ON THE SAMPLE COMPLEXITY OF SUBSPACE CLUSTERING WITH MISSING DATA

D. Pimentel, R. Nowak

University of Wisconsin
Electrical and Computer Engineering
Madison, WI, 53706, USA

L. Balzano

University of Michigan
Electrical Engineering and Computer Science
Ann Arbor, MI, 48109, USA

ABSTRACT

Subspace clustering is a useful tool for analyzing large complex data, but in many relevant applications missing data are common. Existing theoretical analysis of this problem shows that subspace clustering from incomplete data is possible, but that analysis requires the number of samples (i.e., partially observed vectors) to be *super-polynomial in the dimension d*. Such huge sample sizes are unnecessary when no data are missing and uncommon in applications. There are two main contributions in this paper. First, it is shown that if subspaces have rank at most r and the number of partially observed vectors greater than d^{r+1} (times a poly-logarithmic factor), then with high probability the true subspaces are the only subspaces that agree with the observed data. We may conclude that subspace clustering may be possible without impractically large sample sizes and that we can certify the output of *any* subspace clustering algorithm by checking its fit to the observed data. The second main contribution is a novel EM-type algorithm for subspace clustering with missing data. We demonstrate and compare it to several other algorithms. Experiments with simulated and real data show that such algorithm works well in practice.

***Index Terms*—** Matrix Completion, Subspace Clustering

1. INTRODUCTION

Let $\mathbf{X}$ be a $d \times N$ data matrix whose columns lie in the union of several unknown low-dimensional subspaces of $\mathbb{R}^d$. The goal of subspace clustering is to infer the underlying subspaces from $\mathbf{X}$ and to cluster the columns of $\mathbf{X}$ according to the subspaces. In this paper, we suppose that $\mathbf{X}$ is partially observed with entries missing at random and aim at the same goal.

This problem arises in applications ranging from computer vision to network inference. Existing theoretical analysis of this problem shows that subspace clustering from incomplete data is possible, provided the number of samples N is super-polynomial in the dimension of the subspaces [1]. In practice, it is rare to have such huge numbers of samples. Several heuristic algorithms have been proposed for subspace clustering with missing data, see [2, 3] for two examples. These methods sometimes work reasonably well in practice, but lack theoretical justification. Therefore, determining the minimum number of samples necessary for subspace clustering with missing data is an important open question.

This motivates our main theoretical contribution. We assume the observed data are generated from K generic subspaces (in some sense a weaker condition than the usual incoherence requirements), each of dimension $\leq r$ in $\mathbb{R}^d$, $d > r$. We show that if we observe at least order $d^{r+1}(\log d/r + logK)$ columns and at least order $r \log^2 d$ entries in each column, then identification of the subspaces is possible with large probability. Note, that the total number of columns needed is only polynomial in d, in contrast to the super-polynomial requirement of $d^{\log d}$ of the best previously existing bounds [1]. The second main contribution of the paper is a novel EM algorithm for subspace clustering with missing data. Experiments with real and synthetic data show that the EM algorithm performs better than existing methods, and we compare the performance of the algorithm with our theoretical bounds.

2. KEY ASSUMPTIONS AND MAIN RESULTS

We give a high-level proof of our main result to give an idea of our approach. For a fully detailed proof see [4].

Definition 1. *Denote the set of $d \times N$ matrices of rank r by $\mathcal{M}(r, d \times N)$. A* generic *$(d \times n)$-matrix of rank r is a continuous $\mathcal{M}(r, d, n)$-valued random variable. We say a subspace S is* generic *if a matrix whose columns are drawn i.i.d. according to a non-atomic distribution with support on S is generic a.s.*

A1. The columns of our $d \times N$ data matrix $\mathbf{X}$ are drawn according to a non-atomic distribution with support on the union of at most K generic subspaces. The subspaces, denoted by $\mathcal{S} = \{S_k\}$, each has rank exactly $r < d$.

A2. The probability that a column is drawn from subspace k is ρ_k. Let ρ_* be a bound on $\min_k\{\rho_k\}$.

A3. We observe $\mathbf{X}$ only on a set of entries Ω and denote the observation $\mathbf{X}_\Omega$. Each entry in $\mathbf{X}_\Omega$ is sampled independently with probability p.

