity of necessary computations for something as mundane as the mean of a set of affine lines or planes in 3D, let alone fitting flats to flats of different dimensions, is baffling when one compares it to the simplicity of fitting a point or fitting to points.

In this work, we provide a simple framework, based on representing flats as squared distance fields (see Sec. 2 for details on the representation and the relation to other representations). As we derive in Sec. 4, it allows fitting a flat of desired dimension k to a given set of flats of arbitrary and possibly varying dimension. Importantly, we show that it is equivariant under rigid transformations, the isometries of affine space. Unlike Plücker coordinates and related representations, the representation is *unoriented*, which together with equivariance immediately implies that both angles as well as distances are bisected by least-squares fits to two flats. To our knowledge, this is the first method with these very natural properties. In addition, the necessary computations are similar to fitting flats to points with the PCA, with the most complex operation being the eigendecomposition of a symmetric PSD matrix. An interpretation of the procedure as projected means in ambient space leads to extensions to L_p means, enabling more robust fitting in the presence of outliers using the L_1-norm.

We evaluate the properties of our approach in comparison to Riemannian means in the space of affine Grassmannians on synthetic examples in low dimension (Sec. 5). We verify the predicted properties of both classes of methods. In addition, we demonstrate the method at the example of reconstructing line-like objects from multiple views.

The basic approach leaves much room for applying more advanced statistical methods, some of which we discuss in Sec. 6.

2. Background: Representations of flats

In the following, we assume the ambient space is $\mathbb{R}^d$ (unless otherwise noted). A *k-flat* is a k-dimensional affine subspace of the ambient space. In contrast, a *k-plane* is a linear subspace. Every k-plane is a k-flat, but a k-flat is a k-plane only if it passes through the origin. For $d = 3$, flats can be points ($k = 0$), lines ($k = 1$), or planes ($k = 2$). A $(d - 1)$-flat in $\mathbb{R}^d$ is called *hyperplane*. The set of all k-planes in $\mathbb{R}^d$ is the *Grassmannian of linear subspaces* $\mathrm{Gr}(k, d)$, a smooth compact manifold of dimension $k(d - k)$. Similarly, the set of all k-flats is called the *Grassmannian of affine subspaces* or short the *affine Grassmannian* $\mathrm{Graff}(k, d)$ [19], a smooth $(k + 1)(d - k)$-dimensional manifold. While the Grassmannian, representations, and computational methods are well established, the affine Grassmannian is a more recent development [22, 23], but created interest for applications such as image classification [30] or LiDAR registration [24].

The *orthogonal complement* of a k-plane $\mathcal{P} \subseteq \mathbb{R}^d$ is a linear subspace denoted $\mathcal{P}^\perp$ and contains all vectors that are orthogonal to $\mathcal{P}$. The dimension of $\mathcal{P}^\perp$ is the *co-dimension* $\bar{k} = d - k$ of $\mathcal{P}$.

In the following, we describe several representations of flats, i.e., elements of $\mathrm{Graff}(k, d)$, needed for the ensuing discussion of existing and new techniques for least squares fitting flats to flats.

Parameter form A matrix $\mathbf{A} \in \mathbb{R}^{d \times k}$ with full column rank represents a linear k-space in $\mathbb{R}^d$. A flat $\mathcal{F}$ can be represented by additionally specifying a displacement $\mathbf{b} \in \mathbb{R}^d$:

$$\mathcal{F} = \left\{ \mathbf{x} \in \mathbb{R}^d : \mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{b}, \mathbf{y} \in \mathbb{R}^k \right\}. \tag{1}$$

We ask that the columns of $\mathbf{A}$ are orthonormal and that $\mathbf{A}^\mathsf{T}\mathbf{b} = \mathbf{0}$. This implies that $\mathbf{b}$ is the vector to the point closest to the origin and is unique. Under this condition, the representation $\mathbf{y}$ in Eq. (1) has been referred to as *orthogonal affine coordinates* [22, 23]. We may arbitrarily change the basis of these coordinates with an orthogonal transformation $\mathbf{O} \in \mathbb{R}^{k \times k}$ without affecting the flat. So $(\mathbf{A}\mathbf{O}, \mathbf{b})$, $\mathbf{O}^\mathsf{T}\mathbf{O} = \mathbf{I}$ describes the set of equivalent representations of $\mathcal{F}$.

Stiefel and Grassmann-Plücker coordinates The affine Grassmannian in parameter form can be embedded into a (standard) Grassmannian one dimension higher, similar to Euclidean points being represented as lines in homogeneous coordinates: given a k-flat $(\mathbf{A}, \mathbf{b})$, it is treated as a $k + 1$-plane in $\mathbb{R}^{d+1}$ spanned by $(\mathbf{A}^\mathsf{T}, \mathbf{0})^\mathsf{T}$ and $(\mathbf{b}^\mathsf{T}, 1)^\mathsf{T}$. Normalizing the last basis vector leads to the (homogeneous) *Stiefel coordinates* [11, 22] for a k-flat $\mathcal{F}$:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{b}/\sqrt{1 + \|\mathbf{b}\|^2} \\ \mathbf{0}^\mathsf{T} & 1/\sqrt{1 + \|\mathbf{b}\|^2} \end{bmatrix}. \tag{2}$$

This representation of a flat still admits orthogonal transformations that map $\mathrm{span}(\mathbf{Y})$ into $\mathrm{span}(\mathbf{Y})$. To reduce the degrees of freedom, one may represent the parallelotope spanned by the basis of this linear space as the *exterior product* of the basis vectors. This construction is known as the *Plücker embedding* [27]. Intuitively, one may think of the representation as the signed (hyper)areas of the shadows of the parallelotope onto all k-planes spanned by the axes of a fixed coordinate system. These (hyper)areas are independent of the choice of the vectors $\mathbf{A}'$ spanning the parallelotope, with the only degree of freedom being the total signed volume of the parallelotope. This suggests that the resulting representation consisting of $\binom{d}{k}$ elements is unique up to scale for a given flat. In general, $\binom{d}{k} > (k+1)(d-k)$, so only a subset of the coordinates corresponds to flats. This subset is described as the intersection of quadratic surfaces, the *Grassmann-Plücker relations*. The particular case of affine lines ($k=1$) in ($d=3$)-space of this representation is well known as *Plücker coordinates*[1].

