Hilbert Functions and Applications to the Estimation of Subspace Arrangements*

Allen Y. Yang, Shankar Rao, Andrew Wagner, Yi Ma
Coordinated Science Lab, UIUC
1308 W. Main, Urbana, Illinois 61801
{yangyang,srrao,awagner,yima}@uiuc.edu

Robert M. Fossum
Department of Mathematics, UIUC
1409 W. Green, Urbana, Illinois 61801
rmfossum@uiuc.edu

Abstract

This paper develops a new mathematical framework for studying the subspace-segmentation problem. We examine some important algebraic properties of subspace arrangements *that are closely related to the subspace-segmentation problem. More specifically, we introduce an important class of invariants given by the* Hilbert functions*. We show that there exist rich relations between subspace arrangements and their corresponding Hilbert functions. We propose a new subspace-segmentation algorithm, and showcase two applications to demonstrate how the new theoretical revelation may solve subspace segmentation and model selection problems under less restrictive conditions with improved results.*

1. Introduction

There has been an increasing focus on the estimation and segmentation of multiple-subspace structures in imagery or visual data, such as the problems of motion segmentation [4, 16], face recognition [1], and image representation [10]. This has led to the development of many effective algorithms that generalize the classical principal component analysis (PCA) method. PCA fits one subspace to the data, while new scenarios fit multiple subspaces. These algorithms are based on either heuristic methods (K-subspaces [9]), EM [5]), or algebraic methods (GPCA [17]).

Essentially, all these methods address different aspects of a common mathematical problem: *"How do we represent (or approximate) a given set of data points in a high dimensional space by a collection (mixture) of multiple subspaces?"* A complete solution includes answers to the following closely related sub-problems:

1. Estimating the number of subspaces.
2. Segmenting data points into subsets, each of which is modeled by a single subspace.
3. Estimating the dimension of and a basis for each subspace.

These problems are rich with algebraic, geometric, and combinatorial properties. However, these properties have not been sufficiently exploited by the existing methods. For example, in GPCA [17], the dataset is first fit with a set of polynomials and the individual subspaces are retrieved from the derivatives of the polynomials. In the presence of noise, it is difficult to determine robustly the number of fitting polynomials and the dimensions of the individual subspaces. As we will show later, there are important combinatorial relations among the number and dimensions, which may significantly facilitate this task.

In mathematics, a union of multiple subspaces is formally referred to as a *subspace arrangement*. Subspace arrangements constitute a very special but important class of algebraic sets that have been studied in mathematics for centuries [2, 3, 12]. The importance as well as the difficulty of studying subspace arrangements can hardly be exaggerated. Different aspects of their properties have been and are still being investigated and exploited in many mathematical fields, including algebraic geometry & topology, combinatorics and complexity theory, and graph and lattice theory, etc [2]. Although the results about subspace arrangements are extremely rich and deep, only a few special classes have been fully characterized.

Contributions of This Paper. We introduce and examine some important mathematical concepts and properties of subspace arrangements that are closely related to the subspace-segmentation problem. We study an important class of invariants of subspace arrangements given by the *Hilbert function* [8], and propose a new solution to the subspace-segmentation problem. Especially for the problem of multiple-motion segmentation, several specialized algorithms based on special subspace models are very well-known (e.g., [4, 7, 11, 13, 16]), but they do not work equally well on a wide variety of motion sequences. The paper provides a mathematical framework in which systematic com-

*This work was partially supported by UIUC ECE and CSL startup funds, and by grants NSF CAREER IIS-0347456, NSF CRS-EHS-0509151, NSF CCF-TF-0514955, and ONR YIP N00014-05-1-0633.

parison between different classes of subspace arrangements may be conducted.

This paper also serves other purposes: 1. to reveal the subtlety and difficulty associated with identifying the seemingly simple class of subspace arrangements; 2. to provide a rigorous mathematical footing for some of the existing methods, especially the algebraic methods; and 3. to investigate new avenues that may lead to improved solutions for the estimation and segmentation of subspaces in the future.

2. Subspace Arrangement

2.1. Fundamentals

In what follows, the ambient space is a K-dimensional vector space R^K, where R can be $\mathbb{R}$ or $\mathbb{C}$. $R[\boldsymbol{x}]$ denotes either $\mathbb{R}[x_1,\ldots,x_K]$ or $\mathbb{C}[x_1,\ldots,x_K]$, the ring of polynomial functions in K variables. This ring is graded by degree:

$$R[\boldsymbol{x}] = R_0 \oplus R_1 \oplus \cdots, \tag{1}$$

where R_k denotes the vector space of homogeneous polynomials of degree k.

An r-dimensional subspace V can be defined by $K-r$ linearly independent linear forms $\{f_1,\ldots,f_{K-r}\}$:

$$V := \{\boldsymbol{x} \in R^K : f_k(\boldsymbol{x}) = 0,\ k = 1,\ldots,K-r\}. \tag{2}$$

Let D denote the space of linear forms that vanish on V, then $\dim(D) \doteq d = K - r$. The subspace V is also called the zero set of D, i.e., points in the ambient space that vanish on all polynomials in D, which is denoted as $\mathcal{Z}(D)$. We define

$$\mathcal{I}(V) := \{f \in R[\boldsymbol{x}] : f(\boldsymbol{x}) = 0, \forall \boldsymbol{x} \in V\}. \tag{3}$$

Clearly, $\mathcal{I}(V)$ is an ideal, and it contains polynomials of all degrees that vanish on V.

