21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Subspace Segmentation with A Minimal Squared Frobenius Norm Representation

Siming Wei
College of Computer Science
Zhejiang University
tobiawsm@gmail.com

Yizhou Yu
Department of Computer Science
The University of Hong Kong
yizhouy@acm.org

Abstract

We introduce a novel subspace segmentation method called Minimal Squared Frobenius Norm Representation (MSFNR). MSFNR performs data clustering by solving a convex optimization problem. We theoretically prove that in the noiseless case, MSFNR is equivalent to the classical Factorization approach and always classifies data correctly. In the noisy case, we show that on both synthetic and real-word datasets, MSFNR is much faster than most state-of-the-art methods while achieving comparable segmentation accuracy.

1 Introduction

Many computer vision and machine learning problems, such as structure from motion, face recognition and text classification, often assume that the data is drawn from a union of multiple linear subspaces. Exploiting subspaces is often crucial to the succuss of these applications. Therefore, subspace segmentation has attracted considerable attention in recent years.

1.1 Related Work

Existing literature on subspace segmentation can be roughly grouped into four classes: statistical learning based methods ([7, 11]) employ the mixture-of-Gaussians model. In this group, data is recognized as samples from a mixture of Gaussian distributions. The robustness of statistical methods is guaranteed while their computational cost is usually the bottleneck. Factorization based methods ([1, 3, 10]) compute a representation matrix for the data. In noiseless scenarios, such a matrix directly gives rise to the correct segmentation. However, this group of algorithms are too sensitive to noise. Algebraic methods [8] model the subspaces as polynomials. Fitting such polynomials to the data gives rise to the underlying subspace structure. The shortcoming of this group of methods is that it is computationally too expensive when the data dimension or the number of subspaces is relatively large. Sparsity based methods (e.g, SR [2] and LRR [6]) employ convex optimization to robustly handle noise while increasing the sparsity of the representation matrix. Solving the problem is equivalent to computing the representation matrix while taking the noise into consideration. The final segmentation is often accomplished with spectral clustering methods. Sparsity based methods are usually robust. As a representative of methods in this category, LRR [6] has reported the best segmentation accuracy on the Hopkins155 benchmark [9]. However, LRR involves nuclear norm[1] minimization. Solving LRR requires computing multiple SVD's with $O(n^3)$ complexity each. It gives rise to unbearable computational cost when the data set is large.

1.2 Our Contributions

In this work, we propose a new method called Minimal Squared Frobenius Norm Representation (MSFNR). It employs convex optimization to perform subspace clustering. The method minimizes the sum of the squared Frobenius norm of the representation matrix and the (2,1)-norm of a noise term. Our main contributions are summarized as follows:

1. We prove that in the noiseless case, MSFNR has a unique solution which is exactly the shape

[1]The nuclear norm of a matrix is the sum of the singular values of the matrix.

978-4-9906441-0-9

interaction matrix. This indicates the essential equivalence between MSFNR and the Factorization method for the ideal case.

2. The computational complexity of MSFNR is $O(n^2)$ while the complexity of LRR is $O(n^3)$. Nevertheless, the accuracy of MSFNR is highly competitive in comparison to any state-of-the-art methods including LRR.

3. As a noise removal algorithm, MSFNR is also able to accurately recover the original data.

The remainder of this paper is organized as follows. Section 2 studies the relationship between MSFNR and the classical Factorization method. Section 3 solves the convex optimization problem defined by MSFNR to demonstrate its efficiency. Experimental results are shown in Section 4. Conclusions and thoughts are discussed in the final section. And the appendix has the proof of the key theorem proposed in Section 2.

2 The MSFNR Model

2.1 Problem Definition

Let X be a collection of m dimensional data vectors drawn from a union of k linear subspaces $\{\mathcal{S}_i\}_{i=1}^k$. Let $\{r_i\}_{i=1}^k$ be the dimension of these subspaces respectively. The task of subspace segmentation (or clustering) is to cluster the vectors in X so that the vectors inside the same cluster belong to the same subspace.

Without loss of generality, we assume $X = [X_1, X_2, ..., X_k]$, where all vectors in X_i belong to $\mathcal{S}_i$. Note that here each X_i is a submatrix, not a vector. Denote the number of vectors in X_i as d_i. There must be at least one block diagonal matrix $Z^* = diag\{Z_1^*, Z_2^*, ..., Z_k^*\}$ satisfying $X = XZ$, where the size of the i-th block, Z_i, is d_i. Since Z actually has multiple solutions, any one in the solution set is called a representation matrix. Thus, the clustering task is actually equivalent to computing a block diagonal representation matrix Z^*.

In the ideal case, we assume the data set is noiseless and the subspaces are mutually independent, i.e. $\sum_{i=1}^k \mathcal{S}_i = \oplus_{i=1}^k \mathcal{S}_i$. Under such assumptions, the factorization based method correctly obtains the shape interaction matrix (SIM) as the solution of the block diagonal representation matrix [1]. The Shape Interaction Matrix of X is defined as $SIM(X) = V_r V_r'$, where $X = U_r S_r V_r'$ is the skinny singular value decomposition of X.