Normal form We may also represent the flat $\mathcal{F}$ as the intersection of $\bar{k}$ hyperplanes, each defined as $\mathbf{n}_i^\mathsf{T}\mathbf{x} = c_i$. Writing the normal vectors $\mathbf{n}_i$ as rows of a matrix $\mathbf{N} \in \mathbb{R}^{\bar{k} \times d}$ and the offsets as a vector $\mathbf{c} \in \mathbb{R}^{\bar{k}}$, the flat is represented as:

$$\mathcal{F} = \left\{ \mathbf{x} \in \mathbb{R}^d : \mathbf{N}\mathbf{x} = \mathbf{c} \right\}. \tag{3}$$

Note that the row space of $\mathbf{N}$ is the orthogonal complement of the column space of $\mathbf{A}$ in $\mathbb{R}^d$. As before, we ask that the basis is chosen orthonormal. Under this condition, the vector $\mathbf{c}$ contains the signed distances of the hyperplanes to the origin (see Fig. 1 (left) for an illustration). Note that equality in Eq. (3) is preserved for any change of basis $\mathbf{O} \in \mathbb{R}^{\bar{k} \times \bar{k}}$ and that the rows of $\mathbf{O}\mathbf{N}$ remain orthonormal if $\mathbf{O}$ is orthogonal. Thus, in normal form, the set $(\mathbf{O}\mathbf{N}, \mathbf{O}\mathbf{c})$ describes the same flat $\mathcal{F}$.

---

[1]The original formulation of Plücker coordinates starts from two points on the line in homogeneous coordinates – the construction starting from Stiefel coordinates mentioned above is equivalent, as both span the same linear space.
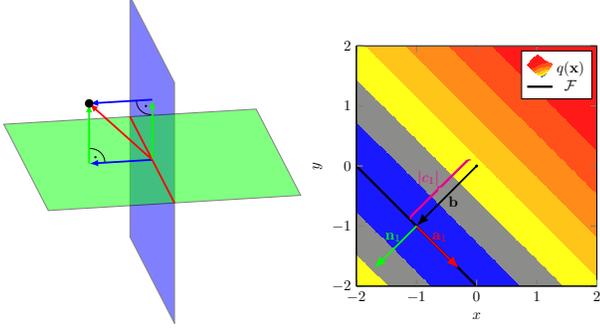
Figure 1. Left: The squared distance between a point (black) and a line (red) corresponds to the sum of squared distances to two orthogonal hyperplanes (blue and green) whose intersection is the line. Right: The connection between the squared distance field q of a line in $\mathbb{R}^2$, its parametric representation $(\mathbf{a}_1, \mathbf{b})$ and normal representation $(\mathbf{n}_1, c_1)$.

Squared distance function The normal representation of $\mathcal{F}$ yields the distances to the hyperplanes for any $\mathbf{x} \in \mathbb{R}^d$ as $\mathbf{Nx} - \mathbf{c}$. We can get the squared distance by computing the the squared norm of this vector:

$$d_{\mathcal{F}}^2(\mathbf{x}) = (\mathbf{Nx}-\mathbf{c})^\top(\mathbf{Nx}-\mathbf{c}) = \mathbf{x}^\top\mathbf{N}^\top\mathbf{Nx}-2\mathbf{c}^\top\mathbf{Nx}+\mathbf{c}^\top\mathbf{c}. \quad (4)$$

Note that the squared distance field is unaffected by starting from a different orthogonal normal frame $\mathbf{ON}$ and corresponding distance vector $\mathbf{Oc}$: the resulting distances are $\mathbf{O}(\mathbf{Nx} - \mathbf{c})$ and the inner product $\mathbf{O}^\top\mathbf{O} = \mathbf{I}$ cancels. This means the symmetric positive semi-definite (PSD) matrix $\mathbf{Q} = \mathbf{N}^\top\mathbf{N}$, the vector $\mathbf{r} = -\mathbf{N}^\top\mathbf{c}$ and the scalar $s = \mathbf{c}^\top\mathbf{c}$ uniquely describe a k-flat as

$$\mathcal{F} = \left\{\mathbf{x} \in \mathbb{R}^d : d_{\mathcal{F}}^2(\mathbf{x}) = \mathbf{x}^\top\mathbf{Qx} + 2\mathbf{r}^\top\mathbf{x} + s = 0\right\}. \quad (5)$$

Any non-zero scalar multiple of the triple $(\mathbf{Q}, \mathbf{r}, s)$ of the squared distance field has the same zero-set, so we may interpret this representation as a homogeneous coordinate for k-flats. Similar to Plücker-Grassmann coordinates, only a subset of symmetric matrices $\mathbf{Q}$, vectors $\mathbf{r}$, and scalars s correspond to k-flats. Note that $\mathbf{Q}$ contains the bases encoded by $\mathbf{A}$ and $\mathbf{N}$ as eigenspaces corresponding to the eigenvalues 0 and 1. So $\mathbf{Q}$ is characterized by the spectrum consisting of the eigenvalues 0 with multiplicity k and 1 with multiplicity $\bar{k}$.

The vector $\mathbf{r}$ has been constructed as $-\mathbf{N}^\top\mathbf{c}$, so it is a linear combination of the normals and has to be an eigenvector of $\mathbf{Q}$ with eigenvalue 1.

Since $\mathbf{Q}$ is PSD, the zero set is equivalent to the local minima of the squared distance field and could be computed by setting the gradient $2\mathbf{Qx} + 2\mathbf{r}$ to zero. This allows recovering s from $(\mathbf{Q}, \mathbf{r})$ by asking that the squared distances evaluate to zero for the minima, yielding

$$s = \mathbf{r}^\top\mathbf{Q}^+\mathbf{r} = \mathbf{r}^\top\mathbf{Qr} = \mathbf{r}^\top\mathbf{r}, \quad (6)$$

where we have exploited that $\mathbf{Q}$ is its own pesudo-inverse under the assumptions on its spectrum and that $\mathbf{r}$ is an eigenvector with eigenvalue one. Since s is redundant, we can represent flats by the pair $(\mathbf{Q}, \mathbf{r})$. We summarize its relation to the standard parameter form $(\mathbf{A}, \mathbf{b})$ below:

$$(\mathbf{Q}, \mathbf{r}) = (\mathbf{I} - \mathbf{AA}^\top, -\mathbf{b}) \qquad (\mathbf{A}, \mathbf{b}) = (\ker(\mathbf{Q}), -\mathbf{r}). \quad (7)$$

Fig. 1 (right) illustrates the connection to both standard and normal forms.