We now consider the concept of subspace arrangement. A more general definition deals with affine subspaces, which may not contain the origin in space, but we only consider linear subspaces.[1]

Definition 1 (Subspace Arrangement) *A subspace arrangement is defined as a finite collection of N linear subspaces in R^K: $\mathcal{A} := \{V_1,\ldots,V_N\}$.*

In this paper, we consider arrangements of subspaces *in general position*. We shall give a more rigorous definition after we introduce the Hilbert function. From a geometric point of view, an arrangement is said to be in general position if all intersections of subspaces in the arrangement and all intersections of their dual spaces have minimal dimensions. For example, three planes intersecting at one point in $\mathbb{R}^3$ are in general position; on the other hand, although three lines that lie on a plane intersect only at the origin, but the intersection of their dual spaces is a line, which is not a minimal intersection. We impose this constraint because in general the degenerate configurations are zero-measure sets in the configuration space. Moderate data noise and machine round off should guarantee that the subspaces lie in general position.[2]

The union of the subspaces in $\mathcal{A}$ is an algebraic set, denoted as $V_{\mathcal{A}} := V_1 \cup \cdots \cup V_N$, whose vanishing ideal is

$$\mathcal{I}(V_{\mathcal{A}}) = \mathcal{I}(V_1) \cap \cdots \cap \mathcal{I}(V_N). \tag{4}$$

The ideal $\mathcal{I}(V_{\mathcal{A}})$ can be graded by the degree of the polynomial

$$\mathcal{I}(V_{\mathcal{A}}) = \mathcal{I}_m \oplus \mathcal{I}_{m+1} \oplus \cdots \oplus \mathcal{I}_N \oplus \cdots. \tag{5}$$

Each $\mathcal{I}_k$ is a vector space that consists of forms of degree k in $\mathcal{I}(V_{\mathcal{A}})$, and $m \geq 1$ is the least degree of the polynomials in $\mathcal{I}(V_{\mathcal{A}})$. Notice that forms that vanish on $V_{\mathcal{A}}$ may have degrees less than N. One example is an arrangement of two lines and one plane in $\mathbb{R}^3$ in general position. Since any two lines lie on a plane, the arrangement can be embedded in a hyperplane arrangement of two planes, and there exist forms of second degree that vanish on the union of the three subspaces.

2.2. Hilbert Functions of Subspace Arrangements

The GPCA approach proposed in [17] depends on the Veronese embedding from the ambient space R^K to a higher dimensional space $R^{M_K^d}$. It is given by $\nu_d : R^K \to R^{M_K^d}$, $[x_1,\ldots,x_K]^T \mapsto [\ldots, x_1^{e_1}\cdots x_K^{e_K},\ldots]^T$, where $e_1+\cdots+e_K = d$. The dimension $M_K^d = \binom{d+K-1}{K-1}$.

Suppose there are N subspaces in an arrangement $\mathcal{A}$, then $\mathcal{I}_N$ contains all the N-forms that vanish on $\mathcal{A}$ in the M_K^N-dimensional vector space, where the vector space is generated by the N-forms $\{\ldots, x_1^{e_1}\cdots x_K^{e_K},\ldots\}_{e_1+\cdots+e_K=N}$. We define $\mathcal{D}_{\mathcal{A}} := \dim(\mathcal{I}_N)$ and call it *characteristic dimension*.

It turns out that $\mathcal{D}_{\mathcal{A}}$ is closely related to the rank of the embedded data matrix L_N in GPCA. Given a set of (noise-free) sample points $\{\boldsymbol{x}_k\}_{k=1}^n$ in $V_{\mathcal{A}}$, the Veronese embedding of the Nth degree gives a data matrix

$$L_N := \begin{bmatrix} \nu_N^T(\boldsymbol{x}_1) \\ \vdots \\ \nu_N^T(\boldsymbol{x}_n) \end{bmatrix} \in R^{n \times M_K^N}. \tag{6}$$

[1] This does not lose any generality: We can always convert an affine subspace to a linear subspace in an ambient space 1-dimensional higher.

[2] This condition is not necessary for all the results followed, we shall explicitly state it where it is used.

