

GPCA with Denoising: A Moments-Based Convex Approach*

Necmiye Ozay¹ Mario Sznaier¹
¹ECE Department, Northeastern University
Boston, MA 02115
{ozay.n,m.sznaier,camps}@neu.edu

Constantino Lagoa² Octavia Camps¹
²Department of EE, Penn State University
University Park, PA 16802
lagoa@engr.psu.edu

Abstract

This paper addresses the problem of segmenting a combination of linear subspaces and quadratic surfaces from sample data points corrupted by (not necessarily small) noise. Our main result shows that this problem can be reduced to minimizing the rank of a matrix whose entries are affine in the optimization variables, subject to a convex constraint imposing that these variables are the moments of an (unknown) probability distribution function with finite support. Exploiting the linear matrix inequality based characterization of the moments problem and appealing to well known convex relaxations of rank leads to an overall semi-definite optimization problem. We apply our method to problems such as simultaneous 2D motion segmentation and motion segmentation from two perspective views and illustrate that our formulation substantially reduces the noise sensitivity of existing approaches.

1. Introduction and Motivation

Many problems of practical interest can be reduced to identifying a combination of an unknown number of subspaces and quadratic surfaces from sample points. Examples include among others image clustering/classification, segmentation of video sequences, motion segmentation under affine and perspective projections, and identification of piecewise affine systems [20, 10, 15, 8, 11]. In the ideal case of noiseless data, the problem can be elegantly solved using an algebraic approach, Generalized Principal Components Analysis (GPCA) [20], that only entails finding the null space of a matrix constructed from the data. In the case where the data is corrupted by noise, pursuing this approach requires first estimating the null space of a matrix whose entries depend polynomially on the noise, a non-trivial problem. If the noise is small, an approximate solution to this problem is given by the subspace associated with the smallest singular value of the matrix [20]. Recently, an improved method has been proposed [12] based on using a lineariza-



Figure 1. An image pair with 3 relocated objects with noisy correspondences superimposed.

tion analysis to reduce the problem to minimizing the Sampson distance from the data points to the zero set of a family of polynomials associated with the algebraic surfaces. In principle this approach leads to a constrained non-linear optimization problem. However, an approximate solution can be obtained by solving a generalized eigenvalue problem and improved using gradient descent methods. While this approach can successfully handle moderate noise levels, its performance degrades substantially as the noise level increases, as illustrated next.

Consider the image pair shown in Fig. 1, taken from [19], where each point coordinate has been corrupted by uniform random noise. The goal here is to estimate the motion and assign each point to a rigid object. As shown in [20] (see also Section 4) this problem can be reduced to that of assigning points to (an unknown number of) subspaces, and hence solved using GPCA. Figures 2 (a) and (b) show a typical outcome when applying the original GPCA method and its robust version (RGPCA) [12], and summarizes the results of 20 random runs with ± 10 pixels noise level. As expected, while robust GPCA outperforms GPCA, it still has a relatively large misclassification rate.

Contributions of the paper:

This paper is motivated by the example above. Our goal is to develop a computationally tractable algorithm for segmenting a mixture of subspaces and quadratic surfaces capable of handling not necessarily small amounts of noise with a-priori unknown statistical properties. Specifically, the main contributions of the paper are:

- **Theoretical:** Our main theoretical contribution shows that the problem of estimating a multivariate poly-

*This work was supported in part by NSF grants IIS-0713003 and ECCS-0901433, and AFOSR grant FA9550-09-1-0253.

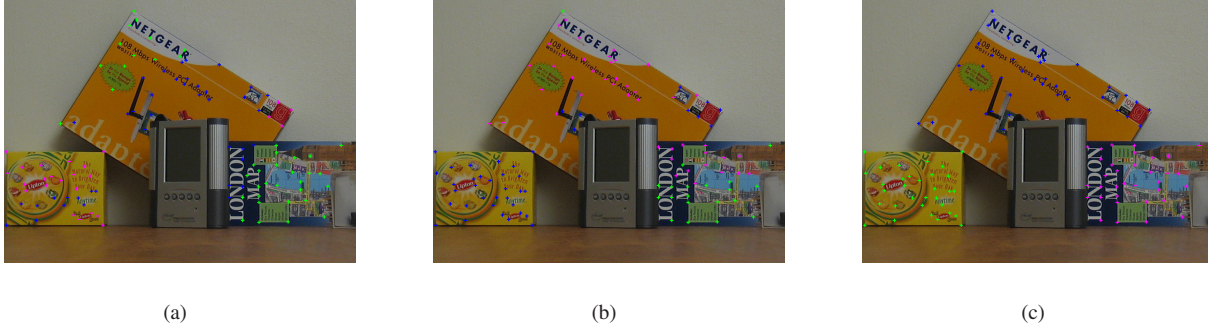


Figure 2. Sample segmentation results and mean misclassification rates. (a): GPCA segmentation (mean misclassification rate: 31.1%). (b): RGPCA segmentation (mean misclassification rate 14.9%). (c): Proposed algorithm (mean misclassification rate 3.7%). The image size is 600×800 and the noise level is ± 10 pixels.

mial $Q(\mathbf{x})$, with a-priori bounded order, from noisy measurements of its zero set can be reduced to a constrained rank minimization problem where all the matrices involved are affine in the decision variables and the constraints are convex. The significance of this result is that it allows for eliminating the polynomial dependency on the (unknown) noise terms that renders the problem of estimating Q (or equivalently the null space of the Veronese matrix in GPCA) difficult, at the price of adding additional (but convex) constraints.

- **Algorithmic:** The theoretical results outlined above allow for recasting the problem of segmenting a mixture of subspaces and quadratic surfaces into a constrained rank minimization problem. Although in principle rank minimization problems are NP hard, this is a very active research area in optimization, and in the past few years a number of efficient convex relaxations have been proposed. Since all matrices involved are affine in the decision variables, the resulting overall problem is convex and can be efficiently solved, leading to a tractable segmentation algorithm. As we illustrate in the paper, this algorithm performs well, even in the presence of substantial noise. An example of this situation is shown in Figure 2 (c).