3. Related work: Riemannian centers of mass

Recall that the Grassmannian is a smooth compact manifold. Let $\{\mathcal{P}_i\}$ represent k-planes, then least-squares fitting a k-plane to this data implies

$$m = \underset{\mathcal{P} \in \mathrm{Gr}(k,d)}{\arg\min} \sum_i d^2(\mathcal{P}, \mathcal{P}_i), \quad (8)$$

where d is a metric. In fact, this is a way of computing a mean in Riemannian manifolds and has been aptly termed *Riemannian center of mass* [17], now often referred to as *Karcher mean*. Note that by embedding $\mathrm{Graff}(k, d)$ into $\mathrm{Gr}(k + 1, d + 1)$, we can use the same approach for fitting a k-flat to given k-flats. In the following, we will first explain the metric we have found to be commonly used in our context, and then how to extend the basic ideas to linear subspaces of varying dimensions.

Principal angles and metric The concept of angles between two lines (in the plane or in space), two planes or a line and a plane in space can be generalized to flats of any dimension. Given a k-plane represented by $\mathbf{A}_1$ and an l-plane represented by $\mathbf{A}_2$ and assuming $k \leq l$, the *principal angles* are a set of k mutual angles $0 \leq \theta_1 \leq \ldots \leq \theta_k \leq \frac{\pi}{2}$ that are defined recursively as:

$$\theta_i := \min\left\{\cos^{-1}\left(\frac{|\mathbf{u}^\top\mathbf{v}|}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) \;\middle|\; \begin{matrix} \mathbf{u}\in\mathrm{span}(\mathbf{A}_1), & \mathbf{u}^\top\mathbf{u}_j=0, \\ \mathbf{v}\in\mathrm{span}(\mathbf{A}_2), & \mathbf{v}^\top\mathbf{v}_j=0, \\ \forall j\in\{1,\ldots,i-1\} \end{matrix}\right\}. \quad (9)$$

The vectors $(\mathbf{u}_i, \mathbf{v}_i)$ forming the principal angles are the *principal vectors*. The principal angles and vectors can be computed via a (reduced) Singular Value Decomposition (SVD) [4]. When decomposing $\mathbf{A}_1^\top\mathbf{A}_2 = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k)$, the principal vectors are the columns of $\mathbf{A}_1\mathbf{U}$ and $\mathbf{A}_2\mathbf{V}$, respectively, and the corresponding principal angles are given by $\theta_i = \cos^{-1}(\sigma_i)$. A commonly used metric on $\mathrm{Gr}(k, d)$ constructed from the principal angles is $\left(\sum_{i=1}^k \theta_i^2\right)^{1/2}$.

While one can compute principal angles between flats in the very same way, the angles lack information about the displacement between the flats if they are not intersecting

(e.g., consider two skew lines in 3-space). The distance due to rotation of the linear subspaces and translation of the points closest to the origin are automatically reconciled by using the Stiefel coordinates introduced above. The resulting principal angles in this embedding have been referred to as *affine principal angles* [22]. Moreover, optimization as necessary here for computing the mean is well-understood in Stiefel manifolds [2], and we detail the computation of means next.

Computing the mean in Stiefel coordinates Let flats $\mathcal{F}_i$ be given and represented in Stiefel coordinates $\mathbf{Y}_i$. For Stiefel coordinate $\mathbf{Y}$, the gradient of the sum of squared distances in Eq. (8) is given by [18]

$$-\sum_{i=1}^{m} \exp_{\mathbf{Y}}^{-1}(\mathbf{Y}_i), \qquad (10)$$

with $\exp_{\mathbf{Y}}^{-1}(\mathbf{X})$ denoting the derivative of the geodesic that connects $\mathbf{Y}$ and $\mathbf{X}$. As the affine principal angles, similar to principal angles for planes as in Eq. (9), are also defined for flats of different dimensions, geodesic distances between flats of different dimensions and their gradient can be computed in a similar manner as described in [22, 36]. The gradient can be exploited to minimize the sum of squared geodesic distances in an iterative minimization scheme such as gradient descent or Newton's method on Stiefel manifolds [1, 2, 8]. This essentially requires re-orthogonalization of $\mathbf{Y}$ after each descent step.

In our implementation, we used a variable step size gradient descent scheme but avoided computing the Hessian. For further information on the gradient computation, our adaptation to flats of different dimensions, and the orthogonalization scheme we found to yield the best results, we refer to the supplementary material.

4. Method

Given a set of flats $\{\mathcal{F}_i\}$, we want to fit a flat $\hat{\mathcal{F}}$ with fixed dimension k in the least-squares sense. We start by revisiting the case of fitting a point, i.e., the case $k = 0$, and then extend the approach to $k > 0$.

Fitting a point The sum of squared distances to flats from an arbitrary point $\mathbf{x}$ can be written as the sum of the squared distance fields to the flats:

$$\sigma^2(\mathbf{x}) = \sum_i d_{\mathcal{F}_i}^2(\mathbf{x}^*) \qquad (11)$$

$$= \mathbf{x}^\mathsf{T} \left(\sum_i \mathbf{Q}_i \right) \mathbf{x} + 2 \left(\sum_i \mathbf{r}_i \right) \mathbf{x} + \sum_i s_i$$

$$= \mathbf{x}^\mathsf{T} \mathbf{Q}^* \mathbf{x} + 2\mathbf{r}^* \mathbf{x} + s^*. \qquad (12)$$

Algorithm 1: Closest Flat (Mean-SDF)