The vectors $\{\boldsymbol{c}\}$ in the null space of L_N correspond to homogeneous polynomials of degree N that vanish on $\{\boldsymbol{x}_k\}_{k=1}^n$, i.e., $\{f(\boldsymbol{x}) = \nu_N^T(\boldsymbol{x}) \cdot \boldsymbol{c}\}$. If n is large enough and the sample points $\{\boldsymbol{x}_k\}$ are in general position in $V_{\mathcal{A}}$, the so-obtained f's span exactly $\mathcal{I}_N$. Hence, we have

$$\mathcal{D}_{\mathcal{A}} = \dim(\mathcal{I}_N) = \dim(\text{Null}(L_N)). \tag{7}$$

$\mathcal{D}_{\mathcal{A}}$ is related to the Hilbert function of $V_{\mathcal{A}}$ [8]:

Definition 2 (Hilbert Function) *The Hilbert function $h_V(m)$ is defined to be the codimension, in the vector space of all homogeneous polynomials of degree m on R^K, of the subspace of those polynomials vanishing on an algebraic set V, i.e.,*

$$h_V(m) = \dim(S(V)_m), \tag{8}$$

where $S(V) = R[\boldsymbol{x}]/\mathcal{I}(V)$ is the homogeneous coordinate ring and the variable m denotes the mth graded piece.

Hence, for the algebraic set $V_{\mathcal{A}}$ consisting of N linear subspaces in R^K, the following relations hold

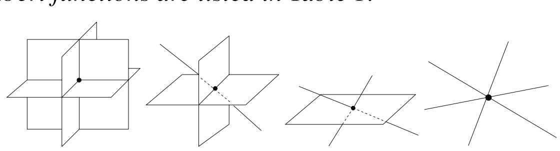
$$h_{V_{\mathcal{A}}}(N) = M_K^N - \mathcal{D}_{\mathcal{A}} = \text{rank}(L_N), \tag{9}$$

where the second equality holds for sufficient sample points $\{\boldsymbol{x}_k\} \subset V_{\mathcal{A}}$. Notice that the definition of the Hilbert function applies to general nonlinear varieties. Despite the importance of the function, no closed-form solution is known even for general subspace arrangements. Results of some special cases can be found in [8].

Using the Hilbert function, we now define the concept of general position for the subspace-segmentation problem.

Definition 3 (General Position of An Arrangement) *Define an arrangement $\mathcal{A}$ of N subspaces $\{V_k\}$ in general position if $h_{V_{\mathcal{A}}}(N) = \max_{W=V_1'\cup\cdots\cup V_N'}\{h_W(N)\}$, where $\{V_k'\}$ are arbitrary subspaces in R^K such that $\dim(V_k') = \dim(V_k)$. That is, an arrangement in general position has the maximal value of the corresponding Hilbert function.*

Example 4 (Three Subspaces in General Position in $\mathbb{R}^3$) *Consider $V_{\mathcal{A}}$ to be an arrangement of three subspaces in $\mathbb{R}^3$. There are in total four possible configurations of $V_{\mathcal{A}}$ shown in Figure 1, and the values of their corresponding Hilbert functions are listed in Table 1.*



(a) $(1,1,1)$ (b) $(1,1,2)$ (c) $(1,2,2)$ (d) $(2,2,2)$

Figure 1. Four configurations of three subspaces in $\mathbb{R}^3$. The number in the parentheses are codimensions.

Given a dataset sampled from one of the above configurations, the Veronese map embeds data samples from $\mathbb{R}^3$ to $\mathbb{R}^{10}$ (as $N = K = 3$). It is clear that the null space of the data matrix L_3 can only assume four possible dimensions, namely, 1, 2, 4, and 7. Furthermore, the dimension $\mathcal{D}_{\mathcal{A}}$ uniquely determines the codimension of the three subspaces in the arrangement.

Table 1. Hilbert functions of the four configurations.

d_1	d_2	d_3	$h_{V_{\mathcal{A}}}(3)$	$\mathcal{D}_{\mathcal{A}}$
1	1	1	9	1
1	1	2	8	2
1	2	2	6	4
2	2	2	3	7

This example suggests that, in general, given the dimensions of individual subspaces, we know the rank of the embedded data matrix; conversely, given the rank of the embedded data matrix, we can determine to a large extent the possible dimensions of the individual subspaces. Therefore, knowing these combinatorial and algebraic properties of subspace arrangements will help us to at least rule out impossible rank values for the data matrix or the impossible subspace dimensions.

One observation of the characteristic dimension $\mathcal{D}_{\mathcal{A}}$ is that it only depends on the dimensions of the subspaces, given they are in general position. It can be shown by considering the maximality property of the Hilbert function in general position.

Proposition 5 *Let $\mathcal{A} = \{V_1, \ldots, V_N\}$ be a subspace arrangement in general position, and the codimension of each V_k is d_k. The characteristic dimension $\mathcal{D}_{\mathcal{A}}$ is a function of the dimensions of the subspaces. We then denote $\mathcal{D}_{\mathcal{A}} = \mathcal{D}(d_1, \ldots, d_N)$, with $d_1 \leq \cdots \leq d_N$.*

Proposition 6 (Arrangement of Two Subspaces) *Let D_1 and D_2 denote the subspaces of linear forms vanishing on V_1 and V_2 in general position with codimensions d_1 and d_2, respectively. Then, $dim(D_1 \cap D_2) = max(d_1 + d_2 - K, 0)$, and*

$$\mathcal{D}(d_1, d_2) = d_1 d_2 - \binom{d_1+d_2-K}{2}. \tag{10}$$

Unfortunately, no general formula of $\mathcal{D}$ is known for arrangements that consist of $N \geq 3$ subspaces. Only in some special cases can the number of subspaces be reduced.