2. Preliminaries

2.1. Notation

- $\mathbf{M} \succeq 0$ matrix \mathbf{M} is positive semidefinite.
- \mathcal{P}_D^n set of n^{th} degree multivariate polynomials in D variables. n and D may be omitted when clear from the context.
- $\nu_n(\mathbf{x})$ Veronese map of degree n :
- $\nu_n(x_1 \dots x_D) \doteq [x_1^n \ x_1^{n-1}x_2 \ \dots \ x_D^n]^T$ (i.e. all possible monomials of order n in D variables, in lexicographical order.)

2.2. Statement of the Problem

Let $\mathcal{A} \doteq S_1 \cup S_2 \cup \dots \cup S_n$ denote an arrangement of subspaces S_k embedded in an ambient space of dimension D . To this arrangement of subspaces one can associate a set $\mathcal{V}_{\mathcal{A}}^n$ of homogeneous multivariate polynomials $Q_j(\mathbf{x})$ of degree n that has \mathcal{A} as its zero set, e.g. $\mathcal{V}_{\mathcal{A}}^n \doteq \{Q_j(\mathbf{x}) \in \mathcal{P}^n : Q_j(\mathbf{x}) = 0 \iff \mathbf{x} \in \mathcal{A}\}$. Following [12], in the sequel we will refer to the set $\mathcal{V}_{\mathcal{A}}^n$ as the set of vanishing polynomials of the arrangement \mathcal{A} . Note that each polynomial $Q_j(\mathbf{x})$ can be written as $Q_j = \nu_n(\mathbf{x})^T \mathbf{c}_j$, where the vector $\mathbf{c}_j \in \mathbb{R}^m$, $m = \binom{n+D-1}{D-1}$, contains the coefficients of the vanishing polynomial in appropriate order. It follows that given N_p noiseless samples $\mathbf{x}_1, \dots, \mathbf{x}_{N_p} \in \mathcal{A}$, the vectors \mathbf{c}_j span the null space \mathcal{N} of the matrix $\mathbf{V}(\mathbf{x}) = [\nu_n(\mathbf{x}_1) \ \nu_n(\mathbf{x}_2) \ \dots \ \nu_n(\mathbf{x}_{N_p})]^T$. Thus these vectors (and hence a basis for $\mathcal{V}_{\mathcal{A}}^n$) can be found via a simple singular value decomposition of $\mathbf{V}(\mathbf{x})$. On the other hand, in the case of noisy samples $\mathbf{x}_i = \hat{\mathbf{x}}_i + \boldsymbol{\eta}_i$ the matrix \mathbf{V} depends polynomially on the noise $\boldsymbol{\eta}_i$. Thus, when $\|\boldsymbol{\eta}\|$ is not small, the procedure outlined above no longer works, even when replacing the null space \mathcal{N} of \mathbf{V} with the subspace associated with its smallest singular values.

Our goal is to develop a computationally tractable algorithm (and the supporting theory) that allows for estimating both the set of vanishing polynomials and the subspaces from samples corrupted by (not necessarily small) noise. Specifically, we address the following problem:

Problem 1. [Polynomial estimation and subspace segmentation] Given a (sufficiently dense) set of noisy samples $\mathbf{x}_i = \hat{\mathbf{x}}_i + \boldsymbol{\eta}_i$ of points $\hat{\mathbf{x}}_i$ drawn from an arrangement of subspaces \mathcal{A} , and a-priori bounds on the number of subspaces, n , and the norm of noise $\|\boldsymbol{\eta}_i\|_2 \leq \epsilon$:

- 1.- Estimate a basis for $\mathcal{V}_{\mathcal{A}}^n$, the set of vanishing polynomials of \mathcal{A} of degree up to n .
- 2.- Estimate a subspace S_k^\perp normal to each subspace S_k .

- 3.- For each noisy sample \mathbf{x}_i find a subspace \mathcal{S}_k , a point $\hat{\mathbf{x}}_i \in \mathcal{S}_k$ and an admissible noise $\boldsymbol{\eta}_i$, $\|\boldsymbol{\eta}_i\| \leq \epsilon$ such that $\mathbf{x}_i = \hat{\mathbf{x}}_i + \boldsymbol{\eta}_i$.

Our main result establishes that the problem above can be reduced to a constrained rank minimization problem where all the matrices involved are affine in the optimization variables. Once this result is established, tractable algorithms can be obtained by appealing to recently introduced convex relaxations of rank minimization problems.

2.3. The Problem of Moments

The key step in establishing our main result is to transform the problem from an optimization over noise sequences to an optimization over probability measures supported in the ϵ -ball. To this effect, we will use several classical results establishing the necessary and sufficient conditions for a given sequence to be the moments of a probability distribution. For completeness, these results are briefly quoted below. Details are given for instance in [4].

Let \mathbf{K} be a closed subset of \mathbb{R}^D and let $\boldsymbol{\alpha}$ be a multi-index (i.e. $\boldsymbol{\alpha} \in \mathbb{N}^D$) representing the powers of a monomial in D variables. Given a sequence of scalars $\{m_{\boldsymbol{\alpha}}\}$, the **K-moment problem** is to determine whether there exists a probability measure μ supported on \mathbf{K} such that it has each $m_{\boldsymbol{\alpha}}$ as its $\boldsymbol{\alpha}^{\text{th}}$ moment. That is:

$$m_{\boldsymbol{\alpha}} = \mathbf{E}_{\mu}(\mathbf{x}^{\boldsymbol{\alpha}}) = \int_{\mathbf{K}} \mathbf{x}^{\boldsymbol{\alpha}} \mu(dx) \quad (1)$$

where $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_D^{\alpha_D}$. In particular, in the sequel we are interested in probability measures that are supported on balls of radius ϵ centered at the origin. Next, we recall a theorem from [13] that provides necessary and sufficient conditions for the existence of such a measure.