Input : Flats (of arbitrary dimension) as squared distance fields $(\mathbf{Q}_1, \mathbf{r}_1), \ldots, (\mathbf{Q}_m, \mathbf{r}_m)$
Output: k-Flat in parameter form $(\mathbf{A}, \mathbf{b})$
begin
    $\mathbf{Q}^* \leftarrow \sum_{i=1}^{m} \mathbf{Q}_i$
    $\mathbf{r}^* \leftarrow \sum_{i=1}^{m} \mathbf{r}_i$
    $\mathbf{U}\mathbf{D}\mathbf{U}^\mathsf{T} \leftarrow \mathbf{Q}^*$ // Eigendecomposition, see Eq. 14 for conventions
    $\mathbf{A} \leftarrow [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ // Eq. 15
    $\mathbf{Q}^{*^+} \leftarrow \mathbf{U}\mathbf{D}^+\mathbf{U}^\mathsf{T}$
    $\mathbf{b} \leftarrow (\mathbf{I} - \mathbf{A}\mathbf{A}^\mathsf{T})\mathbf{Q}^{*^+}\mathbf{r}^*$
end

The minimal value is attained at a critical point and can be found by setting the gradient to zero, yielding the necessary condition $\mathbf{Q}^*\hat{\mathbf{x}} = \mathbf{r}^*$. If this sum of squared distances has a unique minimum, the linear system has a unique solution.

Notice that simply adding the representations of the flats is quite convenient, particularly in applications where flats are being added dynamically: this is commonly exploited in computer graphics for modeling with piecewise planar surfaces. By associating squared distances with the planar polygons, it is easy to measure changes to the shape relative to the original shape, most prominently used in surface simplification [10]. Conversely, the same concept can be used for fitting planes to sample points [38].

Fitting discretely sampled flats Now let us consider the case $k = 1$, an affine line. Naturally, one would ask that the sum of squared distances over all points on the line are minimized. This leads to an indefinite integral that takes on a finite value only in degenerate cases. It is illuminating, however, to start with a finite number m of samples of the line: represent the line in parameter form as $y\mathbf{a} + \mathbf{b}$ with $\|\mathbf{a}\| = 1$ and sample it in parametric locations $\{y_j\}$. We assume that these locations are mean-unbiased, i.e. $\sum_j y_j = 0$. The sum of squared distances summed up over all samples is

$$\sum y_j^2 \mathbf{a}^\mathsf{T} \mathbf{Q}^* \mathbf{a} + m\mathbf{b}^\mathsf{T} \mathbf{Q}^* \mathbf{b} + 2m\mathbf{b}^\mathsf{T} \mathbf{r}^* + ms^*, \qquad (13)$$

where we have already removed all terms with the factor $\sum_j y_j = 0$. We notice that we can independently optimize for $\mathbf{a}$ and $\mathbf{b}$. For $\mathbf{a}$ we are looking to minimize $c\mathbf{a}^\mathsf{T}\mathbf{Q}\mathbf{a}$ with positive $c = \sum_j y_j^2$ subject to $\|\mathbf{a}\| = 1$. This means $\mathbf{a}$ is the eigenvector of $\mathbf{Q}^*$ corresponding to the smallest eigenvalue. The positive constant c has no consequences. For $\mathbf{b}$ we find that it has to satisfy $\mathbf{Q}^*\mathbf{b} = \mathbf{r}^*$ (regardless of the sampling $\{y_j\}$). The resulting pair $(\mathbf{a}, \mathbf{b})$ describes the desired line, but is not yet in standard form, because $\mathbf{b}$ is not necessarily orthogonal to $\mathbf{a}$. This can be achieved by replacing $\mathbf{b}$ with $\mathbf{b} - (\mathbf{a}^\mathsf{T}\mathbf{b})\mathbf{a}$.

It is straightforward to extend this analysis to k-flats for $k > 1$. We find that the basis $\mathbf{A}$ has to be chosen as the eigenvectors corresponding to the smallest k eigenvalues of $\mathbf{Q}^*$. To make this concrete, let

$$\mathbf{Q}^* = \mathbf{UDU}^\mathsf{T}, \quad \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I},$$
$$\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d), \quad 0 \le \lambda_0 \le \ldots \le \lambda_{d-1} \tag{14}$$

then $\mathbf{A}$ consists of the first k columns of $\mathbf{U}$:

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_d] \implies \mathbf{A} = [\mathbf{u}_1, \ldots, \mathbf{u}_k] \tag{15}$$

The condition $\mathbf{Q}^*\mathbf{b} = \mathbf{r}^*$ is independent of k and bringing it into standard parameter form leads to

$$\mathbf{b} = (\mathbf{I} - \mathbf{A}\mathbf{A}^\mathsf{T})\mathbf{Q}^{*^+}\mathbf{r}^*. \tag{16}$$

Note that this solution is independent of the number of samples and their distribution. In fact, it extends to any weighting of the point samples as long as the weighted mean is still zero. With appropriate weighting, we could make sure that the constant c remains finite for infinite sampling, including a dense sampling of the line. This means the choice for $(\mathbf{A}, \mathbf{b})$ described in Eqs. (15) and (16) solve the fitting problem for flats in the least-squares sense. Alg. 1 summarizes the described procedure in pseudocode.

Complexity   Assuming we start from flats given in parameter form $(\mathbf{A}_i, \mathbf{b}_i)$ building the matrices $\mathbf{Q}_i = \mathbf{I} - \mathbf{A}\mathbf{A}^\mathsf{T}$ is $\mathcal{O}(kd^2)$ for each flat. Summing up $\mathbf{Q}^*$ and $\mathbf{r}^*$ is linear in the number of flats and the number of coefficients in the matrices and vectors. The most complex operation is the necessary eigendecomposition of $\mathbf{Q}^*$, which is $\mathcal{O}(d^3)$. We implement this using known closed-form solutions for $d = 2, 3$ [7] and a symmetric QR algorithm [12] for $d > 3$. Overall, the complexity of fitting a k-flat in this way to a set of arbitrary flats is identical to the complexity of fitting flats to points using the PCA. This appears to be quite natural, yet it is remarkable that existing solutions (to our knowledge) are significantly more involved and much slower in practice (see also Sec. 5).

Uniqueness and degeneracies   All steps in the above fitting procedure are uniquely determined, except for selecting the eigenvectors $\mathbf{A}$ corresponding to the smallest k eigenvalues of $\mathbf{Q}^*$. This step assumes that λ_k is strictly smaller than λ_{k+1}.

