Binary Reduced Row Echelon Form Approach for Subspace Segmentation

Akram Aldroubi
Department of Mathematics
Vanderbilt University, Nashville, TN, 37212 USA
Email: akram.aldroubi@vanderbilt.edu

Ali Sekmen
Department of Computer Science
Tennessee State University, Nashville, TN 37209
Email:asekmen@tnstate.edu

***Abstract*—This paper introduces a subspace segmentation and data clustering method for a set of data drawn from a union of subspaces. The proposed method works perfectly in absence of noise, i.e., it can find the number of subspaces, their dimensions, and an orthonormal basis for each subspace. The effect of noise on this approach depends on the noise level and relative positions of subspaces. We provide a performance analysis in presence of noise and outliers.**

I. Introduction

The goal of subspace clustering is to identify all of the subspaces that a set of data $\mathbf{W} = \{w_1, ..., w_N\} \in \mathbb{R}^D$ is drawn from and assign each data point w_i to the subspace it belongs to. The number of subspaces, their dimensions, and a basis for each subspace are to be determined even in presence of noise, missing data, and outliers. In some subspace clustering problems, the number M of subspaces or the dimensions of the subspaces $\{d_i\}_{i=1}^M$ are known. A number of approaches have been devised to solve the problem above or some of its special cases. They are based on sparsity methods [1], [2], [3], [4], algebraic methods [5], [6], iterative and statistical methods [7], [8], [9], [10], [11], [12], and spectral clustering methods [2], [3], [13], [14], [15], [16], [17], [18], [19].

In this work, we develop an algebraic method for solving the general subspace segmentation problem for noiseless data. For the case where all the subspaces are four dimensional, Gear observed, without proof, that the reduced echelon form can be used to segment motions in videos [20]. In this paper, we develop this idea and show that the reduced row echelon form can completely solve the subspace segmentation problem in its most general version for noiseless data. For noisy data, the reduced echelon form method does not work, and a thresholding must be applied. The effect of the noise on the reduced echelon form method depends on the noise level and the relative positions of the subspaces.

A. Non-Linear Approximation Formulation

When M is known, the subspace segmentation problem, for both the finite and infinite dimensional space cases, can be formulated as follows:

Let $\mathcal{B}$ be a Banach space, $\mathbf{W} = \{w_1, \dots, w_N\}$ a finite set of vectors in $\mathcal{B}$. For $i = 1, \dots, M$, let $\mathcal{C} = C_1 \times C_2 \times \dots \times C_M$ be the cartesian product of M family C_i of closed subspaces of $\mathcal{B}$ each containing the trivial subspace $\{0\}$. Thus, an element $\mathbf{S} \in \mathcal{C}$ is a sequence $\{S_1, \dots, S_M\}$ of M subspaces of $\mathcal{B}$ with $S_i \in C_i$. An example for finite dimensions is when $\mathcal{B} = \mathbb{R}^D$ and $\mathcal{C}$ is the family of all subspaces of $\mathbb{R}^D$ of dimensions less than or equal to D. An example for infinite dimensions is when $\mathcal{B} = L^2(\mathbb{R}^D)$ and $\mathcal{C}$ is a family of closed, shift-invariant subspaces of $L^2(\mathbb{R}^D)$ that are generated by finite generators.

Problem 1.

1) *Given a finite set $\mathbf{W} \subset \mathcal{B}$, a fixed p with $0 < p \leq \infty$, and a fixed integer $M \geq 1$, find the infimum of the expression*

$$e(\mathbf{W}, \mathbf{S}) := \sum_{w \in \mathbf{W}} \min_{1 \leq j \leq M} d^p(w, S_j),$$

over $\mathbf{S} = \{S_1, \dots, S_M\} \in \mathcal{C}$, and $d(x, y) := \|x - y\|_{\mathcal{B}}$.

2) *Find a sequence of M-subspaces $\mathbf{S}^o = \{S_1^o, \dots, S_M^o\} \in \mathcal{C}$ (if it exists) such that*

$$e(\mathbf{W}, \mathbf{S}^o) = \inf\{e(\mathbf{W}, \mathbf{S}) : \mathbf{S} \in \mathcal{C}\}. \quad \text{(I.1)}$$

In the presence of outliers, it is shown that $p = 1$ is a good choice [21] and a good choice for light-tailed noise is $p = 2$. The necessary and sufficient conditions for the existence of a solution when $p = 2$ and $\mathcal{B}$ is a Hilbert space can be found in [22].

