

# Algebraic Clustering of Affine Subspaces

Manolis C. Tsakiris and René Vidal, *Fellow, IEEE*

**Abstract**—Subspace clustering is an important problem in machine learning with many applications in computer vision and pattern recognition. Prior work has studied this problem using algebraic, iterative, statistical, low-rank and sparse representation techniques. While these methods have been applied to both linear and affine subspaces, theoretical results have only been established in the case of linear subspaces. For example, algebraic subspace clustering (ASC) is guaranteed to provide the correct clustering when the data points are in *general position* and the union of subspaces is *transversal*. In this paper we study in a rigorous fashion the properties of ASC in the case of affine subspaces. Using notions from algebraic geometry, we prove that the *homogenization trick*, which embeds points in a union of affine subspaces into points in a union of linear subspaces, preserves the general position of the points and the transversality of the union of subspaces in the embedded space, thus establishing the correctness of ASC for affine subspaces.

**Index Terms**—Algebraic Subspace Clustering, Affine Subspaces, Homogeneous Coordinates, Algebraic Geometry.

## 1 INTRODUCTION

Subspace clustering is the problem of clustering a collection of points drawn approximately from a union of linear or affine subspaces. This is an important problem in machine learning with many applications in computer vision and pattern recognition such as clustering faces, digits, images and motions [1]. Over the past 15 years, a variety of subspace clustering methods have appeared in the literature, including iterative [2], [3], probabilistic [4], algebraic [5], spectral [6], [7], and self-expressiveness-based [8], [9], [10], [11], [12] approaches. Among them, the *Algebraic Subspace Clustering* (ASC) method of [5], also known as GPCA, establishes an interesting connection between machine learning and algebraic geometry (see also [13] for another such connection). By describing a union of  $n$  linear subspaces as the zero set of a system of homogeneous polynomials of degree  $n$ , ASC clusters the subspaces in closed form via polynomial fitting and differentiation (or alternatively polynomial factorization [14]).

**Merits of algebraic subspace clustering.** In addition to providing interesting algebraic geometric insights in the problem of subspace clustering, ASC is unique among other methods in that it is guaranteed to provide the correct clustering, under the mild hypothesis that the union of subspaces is *transversal* and the data points are in *general position* (in an algebraic geometric sense). This entails that ASC can handle subspaces of dimension  $d$  comparable to the ambient dimension  $D$  (high relative dimension  $d/D$ ). In contrast, most state-of-the-art methods, such as *Sparse Subspace Clustering* (SSC) [10], [15], require the subspaces to be of sufficiently small relative dimension. Therefore, instances of applications where ASC is a natural candidate, while SSC is in principle inapplicable, are projective motion segmentation [16], 3D point cloud analysis [17] and hybrid system identification [18]. Moreover, it was recently demonstrated in [19] that, using the idea of *filtrations* of unions of subspaces [20], [21], ASC not only can be robustified to noise, but also outperforms state-of-the-art

methods in the popular benchmark dataset Hopkins155 [22] for real world motion segmentation.

**Dealing with affine subspaces.** In several important applications, such as motion segmentation, the underlying subspaces do not pass through the origin, i.e., they are affine. Methods such as K-subspaces [2], [3] and mixtures of probabilistic PCA [4] can trivially handle this case by explicitly learning models of affine subspaces. Likewise, the spectral clustering method of [23] can handle affine subspaces by constructing an affinity that depends on the distance from a point to a subspace. However, these methods do not come with theoretical conditions under which they are guaranteed to give the correct clustering. On the other hand, when data  $\mathcal{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \subset \mathbb{R}^D$  lying in a union of  $n$  distinct affine subspaces are embedded into *homogeneous coordinates*

$$\tilde{\mathcal{X}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix} \subset \mathbb{R}^{D+1}, \quad (1)$$

they lie in a union of  $n$  distinct linear subspaces of  $\mathbb{R}^{D+1}$ . If the *embedded data*  $\tilde{\mathcal{X}}$  satisfy the geometric separation conditions of [15] with respect to the underlying union of linear subspaces, then SSC [10] applied to  $\tilde{\mathcal{X}}$  is guaranteed to yield a *subspace preserving* affinity. However, while the conditions in [15] have a clear geometric interpretation for linear subspaces, it is unclear what these conditions entail when applied to affine subspaces via the embedding in (1). On the other hand, recent work [24], [25] shows that an  $\ell_0$  version of SSC ( $\ell_0$ -SSC) yields the correct clustering under mild conditions of general position (in a linear algebraic sense), and this analysis can be easily extended to affine subspaces. However,  $\ell_0$ -SSC has exponential complexity as opposed to the polynomial complexity  $\ell_1$ -SSC [10]. While the ASC approach discussed here also has exponential complexity in general, under certain conditions it is more efficient than  $\ell_0$ -SSC.<sup>1</sup>

1. The worst-case complexity of ASC is  $\mathcal{O}\left(N \binom{n+D}{n}^2\right)$ , which is linear in the number of data points,  $N$ , and exponential in the number of subspaces,  $n$ , and the dimension of the original data  $D$ . In contrast, the worst-case complexity of  $\ell_0$ -SSC is  $\mathcal{O}\left(N(D+1)(d+1)^2 \binom{N-1}{d+1}\right)$ , where  $d$  is the maximal dimension of the affine subspaces. Hence, when  $n$  and  $D$  are small and  $d \approx \mathcal{O}(D)$ , the complexity of  $\ell_0$ -SSC as a function of  $N$  dominates that of ASC. A detailed comparison of the complexities of ASC and  $\ell_0$ -SSC is beyond the scope of the paper and is thus omitted.

• *The authors are with the Center of Imaging Science, Johns Hopkins University, Baltimore, MD, 21218, USA.  
E-mail: m.tsakiris@jhu.edu, rvidal@cis.jhu.edu*

Returning to ASC, the traditional way to handle points from a union of affine subspaces [26] is to use homogeneous coordinates as in (1), and subsequently apply ASC to the embedded data. We refer to this two-step approach as Affine ASC (AASC). Although AASC has been observed to perform well in practice, it lacks a sufficient theoretical justification. On one hand, while it is true that the embedded points live in a union of associated linear subspaces, it is obvious that they have a very particular structure inside these subspaces. Specifically, even if the original points are *generic*, in the sense that they are sampled uniformly at random from the affine subspaces, the embedded points are clearly *non-generic*, in the sense that they always lie in the zero-measure intersection of the union of the associated linear subspaces with the hyperplane  $x_0 = 1$ .<sup>2</sup> Thus, even in the absence of noise, one may wonder whether this *non-genericity* of the embedded points will affect the behavior of AASC and to what extent. On the other hand, even if the affine subspaces are transversal, there is no guarantee that the associated linear subspaces are also transversal. Thus, it is natural to ask for conditions on the affine subspaces and the data points under which AASC is guaranteed to give the correct clustering.

**Paper contributions.** In this paper we adapt abstract notions from algebraic geometry to the context of unions of affine subspaces in order to rigorously prove the correctness of AASC in the absence of noise. More specifically, we define in a very precise fashion the notion of points being in *general position* in a union of  $n$  linear or affine subspaces. Intuitively, points are in general position if they can be used to uniquely reconstruct the union of subspaces they lie in by means of polynomials of degree  $n$  that vanish on the points. Then we show that the embedding (1) preserves the property of points being in general position, which is one of the two success conditions of ASC. We also show that the second condition, which is the *transversality* of the union of linear subspaces in  $\mathbb{R}^{D+1}$  that is associated to the union of affine subspaces in  $\mathbb{R}^D$  under the embedding (1), is also satisfied, provided that

- 1) the union of subspaces formed by the linear parts of the affine subspaces is transversal, and
- 2) the translation vectors of the affine subspaces do not lie in the zero measure set of a certain algebraic variety.

Our exposition style is for the benefit of the reader unfamiliar with algebraic geometry. We introduce notions and notations as we proceed and give as many examples as space allows. We leave the more intricate details to the various proofs.

## 2 ALGEBRAIC SUBSPACE CLUSTERING REVIEW

This section gives a brief review of the ASC theory ([5], [27], [28], [21]). After defining the subspace clustering problem in Section 2.1, we describe unions of linear subspaces as algebraic varieties in Section 2.2, and give the main theorem of ASC (Theorem 1) in terms of vanishing polynomials in Section 2.3. In Section 2.4 we elaborate on the main hypothesis of Theorem 1, the *transversality* of the union of subspaces. In Section 2.5 we introduce the notion of points *in general position* (Definition 5) and adapt Theorem 1 to the more practical case of a finite set of points (Theorem 9).

### 2.1 Subspace Clustering Problem

Let  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  be a set of points that lie in an unknown union of  $n > 1$  linear subspaces  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$ , where  $\mathcal{S}_i$  a linear

2. Here and in the rest of the paper, we consider only the uniform measure.

subspace of  $\mathbb{R}^D$  of dimension  $d_i < D$ . The goal of subspace clustering is to find the number of subspaces, their dimensions, a basis for each subspace, and cluster the data points based on their subspace membership, i.e., find a decomposition or *clustering* of  $\mathcal{X}$  as  $\mathcal{X} = \mathcal{X}_1 \cup \dots \cup \mathcal{X}_n$ , where  $\mathcal{X}_i = \mathcal{X} \cap \mathcal{S}_i$ .