Theorem 1. Let $p = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} \in \mathcal{P}$ denote a generic (multivariate) polynomial. Given a sequence $\mathbf{m} \doteq \{m_{\boldsymbol{\alpha}}\}$, there exists a linear functional $\mathbf{E}: \mathcal{P} \rightarrow \mathcal{R}$ such that

$$\mathbf{E}(p) = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}} \quad (2)$$

and $\{m_{\boldsymbol{\alpha}}\}$ are the moments of a distribution supported on $\|\mathbf{x}\|_2 \leq \epsilon$, if and only if the following two conditions hold for all $p \in \mathcal{P}$:

$$\mathbf{E}(p^2) \geq 0 \quad (3)$$

$$\mathbf{E}((\epsilon^2 - (x_1^2 + \dots + x_D^2))p^2) \geq 0 \quad (4)$$

Remark 1. The conditions given in the above theorem consist of infinite semidefinite quadratic forms which can be converted into (infinite) linear matrix inequalities (LMIs) in the moment variables $\{m_{\boldsymbol{\alpha}}\}$.

Next, we briefly discuss how to build a matrix representation of a given sequence \mathbf{m} that contains all the moments up to order 2δ . Although the order of the subsequence is immaterial, for the sake of clarity of presentation, we arrange the moments according to a graded reverse lexicographic order (grevlex) of the corresponding monomials so that we have $\mathbf{0} = \boldsymbol{\alpha}^{(1)} < \dots < \boldsymbol{\alpha}^{(M_{\delta})}$, where $M_{\delta} \doteq \binom{\delta + D}{D}$ is the number of moments in \mathbb{R}^D up to order δ . Then, the moment conditions take the form:

$$\begin{aligned} \mathbf{L}^{(\delta)}(\mathbf{m}) &\succeq 0 \\ \mathbf{K}^{(\delta)}(\epsilon, \mathbf{m}) &\succeq 0 \end{aligned} \quad (5)$$

where

$$\begin{aligned} \mathbf{L}^{(\delta)}(i, j) &= m_{\boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)}} \text{ for all } i, j \leq M_{\delta} \\ \mathbf{K}^{(\delta)}(i, j) &= (\epsilon^2 m_{\boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)}} - m_{\boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)} + (2, 0, \dots, 0)} - \dots - m_{\boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)} + (0, \dots, 0, 2)}) \text{ for all } i, j \leq M_{\delta-1} \end{aligned}$$

It can be shown [9] that the linear matrix inequalities in equation (5) are necessary conditions for the existence of a measure μ supported in the ϵ -ball that has the sequence \mathbf{m} as its moments. Moreover, as $\delta \uparrow \infty$, (5) becomes equivalent to conditions (3)–(4) in Theorem 1; hence it is sufficient as well. Thus, it is possible to get progressively better approximations to the infinite dimensional conditions (3)–(4) with finite dimensional LMIs by increasing the size of the moment matrices (i.e. by increasing δ) [9]. It is also worth noting that if as δ increases, the rank of the moment matrices stops increasing, the so-called *flat extension* property ([4]) is satisfied. In this case, the finite dimensional conditions (5) corresponding to this value of δ are necessary and sufficient for the existence of a measure supported in the ϵ ball.

Example 1. This simple example illustrates the structure of moment matrices when $D = 2$, $\delta = 2$:

$$\mathbf{L} = \begin{bmatrix} 1 & m_{(1,0)} & m_{(0,1)} & m_{(2,0)} & m_{(1,1)} & m_{(0,2)} \\ m_{(1,0)} & m_{(2,0)} & m_{(1,1)} & m_{(3,0)} & m_{(2,1)} & m_{(1,2)} \\ m_{(0,1)} & m_{(1,1)} & m_{(0,2)} & m_{(2,1)} & m_{(1,2)} & m_{(0,3)} \\ m_{(2,0)} & m_{(3,0)} & m_{(2,1)} & m_{(4,0)} & m_{(3,1)} & m_{(2,2)} \\ m_{(1,1)} & m_{(2,1)} & m_{(1,2)} & m_{(3,1)} & m_{(2,2)} & m_{(1,3)} \\ m_{(0,2)} & m_{(1,2)} & m_{(0,3)} & m_{(2,2)} & m_{(1,3)} & m_{(0,4)} \end{bmatrix} \quad (6)$$

$$\begin{aligned} \mathbf{K} &= \epsilon^2 \mathbf{L}(1 : 3, 1 : 3) - \begin{bmatrix} m_{(2,0)} & m_{(3,0)} & m_{(2,1)} \\ m_{(3,0)} & m_{(4,0)} & m_{(3,1)} \\ m_{(2,1)} & m_{(3,1)} & m_{(2,2)} \end{bmatrix} \\ &- \begin{bmatrix} m_{(0,2)} & m_{(1,2)} & m_{(0,2)} \\ m_{(1,2)} & m_{(2,2)} & m_{(1,3)} \\ m_{(0,3)} & m_{(1,3)} & m_{(0,4)} \end{bmatrix} \end{aligned} \quad (7)$$

3. Main Results

In this section we present the main theoretical result of the paper that allows for recasting Problem 1 into a constrained rank minimization form that can be further relaxed to a convex semi-definite optimization problem.