Proposition 7 (Arrangement with Hyperplanes) *Let $d_1, \ldots, d_N$ be the codimensions of the subspaces in ascending order. If $d_1 = 1$, i.e., V_1 is a hyperplane, then*

$$\mathcal{D}(1, d_2, \ldots, d_N) = \mathcal{D}(d_2, \ldots, d_N). \tag{11}$$

In other words, the introduction of an additional hyperplane does not affect the value of $\mathcal{D}$.

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

2.3. Relations among Hilbert Functions

In many applications, our task is to identify a subspace arrangement from among a class of possible arrangements $\{\mathcal{A}_i\}_{i\in I}$ (with I being some index set) that best fits a given dataset. Thus, it would be very useful if we know the values of the Hilbert function h or equivalently the characteristic dimension $\mathcal{D}$ in advance (as Example 4).

Numerical Computation of Hilbert Functions. Although closed-form formulae are in general not available for h or $\mathcal{D}$, we can still resort to numeric methods to obtain the values of h or $\mathcal{D}$, especially for the subspace arrangements of the most practical importance. By Proposition 5, all arrangements with identical subspace dimensions have the same Hilbert function if they are in general position. Therefore, for each type of arrangements, we can numerically generate a set of subspaces and a set of noise-free sample points that are in general position, and compute $\mathcal{D}$ using the relation (7): $\mathcal{D}_{\mathcal{A}} = \dim(\text{Null}(L_N))$, or the relation (9): $h_{V_{\mathcal{A}}}(N) = \text{rank}(L_N)$. The numerical solution for each type of configuration needs to be computed only once, and the results can be stored in a table such as Table 1 for each ambient space dimension K and number of subspaces N.

Some Conjectured Formulae. The numerical methods have their limit in terms of the number of subspaces and their dimensions that these methods can handle. Thus, to explore the properties of higher dimensional subspace arrangements, we must resort to algebraic techniques. The numerical results have led to the revelation of many interesting relations of the Hilbert functions among different subspace arrangements. Listed below are a few conjectures derived from the tables of characteristic dimension that we have computed so far.

Conjecture 8 *When $2 \le d_1 \le \cdots \le d_N \le K$,*[3] *the following formulae are true whenever the dimensions of the subspaces are in lexicographic-order. The relation between the two sides is that the codimension of the subspaces are identical except the changes in V_{m+1}.*

$$\begin{aligned} \mathcal{D}(\ldots, d_m, K, \ldots) &= \mathcal{D}(\ldots, d_m, K-1, \ldots) + 1; \\ \mathcal{D}(\ldots, d_m, K-1, \ldots) &= \mathcal{D}(\ldots, d_m, K-2, \ldots) + N; \\ \mathcal{D}(\ldots, d_m, K-2, \ldots) &= \mathcal{D}(\ldots, d_m, K-3, \ldots) + \binom{N+1}{2} - \\ &\qquad \#\{d_i = 2\}_{i=1}^m. \end{aligned}$$

Theoretically, these conjectures reveal deep relations among different arrangements that appear not to have been noticed before our computer simulated results. Numerically, these conjectures, once proved, can greatly speed up the computation by using the previous results recursively. In fact, Table 1 can be derived solely from the second conjecture rule that defines the relation in a change of subspace dimension from $K-2$ to $K-1$.

[3] By Proposition 7, subspaces of codimension 1 can be removed from the arrangement without changing the characteristic dimension.

Arrangements with Trivial Subspaces. The first rule in Conjecture 8 describes the characteristic dimension of an arrangement that has codimension K in R^K, which means the subspace is trivial – consisting of only the origin. The real model for a dataset shall not have such subspaces. However, the introduction of such subspaces represents the situation where the dataset has been over-fit with a larger number of subspaces than what is needed.

Example 9 (Over-Fit Hyperplane Arrangements in $\mathbb{R}^5$) *Consider a dataset sampled from a number of hyperplanes in general position in $\mathbb{R}^5$. Suppose we only know that the number of the hyperplanes is at most 4, and we embed the data via the degree-4 Veronese map anyway. Table 2 shows the possible values of the Hilbert function (hence the rank of the embedded data matrix).*

Table 2. Hilbert functions of hyperplane arrangements in $\mathbb{R}^5$.

d_1	d_2	d_3	d_4	$h_{V_{\mathcal{A}}}(4)$	$\mathcal{D}_{\mathcal{A}}$
1	1	1	1	69	1
1	1	1	5	65	5
1	1	5	5	55	15
1	5	5	5	35	35

The first row shows that if the number of hyperplanes is exactly equal to the degree of the Veronese map, then $\mathcal{D}_{\mathcal{A}} = 1$, i.e., the data matrix L_4 has a rank-1 null space. The following rows show the values of $h(4)$ for $N = 3, 2, 1$, respectively. If the rank of the matrix L_4 matches any of these values, we know exactly the number of hyperplanes in the arrangement. Notice that the model selection method proposed for hyperplane arrangements in [17], which incrementally searches for a co-rank one data matrix, becomes a special use of the Hilbert function. Thus, in general, knowing the values of $h(m)$ even for $m > N$ may significantly help determine the correct number of subspaces.