### 2.2 Unions of Linear Subspaces as Algebraic Varieties

The key idea behind ASC is that a union of  $n$  linear subspaces  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$  of  $\mathbb{R}^D$  is the *zero set* of a finite set of homogeneous<sup>3</sup> polynomials of degree  $n$  with real coefficients in  $D$  indeterminates  $\mathbf{x} := [x_1, \dots, x_D]^\top$ . Such a set is called an *algebraic variety* [29], [30]. For example, a union of  $n$  hyperplanes  $\Phi = \mathcal{H}_1 \cup \dots \cup \mathcal{H}_n$ , where the  $i$ th hyperplane  $\mathcal{H}_i = \{\mathbf{x} : \mathbf{b}_i^\top \mathbf{x} = 0\}$  is defined by its normal vector  $\mathbf{b}_i \in \mathbb{R}^D$ , is the zero set of the polynomial

$$p(\mathbf{x}) = (\mathbf{b}_1^\top \mathbf{x})(\mathbf{b}_2^\top \mathbf{x}) \cdots (\mathbf{b}_n^\top \mathbf{x}), \quad (2)$$

in the sense that a point  $\mathbf{x}$  belongs to the union  $\Phi$  if and only if  $p(\mathbf{x}) = 0$ . Likewise, the union of a plane with normal  $\mathbf{b}$  and a line with normals  $\mathbf{b}_1, \mathbf{b}_2$  in  $\mathbb{R}^3$  is the zero set of the two polynomials

$$p_1(\mathbf{x}) = (\mathbf{b}^\top \mathbf{x})(\mathbf{b}_1^\top \mathbf{x}) \quad \text{and} \quad p_2(\mathbf{x}) = (\mathbf{b}^\top \mathbf{x})(\mathbf{b}_2^\top \mathbf{x}). \quad (3)$$

More generally, for  $n$  subspaces of arbitrary dimensions, these *vanishing polynomials* are homogeneous of degree  $n$ . Moreover, they are factorizable into  $n$  linear forms, with each linear form defined by a vector orthogonal to one of the  $n$  subspaces.<sup>4</sup>

### 2.3 Main Theorem of ASC

The set  $\mathcal{I}_\Phi$  of polynomials that vanish at every point of a union of linear subspaces  $\Phi$  has a special algebraic structure: it is closed under addition and it is closed under multiplication by any element of the *polynomial ring*  $\mathcal{R} = \mathbb{R}[x_1, \dots, x_D]$ . Such a set of polynomials is called an *ideal* [29], [30] of  $\mathcal{R}$ . If we restrict our attention to the subset  $\mathcal{I}_{\Phi,n}$  of  $\mathcal{I}_\Phi$  that consists only of vanishing polynomials of degree  $n$ , we notice that  $\mathcal{I}_{\Phi,n}$  is a finite dimensional real vector space, because it is a subspace of  $\mathcal{R}_n$ , the latter being the set of all homogeneous polynomials of  $\mathcal{R}$  of degree  $n$ , which is a vector space of dimension  $M_n(D) := \binom{n+D-1}{n}$ .

**Theorem 1** (Main Theorem of ASC, [5]). *Let  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$  be a transversal union of linear subspaces of  $\mathbb{R}^D$ . Let  $p_1, \dots, p_s$  be a basis for  $\mathcal{I}_{\Phi,n}$  and let  $\mathbf{x}_i$  be a point in  $\mathcal{S}_i$  such that  $\mathbf{x}_i \notin \bigcup_{i' \neq i} \mathcal{S}_{i'}$ . Then  $\mathcal{S}_i = \text{Span}(\nabla p_1|_{\mathbf{x}_i}, \dots, \nabla p_s|_{\mathbf{x}_i})^\perp$ .*

In other words, we can estimate the subspace  $\mathcal{S}_i$  passing through a point  $\mathbf{x}_i$ , as the orthogonal complement of the span of the gradients of all the degree- $n$  vanishing polynomials evaluated at  $\mathbf{x}_i$ . Observe that the only assumption on the subspaces required by Theorem 1, is that they are *transversal*, a notion explained next.

### 2.4 Transversal Unions of Linear Subspaces

Intuitively, transversality is a notion of general position of subspaces, which entails that all intersections among subspaces are as small as possible, as allowed by their dimensions. Formally:

3. A polynomial in many variables is called homogeneous if each of its monomials has the same degree. For example,  $x_1^2 + x_1x_2$  is homogeneous of degree 2, while  $x_1^2 + x_2$  is non-homogeneous of degree 2.

4. Strictly speaking this is not always true; it is true though in the generic case, for example, if the subspaces are transversal (see Definition 2).

**Definition 2** ([27]). A union  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$  of linear subspaces of  $\mathbb{R}^D$  is transversal, if for any subset  $\mathfrak{J}$  of  $[n] := \{1, 2, \dots, n\}$

$$\text{codim} \left( \bigcap_{i \in \mathfrak{J}} \mathcal{S}_i \right) = \min \left\{ D, \sum_{i \in \mathfrak{J}} \text{codim}(\mathcal{S}_i) \right\}, \quad (4)$$

where  $\text{codim}(\mathcal{S}) = D - \dim(\mathcal{S})$  denotes the codimension of  $\mathcal{S}$ .

To understand Definition 2, let  $\mathbf{B}_i$  be a  $D \times c_i$  matrix containing a basis for  $\mathcal{S}_i^\perp$ , where  $c_i$  is the codimension of  $\mathcal{S}_i$ , and let  $\mathfrak{J}$  be a subset of  $[n]$ , say  $\mathfrak{J} = \{1, \dots, \ell\}$ ,  $\ell \leq n$ . Then a point  $\mathbf{x}$  belongs to  $\bigcap_{i \in \mathfrak{J}} \mathcal{S}_i$  if and only if  $\mathbf{x}^\top \mathbf{B}_{\mathfrak{J}} = \mathbf{0}$ , where  $\mathbf{B}_{\mathfrak{J}} = [\mathbf{B}_1, \dots, \mathbf{B}_\ell]$ . Hence, the dimension of  $\bigcap_{i \in \mathfrak{J}} \mathcal{S}_i$  is equal to the dimension of the left nullspace of  $\mathbf{B}_{\mathfrak{J}}$ , or equivalently,

$$\text{codim} \left( \bigcap_{i \in \mathfrak{J}} \mathcal{S}_i \right) = \text{rank}(\mathbf{B}_{\mathfrak{J}}). \quad (5)$$

Since  $\mathbf{B}_{\mathfrak{J}}$  is a  $D \times \sum_{i \in \mathfrak{J}} c_i$  matrix, we must have that

$$\text{rank}(\mathbf{B}_{\mathfrak{J}}) \leq \min \left\{ D, \sum_{i \in \mathfrak{J}} c_i \right\}. \quad (6)$$

Hence, transversality is equivalent to  $\mathbf{B}_{\mathfrak{J}}$  being full-rank, as  $\mathfrak{J}$  ranges over all subsets of  $[n]$ , and  $\mathbf{B}_{\mathfrak{J}}$  drops rank if and only if all maximal minors of  $\mathbf{B}_{\mathfrak{J}}$  vanish, in which case there are certain algebraic relations between the basis vectors of  $\mathcal{S}_i^\perp$ ,  $i \in \mathfrak{J}$ . Since any set given by algebraic relations has measure zero [31], this shows that a union of subspaces is transversal with probability 1.

**Proposition 3.** Let  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$  be a union of  $n$  linear subspaces in  $\mathbb{R}^D$  of codimensions  $0 < c_i < D$ ,  $i \in [n]$ . Let  $\mathbf{b}_{i1}, \dots, \mathbf{b}_{ic_i}$  be a basis for  $\mathcal{S}_i^\perp$ . If the vectors  $\{\mathbf{b}_{ij}\}_{i=1, \dots, n}^{j=1, \dots, c_i}$  do not lie in the zero-measure set of a (proper) algebraic variety of  $\mathbb{R}^{D \times \sum_{i \in [n]} c_i}$ , then  $\Phi$  is transversal.

**Example 4.** Consider two planes  $\mathcal{S}_1, \mathcal{S}_2$  in  $\mathbb{R}^3$  with normals  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . Then one expects their intersection  $\mathcal{S}_1 \cap \mathcal{S}_2$  to be a line, and hence be of codimension  $2 = \min(3, 1 + 1)$ , unless the two planes coincide, which happens only if  $\mathbf{b}_1$  is colinear with  $\mathbf{b}_2$ . Clearly, if one randomly selects two planes in  $\mathbb{R}^3$ , the probability that they are not transversal is zero. If we consider a third plane  $\mathcal{S}_3$  with normal  $\mathbf{b}_3$  such that every intersection  $\mathcal{S}_1 \cap \mathcal{S}_2$ ,  $\mathcal{S}_1 \cap \mathcal{S}_3$  and  $\mathcal{S}_2 \cap \mathcal{S}_3$  is a line, then the three planes fail to be transversal only if  $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3$  is a line. But this can happen only if the three normals  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  are linearly dependent, which again is a probability zero event if the three planes are randomly selected.

This reveals the important fact that the theoretical conditions for success of ASC (in the absence of noise) are much weaker than those for other methods such as SSC and LRSC, since as we just pointed out ASC will succeed almost surely (Theorem 1).<sup>5</sup>

## 2.5 Points In General Position

In practice, we may not be given the polynomials  $p_1, \dots, p_s$  that vanish on a union of subspaces  $\Phi = \bigcup_{i=1}^n \mathcal{S}_i$ , but rather a finite collection of points  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  sampled from  $\Phi$ . If we want to fully characterize  $\Phi$  from  $\mathcal{X}$ , the least we can ask is that  $\mathcal{X}$  uniquely defines  $\Phi$  as a set, otherwise the problem becomes ill-posed. Since it is known that  $\Phi$  is the zero set of  $\mathcal{I}_{\Phi, n}$  [5], i.e.,  $\Phi = \mathcal{Z}(\mathcal{I}_{\Phi, n})$ , it is natural to require that  $\Phi$  can be recovered

5. Of course, the main disadvantage of ASC with respect to SSC or LRSC is its combinatorial computational complexity, which remains an open problem.

as the zero set of all homogeneous polynomials of degree  $n$  that vanish on  $\mathcal{X}$ .

**Definition 5 (Points in general position).** Let  $\Phi$  be a union of  $n$  linear subspaces of  $\mathbb{R}^D$ , and  $\mathcal{X}$  a finite set of points in  $\Phi$ . We will say that  $\mathcal{X}$  is in general position in  $\Phi$ , if  $\Phi = \mathcal{Z}(\mathcal{I}_{\mathcal{X}, n})$ .

Recall from Theorem 1 that for ASC to succeed, we need a basis  $p_1, \dots, p_s$  for  $\mathcal{I}_{\Phi, n}$ . The next result shows that if  $\mathcal{X}$  is in general position in  $\Phi$ , then we can compute such a basis from  $\mathcal{X}$ .

**Proposition 6.**  $\mathcal{X}$  is in general position in  $\Phi \Leftrightarrow \mathcal{I}_{\mathcal{X}, n} = \mathcal{I}_{\Phi, n}$ .

*Proof.* ( $\Rightarrow$ ) Suppose  $\mathcal{X}$  is in general position in  $\Phi$ , i.e.,  $\Phi = \mathcal{Z}(\mathcal{I}_{\mathcal{X}, n})$ . We will show that  $\mathcal{I}_{\mathcal{X}, n} = \mathcal{I}_{\Phi, n}$ . The inclusion  $\mathcal{I}_{\mathcal{X}, n} \supset \mathcal{I}_{\Phi, n}$  is immediate, since if  $p \in \mathcal{I}_{\Phi, n}$  vanishes on  $\Phi$ , then it will vanish on the subset  $\mathcal{X}$  of  $\Phi$ . Conversely, let  $p \in \mathcal{I}_{\mathcal{X}, n}$ . Since by hypothesis  $\Phi = \mathcal{Z}(\mathcal{I}_{\mathcal{X}, n})$ , we will have that  $p(\mathbf{x}) = 0$ ,  $\forall \mathbf{x} \in \Phi$ , i.e.,  $p$  vanishes on  $\Phi$ , i.e.,  $p \in \mathcal{I}_{\Phi, n}$ , i.e.,  $\mathcal{I}_{\mathcal{X}, n} \subset \mathcal{I}_{\Phi, n}$ .