Theorem 2. *Let $\{Q_j(\mathbf{x})\}$ denote the set of polynomials of an arrangement of subspaces \mathcal{A} and denote by n_q its dimension. Consider a set of measurements corrupted by norm-bounded noise*

$$\mathbf{x}_i = \hat{\mathbf{x}}_i + \boldsymbol{\eta}_i, \quad i = 1, 2, \dots, N_p \quad (8)$$

where $\hat{\mathbf{x}}_i \in \mathcal{A}$ and $\|\boldsymbol{\eta}_i\|_2 \leq \epsilon$. Then:

1. If $n_q > 1$, there exists an admissible noise sequence $\boldsymbol{\eta}_i$ and N_p points $\hat{\mathbf{x}}_i$ satisfying (8) and such that $Q_j(\hat{\mathbf{x}}_i) = 0$ if and only if there exist N_p sequences $\mathbf{m}_i \doteq \{m_\alpha\}$ such that the following conditions hold:

$$\text{rank} \{ \mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p}) \} = h - n_q \quad (9)$$

$$\mathbf{L}_i(\mathbf{m}_i) \succeq 0, \quad i = 1, \dots, N_p \quad (10)$$

$$\mathbf{K}_i(\epsilon, \mathbf{m}_i) \succeq 0, \quad i = 1, \dots, N_p \quad (11)$$

$$\text{rank}(\mathbf{L}_i(\mathbf{m}_i)) = 1, \quad i = 1, \dots, N_p \quad (12)$$

where $h = \binom{n+D-1}{D-1}$; $\mathbf{M} \doteq \mathbf{E}(\mathbf{V})^1$ and \mathbf{L}_i and \mathbf{K}_i are the moment matrices, defined in (5), associated with $\boldsymbol{\eta}_i$ (i.e. the noise affecting the i^{th} sample point).

2. If $n_q = 1$, e.g. when all the subspaces S_k have dimension $D - 1$, then the rank constraint (12) is no longer required (e.g. only (9)-(11) need to be enforced.)

Proof. Given in the Appendix \square

3.1. A Convex Relaxation

Although there are a few methods to solve rank constrained semidefinite programs (see for instance [14]), these are usually computationally intensive and do not have convergence guarantees. For this reason, we proceed by relaxing the conditions in order to obtain a convex program that approximates the original problem and that can be efficiently solved with off-the-shelf solvers. To this end, we first consider the following problem:

$$\begin{aligned} & \text{minimize}_{\mathbf{m}_i} \quad \text{rank} \{ \mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p}) \} \\ & \text{subject to} \quad \mathbf{L}_i^{\lceil \frac{n}{2} \rceil}(\mathbf{m}_i) \succeq 0 \quad i = 1, \dots, N_p \\ & \quad \quad \quad \mathbf{K}_i^{\lceil \frac{n}{2} \rceil}(\epsilon, \mathbf{m}_i) \succeq 0 \quad i = 1, \dots, N_p \end{aligned} \quad (13)$$

where we truncate the moment matrices such that they only contain the moments up to order n ($n + 1$ if n is odd) which

¹Here the expectation operator \mathbf{E} acts elementwise on $\mathbf{V}(\mathbf{x}) = \mathbf{V}(\hat{\mathbf{x}} + \boldsymbol{\eta})$. That is, \mathbf{M} is constructed by replacing all the monomials in the noise terms $\boldsymbol{\eta}$ in $\mathbf{V}(\mathbf{x})$ with the corresponding moments.

is the maximum order of the noise monomials appearing in the embedded data matrix \mathbf{V} .

Equation (13) is an affine matrix rank minimization problem subject to convex constraints. Although rank minimization is an NP-Hard problem, efficient convex relaxations are available. In particular, good approximate solutions can be obtained by using a log-det heuristic [5] that relaxes rank minimization to a sequence of convex problems². Inspired by the adaptive step size defined for weighted ℓ_1 minimization in [3], the following problem is solved at each iteration:

$$\begin{aligned} & \min_{\mathbf{m}_{1:N_p}, \mathbf{Y}, \mathbf{Z}} \quad \text{trace}(\mathbf{W}_y^{(k)} \mathbf{Y}) + \text{trace}(\mathbf{W}_z^{(k)} \mathbf{Z}) \\ & \text{subject to} \quad \begin{bmatrix} \mathbf{Y} & \mathbf{M}(\mathbf{m}_{1:N_p}) \\ \mathbf{M}(\mathbf{m}_{1:N_p})^T & \mathbf{Z} \end{bmatrix} \succeq 0 \\ & \quad \quad \quad \mathbf{L}_i^{\lceil \frac{n}{2} \rceil}(\mathbf{m}_i) \succeq 0, \quad \mathbf{K}_i^{\lceil \frac{n}{2} \rceil}(\mathbf{m}_i) \succeq 0 \quad i = 1, \dots, N_p \end{aligned} \quad (14)$$

where $\mathbf{W}_y^{(k+1)} = (\mathbf{Y}^{(k)} + \lambda_k \mathbf{I})^{-1}$, $\mathbf{W}_z^{(k+1)} = (\mathbf{Z}^{(k)} + \lambda_k \mathbf{I})^{-1}$ are weights with $\mathbf{Y}^{(k)}, \mathbf{Z}^{(k)}$ being the arguments of the optimal solution in the k^{th} iteration; λ_k , the regularization parameter, is set to the $(h - n_q + 1)^{\text{th}}$ largest singular value of current optimal \mathbf{M} in iteration k ; and $\mathbf{W}_y^{(0)}, \mathbf{W}_z^{(0)}$ are initialized with identity matrices³. Note that the matrices \mathbf{M} , \mathbf{L} and \mathbf{K} are affine in the moment variables \mathbf{m}_i as defined in Theorem 2, and \mathbf{Y} and \mathbf{Z} are symmetric positive definite auxiliary variables. Hence, (14) is a convex semidefinite program.

Remark 2. *Although when $n_q > 1$, the relaxation is less tight since we drop the rank-constraints on the moment matrices, in practice the moment matrices found by solving (14) are close to rank 1. Hence approximating them with rank 1 matrices using a singular value decomposition (SVD) gives satisfactory results (see the supplementary material).*

After solving (14), it is possible to extract an admissible noise sequence from the moment matrix \mathbf{L} as described in [7]. In particular, when \mathbf{L}_i has rank 1, one can retrieve the noise from the elements of the first singular vector of \mathbf{L}_i corresponding to first order moments. Once an admissible noise sequence is obtained, we denoise the data and proceed with segmentation using polynomial differentiation as in [20].