Throughout the paper we use $\mathbf{X}_\Omega$ to refer indistinctly to the matrix $\mathbf{X}_\Omega$ as well as the set of columns of the matrix $\mathbf{X}_\Omega$. We split $\mathbf{X}_\Omega$ in two sets: the *search* set, $\widetilde{\mathbf{X}}_\Omega$, with $\widetilde{N} := |\widetilde{\mathbf{X}}_\Omega|$

columns, and the *test* set, $\bar{\mathbf{X}}_\Omega$, with $\bar{N} := |\bar{\mathbf{X}}_\Omega|$ columns, s.t. $N := |\mathbf{X}_\Omega| = \widetilde{N} + \bar{N}$. We use $\widetilde{\mathbf{X}}_\Omega^{[k]}$ to denote the columns of $\widetilde{\mathbf{X}}_\Omega$ corresponding to the k^{th} subspace, and equivalently for $\bar{\mathbf{X}}_\Omega^{[k]}$.

We now present our main result, which we prove in Section 3. The theorem below shows that if we observe at least order $d^{r+1}(\log d/r + \log K)$ columns and at least order $r\ \log^2 d$ entries per column, then identification of the subspaces is possible with large probability. This result can be easily generalized to a relaxed version of assumption **A1** to the case where the dimensions of the subspaces are upper bounded by r.

Theorem 1. *Suppose* **A1-A3** *hold. Let* $\epsilon > 0$ *be given. Assume the number of subspaces* $K \leq \frac{\epsilon}{6} e^{d/4}$*, the total number of columns* $N = \widetilde{N} + \bar{N} \geq (2d + 4M)/\rho_*$*, and*

$$p \geq \frac{1}{d} 128 \mu_1^2 r \beta_0 \log^2(2d),$$

$$\beta_0 = \sqrt{1 + \frac{\log\left(\frac{6K}{\epsilon} 12 \log(d)\right)}{2\log(2d)}},$$

$$M = \left(\frac{de}{r+1}\right)^{r+1} \left((r+1)\log\left(\frac{de}{r+1}\right) + \log\left(\frac{8K}{\epsilon}\right)\right),$$

where $\mu_1^2 := \max_k \frac{d^2}{r} \|U_k V_k^*\|_\infty^2$ *and* $U_k \Sigma_k V_k^*$ *is the singular value decomposition of* $\widetilde{\mathbf{X}}^{[k]}$*. Then with probability at least* $1 - \epsilon$*,* $\mathcal{S}$ *can be uniquely determined from* $\mathbf{X}_\Omega$*.*

3. PROOFS OF MAIN RESULTS

First we make some notational remarks. When we write $x_\omega \in \mathbf{X}_\Omega$, we are referring to one column in the set of columns $\mathbf{X}_\Omega$. When we write $x_\omega \in S_k$ or $\mathbf{X}_\Omega \subset S_k$ we mean that S_k fits the column x_ω in the sense that there exists a completion $\hat{x}$ of x_ω such that $\hat{x} \in S_k$. We also introduce the definition of a *validating set*:

Definition 2. *(Validating set) Consider a collection of columns* $\{x_{i_{\omega_i}}\}_{i=1}^m$*. Consider a graph* $\mathcal{G}$ *with m nodes representing these m columns, where edge* (i, j) *exists if* $|\omega_i \cap \omega_j| > r$*. We say* $\{x_{i_{\omega_i}}\}_{i=1}^m$ *is a validating set if* $\mathcal{G}$ *is connected and* $\bigcup_{i=1}^m \omega_i = \{1, ..., d\}$*.*

We now describe the intuition behind our approach. We consider an exhaustive search over every set of d columns in the search set. We use $\widetilde{\mathbf{X}}_\Omega^{(\ell)}$ to denote the ℓ^{th} combination of d columns of $\widetilde{\mathbf{X}}_\Omega$, where ℓ ranges from 1 to $\binom{\widetilde{N}}{d}$. For each of these combinations, if there is a subspace that uniquely fits all d columns, we validate it by finding a validating set in the test set that fits the subspace. We use $\hat{\mathcal{S}}$ to denote the collection of all subspaces satisfying these conditions.