( $\Leftarrow$ ) Suppose  $\mathcal{I}_{\mathcal{X}, n} = \mathcal{I}_{\Phi, n}$ ; then  $\mathcal{Z}(\mathcal{I}_{\mathcal{X}, n}) = \mathcal{Z}(\mathcal{I}_{\Phi, n})$ . Since  $\Phi = \mathcal{Z}(\mathcal{I}_{\Phi, n})$  [5], we have  $\Phi = \mathcal{Z}(\mathcal{I}_{\mathcal{X}, n})$ .  $\square$

Next, we show that points in general position always exist.

**Proposition 7.** Any union  $\Phi$  of  $n$  linear subspaces of  $\mathbb{R}^D$  admits a finite subset  $\mathcal{X}$  that lies in general position in  $\Phi$ .

*Proof.* This follows from Theorem 2.9 in [28], together with the regularity result of [32], which says that the maximal degree of a generator of  $\mathcal{I}_{\Phi}$  does not exceed  $n$ .  $\square$

**Example 8.** Let  $\Phi = \mathcal{S}_1 \cup \mathcal{S}_2$  be the union of two planes of  $\mathbb{R}^3$  with normal vectors  $\mathbf{b}_1, \mathbf{b}_2$ , and let  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$  be four points of  $\Phi$ , such that,  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}_1 - \mathcal{S}_2$  and  $\mathbf{x}_3, \mathbf{x}_4 \in \mathcal{S}_2 - \mathcal{S}_1$ . Let  $\mathcal{H}_{13}$  and  $\mathcal{H}_{24}$  be the planes spanned by  $\mathbf{x}_1, \mathbf{x}_3$  and  $\mathbf{x}_2, \mathbf{x}_4$  respectively, and let  $\mathbf{b}_{13}, \mathbf{b}_{24}$  be the normals to these planes. Then the polynomial  $q(x) = (\mathbf{b}_{13}^\top x)(\mathbf{b}_{24}^\top x)$  certainly vanishes on  $\mathcal{X}$ . But  $q$  does not vanish on  $\Phi$ , because the only (up to a scalar) homogeneous polynomial of degree 2 that vanishes on  $\Phi$  is  $p(x) = (\mathbf{b}_1^\top x)(\mathbf{b}_2^\top x)$ . Hence  $\mathcal{X}$  is not in general position in  $\Phi$ . The geometric reasoning is that two points per plane are not enough to uniquely define the union of the two planes; instead a third point in one of the planes is required.

In terms of a finite set of points  $\mathcal{X}$ , Theorem 1 becomes:

**Theorem 9.** Let  $\mathcal{X}$  be a finite set of points sampled from a union  $\Phi$  of  $n$  linear subspaces of  $\mathbb{R}^D$ . Let  $p_1, \dots, p_s$  be a basis for  $\mathcal{I}_{\mathcal{X}, n}$ , the vector space of homogeneous polynomials of degree  $n$  that vanish on  $\mathcal{X}$ . Let  $\mathbf{x}_i$  be a point in  $\mathcal{X}_i := \mathcal{X} \cap \mathcal{S}_i$  such that  $\mathbf{x}_i \notin \bigcup_{i' \neq i} \mathcal{S}_{i'}$ . If  $\mathcal{X}$  is in general position in  $\Phi$  (Definition 5), and  $\Phi$  is transversal (Definition 2), then  $\mathcal{S}_i = \text{Span}(\nabla p_1|_{\mathbf{x}_i}, \dots, \nabla p_s|_{\mathbf{x}_i})^\perp$ .

## 3 PROBLEM STATEMENT AND CONTRIBUTIONS

In this section we begin by defining the problem of clustering unions of affine subspaces in Section 3.1. In Section 3.2 we analyze the traditional algebraic approach for handling affine subspaces and point out that its correctness is far from obvious. Finally, in Section 3.3 we state the main findings of this paper.

### 3.1 Affine Subspace Clustering Problem

Let  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  be a finite set of points living in a union  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  of  $n$  affine subspaces of  $\mathbb{R}^D$ . Each affine subspace  $\mathbb{A}_i$  is the translation by some vector  $\boldsymbol{\mu}_i \in \mathbb{R}^D$  of a  $d_i$ -dimensional linear subspace  $\mathcal{S}_i$ , i.e.,  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$ . The affine subspace clustering problem involves clustering the points  $\mathcal{X}$  according to their subspace membership, and finding a parametrization of each affine subspace  $\mathbb{A}_i$  by finding a translation vector  $\boldsymbol{\mu}_i$  and a basis for its linear part  $\mathcal{S}_i$ , for all  $i = 1, \dots, n$ . Note that there is an inherent ambiguity in determining the translation vectors  $\boldsymbol{\mu}_i$ , since if  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$ , then  $\mathbb{A}_i = \mathcal{S}_i + (\mathbf{s}_i + \boldsymbol{\mu}_i)$  for any vector  $\mathbf{s}_i \in \mathcal{S}_i$ . Consequently, the best we can hope for is to determine the unique component of  $\boldsymbol{\mu}_i$  in the orthogonal complement  $\mathcal{S}_i^\perp$  of  $\mathcal{S}_i$ .

### 3.2 Traditional Algebraic Approach

Since the inception of ASC, the standard algebraic approach to cluster points living in a union of affine subspaces has been to embed the points into  $\mathbb{R}^{D+1}$  and subsequently apply ASC [26]. The precise embedding  $\phi_0 : \mathbb{R}^D \hookrightarrow \mathbb{R}^{D+1}$  is given by

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_D) \xrightarrow{\phi_0} \tilde{\boldsymbol{\alpha}} = (1, \alpha_1, \dots, \alpha_D). \quad (7)$$

To understand the effect of this embedding and why it is meaningful to apply ASC to the embedded points, let  $\mathbb{A} = \mathcal{S} + \boldsymbol{\mu}$  be a  $d$ -dimensional affine subspace of  $\mathbb{R}^D$ , with  $\mathbf{u}_1, \dots, \mathbf{u}_d$  being a basis for its linear part  $\mathcal{S}$ . As noted in Section 3.1, we can also assume that  $\boldsymbol{\mu} \in \mathcal{S}^\perp$ . For  $\mathbf{x} \in \mathbb{A}$ , there exists  $\mathbf{y} \in \mathbb{R}^d$  such that

$$\mathbf{x} = \mathbf{U}\mathbf{y} + \boldsymbol{\mu}, \quad \mathbf{U} := [\mathbf{u}_1, \dots, \mathbf{u}_d] \in \mathbb{R}^{D \times d}. \quad (8)$$

Then the embedded point  $\tilde{\mathbf{x}} := \phi_0(\mathbf{x})$  can be written as

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} = \tilde{\mathbf{U}} \begin{bmatrix} 1 \\ \mathbf{y} \end{bmatrix}, \quad \tilde{\mathbf{U}} := \begin{bmatrix} 1 & 0 & \dots & 0 \\ \boldsymbol{\mu} & \mathbf{u}_1 & \dots & \mathbf{u}_d \end{bmatrix}. \quad (9)$$

Equation (9) clearly indicates that the embedded point  $\tilde{\mathbf{x}}$  lies in the linear  $(d+1)$ -dimensional subspace  $\tilde{\mathcal{S}} := \text{Span}(\tilde{\mathbf{U}})$  of  $\mathbb{R}^{D+1}$  and the same is true for the entire affine subspace  $\mathbb{A}$ . From (9) one sees immediately that  $(\mathbf{u}_1, \dots, \mathbf{u}_d, \boldsymbol{\mu})$  can be used to construct a basis of  $\tilde{\mathcal{S}}$ . The converse is also true: given any basis of  $\tilde{\mathcal{S}}$  one can recover a basis for the linear part  $\mathcal{S}$  and the translation vector  $\boldsymbol{\mu}$  of  $\mathbb{A}$ . Hence, the embedding  $\phi_0$  takes a union of affine subspaces  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  into a union of linear subspaces  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  of  $\mathbb{R}^{D+1}$ , in a way that there is a 1-1 correspondence between the parameters of  $\mathbb{A}_i$  (a basis for the linear part and the translation vector) and the parameters of  $\tilde{\mathcal{S}}_i$  (a basis) for every  $i \in [n]$ .

To the best of our knowledge, the correspondence between  $\mathbb{A}_i$  and  $\tilde{\mathcal{S}}_i$  has been the sole theoretical justification so far in the subspace clustering literature for the traditional Affine ASC (AASC) approach for dealing with affine subspaces, which consists of

- 1) applying the embedding  $\phi_0$  to points  $\mathcal{X}$  in  $\Psi$ ,
- 2) computing a basis  $p_1, \dots, p_s$  for the vector space  $\mathcal{I}_{\tilde{\mathcal{X}}, n}$  of homogeneous polynomials of degree  $n$  that vanish on the embedded points  $\tilde{\mathcal{X}} := \phi_0(\mathcal{X})$ ,
- 3) for  $\tilde{\mathbf{x}}_i \in \tilde{\mathcal{X}} \cap \tilde{\mathcal{S}}_i - \bigcup_{i' \neq i} \tilde{\mathcal{S}}_{i'}$ , estimating  $\tilde{\mathcal{S}}_i$  via the formula

$$\tilde{\mathcal{S}}_i = \text{Span}(\nabla p_1|_{\tilde{\mathbf{x}}_i}, \dots, \nabla p_s|_{\tilde{\mathbf{x}}_i})^\perp, \quad (10)$$

- 4) and extracting the translation vector of  $\mathbb{A}_i$  and a basis for its linear part from a basis of  $\tilde{\mathcal{S}}_i$ .

According to Theorem 9, the above process will succeed, if i) the embedded points  $\tilde{\mathcal{X}}$  are in general position in  $\tilde{\Phi}$  (in the sense of

Definition 5), and ii) the union of linear subspaces  $\tilde{\Phi}$  is transversal. Note that these conditions need not be satisfied a-priori because of the particular structure of both the embedded data in (1) and the basis in (9). This gives rise to the following reasonable questions:

**Question 10.** *Under what conditions on  $\mathcal{X}$  and  $\Psi$ , will  $\tilde{\mathcal{X}}$  be in general position in  $\tilde{\Phi}$ ?*

**Question 11.** *Under what conditions on  $\Psi$  will  $\tilde{\Phi}$  be transversal?*

### 3.3 Contributions

The main contribution of this paper is to answer Questions 10-11.