3.2. Extension to Quadratic Surfaces

In this section we briefly present a straightforward extension of our moments-based approach to the segmentation of

²Although there are very recent faster algorithms for rank minimization (e.g. [2]), they currently cannot handle semidefinite constraints.

³The first iteration solves the nuclear norm heuristic. Then each iteration aims to reduce the rank further through the weighting scheme. In our experiments, the convergence is typically achieved within the first 10 iterations.

a mixture of quadratic surfaces discussed in [15]. In particular, we are interested in the class of problems that arises in the context of motion segmentation from two perspective views, given the point correspondences [15, 21, 16]. The main idea is that point correspondences of a single rigid body satisfy the epipolar constraint $\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$ ⁴ where $\mathbf{F} \in \mathbb{R}^{3 \times 3}$ is the fundamental matrix and $\mathbf{x}_j = (x_j, y_j, 1)^T$ for $j = 1, 2$, are the corresponding points in two views in homogeneous coordinates. In the case of n rigid objects, all point correspondences satisfy

$$\prod_{i=1}^n \mathbf{x}_1^T \mathbf{F}_i \mathbf{x}_2 = 0. \quad (15)$$

Let $\mathbf{z} = (x_1, y_1, x_2, y_2)^T$ be a joint vector of corresponding pairs and define the *perspective embedding* of degree $2n$, $\pi_{2n}(\mathbf{z}) : \mathbb{R}^4 \rightarrow \mathbb{R}^m$, $m = \binom{2+n}{n}$ as

$$\pi_{2n}(\mathbf{z}) \doteq \begin{bmatrix} x_1^{\alpha_1^{(1)}} y_1^{\alpha_2^{(1)}} x_2^{\alpha_3^{(1)}} y_2^{\alpha_4^{(1)}} \\ \vdots \\ x_1^{\alpha_1^{(m)}} y_1^{\alpha_2^{(m)}} x_2^{\alpha_3^{(m)}} y_2^{\alpha_4^{(m)}} \end{bmatrix}$$

where the exponents of the monomials satisfy $\alpha_1^{(k)} + \alpha_2^{(k)} \leq n$, $\alpha_3^{(k)} + \alpha_4^{(k)} \leq n$ for all $k = 1, \dots, m$.

As in section 2.2, we can associate with Eq. (15) a polynomial $Q_j(\mathbf{z}) = \mathbf{c}_j^T \pi_{2n}(\mathbf{z})$ where \mathbf{c}_j is the coefficient vector. It follows that, given N_p noiseless corresponding pairs $\mathbf{z}_1, \dots, \mathbf{z}_{N_p}$, the vectors \mathbf{c}_j span the null space \mathcal{N} of the embedded data matrix $\mathbf{P}(\mathbf{z}) = [\pi_{2n}(\mathbf{z}_1) \ \pi_{2n}(\mathbf{z}_2) \ \dots \ \pi_{2n}(\mathbf{z}_{N_p})]^T$. In the noisy case, $\mathbf{P}(\mathbf{z})$ depends polynomially on the noise η . However, as before, we can resort to moments-based rank minimization of $\mathbf{E}(\mathbf{P})$ to find a noise sequence η^* that renders \mathbf{P} rank deficient and obtain its null space (and hence Q_j). Once the data is denoised using η^* , we proceed as in [15]. First, we form the *mutual contraction subspaces* between different pairs $(\mathbf{z}_k, \mathbf{z}_l)$ from the derivatives and Hessians of polynomials Q_j . Applying spectral clustering to the similarity matrix built from the subspace angles between mutual contraction subspaces leads then to the desired segmentation.

4. Experiments

In this section we illustrate the ability of the proposed method to deal with relatively large noise using both synthetic and real data. The data we use here is taken from the literature and thus publicly available.⁵

⁴As shown in [1], this equation can be written as a quadratic form in the joint image space.

⁵A reference implementation is provided at <http://www.coe.neu.edu/~necmiye/cvpr10.htm>

4.1. Synthetic Data

First, we use synthetic data to investigate the performance of the moments-based method for different number of subspaces, different subspace dimensions and different noise levels. For each set of examples, we generated 20 random instances of the problem data with data points lying on a subspace arrangement corrupted by random noise in the direction of subspace normals with uniform random magnitude in $[0.8\epsilon, \epsilon]$. All data points were sampled within the unit hypercube of the ambient space, so that the noise level, ϵ , corresponds roughly to the percent noise level. The segmentation was performed using the convex relaxation described in section 3.1, implemented in Matlab using CVX toolbox [6], and performance was evaluated in terms of the average, over all runs, of the worst case fitting error,

$$\text{err}_f = \max_{i \in [1, N_p]} \min_{k \in [1, n]} \mathbf{b}_k^T \mathbf{x}_i$$

where \mathbf{b}_k 's are the subspace normals found by each algorithm. The results are summarized in Table 1, showing that in all cases the moments-based method outperforms both GPCA and RGPCA.

Figure 3 shows a typical noisy data set and the denoised version obtained by substituting the noise estimates found via moments. As illustrated there, it is possible to align the noisy data points on subspaces before the clustering stage. Hence applying GPCA clustering to these almost noiseless data points avoids the difficulties entailed in using polynomial differentiation in the presence of noise.