The core of the main result lies in Lemmas 5 and 7 below. By Lemma 5, every combination of d columns from a single subspace will have a unique completion and fit a validating subspace with high probability. By Lemma 7 only true subspaces can fit a validating set. Putting together these two results, we get that with high probability $\hat{\mathcal{S}} = \mathcal{S}$. That is, we can identify $\mathcal{S}$ from $\mathbf{X}_\Omega$.

We begin the proof of our main theorem using Theorems 2 of [5] and 2.6 of [6] with some adjustments to our context. We state our versions here as Lemmas 1 and 2.

Lemma 1 (Low-Rank Matrix Completion [5])**.** *Consider a* $d \times d$ *rank-r matrix* $\mathbf{Y}$*. Let the row and column spaces of* $\mathbf{Y}$ *have coherences (as in Definition 1 of [5]) bounded above by some positive* μ_0*. Suppose that every entry of* $\mathbf{Y}$ *has been observed independently with probability p to yield* Y_Ω*, with*

$$p \geq \frac{1}{d} 128 \max\{\mu_1^2, \mu_0\} r \beta_0 \log^2(2d),$$

β_0 *and* μ_1 *as in Theorem 1. Then* $\mathbf{Y}^*$*, the minimizer to the nuclear norm minimization problem (Equation 2 of [5]) is unique and equal to* $\mathbf{Y}$ *with probability at least* $1 - \frac{\epsilon}{3K}$*.*

Lemma 2 (Completion Identifiability [6], Thm 2.6)**.** *Let* Ω *be given. Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two different generic rank-r matrices. Then* $\mathbf{X}_\Omega$ *is completable (i.e.* $\mathbf{X}$ *can be recovered from* $\mathbf{X}_\Omega$*) if and only if* $\mathbf{Y}_\Omega$ *is completable.*

These two lemmas are used to prove Lemma 3, which is a version of Lemma 1 that gives us a probability of low-rank completion of generic matrices.

Lemma 3 (Generic Low-Rank Matrix Completion)**.** *Consider a* $d \times d$ *generic matrix* $\mathbf{X}$*. Suppose that every entry of* $\mathbf{X}$ *has been observed independently with probability p to yield* $\mathbf{X}_\Omega$*, with p and* β_0 *as in Theorem 1. Then* $\mathbf{X}$ *can be recovered with probability at least* $1 - \frac{\epsilon}{3K}$*.*

Proof. In Lemma 1, μ_0 and μ_1 satisfy $1 \leq \mu_0 \leq d/r$ and $\mu_1 \geq 1$. So we can take a generic matrix $\mathbf{Y}$ that satisfies all assumptions of Lemma 1 with $\mu_0 = 1$. Ω satisfies the sample assumptions of Lemma 1 with p as in Theorem 1, so $\mathbf{Y}_\Omega$ satisfies all assumptions of Lemma 1. Then with probability at least $1 - \frac{\epsilon}{3K}$, $\mathbf{Y}_\Omega$ is uniquely completable, and so is $\mathbf{X}_\Omega$ by Lemma 2, as both $\mathbf{X}$ and $\mathbf{Y}$ are generic. □

Now that we can apply low-rank matrix completion results to generic matrices, with enough columns from each subspace our exhaustive approach will find at least one collection of columns where matrix completion will succeed. Next we bound the probability of having a validating set in $\bar{\mathbf{X}}_\Omega^{[k]}$.

Lemma 4. *Assume* $|\bar{\mathbf{X}}_\Omega^{[k]}| \geq 2M$*, with M as in Theorem 1. Then with probability at least* $1 - \frac{\epsilon}{4K}$*,* $\bar{\mathbf{X}}_\Omega^{[k]}$ *has at least one validation set.*

Proof. Consider $\mathbf{Y}_\Omega$ with $|\mathbf{Y}_\Omega| = M$, and whose columns all have exactly $r + 1$ entries sampled uniformly and independently at random. It can be shown with the Chernoff bound that the probability that $\bar{\mathbf{X}}_\Omega^{[k]}$ contains a validating set is larger than the probability that $\mathbf{Y}_\Omega$ does. The latter is larger than the probability that the columns of $\mathbf{Y}_\Omega$ have all the different $\binom{d}{r+1}$ observation sets, which implies it contains a validating set. By the well-known Coupon Collector problem, $\mathbf{Y}_\Omega$ has all the different observation sets w.h.p, giving the lemma. □

Lemma 5 puts together Lemmas 3, 4, and the multiplicative Chernoff bound to show that with high probability the true subspaces will be contained in $\hat{\mathcal{S}}$.