Definition 1. *For $0 < p \leq \infty$, a set of closed subspaces C of a Banach space $\mathcal{B}$ has the Minimum Subspace Approximation Property p-(MSAP) if for every finite subset $\mathbf{W} \subset \mathcal{B}$ there exists an element $S \in C$ that minimizes the expression $e(\mathbf{W}, S) = \sum_{w \in \mathbf{W}} d^p(w, S)$ over all $S \in C$.*

Under the assumption that each family of subspaces C_i satisfies p-(MSAP), problem 1 has a minimizer [23]:

Theorem 1. *If for each $i = 1, \dots M$, C_i satisfies p-(MSAP), then Problem 1 has a minimizing set of subspaces for all finite sets of data.*

Theorem 1 suggests an iterative search algorithm for the optimal solution $\mathbf{S}^o$. Obviously, this solution can be obtained by Algorithm 1. This algorithm will work well if a good initial partition is chosen. Otherwise, the algorithm may terminate in a local minimum instead of the global minimum.

Algorithm 1 Optimal Solution $\mathbf{S}^o$

1: Pick any partition $P \in \mathcal{P}(\mathbf{W})$
2: For each subset $\mathbf{W}_i$ in the partition P find the subspace $S_i^o(P) \in C_i$ that minimizes the expression $e(\mathbf{W}_i, S) = \sum_{w \in \mathbf{W}_i} d^p(w, S)$
3: **while** $\sum_{i=1}^{M} e(\mathbf{W}_i, S_i^o(P)) > e(\mathbf{W}, \mathbf{S}^o(P))$ **do**
4: **for all** i from 1 to M **do**
5: Update $\mathbf{W}_i = \{w \in \mathbf{W} : d(w, S_i^o(P)) \leq d(w, S_k^o(P)),\ k = 1, \ldots, M\}$
6: Update $S_i^o(P) = \operatorname*{argmin}_{S \in C_i} e(\mathbf{W}_i, S)$
7: **end for**
8: Update $P = \{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$
9: **end while**
10: $\mathbf{S}^o = \{S_1^o(P), \ldots, S_M^o(P)\}$

II. Subspace Segmentation - Noiseless Case

In this section we consider the problem in which a set of vectors $\mathbf{W} = \{w_1, \ldots, w_N\}$ are drawn from a union $\mathcal{U} = \bigcup_{i \in I} S_i$ of M subspaces $S_i \in \mathbb{R}^D$ of dimension d_i. In order to find the M subspaces from the data set $\mathbf{W}$ it is clear that we need enough vectors $\mathbf{W} = \{w_1, \ldots, w_N\}$. In particular for the problem of subspace segmentation, it is necessary that the set $\mathbf{W}$ can be partitioned into M sets $\mathbf{W} = \{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$ such that span $\mathbf{W}_i = S_i$, $i = 1, \ldots, M$. Thus, we need to assume that we have enough data for solving the problem. In particular, we assume that any $k \leq d$ vectors drawn from a subspace S of dimension d are linearly independent, and we make the following definition.

Definition 1. *Let S be a linear subspace of $\mathbb{R}^D$ with dimension d. A set of data $\mathbf{W}$ drawn from $S \subset \mathbb{R}^D$ with dimension d is said to be* generic *if (i) $|\mathbf{W}| > d$, and (ii) every d vectors from $\mathbf{W}$ form a basis for S.*

Another assumption that we will make is that the union of subspaces $\mathcal{U} = \bigcup_{i \in I} S_i$ from which the data is drawn consists of independent subspaces:

Definition 2. *(Independent Subspaces) Subspaces $\{S_i \subset \mathbb{R}^D\}_{i=1}^n$ are called independent if $dim(S_1 + \cdots + S_n) = dim(S_1) + \cdots + dim(S_n)$.*

Definition 3. *Matrix R is said to be the* binary reduced row echelon form *of matrix A if all non-pivot column vectors are converted to binary vectors, i.e., non-zero entries are set to one.*

The following theorem suggests a very simple yet effective approach to cluster the data points. The proofs of the following Theorems can be found in [23].