3. Applications and Experiments

In this section, we demonstrate how the study of Hilbert functions of subspace arrangements can help to solve subspace segmentation and model selection problems. The purpose of these examples is to stimulate future research to incorporate these algebraic constraints to other real systems that adopt robust or statistical schemes.

3.1. Subspace Segmentation of Mixture Dimensions

Most of the previous GPCA applications deal only with arrangements of hyperplanes, where $\mathcal{D}_{\mathcal{A}} \equiv 1$. We propose a novel yet simple algorithm for the subspace segmentation problem, which is a variation of the original GPCA

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 IEEE

method [17]. The knowledge of Hilbert functions allows us to consider the whole null space of the data matrix L_N for arbitrary mixtures of subspaces. We will also compare our algorithm with two iterative methods, namely, EM and K-subspaces.

Suppose the number of subspaces N and their codimensions are all given. The Hilbert function constraint provides us with the theoretical rank of the data matrix L_N. We can obtain a set of polynomials $P(\boldsymbol{x})$ with coefficients equal to the eigenvectors in the null space of L_N. Evaluate $DP(\boldsymbol{x})$ at each point, we get a set of vectors normal to the subspace that the point lies in. The original GPCA method relies on one good sample per subspace to segment the dataset. In the presence of noise, no single sample is reliable. However, if we take the average of the normals of all samples in one subspace, we expect that it will smooth out the random noise.

Now the remaining question is how to obtain an average of normals for a subspace without knowing the segmentation of the samples. In this case, we know all the possible codimensions of the subspaces and the number of subspaces having the same codimension. For each sample $\boldsymbol{x}_k$, we invoke a voting technique on the feature space of normal vectors that are perpendicular to the subspaces. The voting technique is widely recognized as an effective model estimation method. We cast votes on the candidate codimension classes based on the dominant directions of the space $DP(\boldsymbol{x})$. Notice that in the presence of noise, $DP(\boldsymbol{x})$ is usually full rank and the number of column vectors is much larger than the largest codimension. For instance, on the third row of Table 1, $\mathcal{D} = 4$, but the largest codimension is only 2. If an arrangement is sampled from three subspaces of codimensions $(1, 2, 2)$, the matrix $DP(\boldsymbol{x}_k)$ evaluated at $\boldsymbol{x}_k$ is a 3×10 matrix. The noise-free rank of $DP(\boldsymbol{x}_k)$ may be 1 or 2, depending on which subspace $\boldsymbol{x}_k$ lies in. When more dominant vectors than its true rank are estimated, the extra vectors will be randomly distributed by the noise. When less vectors are estimated, they form a basis of a subspace with a random orientation in the space perpendicular to the subspace.

In summary, for each sample, we assume it could be sampled from all candidate codimension classes, and vote by the dominant vectors of $DP(\boldsymbol{x}_k)$ as a basis. Finally, the bases associated with the highest votes will be recognized as the normal vectors perpendicular to the subspaces. Algorithm 1 is the pseudo-code of the proposed method. The only parameter is $angleTolerance$, which determines whether the new basis is close to an registered basis in the history, or a new voting candidate will be created.

In the experiment, we compare the performance of our algorithm with EM, K-subspaces and the original algebraic GPCA (AGPCA). We also compare the performance of two methods called GPCA+EM and GPCA+K-subspaces,

Algorithm 1 Generalized Principal Component Analysis

Given a set of samples $\{\boldsymbol{x}_k\}_{k=1}^n$ in R^K, fit a N linear subspace model with codimensions $d_1, \ldots, d_N$:

Set $angleTolerance$, let C be the number of distinct codimensions, and obtain $\mathcal{D}$ by the Hilbert function constraint.
Let $V\{1\}, \ldots, V\{C\}$ be integer arrays as voting counters. Let $U\{1\}, \ldots, U\{C\}$ be matrix arrays for basis candidates.
Construct $L_N = [\nu_N(\boldsymbol{x}_1), \ldots, \nu_N(\boldsymbol{x}_n)]$.
Form the set of polynomials $P(\boldsymbol{x})$ and compute $DP(\boldsymbol{x})$.
for all sample $\boldsymbol{x}_k$ **do**
    for all $1 \leq i \leq C$ **do**
        Assume $\boldsymbol{x}_k$ is from a subspace with the codimension d equal to that of the class i. Find the first d principal components $B \in R^{K \times d}$ in the matrix $DP|_{\boldsymbol{x}_k}$.
        Compare B with all candidates in $U\{i\}$.
        if $\exists j$, subspace_angle$[B, U\{i\}(j)] < angleTolerance$ **then**
            $V\{i\}(j) = V\{i\}(j) + 1$.
            Average the principal directions with the new basis B.
        else
            Add a new entry in $V\{i\}$ and $U\{i\}$.
        end if
    end for
end for
for all $1 \leq i \leq C$ **do**
    $m =$ the number of subspaces in class i.
    Choose the first m highest votes in $V\{i\}$ with their corresponding bases in $U\{i\}$.
    Assign corresponding samples into the subspaces, and cancel their votes in the other classes.
end for
Segment the remaining samples based on these bases.

which use GPCA closed-form solutions as initialization to start the iterations. The results are shown in Table 3.