Lemma 5 (True Positive)**.** *With the same assumptions as in Theorem 1, $S_k \in \hat{\mathcal{S}}$ with probability at least $1 - \frac{\epsilon}{K}$ for every k.*

The last Lemmas before the proof of Theorem 1 show that no subspace other than those in $\mathcal{S}$ will be contained in $\hat{\mathcal{S}}$.

Lemma 6. *Let $x_{\omega_x}, y_{\omega_y} \in \bar{\mathbf{X}}_\Omega$, with $|\omega_x \cap \omega_y| > r$. Suppose $\widetilde{S}$ fits x_{ω_x} and y_{ω_y}. Then x_{ω_x} and y_{ω_y} belong to the same subspace, say S_k, a.s. Furthermore, letting $\omega = \omega_x \cup \omega_y$, $\widetilde{S}_\omega = S_{k_\omega}$ a.s.*

Lemma 7. *If $\widetilde{S}$ fits a validating set, $\widetilde{S} = S_k$ for some k a.s.*

Lemma 6 follows by genericity of our subspaces and column vectors. Lemma 7 follows by induction on Lemma 6, and the definition of a validating set. We now present the proof of our main result, Theorem 1.

Proof. (**Theorem 1**) It suffices to show that w.h.p. $\hat{\mathcal{S}} = \mathcal{S}$. By Lemma 5, $\mathsf{P}(S_k \notin \hat{\mathcal{S}}) \leq \frac{\epsilon}{K}\ \forall\ k$. Notice that $\mathsf{P}(S_\ell \in \hat{\mathcal{S}} \backslash \mathcal{S})$ is equivalent to the probability that S_ℓ fits a validating set given that $S_\ell \neq S_k\ \forall\ k$, and this probability is zero by Lemma 7. A union bound over these probabilities yields the Theorem. □

4. EM ALGORITHM

The problem of subspace clustering with missing data can be posed as fitting a mixture of Gaussians with low-rank covariances to incomplete data. This naturally suggests considering extensions of the EM algorithm in [7] to handle missing data, or alternatively, a generalization of the EM algorithm in [8] to low-rank covariance matrices. We propose the following EM algorithm for this task, based largely on [7]. To begin we assume the data are contaminated with Gaussian noise.

Consider the usual Gaussian mixture framework, and split every $x_i \in \mathbf{X}_\Omega$ into its observed and missing parts:

$$\begin{bmatrix} x_i^o \\ x_i^m \end{bmatrix} = \sum_{k=1}^{K} \mathbb{1}_{\{z_i=k\}} \left(\begin{bmatrix} W_k^{o_i} \\ W_k^{m_i} \end{bmatrix} y_i + \begin{bmatrix} \mu_k^{o_i} \\ \mu_k^{m_i} \end{bmatrix} + \eta_i \right),$$

where $\{1, ..., K\} \ni z_i \overset{iid}{\sim} \rho \perp y_i \overset{iid}{\sim} \mathcal{N}(0, I)$, W_k is a $d \times r$ matrix whose span is S_k, and $\eta_i | z_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_{z_i}^2 I)$ models the noise in the z_i^{th} subspace. We are interested on the Maximum Likelihood Estimate of $\theta = \{W, \mu, \rho, \sigma^2\}$, where $W := \{W_k\}_{k=1}^K$, $\mu := \{\mu_k\}_{k=1}^K$, ρ and $\sigma^2 := \{\sigma_k^2\}_{k=1}^K$.

Let $\mathbf{X}^o := \{x_i^o\}_{i=1}^N$, $\mathbf{X}^m := \{x_i^m\}_{i=1}^N$, $\mathbf{Y} := \{y_i\}_{i=1}^N$, $\mathbf{Z} := \{z_i\}_{i=1}^N$ s.t. $\mathbf{X}^o$ is our data, θ is the parameter of interest, and $\mathbf{X}^m$, $\mathbf{Y}$ and $\mathbf{Z}$ are the *hidden* variables in the EM algorithm. The iterates of the algorithm are computed as follows, where $\mathsf{E}_k \langle \cdot \rangle$ denotes $\mathsf{E}_{\cdot | x_i^o, z_i = k, \hat{\theta}}[\cdot]$.