Theorem 2. *Let $\{S_i\}_{i=1}^M$ be a set of non-trivial linearly independent subspaces of $\mathbb{R}^D$ with corresponding dimensions $\{d_i\}_{i=1}^M$. Let $\mathbf{W} = [w_1 \cdots w_N] \in \mathbb{R}^{D \times N}$ be a matrix whose columns are drawn from $\bigcup_{i=1}^M S_i$. Assume the data is drawn from each subspace and that it is generic. Let Brref$(\mathbf{W})$ be the binary reduced row echelon form of $\mathbf{W}$. Then*

1) *The inner product $\langle e_i, b_j \rangle$ of a pivot column e_i and a non-pivot column b_j in Brref$(\mathbf{W})$ is one, if and only if the corresponding column vectors $\{w_i, w_j\}$ in $\mathbf{W}$ belong to the same subspace S_l for some $l = 1, \ldots, M$.*
2) *Moreover, $\dim(S_l) = \|b_j\|_1$, where $\|b_j\|_1$ is the l_1-norm of b_j.*
3) *Finally, $w_p \in S_l$ if and only if $b_p = b_j$ or $\langle b_p, b_j \rangle = 1$.*

The data $\mathbf{W}$ can be partitioned into M clusters $\{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$, such that span$\mathbf{W}_l = S_l$. The clusters can be formed as follows: Pick a non-pivot element b_j in Brref$(\mathbf{W})$, and group together all columns b_p in Brref$(\mathbf{W})$ such that $\langle b_j, b_p \rangle > 0$. Repeat the process with a different non-pivot column until all columns are exhausted.

III. Subspace Segmentation - Noisy Case

In practice the data $\mathbf{W}$ is corrupted by noise. In this case, the Reduced Row Echelon Form (RREF)-based algorithm cannot work, even under the assumption of Theorem 2, since the noise will have two effects: 1) The rank of the data corrupted by noise $\mathbf{W} + \eta \subset \mathbb{R}^D$ becomes full; i.e., $\text{rank}(\mathbf{W} + \eta) = D$; and 2) Even under the assumption that $r = D$, none of the entries of the non-pivot columns of rref$(\mathbf{W} + \eta)$ will be zero. One way of circumventing this problem, is to use the RREF-based algorithm in combination with thresholding to set to zero those entries that are small. The choice of the threshold depends on the noise characteristics and the position of the subspaces relative to each other.

In general, $dim(\sum_{i=1}^M S_i) = \text{rank}(\mathbf{W}) \leq D$, where D is the dimension of the ambient space $\mathbb{R}^D$. After projection of $\mathbf{W}$, the new ambient space is isomorphic to $\mathbb{R}^r$, where $r = \text{rank}(\mathbf{W})$, and we may assume that $\text{rank}(\mathbf{W}) = D$. Without loss of generality, let us assume that $\mathbf{W} = \begin{bmatrix} A & B \end{bmatrix}$ where the columns of A form basis for $\mathbb{R}^D$, i.e., the columns of A consist of d_i linearly independent vectors from each subspace S_i, $i = 1, \ldots, M$. Let $\widetilde{\mathbf{W}} = \mathbf{W} + \mathbf{N}$ be the data with additive noise. Then the reduced echelon form applied to $\widetilde{\mathbf{W}}$ is given by rref$(\widetilde{\mathbf{W}}) = \begin{bmatrix} I & \widetilde{A}^{-1}\widetilde{B} \end{bmatrix}$. Let b_i and $\tilde{b}_i$ denote the columns of B and $\widetilde{B}$ respectively, $e_i = \widetilde{A}^{-1}\tilde{b}_i - A^{-1}b_i$, $\Delta = \widetilde{A} - A$, and $\nu_i = \tilde{b}_i - b_i$. Let σ_{min} denote the smallest singular value of A, then if $||\Delta|| \leq \sigma_{min}(A)$, we get

$$\|e_i\|_2 \leq \frac{\|\nu_i\|_2}{\sigma_{min}(A)} + \frac{\|\Delta\|}{\sigma_{min}^2(A)} \left(\frac{1}{1 - \frac{\|\Delta\|}{\sigma_{min}(A)}} \right) (\|b_i\|_2 + \|\nu_i\|_2), \tag{III.1}$$

where $\|\cdot\|$ denotes the operator norm $\|\cdot\|_{\ell^2 \to \ell^2}$. Unless specified otherwise, the noise $\mathbf{N}$ will be assumed to consist of entries that are i.i.d. $\mathcal{N}(0, \sigma^2)$ Gaussian noise with zero mean and variance σ^2. For this case, the expected value of $\|\Delta\|$ can be estimated by $\mathbb{E}\|\Delta\| \leq C\sqrt{D}\sigma$ [24]. Note that to estimate the error in (III.1) we still need to estimate $\sigma_{min}(A)$. This singular value depends on the position of the subspaces $\{S_i\}_{i=1}^M$ relative to each other which can be measured by the principle angles between them. The principle angles between two subspaces $\mathcal{F}, \mathcal{G}$, can be obtained using any pair of orthogonal bases for $\mathcal{F}, \mathcal{G}$ as described in the following Lemma [25]:

