From subspace clustering to full-rank matrix completion

Emmanuel Candes, Lester Mackey, and Mahdi Soltanolkotabi
Department of Statistics, Stanford University
{candes, lmackey, mahdisol}@stanford.edu

Abstract

Subspace clustering is the problem of finding a multi-subspace representation that best fits a collection of points taken from a high-dimensional space. This type of structure occurs naturally in many applications ranging from bioinformatics, image/text clustering to semi-supervised learning. The companion paper [3] shows that robust and tractable subspace clustering is possible with minimal requirements on the orientation of the subspaces and number of samples per subspace.

This note summarizes a forthcoming work [1] on subspace clustering when some of the entries in the data matrix are missing. This problem may also be viewed as a generalization of standard low-rank matrix completion to cases where the matrix is of high or potentially full-rank. Synthetic and real data experiments confirm the effectiveness of these methods.

1 Problem formulation and model

Consider a real-valued $n \times N$ dimensional matrix $\boldsymbol{X}$. We assume that the columns of $\boldsymbol{X}$ lie in a union of L unknown linear subspaces, of unknown dimensions. A small subset of the entries of such a matrix is revealed. The goal is two fold: 1) partition the columns into different clusters based on subspace of origin and approximate the underlying subspaces. 2) inpute the missing entries. Throughout we assume that the each entry of $\boldsymbol{X}$ is observed with probability $1 - \delta$.

2 Method

Here we explain our method for subspace clustering with missing data. Upon finding the correct clustering, one can apply any one of the low-rank matrix recovery algorithms on each cluster to complete the missing entries. To introduce our method, we first study the problem when all entries are revealed.

2.1 No missing entries

Most spectral clustering algorithms follow a two-step procedure: I) Construct a weighted graph $\boldsymbol{W}$ that captures the similarity between any pair of points, II) Select clusters by applying spectral clustering techniques to $\boldsymbol{W}$.

Following [2,3,6] we build the affinity graph in Step I by finding the sparsest expansion of each column $\boldsymbol{x}^{(i)}$ of $\boldsymbol{X}$ as a linear combination every other column. Under some generic conditions, one expects that the sparsest representation of $\boldsymbol{x}^{(i)}$ would only select vectors from the subspace in which $\boldsymbol{x}^{(i)}$ happens to lie in. This leads to the following sequence of optimization problems

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^N} \ \|\boldsymbol{\beta}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{x}^{(i)} \text{ and } \beta_i = 0. \tag{2.1}$$

One then collects the outcome of these N optimization problems as columns of a matrix $\boldsymbol{B}$ and then sets the weighted graph to $\boldsymbol{W} = |\boldsymbol{B}| + \left|\boldsymbol{B}^T\right|$.

2.2 Bias-corrected Dantzig selector

We will use $\Omega_i \subset \{1, 2, \ldots, n\}$ to denote the observations we have from the i-th column of $\boldsymbol{X}$ and $\boldsymbol{X}_{\Omega_i}$ to denote the submatrix of $\boldsymbol{X}$ with rows selected by Ω_i. We build a matrix $\boldsymbol{Y}$ based on the observed entries of $\boldsymbol{X}$ as follows

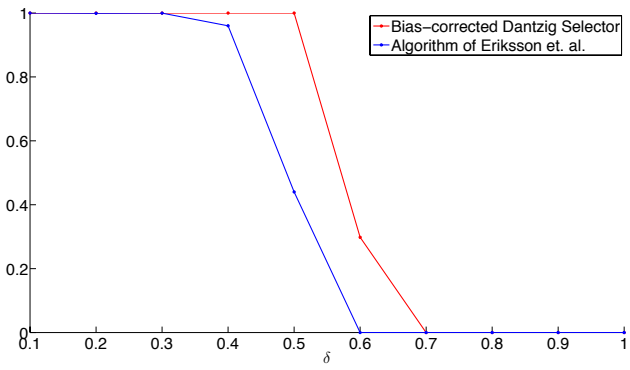

Figure 1: The fraction of of correctly completed columns (with a tolerance of 10^{-5}), versus the fraction of missing entries δ for the bias-corrected Dantzig Selector and the algorithm suggested in [4].