The dependence of the spectrum of the sum of matrices on the spectra of the summands is generally quite involved [9, 20], but it is worth pointing out some special cases in our context: (1) If all input flats are points, then $\mathbf{Q}^*$ also represents a point, i.e., all its eigenvalues are identical. So while our approach works for arbitrary k-flats, it fails for the special case $k = 0$ for all flats (which is exactly

what the PCA solves). (2) Two orthogonal lines in $\mathbb{R}^2$, three mutually orthogonal planes in $\mathbb{R}^3$, or more generally d mutually orthogonal $d - 1$ flats in $\mathbb{R}^d$ result in an isotropic squared distance field, so all eigenvalues of $\mathbf{Q}^*$ are identical. (3) Similar statements can be made for subspaces, leading to parts of the spectrum being isotropic. For example, two non-intersecting lines in $\mathbb{R}^3$ with orthogonal directions uniquely define a plane, but every line in the plane spanned by the directions is an equally good fit.

Equivariance   It seems very natural to ask that least squares fitting is *isometry invariant*. The isometries of affine spaces are rigid transformations and reflections. This means we expect that if the input is rigidly transformed and reflected, the fitted flat undergoes the same transformation – it is *equivariant* to these transformations.

This property is not difficult to show for the procedure above. Instead of $\mathbf{x}$, we plug in the transformed point $\mathbf{Rx} + \mathbf{t}$, where $\mathbf{R}$ is an orthogonal transformation in $\mathbb{R}^d$ and $\mathbf{t} \in \mathbb{R}^d$ a translation. What we find is that in the sum of squared distances Eq. (13), the term $c\mathbf{a}^\mathsf{T}\mathbf{Q}^*\mathbf{a}$ transforms into $c\mathbf{a}^\mathsf{T}\mathbf{R}^\mathsf{T}\mathbf{Q}^*\mathbf{Ra}$, indicating that the direction vector $\mathbf{a}$ has to be rotated by $\mathbf{R}$, as expected. For the terms involving $\mathbf{b}$ we get $(\mathbf{b} + \mathbf{t})^\mathsf{T}\mathbf{Q}^*(\mathbf{b} + \mathbf{t}) + 2(\mathbf{b} + \mathbf{t})^\mathsf{T}\mathbf{r}^*$ and setting the gradient w.r.t. $\mathbf{b}$ to zero yields $\mathbf{Q}^*(\mathbf{b} + \mathbf{t}) = -\mathbf{r}^*$, again, as expected. Note that the standard form for $\mathbf{b}$ as computed in Eq. (16) is not necessarily translated by $\mathbf{t}$, as the projection onto $\mathbf{A}$ is removed.

It seems worth pointing out that, as natural as this property may appear, the Riemannian center on the affine Grassmannian varies with the translation of the coordinate system. We demonstrate this with examples in the supplementary material.

Elementary properties   At least for two k-flats, we can formulate what we expect for their mean: if the two flats intersect, the mean should be the bisector; and if they are parallel, the mean should be parallel and have the same distance to both of them.

In fact, both properties directly follow from the fact that the computations are symmetric in the inputs and equivariant to translations and reflections. To see this, translate the flats so that they are symmetric w.r.t. the origin. Then reflection at the origin has no effect on the input, so the output has to respect the symmetry as well.

Interpretation as projected mean   Consider the fitting as starting from a set of flats represented by the pairs $(\mathbf{Q}_i, \mathbf{r}_i)$ and generating the mean in this form as $(\hat{\mathbf{Q}}, \hat{\mathbf{r}})$. In this setting, we may interpret the procedure as: first, compute the mean in the ambient space of symmetric PSD matrices and Euclidean vectors and, second, project onto the manifold of matrices and vectors representing k-flats.

Clearly, $(\mathbf{Q}^*, \mathbf{r}^*)$ arise from the set of $\mathbf{Q}_i, \mathbf{r}_i$ as a mean (modulo irrelevant scale factors). Given $\mathbf{Q}^*$, we constructed $\mathbf{A}$ based on the eigenvector corresponding to the smallest eigenvectors. This suggests that the normal space $\mathbf{N}$ contains the eigenvectors corresponding to the largest eigenvectors. Or, in other words, the fitted flat is represented by $\hat{\mathbf{Q}}$ with the same eigenspace as $\mathbf{Q}^*$, but the k smallest eigenvalues mapped to zero and the remaining eigenvalues mapped to one. Recall the decomposition of $\mathbf{Q}^*$ in Eq. (14), then we get

$$\hat{\mathbf{Q}} = \mathbf{U}\operatorname{diag}(\underbrace{0, \ldots, 0}_{k \text{ times}}, 1, \ldots, 1)\mathbf{U}^\mathsf{T}. \qquad (17)$$

And $\hat{\mathbf{r}}$ is related to $\mathbf{b}$ in Eq. (16) simply by $\hat{\mathbf{r}} = -\mathbf{b} = -\hat{\mathbf{Q}}\mathbf{Q}^{*+}\mathbf{r}^*$.

We want to show that this mapping is an orthogonal projection. For the PSD matrix, we claim that any unitarily invariant norm in the space of symmetric PSD matrices is minimized. To see this, consider minimizing $\|\mathbf{Q}^* - \mathbf{X}\|$ among symmetric matrices $\mathbf{X}$ having the desired spectrum $(\mathbf{0}_k^\mathsf{T}, \mathbf{1}_{\bar{k}}^\mathsf{T})$. If $\mathbf{X}$ has this spectrum then so does $\mathbf{X}' = \mathbf{U}\mathbf{X}\mathbf{U}^\mathsf{T}$ so we get the equivalent minimization problem

$$\underset{\lambda(\mathbf{X}')=(\mathbf{0}_k^\mathsf{T}, \mathbf{1}_{\bar{k}}^\mathsf{T})}{\arg\min} \|\mathbf{D} - \mathbf{X}'\|, \quad \mathbf{D} = \mathbf{U}^\mathsf{T}\mathbf{Q}^*\mathbf{U}. \qquad (18)$$

Mirsky [28] has shown that for any matrix the norm of the difference is not smaller than the norm difference of the sorted singular values. For symmetric PSD matrices, the singular values are the eigenvalues. So we have

$$\|\mathbf{D} - \mathbf{X}'\| \geq \|\mathbf{D} - \operatorname{diag}(\mathbf{0}_k^\mathsf{T}, \mathbf{1}_{\bar{k}}^\mathsf{T})\|, \qquad (19)$$

showing that the best solution for $\mathbf{X}'$ is the diagonal matrix, and we find $\mathbf{X} = \mathbf{U}^\mathsf{T}\mathbf{X}'\mathbf{U}$ as constructed.