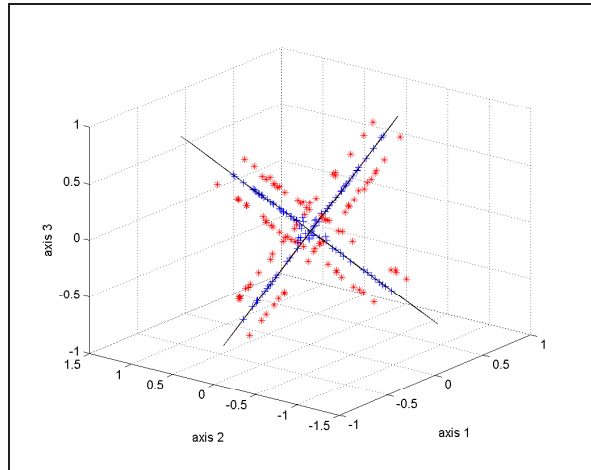


Figure 3. Example using synthetic data laying on two planes (in \mathbb{R}^3). Here the red stars and blue plus signs indicate the original (noisy) and the denoised data points, respectively.

4.2. 2-D motion estimation and segmentation

Next, we consider the problem of simultaneous multiple 2-D motion estimation and clustering from two images. Let

D	d_k	N	ϵ	Moments err_f	GPCA err_f	RGPCA err_f
4	[3, 3]	[50, 50]	0.15	0.250 (0.183)	0.488 (0.477)	0.253 (0.234)
3	[2, 2]	[50, 50]	0.10	0.101 (0.100)	0.393 (0.334)	0.192 (0.134)
3	[2, 2]	[50, 50]	0.15	0.154 (0.151)	0.488 (0.443)	0.289 (0.225)
3	[2, 2]	[50, 50]	0.20	0.227 (0.205)	0.543 (0.539)	0.370 (0.329)
3	[2, 2, 2]	[40, 40, 40]	0.15	0.259 (0.206)	0.499 (0.477)	0.421 (0.398)

Table 1. Synthetic Data Results. D and d_k denote the dimension of the ambient space and subspaces, respectively. N shows the number of samples per subspace. ϵ denotes the true noise level. The last three columns show the mean and median (in parenthesis) fitting errors.

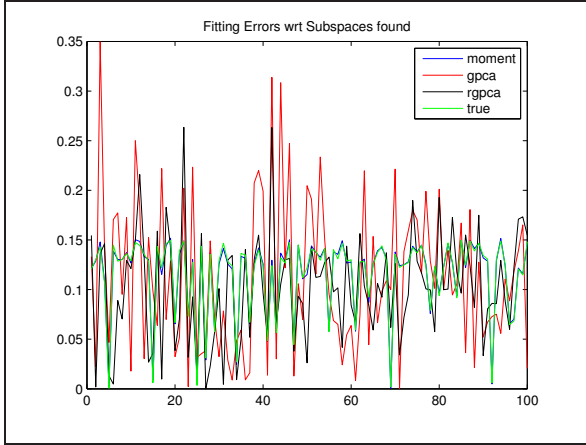


Figure 4. Fitting errors for GPCA, RGPCA and moments-based methods for the example in Fig. 3.

$(\mathbf{x}_i^{(n)}, \mathbf{y}_i^{(n)})$ be the coordinates of the n^{th} feature point in the i^{th} image. The coordinates of this point across frames are related through

$$\begin{bmatrix} \mathbf{x}_1^{(n)} \\ \mathbf{y}_1^{(n)} \end{bmatrix} = \mathbf{R}_j \begin{bmatrix} \mathbf{x}_2^{(n)} \\ \mathbf{y}_2^{(n)} \end{bmatrix} + \mathbf{T}_j$$

where $(\mathbf{R}_j, \mathbf{T}_j)$ are the rotation and translation matrices of the motion of the j^{th} object. Rearranging shows that the vectors, $\mathbf{f}^{(n)} = [\mathbf{x}_1^{(n)}, \mathbf{y}_1^{(n)}, \mathbf{x}_2^{(n)}, \mathbf{y}_2^{(n)}, 1]^T$ corresponding to points belonging to the same object j lay on a 3 dimensional subspace in \mathbb{R}^5 . If multiple objects with different motions are present, each lays in a different subspace and thus, in the noiseless case, can be segmented using GPCA [20].

The image pair shown in Fig. 1 is taken from [19]. We manually marked the feature correspondences and then added uniform random noise to each feature point coordinate in both images. We formed the vectors $\mathbf{f}^{(n)}$, and projected them to \mathbb{R}^4 using a 5×4 matrix with orthonormal columns. Note that such a projection preserves the subspace clusters with probability 1. Specifically, we performed an SVD of the matrix $\mathbf{F} = [\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(N)}] = \mathbf{U}\mathbf{D}\mathbf{V}^T$ and projected the data using the first 4 columns of \mathbf{U} (i.e. $\mathbf{X} = \mathbf{U}_{1:4}^T \mathbf{F}$). Finally, we applied our method, GPCA and RGPCA to cluster these 4-D vectors in \mathbf{X} . The results, along with a typical segmentation, are summarized

in Fig. 2, showing that the moments-based algorithm yields a substantially lower misclassification rate.

4.3. Two view perspective motion segmentation

In this section, we demonstrate the performance of our method in the problem of motion segmentation from two perspective views and compare against hybrid quadratic surface analysis (HQSA) [15], using both synthetic and real data.

First, we artificially generated two teapots and projected 95 points from each teapot surface to the image plane using a perspective camera model. Then we moved both teapots to different locations and generated a second image proceeding in the same way. The two views used are shown in Fig. 5. Then, we ran 20 random trials by adding iid zero mean Gaussian noise with variance 0.15 to each image point in both images. The mean misclassification error for moments-based method is around 2% whereas the mean misclassification error for HQSA is 23%.

Next we consider two real examples. Figure 6 shows two piles of boxes that are moved in two different directions. This example is taken from [17]. We used the first and ninth frames of the sequence as our two perspective views. The missing tracks were ignored. In the resulting segmentation HQSA misclassifies 29 points whereas the moments-based approach misclassifies only 10 points. Since our model takes into account only additive noise and this is not the only source of error, the final classification is not perfect. However, the moments-based approach still outperforms HQSA. For the next example, we used as perspective views the first and last frames of the “truck2” sequence from the Hopkins 155 dataset [18]. In this case, the misclassification rate for the moments-based approach is 9.06% (it only misclassifies the left rear wheel as background). On the other hand, HQSA has a misclassification rate of 42.90%, clustering half of the points on the truck as background.