Regarding Question 10, one may be tempted to conjecture that  $\tilde{\mathcal{X}}$  is in general position in  $\tilde{\Phi}$ , if the components of the points  $\mathcal{X}$  along the union  $\Phi := \bigcup_{i=1}^n \mathcal{S}_i$  of the linear parts of the affine subspaces are in general position inside  $\Phi$ . However, this conjecture is not true, as illustrated by the next example.

**Example 12.** *Suppose that  $\Psi = \mathbb{A}_1 \cup \mathbb{A}_2$  is a union of two affine planes  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$  of  $\mathbb{R}^3$ . Then  $\Phi = \mathcal{S}_1 \cup \mathcal{S}_2$  is a union of 2 planes in  $\mathbb{R}^3$  and as argued in Example 8, we can find 5 points in general position in  $\Phi$ . However,  $\tilde{\Phi} = \tilde{\mathcal{S}}_1 \cup \tilde{\mathcal{S}}_2$  is a union of 2 hyperplanes in  $\mathbb{R}^4$  and any subset of  $\tilde{\Phi}$  in general position must consist of at least  $\mathcal{M}_2(4) - 1 = \binom{2+3}{2} - 1 = 9$  points.<sup>6</sup>*

To state the precise necessary and sufficient condition for  $\tilde{\mathcal{X}}$  to be in general position in  $\tilde{\Phi}$ , we first show that  $\Psi$  is the zero-set of non-homogeneous polynomials of degree  $n$ .

**Proposition 13.** *Let  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  be a union of affine subspaces of  $\mathbb{R}^D$ , where each affine subspace  $\mathbb{A}_i$  is the translation of a linear subspace  $\mathcal{S}_i$  of codimension  $c_i$  by a translation vector  $\boldsymbol{\mu}_i$ . For each  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$ , let  $\mathbf{b}_{i1}, \dots, \mathbf{b}_{ic_i}$  be a basis for  $\mathcal{S}_i^\perp$ . Then  $\Psi$  is the zero set of all degree- $n$  polynomials of the form*

$$\prod_{i=1}^n (\mathbf{b}_{ij_i}^\top \mathbf{x} - \mathbf{b}_{ij_i}^\top \boldsymbol{\mu}_i) : (j_1, \dots, j_n) \in [c_1] \times \dots \times [c_n]. \quad (11)$$

Thanks to Proposition 13 we can define points  $\mathcal{X}$  to be in general position in  $\Psi$ , in analogy to Definition 5.

**Definition 14.** *Let  $\Psi$  be a union of  $n$  affine subspaces of  $\mathbb{R}^D$  and  $\mathcal{X}$  a finite subset of  $\Psi$ . We will say that  $\mathcal{X}$  is in general position in  $\Psi$ , if  $\Psi$  can be recovered as the zero set of all polynomials of degree  $n$  that vanish on  $\mathcal{X}$ . Equivalently, a polynomial of degree  $n$  vanishes on  $\Psi$  if and only if it vanishes on  $\mathcal{X}$ .*

We are now ready to answer our Question 10.

**Theorem 15.** *Let  $\mathcal{X}$  be a finite subset of a union of  $n$  affine subspaces  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  of  $\mathbb{R}^D$ , where  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$ , with  $\mathcal{S}_i$  a linear subspace of  $\mathbb{R}^D$  of codimension  $0 < c_i < D$ . Let  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  be the union of  $n$  linear subspaces of  $\mathbb{R}^{D+1}$  induced by the embedding  $\phi_0 : \mathbb{R}^D \hookrightarrow \mathbb{R}^{D+1}$  in (7). Denote by  $\tilde{\mathcal{X}} \subset \tilde{\Phi}$  the image of  $\mathcal{X}$  under  $\phi_0$ . Then  $\tilde{\mathcal{X}}$  is in general position in  $\tilde{\Phi}$  if and only if  $\mathcal{X}$  is in general position in  $\Psi$ .*

Our Second Theorem answers Question 11.

**Theorem 16.** *Let  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  be a union of  $n$  affine subspaces of  $\mathbb{R}^D$ , with  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$  and  $\boldsymbol{\mu}_i = \mathbf{B}_i \mathbf{a}_i$ , where  $\mathbf{B}_i \in \mathbb{R}^{D \times c_i}$  is a basis for  $\mathcal{S}_i^\perp$  with  $c_i = \text{codim } \mathcal{S}_i$ . If  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  is transversal*

<sup>6</sup> Otherwise one can fit a polynomial of degree 2 to the points, which does not vanish on  $\tilde{\Phi}$ .

and  $\mathbf{a}_1, \dots, \mathbf{a}_n$  do not lie in the zero-measure set of a proper algebraic variety<sup>7</sup> of  $\mathbb{R}^{c_1} \times \dots \times \mathbb{R}^{c_n}$ , then  $\tilde{\Phi}$  is transversal.

One may wonder if some of the  $\mu_i$  can be zero and  $\tilde{\Phi}$  still be transversal. This depends on the  $c_i$  as the next example shows.

**Example 17.** Let  $\mathbb{A}_1 = \text{Span}(\mathbf{b}_{11}, \mathbf{b}_{12})^\perp + \mu_1$  be an affine line and  $\mathbb{A}_2 = \text{Span}(\mathbf{b}_2)^\perp + \mu_2$  an affine plane of  $\mathbb{R}^3$ . Suppose that  $\tilde{\Phi} = \text{Span}(\mathbf{b}_{11}, \mathbf{b}_{12})^\perp \cup \text{Span}(\mathbf{b}_2)^\perp$  is transversal. Then  $\tilde{\Phi} = \tilde{\mathcal{S}}_1 \cup \tilde{\mathcal{S}}_2$  is transversal if and only if the matrix

$$\tilde{\mathbf{B}}_{[3]} = \begin{bmatrix} -\mathbf{b}_{11}^\top \mu_1 & -\mathbf{b}_{12}^\top \mu_1 & -\mathbf{b}_2^\top \mu_2 \\ \mathbf{b}_{11} & \mathbf{b}_{12} & \mathbf{b}_2 \end{bmatrix} \in \mathbb{R}^{4 \times 3} \quad (12)$$

has rank 3. But  $\text{rank}(\tilde{\mathbf{B}}_{[3]}) = 3$ , irrespectively of what the  $\mu_i$  are, simply because the matrix  $\mathbf{B}_{[3]} = [\mathbf{b}_{11} \ \mathbf{b}_{12} \ \mathbf{b}_2]$  is full rank (by the transversality assumption on  $\tilde{\Phi}$ ). Now let us replace the affine plane  $\mathbb{A}_2$  with a second affine line  $\mathbb{A}_2 = \text{Span}(\mathbf{b}_{21}, \mathbf{b}_{22})^\perp + \mu_2$ . Then  $\tilde{\Phi}$  is transversal if and only if

$$\tilde{\mathbf{B}}_{[3]} = \begin{bmatrix} -\mathbf{b}_{11}^\top \mu_1 & -\mathbf{b}_{12}^\top \mu_1 & -\mathbf{b}_{21}^\top \mu_2 & -\mathbf{b}_{22}^\top \mu_2 \\ \mathbf{b}_{11} & \mathbf{b}_{12} & \mathbf{b}_{21} & \mathbf{b}_{22} \end{bmatrix} \in \mathbb{R}^{4 \times 4} \quad (13)$$

has rank 4, which is impossible if both  $\mu_1, \mu_2$  are zero.

As a corollary of Theorems 9, 15 and 16, we get the correctness Theorem of ASC for the case of affine subspaces.

**Theorem 18.** Let  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  be a union of affine subspaces of  $\mathbb{R}^D$ , with  $\mathbb{A}_i = \mathcal{S}_i + \mu_i$  and  $\mu_i = \mathbf{B}_i \mathbf{a}_i$ , where  $\mathbf{B}_i \in \mathbb{R}^{D \times c_i}$  is a basis for  $\mathcal{S}_i^\perp$  with  $c_i = \text{codim } \mathcal{S}_i$ . Let  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  be the union of  $n$  linear subspaces of  $\mathbb{R}^{D+1}$  induced by the embedding  $\phi_0 : \mathbb{R}^D \hookrightarrow \mathbb{R}^{D+1}$  of (7). Let  $\mathcal{X}$  be a finite subset of  $\Psi$  and denote by  $\tilde{\mathcal{X}} \subset \tilde{\Phi}$  the image of  $\mathcal{X}$  under  $\phi_0$ . Let  $p_1, \dots, p_s$  be a basis for  $\mathcal{I}_{\tilde{\mathcal{X}, n}}$ , the vector space of homogeneous polynomials of degree  $n$  that vanish on  $\tilde{\mathcal{X}}$ . Let  $\mathbf{x} \in \mathcal{X} \cap \mathbb{A}_1 - \bigcup_{i>1} \mathbb{A}_i$ , and denote  $\tilde{\mathbf{x}} = \phi_0(\mathbf{x})$ . Define

$$\tilde{\mathbf{b}}_k := \nabla p_k|_{\tilde{\mathbf{x}}} \in \mathbb{R}^{D+1}, \quad k = 1, \dots, s, \quad (14)$$

and without loss of generality, let  $\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_\ell$  be a maximal linearly independent subset of  $\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_s$ . Define further  $(\gamma_k, \mathbf{b}_k) \in \mathbb{R} \times \mathbb{R}^D$  and  $(\gamma_1, \mathbf{B}_1) \in \mathbb{R}^\ell \times \mathbb{R}^{D \times \ell}$  as

$$\tilde{\mathbf{b}}_k =: \begin{bmatrix} \gamma_k \\ \mathbf{b}_k \end{bmatrix}, \quad k = 1, \dots, \ell \quad (15)$$

$$\gamma_1 := [\gamma_1, \dots, \gamma_\ell]^\top, \quad \mathbf{B}_1 := [\mathbf{b}_1, \dots, \mathbf{b}_\ell]. \quad (16)$$

If  $\mathcal{X}$  is in general position in  $\Psi$ ,  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  is transversal, and  $\mathbf{a}_1, \dots, \mathbf{a}_n$  do not lie in the zero-measure set of a proper algebraic variety of  $\mathbb{R}^{c_1} \times \dots \times \mathbb{R}^{c_n}$ , then

$$\mathbb{A}_1 = \text{Span}(\mathbf{B}_1)^\perp - \mathbf{B}_1 \left( \mathbf{B}_1^\top \mathbf{B}_1 \right)^{-1} \gamma_1. \quad (17)$$

**Remark 19.** The acute reader may notice that we still need to answer the question of whether  $\Psi$  admits a finite subset  $\mathcal{X}$  in general position, to begin with. This answer is affirmative: If  $\Psi$  satisfies the hypothesis of Theorem 16, then  $\tilde{\Phi}$  will be transversal, and so by Proposition 31  $\mathcal{I}_{\tilde{\Psi}}$  is generated in degree  $\leq n$ , in which case the existence of  $\mathcal{X}$  follows from Theorem 2.9 in [28].