$$Y_{ij} = \begin{cases} \frac{X_{ij}}{(1-\delta)} & \text{if observed} \\ 0 & \text{if missing} \end{cases}. \tag{2.2}$$

Set $\widehat{\mathbf{\Gamma}}_i$ equal to $\boldsymbol{Y}_{\Omega_i}^T \boldsymbol{Y}_{\Omega_i} - \delta \text{diag}(\boldsymbol{Y}_{\Omega_i}^T \boldsymbol{Y}_{\Omega_i})$ with the ith row and column set to zero. Similarly, set $\widehat{\boldsymbol{\gamma}}_i$ equal to $\boldsymbol{Y}_{\Omega_i}^T \boldsymbol{y}_{\Omega_i}^{(i)}$ with the ith row set to zero. A simple calculation shows that when the observed entries are revealed at random $(\widehat{\mathbf{\Gamma}}_i, \widehat{\boldsymbol{\gamma}}_i)$ is an unbiased estimator for $(\boldsymbol{X}_{\Omega_i}^T \boldsymbol{X}_{\Omega_i}, \boldsymbol{X}_{\Omega_i}^T \boldsymbol{x}_{\Omega_i})$. This motivates the following bias-corrected Dantzig selector

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^N} \|\boldsymbol{\beta}\|_{\ell_1} \quad \text{subject to} \quad \left\|\widehat{\boldsymbol{\gamma}}_i - \widehat{\mathbf{\Gamma}}_i \boldsymbol{\beta}\right\|_{\ell_\infty} \leq \lambda \text{ and } \beta_i = 0. \tag{2.3}$$

3 Numerical experiments

Due to lack of space here we present a single synthetic experiment. Real experiments on cancer data will be presented in the accompanying poster/talk. We pick $L = 10$ subspace of dimension $d = 5$ uniformly at random in $\mathbb{R}^{n=100}$. For each subspace, we generate 500 points drawn from a $\mathcal{N}(0, \boldsymbol{U}\boldsymbol{U}^T)$ distribution, where $\boldsymbol{U} \in \mathbb{R}^{n \times d}$ is an orthonormal basis for that subspace. We note that such a matrix has rank 50 and is half-way to being full rank. For the clustering step we used the bias corrected Dantzig selector with the choice of $\lambda = \frac{\sqrt{2 \log N}}{\sqrt{n}} \sqrt{\frac{\delta}{1-\delta}}$ based on theoretical insights in [1]. After identifying the clusters we used OPTSPACE [5] to complete the matrix associated with each cluster. We ran 50 independent trials of our procedure and compared it to the procedure reported in [4]. The results are summarized in Figure 1. This figure indicates that the bias-corrected Dantzig selector can handle a much higher fraction of missing entries. Indeed, while the performance of the procedure in begins to break down at 30% missingness, the bias-corrected Dantzig selector exactly recovers all matrices with up to 50% missingness. We note that in our procedure we did not have to tune any parameters, while the algorithm of [4] requires tunning of 5 different parameters (please see Algorithm 1 in [4] for further details).

References

[1] E. J. Candés, L. Mackey, and M. Soltanolkotabi. From robust subspace clustering to full-rank matrix completion. *In preparation*, 2013.

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

[3] B. Eriksson, L. Balzano, and R. Nowak. High-rank matrix completion and subspace clustering with missing data. *Arxiv preprint arXiv:1112.5629*, 2011.

[4] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *Information Theory, IEEE Transactions on*, 56(6):2980–2998, 2010.

[5] M. Soltanolkotabi and E. J. Candès. A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4):2195–2238, 2012.

[6] M. Soltanolkotabi, E. Elhamifar, and E. J. Candès. Robust subspace clustering. *arXiv:1301.2603v2*, Jan. 2013.