5. Concluding Remarks

In this paper, we propose a new method for segmentation of multiple algebraic surfaces from noisy sample data points. Although particular attention was given to linear and quadratic surfaces, our method easily extends to surfaces

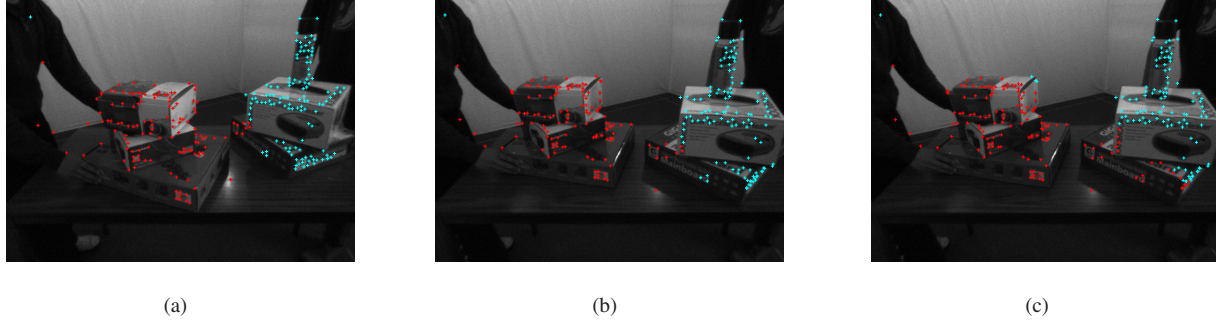


Figure 6. (a)-(b) First and second images with moments-based segmentation superimposed, 10 misclassified points. (c) Segmentation with the method in [15], 29 misclassified points.

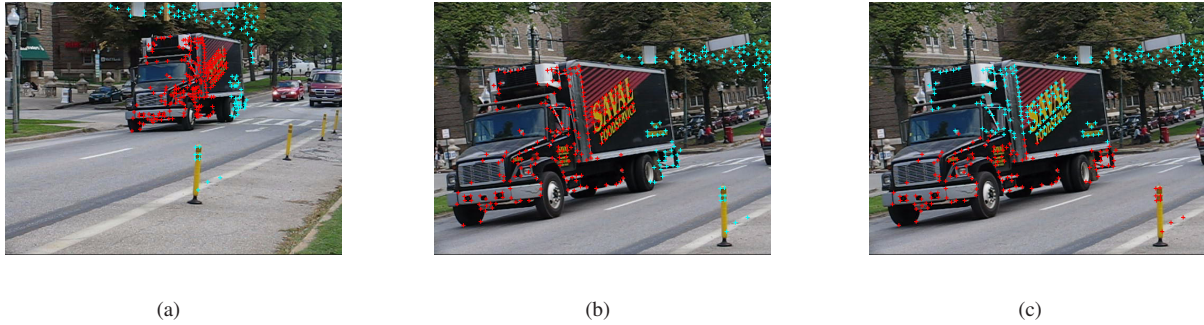


Figure 7. (a)-(b) First and second images with moments-based segmentation superimposed (9.06% misclassification rate). (c) Segmentation with the method in [15] (42.90% misclassification rate).

Sequence	Moments-based	HQSA
teapot	2.1%	23.42%
boxes	4.05%	11.74%
truck2	9.06%	42.90%

Table 2. Misclassification rates for perspective motion segmentation examples.

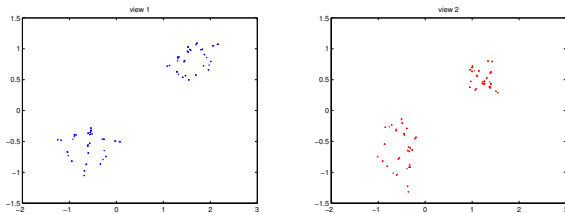


Figure 5. Two perspective images of points on teapot surfaces.

generated by higher order polynomials. The main motivation for the approach discussed in this paper is the so-called generalized principal component analysis (GPCA) proposed by Vidal et al., which addresses the segmentation problem for the case of noise free sample data points. Building on this, the problem of segmentation in the presence of noisy data is reformulated as the problem of finding a point in a set defined by linear matrix inequalities and rank con-

straints on matrices affine in the decision variables. Convex relaxations of rank constraints were then used to obtain a tractable convex optimization problem. The effectiveness of the proposed method is illustrated by several examples, including simultaneous 2D motion segmentation and motion segmentation from two perspective views. The results obtained show that our formulation substantially reduces the noise sensitivity of GPCA.

Effort is now being put in improving the computational performance of this approach. Also, as further research, it would be of interest to study how to extend the proposed approach to non-algebraic surfaces.

References

- [1] P. Anandan and S. Avidan. Integrating local affine into global projective images in the joint image space. In *ECCV (1)*, pages 907–921, 2000.
- [2] J. F. Cai, E. J. Candes, and Z. Shen. A singular value thresholding algorithm for matrix completion. preprint.
- [3] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.
- [4] R. E. Curto and L. A. Fialkow. Truncated k -moment problems in several variables. *Journal of Operator Theory*, 54(1):189–226, 2005.