Algorithm 1 MSFNR Solver

Input: data matrix X, parameter λ.
Initialize : Z,J,E,Y_1,Y_2,μ,$\max_\mu$,ρ,ϵ
while not converged **do**
1. Update J by: $J = (Z\mu + Y_2)/(2 + \mu)$.
2. Update Z by: $Z = (I + X^t X)^{-1}(X^t X - X^t E + J + (X^t Y_1 - Y_2)/\mu)$.
3. Update E by solving: $\arg\min \frac{\lambda}{\mu}\|E\|_{2,1} + \frac{1}{2}\|E - (X - XZ + Y_1/\mu)\|_F^2$. (Lemma 3.3 in [6])
4. Update Y_1, Y_2 by: $Y_1 = Y_1 + \mu(X - XZ - E)$ and $Y_2 = Y_2 + \mu(Z - J)$.
5. Update μ by: $\mu = min(\rho\mu, \max_\mu)$.
6. Check $\|X - XZ - E\|_\infty < \epsilon$ and $\|Z - J\| < \epsilon$
end while .

2.2 A Convex Formulation

Real-word datasets are often heavily polluted by noise. Because it is very sensitive to noise, the Factorization method loses its power in this case. Nevertheless, we have the following observation:

Theorem 2.1. *The shape interaction matrix SIM*(X) *is the unique solution of the optimization problem:* $\min_Y \ |Y|_F^2 \ \ s.t. \ \ X = XY$*. Thus the optimal solution of this problem equals to rank*(X).

Proof. See Appendix. □

Motivated by the above theorem, we formulate a convex optimization problem to take noise into consideration. By adding a noise term, we propose the minimal squared Frobenius norm representation:

$$\min_{Z,E} ||Z||_F^2 + \lambda||E||_{2,1} \quad s.t. \ X = XZ + E, \quad (1)$$

where E denotes the noise term. After solving (1), it is straightforward to apply spectral clustering to Z to obtain the final segmentation.

3 A Solver for MSFNR

MSFNR gives rise to a convex optimization problem. By adding a constraint $Z = J$ to (1), we can solve it by inexact ALM [5]. The pseudo-code is presented in Algorithm 1.

Recall the LRR model [6], which solves:

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_{2,1}, \quad s.t. \ X = XZ + E. \quad (2)$$

It is straightforward to see that the complexity of Algorithm 1 is $O(n^2)$. So far as we know, nuclear norm minimization problems, including LRR, can hardly avoid the computation of multiple SVD's each of which has $O(n^3)$ cost. Therefore, nuclear norm minimizations are not scalable to large datasets. Thus compared with existing sparsity based methods such as LRR, MSFNR has a natural advantage on computational cost.

4 Experimental Results

In this section, we verify the performance of MSFNR on both synthetic and real-word datasets. All the experiments were run on an Intel Pentium Dual core 2.0GHz processor.

Synthetic Data

We first compared the robustness and speed of MSFNR and LRR on synthetic data. We constructed 5 independent linear subspaces with dimensionality equal to 4. Twenty 100-dimensional data vectors were sampled from each subspace. We randomly chose data vectors, and added noise with zero mean and variance $0.3\|x\|$ to them. The symbol $\|\cdot\|$ denotes the 2-norm of a data vector. The optimal parameters of each method were used during the comparison. We repeated the experiment 50 times for every percentage of corruption and then recorded the average accuracy. As shown in Figure 1, MSFNR and LRR achieve a comparable segmentation accuracy.

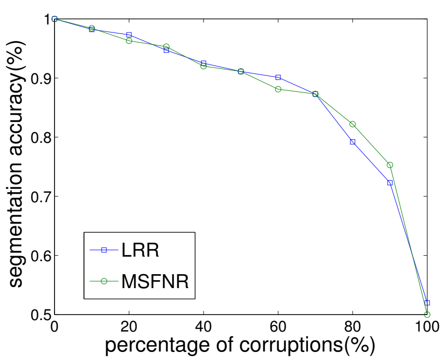

Figure 1. Segmentation accuracy of LRR and MSFNR. Parameter setting: $\lambda_{LRR} = 0.12$ **and** $\lambda_{MSFNR} = 0.37$**.**

We also compared the speed of MSFNR with that of LRR. In this experiment, the data size ranged from 50 to 500. We generated data in the same way as the last experiment. The percentage of polluted vectors is set to a random number. For each data size, we ran both LRR and MSFNR 10 times, and recorded the average computation time. According to Figure 2, it is obvious that MSFNR is much faster than LRR, especially for large data sizes. In particular, MSFNR is almost an order of magnitude faster for the largest data size in our experiments.