From the experiment, we see that our algorithm significantly outperforms the algebraic GPCA method. The statistical methods also perform poorly when the structures have a mixture of subspaces with different dimensions, i.e., the first and second cases. This result is expected because without geometric constraints, we can overfit a lower dimensional subspace with a higher dimensional one (e.g., a 1-D subspace is a special 2-D subspace); or two subspaces combined may be grouped into a single subspace (e.g., two 2-D subspaces form a 4-D subspace). These locally optimal solutions provided by EM and K-subspaces may not be globally optimal. On the other hand, GPCA successfully provides us better segmentation results, regardless of the dimension differences of the subspaces. Finally, GPCA+EM and GPCA+K-subspaces consistently perform the best throughout all the examples. It suggests that we may combine the virtues of both algebraic and statistical methods in real applications in search for globally optimal subspace fittings. The au-

Table 3. Segmentation error of three arrangements. 2% Gaussian noise is added to each sample point.

Subspace Dimensions	EM	K-subspaces	AGPCA	GPCA	GPCA+EM	GPCA+K-subspaces
$(2,2,1)$ in $\mathbb{R}^3$	29%	27%	13.2%	10%	7%	6 %
$(4,2,2,1)$ in $\mathbb{R}^5$	53%	57%	52.6%	8.7%	11.3%	8%
$(4,4,4,4)$ in $\mathbb{R}^5$	20%	25%	67.4%	24%	20%	20%

thors have made all the codes available for download at `http://decision.csl.uiuc.edu/~yangyang/softwarepage.html`. The reader is encouraged to test these methods in applications he/she is interested in.

3.2. Estimating the Number of Hyperplanes

One important type of arrangements is the one consisting of subspaces of equal dimensions, which is called a *pure arrangement*. This type of model has been used to describe a wide range of applications, such as rigid-body motions [11, 17], face recognition [1], and hybrid control systems [14].

Given an arrangement of N subspaces in R^K with dimension r identically the same, we can first project the dataset to $r+1$ dimensional space, and the subspaces in the projected space become all hyperplanes. If we know the number of subspaces, the segmentation of these hyperplanes are already well-studied in the previous work. The knowledge about Hilbert functions comes in handy in determining the number of subspaces.

In this experiment, we want to estimate the number of general rigid-body motions in an affine camera sequence. Features of an independent moving object under affine projection sit in a 4-D subspace [11]. The image coordinates of all views of a feature is lumped to a long vector, and projected to 5-D space by PCA. Therefore, independent motions become hyperplanes in the projected space. If we have an upper bound for the number of motions, say N', we may embed the data by a single Veronese map of degree N'. Table 2 provides exactly the rank of the data matrix for $N'=4$. However, under moderate data noise, it is difficult to directly determine the rank from the singular values.

The geometric-MDL (GMDL) [11] was designed to estimate the rank of a noisy data matrix. For each rank γ, its GMDL value is given by

$$\mathrm{GMDL}(\gamma) = \sum_{l=\gamma+1}^{m} \sigma_l^2 - \gamma(n+m-\gamma)\epsilon^2 \log\left(\frac{\epsilon}{L}\right)^2, \quad (12)$$

where in our context σ_l is the lth singular value of the data matrix L_d, m is the dimension of the Veronese map M_5^d, n is the number of samples, and ϵ^2 is the noise variance of the column vectors of the data matrix, which is difficult to estimate.

We embed the data by the Veronese map of the same degree $N' \geq N$. By doing so, we over-fit a hyperplane arrangement in $\mathbb{R}^5$. We propose the following heuristic modification to the GMDL criterion to estimate the number of motions:

$$N^* = \operatorname*{argmin}_N \Big\{ \sum_{l=h_N(N')+1}^{M_5^{N'}} \sigma_l^2 + \kappa h_N(N')(n+M_5^{N'}) \Big\}, \quad (13)$$

where $h_N(N')$ is the value of the Hilbert function for N hyperplanes and $N'-N$ trivial subspaces (the origin), n is the number of samples, and κ is a small positive constant. When $N'=4$, according to Table 2, we have $h_N(N') = [35, 55, 65, 69]$ for $N=[1,2,3,4]$. Figure 2 shows a superimposed plot of the singular values of L_4 from four noise-free synthetic sequences.

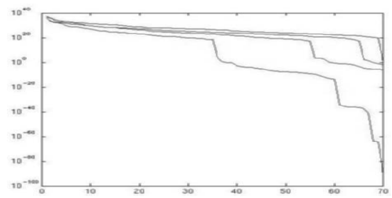


Figure 2. A super-imposed semi-log plot of the singular values when $N=1,\ldots,4$. The ranks drop at position $35, 55, 65, 69$, which match the theoretical values of the Hilbert function.