- [5] M. Fazel, H. Hindi, and S. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *American Control Conference*, 2003.
- [6] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming. (web page and software), June 2009. <http://stanford.edu/~boyd/cvx>.
- [7] D. Henrion and J.-B. Lasserre. *Positive Polynomials in Control*, chapter Detecting global optimality and extracting solutions in GloptiPoly, pages 293–310. Springer, 2005.
- [8] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. J. Kriegman. Clustering appearances of objects under varying illumination conditions. In *CVPR (1)*, pages 11–18, 2003.
- [9] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Opt.*, 11:796–817, 2001.
- [10] L. Lu and R. Vidal. Combined central and subspace clustering for computer vision applications. In *International Conference on Machine Learning*, pages 593–600, 2006.
- [11] Y. Ma and R. Vidal. A closed form solution to the identification of hybrid arx models via the identification of algebraic varieties. In *HSCC*, pages 449–465, March 2005.
- [12] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Rev.*, 50:413–458, 2008.
- [13] J. L. McGregor. Solvability criteria for certain n-dimensional moment problems. *Journal of Approximation Theory*, 30:315–333, 1980.
- [14] R. Orsi, U. Helmke, and J. B. Moore. A newton-like method for solving rank constrained linear matrix inequalities. *Automatica*, 42:1875–1882, 2006.
- [15] S. Rao, A. Y. Yang, A. Wagner, and Y. Ma. Segmentation of hybrid motions via hybrid quadratic surface analysis. In *ICCV*, pages 2–9, 2005.
- [16] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(6):983–995, 2006.
- [17] K. Schindler, D. Suter, and H. Wang. A model selection framework for multibody structure-and-motion of image sequences. *IJCV*, 79(2):159–177, 2008.
- [18] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *CVPR*, 2007.
- [19] O. Tuzel, R. Subbarao, and P. Meer. Simultaneous multiple 3d motion estimation via mode finding on lie groups. In *ICCV*, pages 18–25, 2005.
- [20] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *PAMI*, 27(12):1945–1959, December 2005.
- [21] R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two-view multibody structure from motion. *IJCV*, 68(1):7–25, 2006.

A. Proof of Theorem 2

First, if there exists a noise sequence $\boldsymbol{\eta}_i^*$ such that

$$Q_j(\mathbf{x}_i - \boldsymbol{\eta}_i^*) = \nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^*)^T \mathbf{c}_j = 0 \text{ for all } i, j,$$

then, the moments \mathbf{m}_i of the atomic probability measures μ_i with

$$\text{Prob}_{\mu_i}(\boldsymbol{\eta}_i = \boldsymbol{\eta}_i^*) = 1, \text{ Prob}_{\mu_i}(\boldsymbol{\eta}_i \neq \boldsymbol{\eta}_i^*) = 0.$$

trivially satisfy (9) through (12) and the vectors \mathbf{c}_j span the null space of the matrix $\mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p})$.

To prove the converse, let's first look at part 1. We start by noting that $\text{rank}(\mathbf{L}_i) = 1$ is equivalent to the existence of a unique atomic measure μ_i with one atom whose moments are equal to \mathbf{m}_i (see [4]) and, hence, there exist $\boldsymbol{\eta}_i^*$ such that

$$\text{Prob}_{\mu_i}(\boldsymbol{\eta}_i = \boldsymbol{\eta}_i^*) = 1, \text{ Prob}_{\mu_i}(\boldsymbol{\eta}_i \neq \boldsymbol{\eta}_i^*) = 0.$$

Therefore, for this measure, one has

$$\mathbf{V}(\mathbf{x} - \boldsymbol{\eta}^*) = \mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p}).$$

Note that (11) implies that $\|\boldsymbol{\eta}_i^*\|_2 \leq \epsilon$. Given this, let \mathbf{c}_j be linearly independent vectors which span the null space of the matrix $\mathbf{V}(\mathbf{x} - \boldsymbol{\eta}^*)$. To conclude this part of the proof, define $Q_j(\mathbf{x}_i) = \nu_n(\mathbf{x}_i)^T \mathbf{c}_j$ and $\hat{\mathbf{x}}_i \doteq \mathbf{x}_i - \boldsymbol{\eta}_i^*$. From the reasoning above it follows that

$$Q_j(\hat{\mathbf{x}}_i) = \nu_n(\hat{\mathbf{x}}_i)^T \mathbf{c}_j = \nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^*)^T \mathbf{c}_j = 0 \text{ for all } i, j.$$

We now turn our attention to part 2. Since condition (12) might not be satisfied, the measures μ_i compatible with the moment sequences \mathbf{m}_i are not necessarily atomic measures. However, equation (9) implies that \mathcal{N} , the null space of the matrix $\mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p})$, has dimension one. Let \mathbf{c} span \mathcal{N} , i.e.,

$$\mathbf{M}(\mathbf{x}, \mathbf{m}_1, \dots, \mathbf{m}_{N_p})\mathbf{c} = 0.$$

Linearity of expectation implies that

$$\mathbf{E}_{\mu_i}[\nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i)^T \mathbf{c}] = 0.$$

Therefore, there exist $\boldsymbol{\eta}_i^+$ and $\boldsymbol{\eta}_i^-$ within the noise bounds such that

$$\nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^+)^T \mathbf{c} \geq 0 \text{ and } \nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^-)^T \mathbf{c} \leq 0.$$

Since $\nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i)^T \mathbf{c}$ is a polynomial and, hence, a continuous function of $\boldsymbol{\eta}_i$, there exists a $\boldsymbol{\eta}_i^*$ such that

$$\nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^*)^T \mathbf{c} = 0.$$

The proof is concluded by defining the unique (up to a multiplying constant) vanishing polynomial $Q(\hat{\mathbf{x}}_i) = \nu_n(\hat{\mathbf{x}}_i)^T \mathbf{c}$ which, given the reasoning above, satisfies

$$Q(\hat{\mathbf{x}}_i) = \nu_n(\hat{\mathbf{x}}_i)^T \mathbf{c} = \nu_n(\mathbf{x}_i - \boldsymbol{\eta}_i^*)^T \mathbf{c} = 0 \text{ for all } i.$$

Remark 3. Note that, without the rank constraint (12), a finite order approximation in (10) and (11) only leads to necessary conditions for \mathbf{m}_i to be moments of a probability distribution. Hence, implementing the approach in part 2 with these truncated matrices only provides an approximation of the true $Q(\cdot)$. However, the approximation converges to the true solution as the moment approximation order increases.