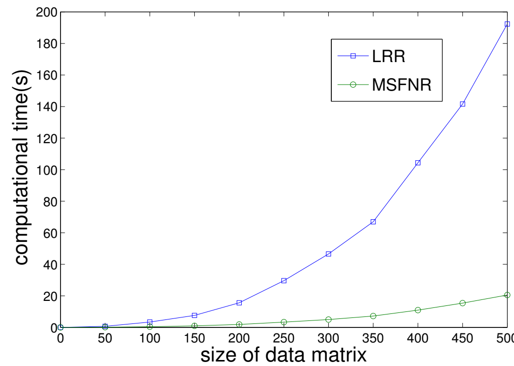

Figure 2. Computational speed of LRR and MSFNR for different data sizes.

Real-World Data

We also tested the performance of MSFNR on the Hopkins155 motion database [9]. The Hopkins155 database has 156 video sequences and each of them is a separate clustering task. Within each sequence, there are $39 \sim 550$ data vectors belonging to two or three motions, and each motion corresponds to a subspace. Table 1 shows the segmentation error rate of MSFNR in comparison to state-of-the-art methods. On this motion dataset, MSFNR is again stable and competitive with respect to LRR.

METHODS	GPCA	SR	LRR	MSFNR
MEAN	30.41	3.44	3.21	3.11
MEDIAN	32.85	0.35	0.41	0.61
STD	11.71	7.55	5.55	4.81

Table 1. Segmentation error rates ($\%$**) on the Hopkin155 database.**

We finally tested the Extended Yale Database B [4]. This database consists of 640 frontal face images of 10 subjects. Each subject has about 64 images. More than half of the face images are polluted with shadows or specular lights. We resized the images to 48×42 pixels. The purpose of this experiment is to show the robustness of MSFNR on heavily corrupted data. Some results on this database are shown in Figure 3. Note that the result on the first image is a very impressive example.

Figure 3. Results of MSFNR on Extended Yale Database B. For each of the eight pairs of images, the left one is the original, and the right one is the result after noise removal. λ is always set to 1.0 here.

5 Conclusions and Future Work

In this paper, we have developed a solution for the subspace segmentation problem with a minimal squared Frobenius norm representation. The equivalence between MSFNR and the Factorization method for the noiseless case guarantees its performance. The complexity of our solution is only $O(n^2)$ and thus is much less expensive than the state-of-the-art method called LRR. In addition to being highly accurate and efficient, MSFNR is also able to successfully perform noise removal on the input data.

We notice that the mechanism to solve Algorithm 1 can naturally be made online. A corresponding online solver may be more flexible and efficient. This potential extension of our method will be studied in future.

References

[1] J. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, 1998.

[2] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2790–2797, 2009.

[3] A. Gruber and Y. Weiss. Multibody factorization with uncertainty and missing data using the EM algorithm. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 707–714, 2004.

[4] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[5] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrix. *submitted to Mathematical Programming*, 2009.

[6] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *International Conference of Machine Learning*, 2010.

[7] Y. Ma, H. Derksen, W. Hong, and J. Wright. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1546–1562, 2007.

[8] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, 2008.

[9] R. Tron and R. Vidal. A benchmark for the comparison of 3-D motion segmentation algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[10] R. Vidal, R. Tron, and R. Hartley. Multiframe motion segmentation with missing data using powerfactorization and GPCA. *International Journal of Computer Vision*, 79(1):85–105, 2008.

[11] A. Yang, S. Rao, and Y. Ma. Robust statistical estimation and segmentation of multiple subspaces. In *Computer Vision and Pattern Recognition Workshop on 25 Years on RANSAC*, pages 99–107, 2006.

Appendix: Proof of Theorem 2.1

Proof. Let $[U_x, S_x, V_x]$ and $[U_y, S_y, V_y]$ be the full SVD of X and Y respectively. Denote $M = V_x' V_y$ and $N = V_x' U_y$. Then $X = XY$ is equivalent to $S_x M = S_x N S_y$. Suppose X is of rank r. We have $M_r = N_r S_y$, where M_r and N_r are matrices formed by the first r rows of M and N respectively.

For the optimization problem in (2.1), we have $r = |N_r S_y M_r'|_F^2 \leq |N S_y M'|_F^2 = |Y|_F^2$. Note that $SIM(X)$ belongs to the optimal set. (5 becomes an equation only if the rank of Y is r. Suppose Y_0 is a solution. According to the proof of Theorem 3.1 in [6], the rank of Y_0 must be r. Thus $M_r = N_r S_y$ implies that the first r columns of M_r form a $r \times r$ orthonormal matrix. If the smallest nonzero singular value of Y_0 is less than 1, the norm of N_r's r-th column must be greater than 1, which is a contradiction. This is to say Y_0's smallest nonzero singular value should not be less than 1. Then all of Y_0's nonzero singular values are equal to 1. For both N and M, the first $r \times r$ block forms an orthonormal matrix while the remaining elements of those r columns and r rows must be zero. Then $Y_0 = (V_x N) S_y (V_x M)' = V_r V_r' = \mathrm{SIM}(X)$. $\square$