In the experiment, we randomly generate 2 to 4 rigid-objects in space, and project the points by an affine camera, whose parameters are also randomly generated at each time. Each sequences consists of 50 frames, and different levels of noise are added to the image coordinates, but no outliers are present in the simulation. We run the simulation at each noise level 200 times. The result is shown in Figure 3.

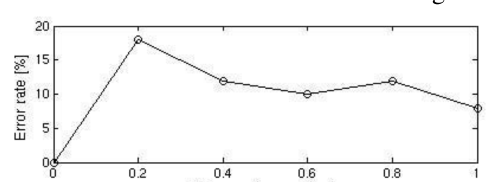


Figure 3. The error rates of estimating the number of 3-D motions. The Gaussian noise variance is between 0-1 pixel.

3.3. Estimating the Dimensions of Subspaces

The advantage of the method shown in the previous section is that the target quantity can be estimated without knowing the segmentation of the sample points. However, if the subspaces in an arrangement have different dimensions instead of all being hyperplanes, detecting the rank alone cannot distinguish different dimensions among the subspaces.

When the dimensions of the subspaces are unknown and possibly different, the algebraic GPCA algorithm tackles the segmentation by fitting a single polynomial for all subspaces regardless of their dimensions, which is equivalent to choosing one containing hyperplane for each subspace, and perform the segmentation on the hyperplanes [15]. The dimension of each segmented subspace is determined separately by applying PCA with a heuristic threshold. This process fails to recognize the constraint between the rank of the embedded data matrix and the individual subspaces as we discussed in Section 2.2. More importantly, no justification can be given to why one heuristic parameter works for all subspaces with different dimensions.

In this section, we outline a method to estimate different dimensions of the subspaces in an arrangement by explicitly harnessing the relations between Hilbert functions and the dimensions of individual subspaces. The basic idea is that given a class of possible models, we should perform the classification for all the candidates, and the model selection should be made based on how well these models fit the sample set.

To elaborate on this idea more clearly, we again use affine camera sequences as an example, although the method is general for other applications. Suppose that we know there are exactly four independent 3-D motions in an image sequence, i.e., $N = 4$. For a general motion, the sample points span a 4-D subspace, and for a (degenerate) planar motion, the sample points span a 3-D subspace [11]. The motion segmentation problem requires not only correctly segmenting the dataset, but also labeling each motion with the correct dimension. Table 4 shows the values of the Hilbert functions of the 5 possible categories of subspace arrangements associated with the 4 motions.

Table 4. Hilbert functions with 3- or 4-D (codimension 1 or 2, respectively) subspaces in $\mathbb{R}^5$.

d_1	d_2	d_3	d_4	$h_{V_\mathcal{A}}(4)$	$\mathcal{D}_\mathcal{A}$
1	1	1	1	69	1
1	1	1	2	68	2
1	1	2	2	66	4
1	2	2	2	62	8
2	2	2	2	54	16

To determine the best subspace arrangement that segments and fits the noisy motion data, we decompose the overall modeling process into two related steps:

1. **Segmentation via GPCA.** For each category, we use Algorithm 1 to segment the dataset and label the subspaces with dimensions. A K-subspaces optimization can be applied to further improve the estimates of the bases. We can immediately reject a category if the K-subspaces iteration does not converge.

2. **Optimal Model Class Selection.** It remains to determine which category gives the globally optimal arrangement model for the data. The value of the Hilbert function h is a measure of the complexity (or dimensionality) of the subspace arrangement. We adopt the GMDL criterion that we proposed in Section 3.2, but replace $\sum \sigma^2$ with the residual found in Step 1 for each category:
$$i^* = \operatorname*{argmin}_i \Big\{ \frac{1}{n} \sum_{k=1}^{n} \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\|^2 + \kappa h_i(4)(n + M_5^4) \Big\}, \tag{14}$$
where $h(4) = [69, 68, 66, 62, 54]$ for row $i = [1, \ldots, 5]$, n is the number of samples, and κ is a small positive weight parameter.

Example 10 (Segmentation via GPCA) *We choose the second row of Table 4 as the ground-truth model, and randomly generate a sequence of three general motions and one planar motion. Each sequence contains 400 frames, and we track 50 points on each object. Various noise levels are added to the image points, and the experiment is repeated 400 times at each noise level. Figure 4 shows a snapshot of four typical images.*

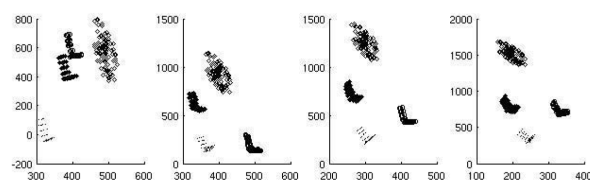


Figure 4. Four frames of a typical scene of four moving objects.

As we have discussed in Section 3.1, A GPCA+K-subspaces method will most likely outperform both GPCA and K-subspaces algorithms in real applications. Figure 5 shows the improvement of the error rates of the segmentation due to the nonlinear K-subspaces optimization. In addition, the optimization also improves the result for the combinatorial problem. Figure 6 shows the difference in the model selection results with or without K-subspaces. A model is correct when 4 subspaces are labeled with correct dimensions. With various noise level, the algorithm achieves the model selection errors of less than 15%.