The rest of the paper is organized as follows: in Section 4 we establish the fundamental algebraic-geometric properties of a union of affine subspaces. Then using these tools, we prove in Section 5 Theorems 15 and 16. The proof of Theorem 18 is straightforward is thus omitted.

7. The precise description of this algebraic variety is given in the proof of the Theorem in Section 5.2.

## 4 ALGEBRAIC GEOMETRY OF UNIONS OF AFFINE SUBSPACES

In Section 4.1 we describe the basic algebraic geometry of affine subspaces and unions thereof, in analogy to the case of linear subspaces. In particular, we show that a single affine subspace is the zero-set of polynomial equations of degree 1, and a union  $\Psi$  of affine subspaces is the zero-set of polynomial equations of degree  $n$ . In Section 4.2 we study more closely the embedding  $\mathbb{A} \xrightarrow{\phi_0} \tilde{\mathcal{S}}$  of an affine subspace  $\mathbb{A} \subset \mathbb{R}^D$  into its associated linear subspace  $\tilde{\mathcal{S}} \subset \mathbb{R}^{D+1}$  (see Section 3.2), which will lead to a deeper understanding of the embedding  $\Psi \xrightarrow{\phi_0} \tilde{\Phi}$  of a union of affine subspaces  $\Psi \subset \mathbb{R}^D$  into its associated union of linear subspaces  $\tilde{\Phi} \subset \mathbb{R}^{D+1}$ . As we will see,  $\Psi$  is dense in  $\tilde{\Phi}$  in a very precise sense, and the algebraic manifestation of this relation (Proposition 31) will be used later in Section 5.1, to prove our Theorem 15.

### 4.1 Affine Subspaces as Affine Varieties

Let  $\mathbb{A} = \mathcal{S} + \mu$  be an affine subspace of  $\mathbb{R}^D$  and let  $\mathbf{b}_1, \dots, \mathbf{b}_c$  be a basis for the orthogonal complement  $\mathcal{S}^\perp$  of  $\mathcal{S}$ . The first important observation is that a vector  $\mathbf{x}$  belongs to  $\mathcal{S}$  if and only if  $\mathbf{x} \perp \mathbf{b}_k$ ,  $\forall k = 1, \dots, c$ . In the language of algebraic geometry this is the same as saying that  $\mathcal{S}$  is the zero set of  $c$  linear polynomials:

$$\mathcal{S} = \mathcal{Z} \left( \mathbf{b}_1^\top \mathbf{x}, \dots, \mathbf{b}_c^\top \mathbf{x} \right), \quad \mathbf{x} := [x_1, \dots, x_D]^\top. \quad (18)$$

**Definition 20.** Let  $\mathcal{Y}$  be a subset of  $\mathbb{R}^D$ . The set  $\mathcal{I}_{\mathcal{Y}}$  of polynomials  $p(x_1, \dots, x_D)$  that vanish on  $\mathcal{Y}$ , i.e.,  $p(y_1, \dots, y_D) = 0$ ,  $\forall [y_1, \dots, y_D]^\top \in \mathcal{Y}$ , is called the vanishing ideal of  $\mathcal{Y}$ .

One may wonder if the linear polynomials  $\mathbf{b}_i^\top \mathbf{x}$ ,  $i = 1, \dots, c$ , form some sort of basis for the vanishing ideal  $\mathcal{I}_{\mathcal{S}}$  of  $\mathcal{S}$ . In fact this is true (see the appendix in [21] for a proof) and can be formalized by saying that these linear polynomials are generators of  $\mathcal{I}_{\mathcal{S}}$  over the polynomial ring  $\mathcal{R} = \mathbb{R}[x_1, \dots, x_D]$ . This means that every polynomial that belongs to  $\mathcal{I}_{\mathcal{S}}$  can be written as a linear combination of  $\mathbf{b}_1^\top \mathbf{x}, \dots, \mathbf{b}_c^\top \mathbf{x}$  with polynomial coefficients, i.e.,

$$p(\mathbf{x}) = p_1(\mathbf{x})(\mathbf{b}_1^\top \mathbf{x}) + \dots + p_c(\mathbf{x})(\mathbf{b}_c^\top \mathbf{x}) \quad (19)$$

where  $p_1, \dots, p_c$  are some polynomials in  $\mathcal{R}$ . More compactly

$$\mathcal{I}_{\mathcal{S}} = (\mathbf{b}_1^\top \mathbf{x}, \dots, \mathbf{b}_c^\top \mathbf{x}), \quad (20)$$

which reads as  $\mathcal{I}_{\mathcal{S}}$  is the ideal generated by the polynomials  $\mathbf{b}_1^\top \mathbf{x}, \dots, \mathbf{b}_c^\top \mathbf{x}$  as in (19). The following important fact<sup>8</sup> will be used in Section 5.1 to prove our Theorem 15.

**Proposition 21.** The vanishing ideal  $\mathcal{I}_{\mathcal{S}}$  of a linear subspace  $\mathcal{S}$  is always a prime ideal, i.e., if  $p, q$  are polynomials such that  $pq \in \mathcal{I}_{\mathcal{S}}$ , then either  $p \in \mathcal{I}_{\mathcal{S}}$  or  $q \in \mathcal{I}_{\mathcal{S}}$ .

Moving on, the second important observation is that  $\mathbf{x} \in \mathbb{A}$  if and only if  $\mathbf{x} - \mu \in \mathcal{S}$ . Equivalently,

$$\mathbf{x} \in \mathbb{A} \Leftrightarrow \mathbf{b}_k \perp \mathbf{x} - \mu, \quad \forall k = 1, \dots, c \quad (21)$$

or in algebraic geometric terms

$$\mathbb{A} = \mathcal{Z} \left( \mathbf{b}_1^\top \mathbf{x} - \mathbf{b}_1^\top \mu, \dots, \mathbf{b}_c^\top \mathbf{x} - \mathbf{b}_c^\top \mu \right). \quad (22)$$

In other words, the affine subspace  $\mathbb{A}$  is an algebraic variety of  $\mathbb{R}^D$ . In fact, we say that  $\mathbb{A}$  is an *affine variety*, since it is defined by non-homogeneous polynomials. To describe the *vanishing ideal*

8. For a proof see Appendix C in [21].

$\mathcal{I}_{\mathbb{A}}$  of  $\mathbb{A}$ , note that a polynomial  $p(x)$  vanishes on  $\mathbb{A}$  if and only if  $p(x + \mu)$  vanishes on  $\mathcal{S}$ . This, together with (20), give

$$\mathcal{I}_{\mathbb{A}} = \left( \mathbf{b}_1^\top x - \mathbf{b}_1^\top \mu, \dots, \mathbf{b}_c^\top x - \mathbf{b}_c^\top \mu \right). \quad (23)$$

Next, we consider a union  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  of affine subspaces  $\mathbb{A}_i = \mathcal{S}_i + \mu_i$ ,  $i \in [n]$ , of  $\mathbb{R}^D$ . We will prove Proposition 13, which describes  $\Psi$  as the zero-set of non-homogeneous polynomials of degree  $n$ , showing that  $\Psi$  is an affine variety of  $\mathbb{R}^D$ .

*Proof.* Denote the set of all polynomials of the form (11) by  $\mathcal{P}$ .

First, we show that  $\Psi \subset \mathcal{Z}(\mathcal{P})$ . Take  $\mathbf{x} \in \Psi$ ; we will show that  $\mathbf{x} \in \mathcal{Z}(\mathcal{P})$ . Since  $\Psi = \mathbb{A}_1 \cup \dots \cup \mathbb{A}_n$ ,  $\mathbf{x}$  belongs to at least one of the affine subspaces, say  $\mathbf{x} \in \mathbb{A}_i$ , for some  $i$ . For every polynomial  $p$  of  $\mathcal{P}$ , there is a linear factor  $\mathbf{b}_{ij}^\top x - \mathbf{b}_{ij}^\top \mu_i$  of  $p$  that vanishes on  $\mathbb{A}_i$  and thus on  $\mathbf{x}$ . Hence  $p$  itself will vanish on  $\mathbf{x}$ . Since  $p$  was an arbitrary element of  $\mathcal{P}$ , this shows that every polynomial of  $\mathcal{P}$  vanishes on  $\mathbf{x}$ , i.e.,  $\mathbf{x} \in \mathcal{Z}(\mathcal{P})$ .

Next, we show that  $\mathcal{Z}(\mathcal{P}) \subset \Psi$ . Let  $\mathbf{x} \in \mathcal{Z}(\mathcal{P})$ ; we will show that  $\mathbf{x} \in \Psi$ . If  $\mathbf{x}$  is a root of all polynomials  $p_{1j}(x) = \mathbf{b}_{1j}^\top x - \mathbf{b}_{1j}^\top \mu_1$ , then  $\mathbf{x} \in \mathbb{A}_1$  and we are done. Otherwise, one of these linear polynomials does not vanish on  $\mathbf{x}$ , say  $p_{11}(\mathbf{x}) \neq 0$ . Now suppose that  $\mathbf{x} \notin \Psi$ . By the above argument, for every affine subspace  $\mathbb{A}_i$  there must exist some linear polynomial  $\mathbf{b}_{i1}^\top x - \mathbf{b}_{i1}^\top \mu_i$ , which does not vanish on  $\mathbf{x}$ . As consequence, the polynomial

$$p(\mathbf{x}) = \prod_{i=1}^n (\mathbf{b}_{i1}^\top x - \mathbf{b}_{i1}^\top \mu_i) \quad (24)$$

does not vanish on  $\mathbf{x}$ , i.e.,  $p(\mathbf{x}) \neq 0$ . But because of the definition of  $\mathcal{P}$ , we must have that  $p \in \mathcal{P}$ . Since  $\mathbf{x}$  was selected to be an element of  $\mathcal{Z}(\mathcal{P})$ , we must have that  $p(\mathbf{x}) = 0$ , which is a contradiction, as we just saw that  $p(\mathbf{x}) \neq 0$ . Consequently, the hypothesis that  $\mathbf{x} \notin \Psi$ , must be false, i.e.,  $\mathcal{Z}(\mathcal{P}) \subset \Psi$ , and the proof is concluded.  $\square$