Lemma 1. *Let $\mathcal{F}$ and $\mathcal{G}$ be two subspaces of $\mathbb{R}^D$ with $p = dim(\mathcal{F}) \leq dim(\mathcal{G}) = q$. Assume that $Q_{\mathcal{F}} \in \mathbb{R}^{D\times p}$ and $Q_{\mathcal{G}} \in \mathbb{R}^{D\times q}$ are matrices whose columns form orthonormal bases for the subspaces $\mathcal{F}$ and $\mathcal{G}$. If $1 \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ are the singular values of $Q_{\mathcal{F}}^t Q_{\mathcal{G}}$, then the principle angles are given by*

$$\theta_k = \arccos(\sigma_k) \quad k = 1, \ldots, p. \tag{III.2}$$

The dependence of the minimum singular value $\sigma_{\min}(A)$ on the principle angles between the subspaces $\{S_i\}_{i=1}^M$ is given in the theorem below, which is one of the two main theorems of this section. The proofs are provided in [23].

Theorem 3. *Assume that $\{S_i\}_{i=1}^M$ are independent subspaces of $\mathbb{R}^D$ with corresponding dimensions $\{d_i\}_{i=1}^M$ such that $\sum_{i=1}^M d_i = D$. Let $\{\theta_j(S_i)\}_{j=1}^{\min(d_i,D-d_i)}$ be the principle angles between S_i and $\sum_{\ell\neq i} S_\ell$. If $A = \begin{bmatrix} a_1 & \ldots & a_D \end{bmatrix}$ is a matrix whose columns $\{a_1, \ldots, a_D\} \subset \cup_{i=1}^M S_i$ form a basis for $\mathbb{R}^D$, with $\|a_i\|_2 = 1$, $i = 1, \ldots, D$, then*

$$\sigma_{\min}^2(A) \leq \min_i \left(\prod_{j=1}^{\min(d_i,D-d_i)} \left(1 - \cos^2(\theta_j(S_i))\right) \right)^{1/D}, \tag{III.3}$$

where $\sigma_{\min}(A)$ is the smallest singular value of A.

Corollary 1. *Under the same conditions of Theorem 3, a simpler but possibly larger upper bound is given by:*

$$\sigma_{\min}^2(A) \leq \min_i (1 - \cos(\theta_1(S_i))^{1/D} 4^{1/D}, \tag{III.4}$$

where $\theta_1(S_i)$ is the minimum angle between S_i and $\sum_{\ell\neq i} S_\ell$.

Theorem 4. *Assume that $\{S_i\}_{i=1}^M$ are independent subspaces of $\mathbb{R}^D$ with corresponding dimensions $\{d_i\}_{i=1}^M$ such that $\sum_{i=1}^M d_i = D$. Let $\{\theta_j(S_i)\}_{j=1}^{\min(d_i,D-d_i)}$ be the principle angles between S_i and $\sum_{\ell\neq i} S_\ell$. Assume that $\mathbf{W} = [w_1 \cdots w_N] \in \mathbb{R}^{D\times N}$ is a matrix whose columns are drawn from $\bigcup_{i=1}^M S_i$ and the data is generic for each subspace S_i. If P is a permutation matrix such that $\mathbf{W}P = \begin{bmatrix} A_P & B_P \end{bmatrix}$, and A_P is invertible, then*

$$\sup_P \{\sigma_{\min}^2(A_P)\} \leq \min_i \left(\prod_{j=1}^{\min(d_i,D-d_i)} \left(1 - \cos^2(\theta_j(S_i))\right) \right)^{1/D} \tag{III.5}$$

In particular,

$$\sup_P \{\sigma_{\min}^2(A_P)\} \leq \min_i (1 - \cos(\theta_1(S_i))^{1/D} 4^{1/D}, \tag{III.6}$$

where $\theta_1(S_i)$ is the minimum angle between S_i and $\sum_{\ell\neq i} S_\ell$.