For $\hat{\mathbf{r}}$, we see that it minimizes $\|\hat{\mathbf{r}} + \hat{\mathbf{Q}}\mathbf{Q}^{*+}\mathbf{r}^*\|$.

Means in other norms The interpretation of the least-squares fitting as first computing the L_2-mean in the space of matrix and vector coefficients of the representation $(\mathbf{Q}, \mathbf{r})$ and then projecting back onto the manifold of k-flats suggest a simple approximation of other means: we can compute any L_p mean in the space of the matrix and vector computations and then project. This reduces the problem of computing L_p means to the well-understood problem of doing so in Euclidean spaces.

We particularly consider the case of the L_1-norm, or geometric median, as it is known to be robust against outliers. For this, we use Weiszfeld's algorithm [34], which is essentially an iterative re-weighted least squares method. As we show in Sec. 5, it indeed shows robustness to outliers, and is not only significantly simpler to implement than Riemannian L_p centers of mass [3], but also preserves the equivariance properties that are missing from the methods based on Stiefel coordinates.

5. Experiments

We demonstrate in the following experiments that our method is both faster and yields more useful results compared to optimization in the affine Grassmannian [22], and provide a brief experiment in a realistic application scenario for reconstructing line-like objects from more than two views. All results reported are based on implementations in C++ using the Eigen library [15] for numerical linear algebra. Running times were gathered on a computer with an Intel i5-13600K CPU and 32GB RAM.

Data generation We randomly generate a target k-flat $\mathcal{F}^*$ in standard form $(\mathbf{A}^*, \mathbf{b}^*)$. Then, we generate m flats of dimension l by sampling $n \geq d$ random points $\mathbf{x}_i = \mathbf{b}^* + \mathbf{A}^*\mathbf{y}_i + \xi_i$ in a sample interval of $\mathbf{y}_i \in [-10, 10]^k$ and displacing them with additive, zero-mean Gaussian noise $\xi_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Subsequently, PCA is used to fit an l-flat to the n sampled points. The noise on the sample points is intended to simulate measurement and calibration errors from the real world. This procedure is repeated until m flats $\{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ are generated as observations to reconstruct $\mathcal{F}^*$.

Iterative approach on the Grassmannian As a comparison to our method, we optimize the objective in Eq. (8) using gradient descent on the Stiefel manifold. We use a combination of the Frobenius norm of the gradient, the difference between two consecutive iterates, and the number of iterations as the stopping criteria.

Our method We use our method from Eq. (15) and Eq. (16) in two variations. In the first one, we compute $\mathbf{Q}^*$ and $\mathbf{r}^*$ as the arithmetic mean of the respective matrices $\{\mathbf{Q}_i\}$ and $\{\mathbf{r}_i\}$ (Mean-SDF). In the second one, we compute $\mathbf{Q}^*$ and $\mathbf{r}^*$ as the median of the respective matrices using Weiszfeld's algorithm, where we use the Frobenius norm for computing $\mathbf{Q}^*$ and the Euclidean norm for $\mathbf{r}^*$ (Median-SDF). Although the projection is identical to the previous case, with this method, we may get more stable results in the presence of outliers.

Metrics To compare the reconstructed flat $\mathcal{F}'$ to the original $\mathcal{F}*$, we use their principal angle(s) $\theta_1, \ldots, \theta_k$ and their least-squares distance [14]

$$d_{\min}(\mathcal{F}^*, \mathcal{F}') = \min_{\mathbf{x}^* \in \mathcal{F}^*, \mathbf{x}' \in \mathcal{F}'} \|\mathbf{x}^* - \mathbf{x}'\| \qquad (20)$$

as separate metrics.

Efficiency Fig. 2 shows the average running time of our method and the method from [22] over ten trials on a log scale. For each dimension of the ambient space $d \in$
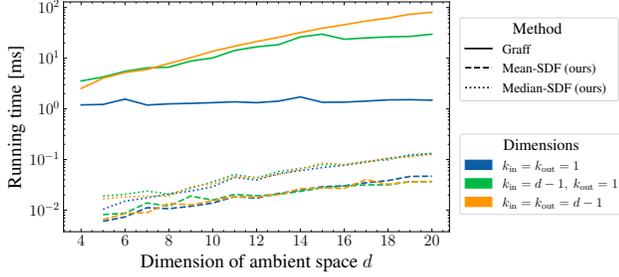
Figure 2. Running time comparison between our methods and the iterative optimization on the affine Grassmannian (Graff) for different dimensions of flats and ambient space but constant input size $m = 20$ and noise level $\sigma = 0.2$.
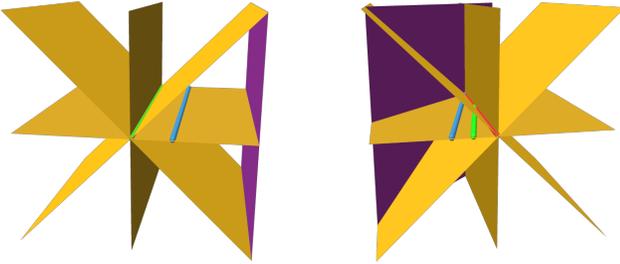

Figure 3. Comparison between the mean and the median estimation of $\mathbf{Q}^*$ and $\mathbf{r}^*$ using our method. A line (red) that is the intersection of four planes (yellow) is reconstructed while having a fifth plane as an outlier (purple). On the left, only the displacement of the outlier is off. On the right, both orientation and displacement are off. The Mean-SDF method yields the blue line, while the green line is the output of the Median-SDF method.