The reader may wonder what the vanishing ideal  $\mathcal{I}_{\Psi}$  of  $\Psi$  is and what its relation is to the linear polynomials whose products generate  $\Psi$ , as in Proposition 13. In fact, this question is still partially open even in the simpler case of a union of linear subspaces [33], [32], [27]. As it turns out,  $\mathcal{I}_{\Psi}$  is intimately related to  $\tilde{\mathcal{I}}_{\tilde{\Phi}}$ , where  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  is the union of linear subspaces associated to  $\Psi$  under the embedding  $\hat{\phi}_0$  of (7). It is precisely this relation that will enable us to prove Theorem 15, and to elucidate it we need the notion of *projective closure* that we introduce next.<sup>9</sup>

## 4.2 The Projective Closure of Affine Subspaces

Let  $\hat{\phi}_0(\mathbb{A})$  be the image of  $\mathbb{A} = \mathcal{S} + \mu$  under the embedding  $\hat{\phi}_0 : \mathbb{R}^D \hookrightarrow \mathbb{R}^{D+1}$  in (7). Let  $\tilde{\mathcal{S}}$  be the  $(d+1)$ -dimensional linear subspace of  $\mathbb{R}^{D+1}$  spanned by the columns of  $\tilde{\mathbf{U}}$  (see (9)). A basis for the orthogonal complement of  $\tilde{\mathcal{S}}$  in  $\mathbb{R}^{D+1}$  is

$$\tilde{\mathbf{b}}_1 := \begin{bmatrix} -\mathbf{b}_1^\top \mu \\ \mathbf{b}_1 \end{bmatrix}, \dots, \tilde{\mathbf{b}}_c := \begin{bmatrix} -\mathbf{b}_c^\top \mu \\ \mathbf{b}_c \end{bmatrix}, \quad (25)$$

since  $\text{codim}(\tilde{\mathcal{S}}_i) = \text{codim}(\mathcal{S})$ , and the  $\tilde{\mathbf{b}}_i$  are linearly independent because the  $\mathbf{b}_i$  are. In algebraic geometric terms

$$\begin{aligned} \tilde{\mathcal{S}} &= \mathcal{Z} \left( \mathbf{b}_1^\top x - (\mathbf{b}_1^\top \mu)x_0, \dots, \mathbf{b}_c^\top x - (\mathbf{b}_c^\top \mu)x_0 \right) \\ &= \mathcal{Z} \left( \tilde{\mathbf{b}}_1^\top \tilde{x}, \dots, \tilde{\mathbf{b}}_c^\top \tilde{x} \right), \quad \tilde{x} := [x_0, x_1, \dots, x_D]^\top. \end{aligned} \quad (26)$$

9. Of course, the notion of *projective closure* is a well-known concept in algebraic geometry; here we introduce it in a self-contained fashion in the context of unions of affine subspaces, dispensing with unnecessary abstractions.

By inspecting equations (22) and (26), we see that every point of  $\hat{\phi}_0(\mathbb{A})$  satisfies the equations (26) of  $\tilde{\mathcal{S}}$ . Since these equations are homogeneous, it will in fact be true that for any point  $\tilde{\mathbf{x}} \in \hat{\phi}_0(\mathbb{A})$  the entire line of  $\mathbb{R}^{D+1}$  spanned by  $\tilde{\mathbf{x}}$  will still lie in  $\tilde{\mathcal{S}}$ . Hence, we may as well think of the embedding  $\hat{\phi}_0$  as mapping a point  $\mathbf{x} \in \mathbb{R}^D$  to a line of  $\mathbb{R}^{D+1}$ . To formalize this concept, we need the notion of *projective space* [30], [34]:

**Definition 22.** *The real projective space  $\mathbb{P}^D$  is defined to be the set of all lines through the origin in  $\mathbb{R}^{D+1}$ . Each non-zero vector  $\alpha$  of  $\mathbb{R}^{D+1}$  defines an element  $[\alpha]$  of  $\mathbb{P}^D$ , and two elements  $[\alpha], [\beta]$  of  $\mathbb{P}^D$  are equal in  $\mathbb{P}^D$ , if and only if there exists a nonzero  $\lambda \in \mathbb{R}$  such that we have an equality  $\alpha = \lambda\beta$  of vectors in  $\mathbb{R}^{D+1}$ . For each point  $[\alpha] \in \mathbb{P}^D$ , we call the point  $\alpha \in \mathbb{R}^{D+1}$  a representative of  $[\alpha]$ .*

Now we can define a new embedding  $\hat{\phi}_0 : \mathbb{R}^D \rightarrow \mathbb{P}^D$ , that behaves exactly as  $\hat{\phi}_0$  in (7), except that it now takes points of  $\mathbb{R}^D$  to lines of  $\mathbb{R}^{D+1}$ , or more precisely, to elements of  $\mathbb{P}^D$ :

$$(\alpha_1, \alpha_2, \dots, \alpha_D) \xrightarrow{\hat{\phi}_0} [(1, \alpha_1, \alpha_2, \dots, \alpha_D)]. \quad (27)$$

A point  $\mathbf{x}$  of  $\mathbb{A}$  is mapped by  $\hat{\phi}_0$  to a line inside  $\tilde{\mathcal{S}}$ , or more specifically, to the point  $[\tilde{\mathbf{x}}]$  of  $\mathbb{P}^D$ , whose representative  $\tilde{\mathbf{x}}$  satisfies the equations (26) of  $\tilde{\mathcal{S}}$ . The set of all lines of  $\mathbb{R}^{D+1}$  that live in  $\tilde{\mathcal{S}}$ , viewed as elements of  $\mathbb{P}^D$ , is denoted by  $[\tilde{\mathcal{S}}]$ , i.e.,

$$[\tilde{\mathcal{S}}] = \left\{ [\alpha] \in \mathbb{P}^D : \alpha \in \tilde{\mathcal{S}} \right\}. \quad (28)$$

The representative  $\alpha$  of every element  $[\alpha] \in [\tilde{\mathcal{S}}]$  satisfies by definition the equations (26) of  $\tilde{\mathcal{S}}$ , and so  $[\tilde{\mathcal{S}}]$  has naturally the structure of an algebraic variety of  $\mathbb{P}^D$ , which is called a *projective variety*. We emphasize that even though the varieties  $\tilde{\mathcal{S}}$  and  $[\tilde{\mathcal{S}}]$  live in different spaces,  $\mathbb{R}^{D+1}$  and  $\mathbb{P}^D$  respectively, they are defined by the same equations. In fact, every algebraic variety  $\mathcal{Y}$  of  $\mathbb{R}^{D+1}$  that is the unions of lines, which is true if and only if  $\mathcal{Y}$  is defined by homogeneous equations, gives rise to a projective variety  $[\mathcal{Y}]$  of  $\mathbb{P}^D$  defined by the same equations.

**Example 23.** *Recall from Section 2.2 that a union  $\tilde{\Phi}$  of linear subspaces is defined as the zero-set of homogeneous polynomials. Then  $\tilde{\Phi}$  gives rise to a projective variety  $[\tilde{\Phi}]$  of  $\mathbb{P}^D$  defined by the same equations as  $\tilde{\Phi}$ , which can be thought of as the set of lines through the origin in  $\mathbb{R}^{D+1}$  that live in  $\tilde{\Phi}$ .*

Returning to our embedding  $\hat{\phi}_0$ , to describe the precise connection between  $\hat{\phi}_0(\mathbb{A})$  and  $[\tilde{\mathcal{S}}]$  we need to resort to the kind of topology that is most suitable for the study of algebraic varieties [30], [34]:

**Definition 24 (Zariski Topology).** *The real vector space  $\mathbb{R}^D$  and the projective space  $\mathbb{P}^D$  can be made into topological spaces, by defining the closed sets of their associated topology to be all the algebraic varieties in  $\mathbb{R}^D$  and  $\mathbb{P}^D$  respectively.*

We are finally ready to state without proof the formal algebraic geometric relation between  $\hat{\phi}_0(\mathbb{A})$  and  $\tilde{\mathcal{S}}$ :

**Proposition 25.** *In the Zariski topology, the set  $\hat{\phi}_0(\mathbb{A})$  is open and dense in  $[\tilde{\mathcal{S}}]$ , in particular  $[\tilde{\mathcal{S}}]$  is the closure<sup>10</sup> of  $\hat{\phi}_0(\mathbb{A})$  in  $\mathbb{P}^D$ .*

The projective variety  $[\tilde{\mathcal{S}}]$  is called the *projective closure* of  $\mathbb{A}$ : it is the smallest projective variety that contains  $\hat{\phi}_0(\mathbb{A})$ . We now characterize the projective closure of a union of affine subspaces.

10. It can further be shown that  $[\tilde{\mathcal{S}}] = \hat{\phi}_0(\mathbb{A}) \cup [\mathcal{S}]$ : intuitively, the set that we need to add to  $\hat{\phi}_0(\mathbb{A})$  to get a closed set is the *slope*  $[\mathcal{S}]$  of  $\mathbb{A}$ .

**Proposition 26.** Let  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  be a union of affine subspaces of  $\mathbb{R}^D$ . Then the projective closure of  $\Psi$  in  $\mathbb{P}^D$ , i.e., the smallest projective variety that contains  $\hat{\phi}_0(\Psi)$ , is

$$\bigcup_{i=1}^n [\tilde{\mathcal{S}}_i] = \left[ \bigcup_{i=1}^n \tilde{\mathcal{S}}_i \right] = [\tilde{\Phi}], \quad (29)$$

where  $\tilde{\mathcal{S}}_i$  is the linear subspace of  $\mathbb{R}^{D+1}$  corresponding to  $\mathbb{A}_i$  under the embedding  $\hat{\phi}_0$  of (7).

The geometric fact that  $[\tilde{\Phi}] \subset \mathbb{P}^D$  is the smallest projective variety of  $\mathbb{P}^D$  that contains  $\hat{\phi}_0(\Psi)$ , manifests itself algebraically in  $\mathcal{I}_\Psi$  being uniquely defined by  $\mathcal{I}_{\tilde{\Phi}}$  and vice versa, in a very precise fashion. To describe this relation, we need a definition.

**Definition 27** (Homogenization - Dehomogenization). Let  $p \in \mathcal{R} = \mathbb{R}[x_1, \dots, x_D]$  be a polynomial of degree  $n$ . The homogenization of  $p$  is the homogeneous polynomial

$$p^{(h)} = x_0^n p \left( \frac{x_1}{x_0}, \frac{x_2}{x_0}, \dots, \frac{x_D}{x_0} \right) \quad (30)$$

of  $\tilde{\mathcal{R}} = \mathbb{R}[x_0, x_1, \dots, x_D]$  of degree  $n$ . Conversely, if  $P \in \tilde{\mathcal{R}}$  is homogeneous of degree  $n$ , its dehomogenization is  $P_{(d)} = P(1, x_1, \dots, x_D)$ , which is a polynomial of  $\mathcal{R}$  of degree  $\leq n$ .