Remark 1. *The value $\sigma_{\min}(A_P)$ can be arbitrarily close to zero, thus, one of the goals is to find D columns of $\mathbf{W}$ that form a basis such that $\sigma_{\min}(A_P)$ is as close to the upper bound as possible without an exhaustive search.*

IV. Subspace Segmentation Algorithm for Noisy Data

Algorithm 1 works perfectly in noiseless data. For noisy data, the success of the algorithm depends on finding a good initial partition. Otherwise, the algorithm may terminate at a local minimum. Theorem 2 works perfectly for *noiseless* data (it determines a basis for each subspace and it correctly clusters all of the data points). An algorithm for implementing Theorem 2 is given in [23]. However, it does not perform very well when sufficiently large noise is present because any threshold value will keep some of the values that need to be zeroed out and will zero out some of the values that need to be kept. However, the thresholded reduced echelon form can be used to determine a set of clusters that can in turn be used to determine a good initial set of subspaces in Algorithm 1.

For example, if the number of subspaces is known and the subspaces have equal and known dimensions (assume that there are M subspaces and each subspace has dimension d), then Algorithm 2 below combines Algorithm 1 and Theorem 2 as follows: First, the reduced row echelon form rref($\mathbf{W}$) of $\mathbf{W}$ is computed. Since the data is noisy, the non-pivot columns of rref($\mathbf{W}$) will most likely have all non-zero entries. The error in those entries will depend on the noise and the positions of the subspaces as in (3). Since each subspace is d-dimensional, the highest d entries of each non-pivot column is set to 1 and all other entries are set to 0. This determines the binary reduced row echelon form Brref($\mathbf{W}$) of $\mathbf{W}$ (note that, according to Theorem 2, each non-pivot column of Brref($\mathbf{W}$) is supposed to have d entries). M groups of the equivalent columns of Brref($\mathbf{W}$) are determined and used as the initial partition for Algorithm 1. This process is described in Algorithm 2. Note that a dimensionality reduction is also performed to speed up the process.

Remark 2. *In Step-5 of Algroithm 2, Brref($\mathbf{W}$) is computed by setting the highest d entries of each non-pivot columns to 1 and the others to 0. If we do not know the dimensions of the subspaces, we may need to determine a threshold from the noise characteristics and a priori knowledge of the relative position of subspaces using* (III.1) *and* (III.3).

Remark 3. *In Step-7 of Algorithm 2, we find the subspace $S_i^o(P)$ that minimizes the expression $e(\mathbf{W}_i, S) = \sum_{w\in\mathbf{W}_i} d^p(w, S)$ for each subset $\mathbf{W}_i$ in the partition P. For data with light-tailed noise (e.g. Gausian distributed noise) $p = 2$ is optimal and the minimum in Step-7 can be found using SVD. For heavy-tailed noise (e.g. Laplacian distributed noise), $p = 1$ is the better choice as described in the simulations section.*

Remark 4. *In order to reduce the dimensionality of the problem, we compute the SVD of $\mathbf{W} = U\Sigma V^t$. In Algorithm 2, each subspace is d-dimensional and there are M subspaces. Therefore, it replaces $\mathbf{W}$ by $(V^t)_r$, where $r = M \times d$ is known or estimated rank of $\mathbf{W}$.*

Algorithm 2 Combined Algorithm - Optimal Solution $\mathbf{S}^o$

Require: Normalized data matrix $\mathbf{W}$.

Set $r = M \times d$.
Compute the SVD of $\mathbf{W}$ and find $(V^t)_r$.
Replace the data matrix $\mathbf{W}$ with $(V^t)_r$.
Compute rref($\mathbf{W}$)
Compute Brref($\mathbf{W}$) by setting the highest d entries of each non-pivot column to 1 and all the others to 0.
Group the non-pivot equivalent columns of Brref($\mathbf{W}$) into M largest clusters $\{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$ and set the initial partition $P = \{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$.
For each subset $\mathbf{W}_i$ in the partition P find the subspace $S_i^o(P)$ that minimizes the expression $e(\mathbf{W}_i, S) = \sum_{w \in \mathbf{W}_i} d^p(w, S)$.
while $\sum\limits_{i=1}^{M} e(\mathbf{W}_i, S_i^o(P)) > e(\mathbf{W}, \mathbf{S}^o(P))$ **do**
  for all i from 1 to M **do**
    Update $\mathbf{W}_i = \{w \in \mathbf{W} : d(w, S_i^o(P)) \leq d(w, S_k^o(P)),\ k = 1, \ldots, M\}$
    Update $S_i^o(P) = \underset{S}{\operatorname{argmin}}\, e(\mathbf{W}_i, S)$
  end for
  Update $P = \{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$
end while
$\mathbf{S}^o = \{S_1^o(P), \ldots, S_M^o(P)\}$

V. Experimental Results

We used the Hopkins 155 Dataset [6] to evaluate our algorithm. The RREF-based algorithm is extremely fast and works well with two-motion video sequences. The average and median errors for all two-motion sequences are 11.45% and 6.78%, respectively (8.81% and 5.44% for checker, 16.04% and 11.94% for traffic, and 17.25% and 12.69% for articulated motion). However, the error is very high for three-motion sequences and obviously it does not work well with such video sequences. We believe that this is due to the fact that the noise is correlated, and the minimum of Problem 1 does not give the correct clustering for this case. The best clustering method to date for clustering in this case is based on similarity between trajectory vectors computed from local subspace estimations [26].