$\{4, \dots, 20\}$, three different experiments were conducted: (1) reconstructing a line from other lines ($k_{in} = k_{out} = 1$), (2) reconstructing a line from hyperplanes ($k_{in} = d - 1$, $k_{out} = 1$), and lastly (3) reconstructing a hyperplane from hyperplanes ($k_{in} = k_{out} = d - 1$). In every case, we chose the number of input flats as $m = d$. A limit of 200 iterations was set for the iterative optimization. It is to be expected that our method is significantly faster than the iterative computation of the Riemannian center, and our results clearly show that this is the case. The running time of our method is similar for all cases irrespective of the dimension of the input and output flats, which is, to some degree, expected, as our method mainly involves the diagonalization of a $d \times d$ matrix whose dimensionality only depends on the dimension of the ambient space. We note that the dimension of the ambient space d is likely to be less pronounced for a larger number of samples m, as then the assembly of $\mathbf{Q}^*, \mathbf{r}^*$ might dominate the computation.

Accuracy Tab. 1 shows distances of a reconstructed affine line from planes in $\mathbb{R}^3$. While we are aware that, in this par-

Table 1. The minimum distance $d_{\min}$ and (principal) angle θ between a target line $\mathcal{F}^*$ and its reconstruction $\mathcal{F}'$ in $\mathbb{R}^3$ from $m = 20$ planes, averaged over ten trials. Top: varying noise levels σ with no outliers. Bottom: varying percentage of outliers with constant noise level $\sigma = 0.2$. The best result for each condition is highlighted in **bold**, the second best is underlined.

	Graff		Mean-SDF		Median-SDF	
	$\downarrow d_{\min}$	$\downarrow \theta$	$\downarrow d_{\min}$	$\downarrow \theta$	$\downarrow d_{\min}$	$\downarrow \theta$
Noise σ						
0.5	0.2174	_0.0402_	_0.1292_	**0.0381**	**0.1192**	0.0434
1.0	0.3411	0.1112	_0.2417_	_0.0913_	**0.1870**	**0.0817**
1.5	0.3953	0.3240	_0.3297_	_0.1357_	**0.2697**	**0.1252**
2.0	_0.4167_	0.3914	0.4843	**0.1045**	**0.2854**	_0.1095_
2.5	0.5946	0.4817	_0.4599_	_0.1374_	**0.3200**	**0.1339**
Outlier %						
0.0	0.2325	0.0212	**0.0227**	_0.0190_	_0.0929_	**0.0177**
0.1	_0.1801_	**0.0242**	2.6684	0.0438	**0.1274**	_0.0423_
0.2	_0.1892_	**0.0189**	6.6070	0.0696	**0.0919**	_0.0518_
0.3	_0.2944_	**0.0387**	8.2614	0.0951	**0.1650**	_0.0822_
0.4	_0.2891_	**0.0325**	12.4167	0.1880	**0.2195**	_0.1764_
0.5	_0.3917_	**0.0568**	15.0596	0.2536	**0.3857**	_0.2232_

ticular scenario, other techniques could have been used, it still serves to illustrate the difference between our method and computations in the affine Grassmannian, which we have found to be consistent across dimensions of flats and ambient dimension.

The upper half shows results for experiments with only Gaussian noise added. For the remaining rows in Tab. 1, we have added outliers to the data (see Fig. 3). It is evident that our Mean-SDF variant is not robust to outliers, especially in terms of displacement, in which the results appear to be worse than the iterative method. Yet the Median-SDF variant recovers from this loss of robustness, especially in terms of the translation distance, even with a high proportion of outliers, and provides the best overall results. The difference between our two variants is visualized in Fig. 3.

Application: Archery A possible application is line reconstruction in multiple-view geometry. Imagine a line-like object being extracted in more than 2 registered cameras. For each camera, the line in screen space creates an instance of a plane in world-space, containing the line in 3-space. These planes are expected to intersect in the common line, but due to noise in calibration and registration of the camera as well as in the feature extraction in the discrete images this will not be the case. The standard approach to this problem is to represent the planes as a matrix $\mathbf{M}$ by stacking the normal representation $\mathbf{N}_i$ (see Sec. 2) in rows and apply the SVD to find the best rank-2 approximation with respect to the Frobenius norm. The kernel of the resulting rank-reduced matrix represents the line. We note that the

Figure 4. Multiple views of an archery target. Given the camera parameters, the re-projected arrows form planes in world space that do not intersect due to detection and calibration errors. Our method can be used to find a good representative. The original arrow is colored in red, given a bad calibrated camera our result is colored in blue and the rank minimization approach [16] is colored in green.

properties of this approach in the presence of noise are not clear, because the normal representation is not unique and, as a consequence, it is not clear what is being minimized. In our method, it is clear that the least-squares distance of the reconstructed line to the planes is being minimized, as we showed that this happens for any sampling of the liner.

We compare the two approaches using a synthetic archery example, where we try to reconstruct the arrows in 3D space. A scene is rendered from four different perspectives and the cameras are calibrated with moderate noise being introduced. We then 'shoot' random arrows onto the target and detect the features in the image spaces of the cameras. As expected, the results for the rank-minimization [16] depends on the choice of representation of the planes. Fig. 4 shows the results for a typical case, comparing to our results. While our method naturally yields consistent reconstruction very close to ground truth, rank-minimization is significantly off in some cases.

6. Discussion

The method for least-squared fitting of flats to flats based on squared distance functions, as simple as it is, exhibits various nice properties that are not present in other, seemingly more principled approaches such as the manifold of affine Grassmannians. We are unaware of any construction that offers the desired equivariance to the isometries of affine space for means or fitting procedures of *flats in arbitrary dimensions* (without shifting the input to compensate for the missing equivariance to translations [24]). We briefly comment below on a possible similar construction using Grassmann-Plücker coordinates, and then mention some further use cases and investigations.

Interpolation in Grassmann-Plücker coordinates Since Plücker coordinates are unique up to scale, one might want to try, similar to the interpretation of our approach, computing the mean of the coefficients and then projecting back onto the space described by the Grassmann-Plücker relations.