**Example 28.** Let  $P = x_0^2 x_1 + x_0 x_2^2 + x_1 x_2 x_3$  be a homogeneous polynomial of degree 3. Its dehomogenization is the degree-3 polynomial  $P_{(d)} = x_1 + x_2^2 + x_1 x_2 x_3$ , and the homogenization of  $P_{(d)}$  is  $(P_{(d)})^{(h)} = x_0^3 \left( \frac{x_1}{x_0} + \frac{x_2^2}{x_0} + \frac{x_1 x_2 x_3}{x_0^3} \right) = P$ .

The next result from algebraic geometry is crucial for our purpose.

**Theorem 29** (Chapter 8 in [30]). Let  $\mathcal{Y}$  be an affine variety of  $\mathbb{R}^D$  and let  $\tilde{\mathcal{Y}}$  be its projective closure in  $\mathbb{P}^D$  with respect to the embedding  $\hat{\phi}_0$  of (27). Let  $\mathcal{I}_{\mathcal{Y}}, \mathcal{I}_{\tilde{\mathcal{Y}}}$  be the vanishing ideals of  $\mathcal{Y}, \tilde{\mathcal{Y}}$  respectively. Then  $\mathcal{I}_{\tilde{\mathcal{Y}}} = \mathcal{I}_{\mathcal{Y}}^{(h)}$ , i.e., every element of  $\mathcal{I}_{\tilde{\mathcal{Y}}}$  arises as a homogenization of some element of  $\mathcal{I}_{\mathcal{Y}}$ , and every element of  $\mathcal{I}_{\mathcal{Y}}$  arises as the dehomogenization of some element of  $\mathcal{I}_{\tilde{\mathcal{Y}}}$ .

We have already seen that  $\tilde{\Phi}$  and  $[\tilde{\Phi}]$  are given as algebraic varieties by identical equations. It is also not hard to see that the vanishing ideals of these varieties are identical as well.

**Lemma 30.** Let  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  be a union of linear subspaces of  $\mathbb{R}^{D+1}$ , and let  $[\tilde{\Phi}] = \bigcup_{i=1}^n [\tilde{\mathcal{S}}_i]$  be the corresponding projective variety of  $\mathbb{P}^D$ . Then  $\mathcal{I}_{[\tilde{\Phi}],k} = \mathcal{I}_{\tilde{\Phi},k}$ , i.e., a degree- $k$  homogeneous polynomial vanishes on  $[\tilde{\Phi}]$  if and only if it vanishes on  $\tilde{\Phi}$ .

As a Corollary of Theorem 29 and Lemma 30, we obtain the key result of this Section, which we will use in Section 5.1.

**Proposition 31.** Let  $\Psi = \bigcup_{i=1}^n \mathbb{A}_i$  be a union of affine subspaces of  $\mathbb{R}^D$ . Let  $\tilde{\Phi} = \bigcup_{i=1}^n \tilde{\mathcal{S}}_i$  be the union of linear subspaces of  $\mathbb{R}^{D+1}$  associated to  $\Psi$  under the embedding  $\hat{\phi}_0$  of (7). Then  $\mathcal{I}_{\tilde{\Phi}}$  is the homogenization of  $\mathcal{I}_\Psi$ .

## 5 PROOFS OF MAIN THEOREMS

### 5.1 Proof of Theorem 15

( $\Rightarrow$ ) Suppose that  $\mathcal{X}$  is in general position in  $\Psi$ . We need to show that  $\tilde{\mathcal{X}}$  is in general position in  $\tilde{\Phi}$ . In view of Proposition 6, and the fact that  $\mathcal{I}_{\tilde{\Phi},n} \subset \mathcal{I}_{\tilde{\mathcal{X}},n}$ , it is sufficient to show that  $\mathcal{I}_{\tilde{\Phi},n} \supset \mathcal{I}_{\tilde{\mathcal{X}},n}$ . To that end, let  $P$  be a homogeneous polynomial of degree  $n$  in

$\mathbb{R}[x_0, x_1, \dots, x_D]$  that vanishes on the points  $\tilde{\mathcal{X}}$ , i.e.,  $P \in \mathcal{I}_{\tilde{\mathcal{X}},n}$ . Then for every point  $\tilde{\alpha} = (1, \alpha_1, \dots, \alpha_D)$  of  $\tilde{\mathcal{X}}$ , we have

$$P(\tilde{\alpha}) = P(1, \alpha_1, \dots, \alpha_D) = P_{(d)}(\alpha_1, \dots, \alpha_D) = 0, \quad (31)$$

that is, the dehomogenization  $P_{(d)}$  of  $P$  vanishes on all points of  $\mathcal{X}$ , i.e.,  $P_{(d)} \in \mathcal{I}_{\mathcal{X}}$ . Now there are two possibilities: either  $P_{(d)}$  has degree  $n$ , in which case  $P = (P_{(d)})^{(h)}$ , or  $P_{(d)}$  has degree strictly less than  $n$ , say  $n - k$ ,  $k \geq 1$ , in which case  $P = x_0^k (P_{(d)})^{(h)}$ . If  $P_{(d)}$  has total degree  $n$ , by the general position assumption on  $\mathcal{X}$ ,  $P_{(d)}$  must vanish on  $\Psi$ . Then by Proposition 31,  $(P_{(d)})^{(h)} \in \mathcal{I}_{\Phi,n}$ , and so  $P \in \mathcal{I}_{\Phi,n}$ . If  $\deg P_{(d)} = n - k$ ,  $k \geq 1$ , suppose we can find a linear form  $G = \tilde{\zeta}^\top \tilde{x}$ , that does not vanish on any of the  $\tilde{\mathcal{S}}_i$ ,  $i \in [n]$ , and it is not divisible by  $x_0$ . Then  $G_{(d)}$  will have degree 1 and will not vanish on any of the  $\mathbb{A}_i$ ,  $i \in [n]$ . Also  $(G_{(d)})^k P_{(d)}$  has degree  $n$  and vanishes on  $\mathcal{X}$ . Since  $\mathcal{X}$  is in general position in  $\Psi$ , we will have that  $(G_{(d)})^k P_{(d)}$  vanishes on  $\Psi$ . Then by Proposition 31,  $G^k (P_{(d)})^{(h)} \in \mathcal{I}_{\tilde{\Phi},n}$ . Since  $\mathcal{I}_{\tilde{\Phi}} = \bigcap_{i=1}^n \mathcal{I}_{\tilde{\mathcal{S}}_i}$  we must have that  $G^k (P_{(d)})^{(h)} \in \mathcal{I}_{\tilde{\mathcal{S}}_i}$ ,  $\forall i \in [n]$ . Since  $\mathcal{I}_{\tilde{\mathcal{S}}_i}$  is a prime ideal (Proposition 21) and  $G \notin \mathcal{I}_{\tilde{\mathcal{S}}_i}$ , it must be the case that  $(P_{(d)})^{(h)} \in \mathcal{I}_{\tilde{\mathcal{S}}_i}$ ,  $\forall i \in [n]$ , i.e.,  $(P_{(d)})^{(h)} \in \mathcal{I}_{\tilde{\Phi}}$ . But  $P = x_0^k (P_{(d)})^{(h)}$ , which shows that  $P \in \mathcal{I}_{\Phi,n}$ .

It remains to be shown that there exists a linear form  $G$  non-divisible by  $x_0$ , that does not vanish on any of the  $\tilde{\mathcal{S}}_i$ . Suppose this is not true; thus if  $G = \mathbf{b}^\top x + \alpha x_0$  is a linear form non-divisible by  $x_0$ , i.e.,  $\mathbf{b} \neq \mathbf{0}$ , then  $G$  must vanish on some  $\tilde{\mathcal{S}}_i$ . In particular, for any non-zero vector  $\mathbf{b}$  of  $\mathbb{R}^D$ ,  $\mathbf{b}^\top x = \mathbf{b}^\top x + 0x_0$  must vanish on some  $\tilde{\mathcal{S}}_i$ . Recall from Section 3.2, that if  $\mathbf{u}_{i1}, \dots, \mathbf{u}_{id_i}$  is a basis for  $\mathcal{S}_i$ , the linear part of  $\mathbb{A}_i = \mathcal{S}_i + \boldsymbol{\mu}_i$ , then

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ \boldsymbol{\mu}_i & \mathbf{u}_{i1} & \cdots & \mathbf{u}_{id_i} \end{bmatrix} \quad (32)$$

is a basis for  $\tilde{\mathcal{S}}_i$ . Since  $\mathbf{b}^\top x$  vanishes on  $\tilde{\mathcal{S}}_i$ , it must vanish on each basis vector of  $\tilde{\mathcal{S}}_i$ . In particular,  $\mathbf{b}^\top \mathbf{u}_{i1} = \cdots = \mathbf{b}^\top \mathbf{u}_{id_i} = 0$ , which implies that the linear form  $\mathbf{b}^\top x$ , now viewed as a function on  $\mathbb{R}^D$ , vanishes on  $\mathcal{S}_i$ , i.e.,  $\mathbf{b}^\top x \in \mathcal{I}_{\mathcal{S}_i}$ . To summarize, we have shown that for every  $\mathbf{0} \neq \mathbf{b} \in \mathbb{R}^D$ , there exists an  $i \in [n]$  such that  $\mathbf{b}^\top x \in \mathcal{I}_{\mathcal{S}_i}$ . Taking  $\mathbf{b}$  equal to the standard vector  $\mathbf{e}_1$  of  $\mathbb{R}^D$ , we see that the linear form  $x_1$  must vanish on some  $\mathcal{S}_i$ , and similarly for the linear forms  $x_2, \dots, x_D$ . This in turn means that the ideal  $\mathfrak{m} := (x_1, \dots, x_D)$  generated by the linear forms  $x_1, \dots, x_D$ , must lie in the union  $\bigcup_{i=1}^n \mathcal{I}_{\mathcal{S}_i}$ . But it is known from Proposition 1.11(i) in [29], that if an ideal  $\mathfrak{a}$  lies in the union of finitely many prime ideals, then the  $\mathfrak{a}$  must lie in one of these prime ideals. Applying this result to our case, we see that, since the  $\mathcal{I}_{\mathcal{S}_i}$  are prime ideals,  $\mathfrak{m} \subset \mathcal{I}_{\mathcal{S}_i}$  for some  $i \in [n]$ . But this says that for any vector in  $\mathcal{S}_i$  all of its coordinates must be zero, i.e.,  $\mathcal{S}_i = 0$ , which violates the assumption  $d_i > 0, \forall i \in [n]$ . This contradiction proves the existence of our linear form  $G$ .