VI. Conclusion

This paper introduces a simple and very fast approach for subspace segmentation for data drawn from a union of subspaces. In absence of noise, our approach can find the number of subspaces, their dimensions, and an orthonormal basis for each subspace. We provide an analysis of our theory and determine its limitations and strengths in presence of outliers and noise.

Acknowledgment

The research of Akram Aldroubi is supported in part by NSF Grant DMS-1108631. The research of Ali Sekmen is supported in part by NASA Grant NNX12AI14A.

References

[1] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 11 2009.
[2] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2790–2797.
[3] R. Vidal and E. Elhamifar, "Clustering disjoint subspaces via sparse representation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010, pp. 1926–1929.
[4] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *International Conference on Machine Learning*, 2010, pp. 663–670.
[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. 1945–1959, 12 2005.
[6] R. Tron and R. Vidal, "A benchmark for the comparison of 3-d motion segmentation algorithms," in *Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
[7] K. Kanatani and Y. Sugaya, "Multi-stage optimization for multi-body motion segmentation," in *IEICE Trans. Inf. and Syst*, 2003, pp. 335–349.
[8] A. Aldroubi and K. Zaringhalam, "Nonlinear least squares in $\mathbb{R}^n$," *Acta Applicandae Mathematicae*, vol. 107, no. 1-3, pp. 325–337, 7 2009.
[9] A. Aldroubi, C. Cabrelli, and U. Molter, "Optimal non-linear models for sparsity and sampling," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 793–812, 12 2009.
[10] P. Tseng, "Nearest q-flat to m points," *Journal of Optimization Theory and Applications*, vol. 105, no. 1, pp. 249–252, 2000.
[11] 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*, vol. 24, no. 6, pp. 381–395, 6 1981.
[12] N. Silva and J. Costeira, "Subspace segmentation with outliers: a grassmannian approach to the maximum consensus subspace," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
[13] U. V. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, pp. 395–416, 2007.
[14] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *International Journal of Computer Vision*, vol. 81, pp. 317–330, 2009.
[15] F. Lauer and C. Schnorr, "Spectral clustering of linear subspaces for motion segmentation," in *IEEE International Conference on Computer Vision*, 2009.
[16] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and nondegenerate," in *9th European Conference on Computer Vision*, 2006, pp. 94–106.
[17] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *Computer Vision and Pattern Recognition*, 6 2007, pp. 1–6.
[18] R. Vidal, "A tutorial on subspace clustering," *IEEE Signal Processing Magazine*, vol. 79, no. 1, pp. 52–68, 2010.
[19] G. Chen and G. Lerman, "Foundations of a multi-way spectral clustering framework for hybrid linear modeling," vol. 9, no. 5, pp. 517–558, 2009.
[20] C. Gear, "Multibody grouping from motion images," *International Journal of Computer Vision*, vol. 29, no. 2, pp. 133–150, 1998.
[21] G. Lerman and T. Zhang, "Robust recovery of multiple subspaces by geometric lp minimization," vol. 39, no. 5, pp. 2686–2715, 2011.
[22] A. Aldroubi and R. Tessera, "On the existence of optimal unions of subspaces for data modeling and clustering," *Foundation of Computational Mathematics*, vol. 11, no. 3, pp. 363–379, 6 2011.
[23] A. Aldroubi and A. Sekmen, "Reduced row echelon form and non-linear approximation for subspace segmentation and high-dimensional data clustering," *Preprint*.
[24] R. Latala, "Some estimates of norms of random matrices," *Proc. Amer. Math. Soc.*, vol. 133, 2005.
[25] G. H. Golub and C. F. V. Loan, *Matrix computations, 3rd Edition*. Johns Hopkins University Press, 1996.
[26] A. Aldroubi and A. Sekmen, "Nearness to local subspace algorithm for subspace and motion segmentation," *arXiv.org*, 2010. [Online]. Available: http://arxiv.org/abs/1010.2198