Taking the mean of the coefficient requires selecting an appropriate scale. In general, this might be difficult [21]

and typically introduces bias for noisy data. In our context, it is not difficult to see that taking the mean leads to consistent results if the basis taken on the flat is orthonormal, i.e., if the coefficients are constructed from the Stiefel coordinates as described in Sec. 2. This method still suffers from several drawbacks: (1) Grassmann-Plücker coordinates represent *oriented* spaces [31]. This means the line constructed from point $\mathbf{b}$ and direction $\mathbf{a}$ is not the same as the one constructed from $\mathbf{b}$ and $-\mathbf{a}$. Taking the mean depends on the choice of orientation. (2) The 'projection' onto coefficients satisfying the Grassmann-Plücker relations is, in general, significantly more involved than computing the eigendecomposition [13]. (3) The Grassmann-Plücker coordinates naturally arise as antisymmetric rank-k tensors. It is unclear (to us) how one could perform computations involving different dimensions, although admittedly Schubert varieties may be used [26], similar to the construction for affine Grassmannians.

Outlook The PCA fitting flats to points may be considered the simplest statistical analysis of a set of point-like observations. This can be extended to fitting several flats such as in clustering methods or, more generally, describing data using mixture models [6, 32, 33, 37]. All of these methods may be generalized to work with flats as the observed samples, using the methods we have introduced here.

As described, our method can be re-interpreted as taking the mean in ambient space and projecting, and in this way used similar to methods in Euclidean spaces. This suggests other methods for improving robustness to outliers could be used, as well as considering the L_∞-norm, whose minimization yields the center of the smallest enclosing ball [35]. This approach would be considerably simpler than computing geodesic disks enclosing a set of flats [25].

Acknowledgements

References

[1] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80:199–220, 2004. 1, 4

[2] Pierre-Antoine Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008. 1, 4

[3] Bijan Afsari. Riemannian lp center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011. 6

[4] Åke Björck and Gene H Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973. 3

[5] James F. Blinn. A homogeneous formulation for lines in 3 space. *SIGGRAPH Comput. Graph.*, 11(2):237–241, jul 1977. 1

[6] Paul S Bradley and Olvi L Mangasarian. K-plane clustering. *Journal of Global optimization*, 16:23–32, 2000. 8

[7] Charles-Alban Deledalle, Loic Denis, Sonia Tabti, and Florence Tupin. Closed-form expressions of the eigen decomposition of 2 x 2 and 3 x 3 hermitian matrices. Technical report, 2017. 5

[8] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998. 1, 4

[9] William Fulton. Eigenvalues of sums of hermitian matrices. *Séminaire Bourbaki*, 40:255–269, 1998. 5

[10] Michael Garland and Paul S Heckbert. Surface simplification using quadric error metrics. In *SIGGRAPH '97*, pages 209–216, 1997. 4

[11] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Birkhäuser Boston, 1994. 2

[12] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013. 5

[13] P. Griffiths and J. Harris. *Principles of Algebraic Geometry*. Wiley Classics Library. Wiley, 2014. 8

[14] Jürgen Gross and Götz Trenkler. On the least squares distance between affine subspaces. *Linear Algebra and its Applications*, 237-238:269–276, 1996. 6

[15] Gaël Guennebaud, Benoît Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010. 6

[16] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, USA, 2 edition, 2003. 1, 8

[17] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977. 3

[18] Hermann Karcher. Riemannian center of mass and so called karcher mean, 2014. 4

[19] Daniel A Klain and Gian-Carlo Rota. *Introduction to geometric probability*. Cambridge University Press, 1997. 2

[20] Allen Knutson and Terence Tao. Honeycombs and sums of hermitian matrices. *Notices Amer. Math. Soc*, 48(2), 2001. 5

[21] Vincent Lesueur and Vincent Nozick. Least square for grassmann-cayley agelbra in homogeneous coordinates. In Fay Huang and Akihiro Sugimoto, editors, *Image and Video Technology – PSIVT 2013 Workshops*, pages 133–144, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. 8

[22] Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye. Numerical algorithms on the affine grassmannian. *SIAM Journal on Matrix Analysis and Applications*, 40(2):371–393, 2019. 1, 2, 4, 6

[23] Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye. The grassmannian of affine subspaces. *Foundations of Computational Mathematics*, 21(2):537–574, 2021. 1, 2

[24] Parker C. Lusk, Devarth Parikh, and Jonathan P. How. Graffmatch: Global matching of 3d lines and planes for wide baseline lidar registration. *IEEE Robotics and Automation Letters*, 8(2):632–639, 2023. 2, 8

[25] Tim Marrinan, P-A Absil, and Nicolas Gillis. On a minimum enclosing ball of a collection of linear subspaces. *Linear Algebra and its Applications*, 625:248–278, 2021. 8

[26] Ezra Miller and Bernd Sturmfels. Matrix schubert varieties. In *Combinatorial Commutative Algebra*, pages 289–310. Springer New York, New York, NY, 2005. 8

[27] Ezra Miller and Bernd Sturmfels. Plücker coordinates. In *Combinatorial Commutative Algebra*, pages 273–288. Springer New York, New York, NY, 2005. 1, 2

[28] L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *The Quarterly Journal of Mathematics*, 11(1):50–59, 01 1960. 6

[29] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901. 1

[30] Krishan Sharma and Renu Rameshan. Image set classification using a distance-based kernel over affine grassmann manifold. *IEEE Transactions on Neural Networks and Learning Systems*, 32(3):1082–1095, 2021. 2

[31] Jorge Stolfi. *Oriented projective geometry: a framework for geometric computations*. PhD thesis, Stanford University, 1995. 8

[32] Paul Tseng. Nearest q-flat to m points. *Journal of Optimization Theory and Applications*, 105:249–252, 2000. 8

[33] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(12):1945–1959, 2005. 8

[34] Endre Weiszfeld and Frank Plastria. On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, 167(1):7–41, 2009. 6

[35] Emo Welzl. Smallest enclosing disks (balls and ellipsoids). In *New Results and New Trends in Computer Science: Graz, Austria, June 20–21, 1991 Proceedings*, pages 359–370. Springer, 2005. 8

[36] Ke Ye and Lek-Heng Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197, 2016. 4

[37] Teng Zhang, Arthur Szlam, and Gilad Lerman. Median k-flats for hybrid linear modeling with many outliers. In *12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 234–241. IEEE, 2009. 8

[38] Tong Zhao, Laurent Busé, David Cohen-Steiner, Tamy Boubekeur, Jean-Marc Thiery, and Pierre Alliez. Variational shape reconstruction via quadric error metrics. In *ACM SIGGRAPH 2023 Conference Proc.*, pages 1–10, 2023. 4