( $\Leftarrow$ ) Now suppose that  $\tilde{\mathcal{X}}$  is in general position in  $\tilde{\Phi}$ . We need to show that  $\mathcal{X}$  is in general position in  $\Psi$ . To that end, let  $p$  be a vanishing polynomial of  $\Psi$  of degree  $n$ , then clearly  $p \in \mathcal{I}_{\mathcal{X}}$ . Conversely, let  $p \in \mathcal{I}_{\mathcal{X}}$  of degree  $n$ . Then for each point  $\alpha \in \mathcal{X}$

$$\begin{aligned} 0 &= p(\alpha) = p(\alpha_1, \dots, \alpha_D) \\ &= p^{(h)}(1, \alpha_1, \dots, \alpha_D) = p^{(h)}(\tilde{\alpha}), \end{aligned} \quad (33)$$

i.e., the homogenization  $p^{(h)}$  vanishes on  $\tilde{\mathcal{X}}$ . By hypothesis  $\tilde{\mathcal{X}}$  is in general position in  $\tilde{\Phi}$ , hence  $p^{(h)} \in \mathcal{I}_{\tilde{\Phi},n}$ . Then by Proposition

31, the dehomogenization of  $p^{(h)}$  must vanish on  $\Psi$ . But notice that  $(p^{(h)})_{(d)} = p$ , and so  $p$  vanishes on  $\Psi$ .

## 5.2 Proof of Theorem 16

Let  $\mathbf{b}_{i_1}, \dots, \mathbf{b}_{i_{c_i}}$  be an orthonormal basis for  $\mathcal{S}_i^\perp$ , then

$$\tilde{\mathbf{b}}_{i_1}, \dots, \tilde{\mathbf{b}}_{i_{c_i}}, \tilde{\mathbf{b}}_{i_{j_i}} := [\mathbf{b}_{i_{j_i}}^\top - \mathbf{b}_{i_{j_i}}^\top \mathbf{B}_i \mathbf{a}_i]^\top, \quad (34)$$

is a basis for  $\tilde{\mathcal{S}}_i^\perp$ . Suppose that  $\tilde{\Phi}$  is not transversal. Then there exists some index set  $\mathfrak{J} \subset [n]$ , say without loss of generality  $\mathfrak{J} = \{1, \dots, \ell\}$ ,  $\ell \leq n$ , such that (see also Section 2.4)

$$\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) < \min \left\{ D + 1, \sum_{i \in \mathfrak{J}} c_i \right\}, \quad (35)$$

$$\tilde{\mathbf{B}}_{\mathfrak{J}} := [\tilde{\mathbf{B}}_1, \dots, \tilde{\mathbf{B}}_\ell], \quad \tilde{\mathbf{B}}_i := [\tilde{\mathbf{b}}_{i_1}, \dots, \tilde{\mathbf{b}}_{i_{c_i}}], \quad (36)$$

where we have used the fact that  $\text{codim} \tilde{\mathcal{S}}_i = \text{codim} \mathcal{S}_i = c_i$ ,  $\forall i \in [n]$ . Since  $\tilde{\Phi}$  is transversal, we must have either  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = D$  or  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = \sum_{i \in \mathfrak{J}} c_i$ . Suppose the latter condition is true, then  $\sum_{i \in \mathfrak{J}} c_i \leq D$ . Then all columns of  $\tilde{\mathbf{B}}_{\mathfrak{J}}$  are linearly independent, which implies that the same will be true for the columns of  $\tilde{\mathbf{B}}_{\mathfrak{J}}$ , and so  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = \sum_{i \in \mathfrak{J}} c_i$ . Since by hypothesis  $\sum_{i \in \mathfrak{J}} c_i \leq D$ , we must have

$$\text{codim} \bigcap_{i \in \mathfrak{J}} \tilde{\mathcal{S}}_i = \text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = \min \left\{ D + 1, \sum_{i \in \mathfrak{J}} c_i \right\}, \quad (37)$$

and so the transversality condition is satisfied for  $\mathfrak{J}$ , which is a contradiction on the hypothesis (35). Consequently, it must be the case that  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = D < \sum_{i \in \mathfrak{J}} c_i$ . Since  $\tilde{\mathbf{B}}_{\mathfrak{J}}$  is a submatrix of  $\tilde{\mathbf{B}}_{\mathfrak{J}}$ , we must have that  $\text{rank} \tilde{\mathbf{B}}_{\mathfrak{J}} \geq D$ . On the other hand, because of (35) we must have  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) \leq D$ , i.e.,  $\text{rank}(\tilde{\mathbf{B}}_{\mathfrak{J}}) = D$ . Now  $\tilde{\mathbf{B}}_{\mathfrak{J}}$  is a  $(D+1) \times (\sum_{i \in \mathfrak{J}} c_i)$  matrix, with the smaller dimension being  $(D+1)$ . Since its rank is  $D$ , it must be the case that all  $(D+1) \times (D+1)$  minors of  $\tilde{\mathbf{B}}_{\mathfrak{J}}$  vanish. The vanishing of these minors defines an algebraic variety  $\mathcal{W}_{\mathfrak{J}}$  of the parametric space  $\prod_{i=1}^n \mathbb{R}^{c_i}$ , and  $\tilde{\Phi}$  is non-transversal if and only if  $(\mathbf{a}_1, \dots, \mathbf{a}_n) \in \mathcal{W} := \bigcup_{\mathfrak{J} \subset [n]} \mathcal{W}_{\mathfrak{J}}$ . Since  $\mathcal{W}$  is a finite union of algebraic varieties it must be an algebraic variety itself, i.e., defined by a set of polynomial equations in the variables  $\mathbf{a}_1, \dots, \mathbf{a}_n$ .

## 6 CONCLUSIONS

We established in a rigorous fashion the correctness of ASC in the case of affine subspaces. Using the technical framework of algebraic geometry, we showed that the embedding of points lying in general position inside a union of affine subspaces preserves the general position. Moreover, the embedding of a transversal union of affine subspaces will almost surely give a transversal union of linear subspaces. Future research will aim at optimal realizations of the embedding in the presence of noise, analyzing SSC for affine subspaces, and reducing the complexity of ASC.

## REFERENCES

- [1] R. Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 52–68, March 2011.
- [2] P. S. Bradley and O. L. Mangasarian, "k-plane clustering," *Journal of Global Optimization*, vol. 16, no. 1, pp. 23–32, 2000.
- [3] P. Tseng, "Nearest  $q$ -flat to  $m$  points," *Journal of Optimization Theory and Applications*, vol. 105, no. 1, pp. 249–252, 2000.
- [4] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

- [5] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1–15, 2005.
- [6] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *International Journal of Computer Vision*, vol. 81, no. 3, pp. 317–330, 2009.
- [7] R. Heckel and H. Bölcskei, "Robust subspace clustering via thresholding," *CoRR*, vol. abs/1307.4891, 2013.
- [8] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recognition Letters*, vol. 43, pp. 47–61, 2014.
- [9] G. Liu, Z. Lin, S. Yan, J. Sun, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, Jan 2013.
- [10] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.
- [11] C. You, D. Robinson, and R. Vidal, "Scalable sparse subspace clustering by orthogonal matching pursuit," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [12] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *European Conference on Computer Vision*, 2012.
- [13] R. Livni, D. Lehari, S. Schein, H. Nachliely, S. Shalev-shwartz, and A. Globerson, "Vanishing component analysis," in *International Conference on Machine Learning*, vol. 28, no. 1, 2013, pp. 597–605.
- [14] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis (GPCA)," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. 621–628.
- [15] M. Soltanolkotabi, E. Elhamifar, and E. J. Candès, "Robust subspace clustering," *Annals of Statistics*, vol. 42, no. 2, pp. 669–699, 2014.
- [16] R. Vidal, Y. Ma, S. Soatto, and S. Sastry, "Two-view multibody structure from motion," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 7–25, 2006.
- [17] A. Sampath and J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1554–1567, 2010.
- [18] R. Vidal, "Recursive identification of switched ARX systems," *Automatica*, vol. 44, no. 9, pp. 2274–2287, September 2008.
- [19] M. C. Tsakiris and R. Vidal, "Filtrated spectral algebraic subspace clustering," in *ICCV Workshop on Robust Subspace Learning and Computer Vision*, 2015, pp. 28–36.
- [20] —, "Abstract algebraic-geometric subspace clustering," in *Asilomar Conference on Signals, Systems and Computers*, 2014.
- [21] —, "Filtrated algebraic subspace clustering," *SIAM Journal on Imaging Sciences*, 2017 (to appear).
- [22] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [23] T. Zhang, A. Szlam, Y. Wang, and G. Lerman, "Hybrid linear modeling via local best-fit flats," *International Journal of Computer Vision*, vol. 100, no. 3, pp. 217–240, 2012.
- [24] Y. Wang, Y.-X. Wang, and A. Singh, "Graph connectivity in noisy sparse subspace clustering," *International Conference on Artificial Intelligence and Statistics*, pp. 538–546, 2016.
- [25] Y. Yang, J. Feng, N. Jovic, J. Yang, and T. S. Huang, " $\ell_0$ -sparse subspace clustering," *European Conference on Computer Vision*, pp. 731–747, 2016.
- [26] R. Vidal, "Generalized principal component analysis (gpc): an algebraic geometric approach to subspace clustering and motion segmentation," Ph.D. dissertation, University of California, Berkeley, August 2003.
- [27] H. Derksen, "Hilbert series of subspace arrangements," *Journal of Pure and Applied Algebra*, vol. 209, no. 1, pp. 91–98, 2007.
- [28] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Review*, vol. 50, no. 3, pp. 413–458, 2008.
- [29] M. Atiyah and I. MacDonald, *Introduction to Commutative Algebra*. Westview Press, 1994.
- [30] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*. Springer, 2007.
- [31] P. Bürgisser and F. Cucker, "The geometry of numerical algorithms," *Springer Science & Business Media*, vol. 349, 2013.
- [32] H. Derksen and J. Sidman, "A sharp bound for the castelnuovo-mumford regularity of subspace arrangements," *Advances in Mathematics*, vol. 172, pp. 151–157, 2002.
- [33] A. Conca and J. Herzog, "Castelnuovo-mumford regularity of products of ideals," *Collectanea Mathematica*, vol. 54, no. 2, pp. 137–152, 2003.
- [34] R. Hartshorne, *Algebraic Geometry*. Springer, 1977.