Example 11 (Total Estimation Error) *To demonstrate the performance of the overall modeling process, we randomly generate four objects in space undergoing either general motion or planar motion, and the feature points are projected by an affine camera model. For the five*

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

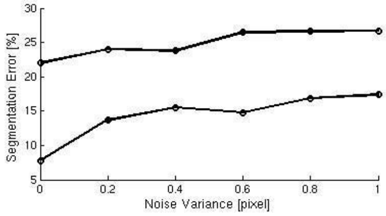


Figure 5. The segmentation error with or without K-subspaces. The lower error curve represents the results with K-subspaces.

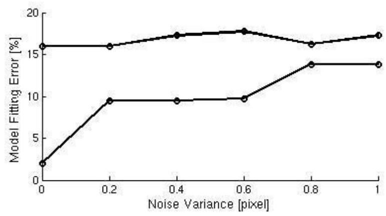


Figure 6. Model selection error. The lower error curve represents the results with K-subspaces.

model categories in Table 4, the number of combinatorial configurations are 1, 4, 6, 4, 1, respectively. Hence in total the algorithm is selecting one out of 16 possible models to fit a dataset.

We run the experiment at each noise level for 400 times, and the total estimation error is shown in Figure 7. In defining the error rate, an estimation is said to be correct only when its estimates of the category and the dimension combination are both correct. The total fitting error is lower than the result in Figure 6 because there is no combinatorial problem for the first and last categories.

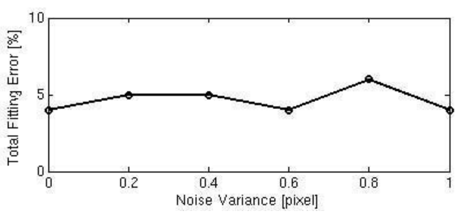


Figure 7. Total model selection error.

4. Discussions and Open Issues

One of the main open problems in the estimation of subspace arrangement is how to systematically combine the algebraic properties of subspace arrangements with robust statistical model-selection techniques. The main difficulty is how to deal with outliers. Notice that one cannot directly apply techniques such as RANSAC [6] to subspace arrangements and expect to always get better results: If one chooses the model in RANSAC as each subspace, then all points on the other subspaces become outliers, and dealing with such large amount of outliers is very difficult and inefficient. If we choose the model to be the subspace arrangement, the minimal sample set needed to determine the model can be very large (in fact the exact bound is not known). Furthermore, random sampling scheme becomes rather inefficient in dealing with points from subspaces of different dimensions: a line is of zero-measure compared to a plane. Thus, random sub-sampling may generate inconsistent subspace arrangements, as the majority of a sub-sampled set may come from only some of the subspaces. So far, these questions have defied systematic and principled solutions, despite many ingenious practical designs and applications.

References

[1] P. Belhumeur, J. Hespanda, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997.

[2] A. Björner. Subspace arrangements. In *Progress in Mathematics series*, volume 119, pages 321–370, 1994.

[3] A. Björner, I. Peeva, and J. Sidman. Subspace arrangements defined by products of linear forms. In *Journal of the London Math. Society (to appear)*, preprint 2003.

[4] J. Costeira and T. Kanade. A multibody factorization method of independently moving objects. *IJCV*, 29(3):159–179, 1998.

[5] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, second edition, 2001.

[6] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–85, 1981.

[7] M. Han and T. Kanade. Reconstruction of a scene with multiple linearly moving objects. In *CVPR*, volume 2, pages 542–549, 2000.

[8] J. Harris. *Algebraic Geometry: A First Course.* Springer, 1992.

[9] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *CVPR*, 2003.

[10] W. Hong, J. Wright, K. Huang, and Y. Ma. A multi-scale hybrid linear model for lossy image representation. In *ICCV*, 2005.

[11] K. Kanatani. Motion segmentation by subspace separation: model selection and reliability evaluation. *International Journal of Image and Graphics*, 2(2):179–197, 2002.

[12] P. Orlik. *Introduction to Arrangements*, volume 72 of *conference board of the mathematical sciences regional conference series in math.* American Mathematics Society, 1989.

[13] A. Shashua and A. Levin. Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In *ICCV*, volume 2, pages 592–599, 2001.

[14] P. van Overschee and B. de Moor. *Subspace Identification for Linear Systems.* Kluwer Academic, 1996.

[15] R. Vidal and R. Hartley. Motion segmentation with missing data using PowerFactorization and GPCA. In *CVPR*, 2004.

[16] R. Vidal and Y. Ma. A unified approach to 2-D and 3-D motion segmentation via polynomial fitting and differentiating. In *ECCV*, 2004.

[17] R. Vidal, Y. Ma, and J. Piazzi. A new GPCA algorithm for clustering subspaces by fitting, differentiating and dividing polynomials. In *CVPR*, pages 510–517, 2004.

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**