$$\widetilde{W}_k = \left[\sum_{i=1}^{N} \mathsf{p}_{i,k} \mathsf{E}_k \left\langle x_i y_i^T \right\rangle - \frac{\left(\sum_{i=1}^N \mathsf{p}_{i,k} \mathsf{E}_k \langle x_i \rangle\right)\left(\sum_{i=1}^N \mathsf{p}_{i,k} \mathsf{E}_k \langle y_i \rangle^T\right)}{\sum_{i=1}^N \mathsf{p}_{i,k}} \right]$$

$$\left[\sum_{i=1}^{N} \mathsf{p}_{i,k} \mathsf{E}_k \left\langle y_i y_i^T \right\rangle - \frac{\left(\sum_{i=1}^N \mathsf{p}_{i,k} \mathsf{E}_k \langle y_i \rangle\right)\left(\sum_{i=1}^N \mathsf{p}_{i,k} \mathsf{E}_k \langle y_i \rangle^T\right)}{\sum_{i=1}^N \mathsf{p}_{i,k}} \right]^{-1},$$

$$\widetilde{\mu}_k = \frac{\sum_{i=1}^N \mathsf{p}_{i,k} \left(\mathsf{E}_k \langle x_i \rangle - \widetilde{W}_k \mathsf{E}_k \langle y_i \rangle \right)}{\sum_{i=1}^N \mathsf{p}_{i,k}},$$

$$\widetilde{\sigma}_k^2 = \frac{1}{d \sum_{i=1}^N \mathsf{p}_{i,k}} \left[\sum_{i=1}^{N} \mathsf{p}_{i,k} \Big(tr\left(\mathsf{E}_k \left\langle x_i x_i^T \right\rangle\right) - 2\widetilde{\mu}_k^T \mathsf{E}_k \langle x_i \rangle + \widetilde{\mu}_k^T \widetilde{\mu}_k \right.$$

$$\left. -2tr\left(\mathsf{E}_k \left\langle y_i x_i^T \right\rangle \widetilde{W}_k\right) + 2\widetilde{\mu}_k^T \widetilde{W}_k \mathsf{E}_k \langle y_i \rangle + tr\left(\mathsf{E}_k \left\langle y_i y_i^T \right\rangle \widetilde{W}_k^T \widetilde{W}_k\right) \Big) \right],$$

$$\widetilde{\rho}_k = \frac{1}{N} \sum_{i=1}^{N} \mathsf{p}_{i,k}\ , \qquad \mathsf{p}_{i,k} := \mathsf{P}_{z_i | x_i^o, \hat{\theta}}(k) = \frac{\hat{\rho}_k \mathsf{P}_{x_i^o | z_i = k, \hat{\theta}}(x_i^o)}{\sum_{j=1}^K \hat{\rho}_j \mathsf{P}_{x_i^o | z_i = j, \hat{\theta}}(x_i^o)}.$$

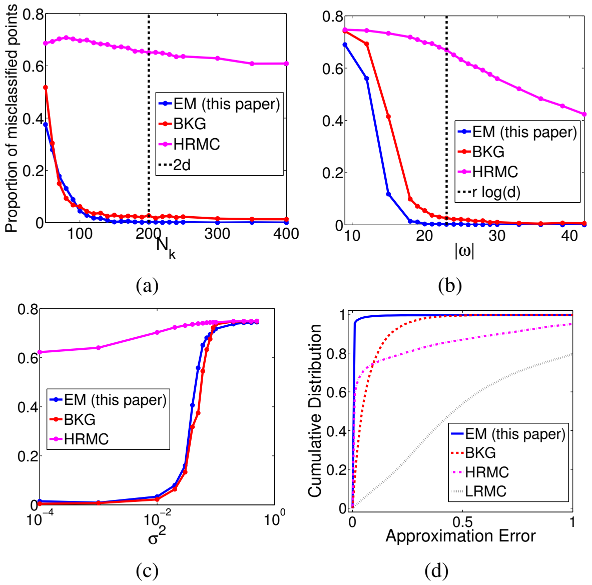
The computations of the expected means and covariances have the highest computational complexity in the noiseless and noisy case respectively, with $|\omega_i^c| r$ and $|\omega_i^c|^2 r$ operations per column per subspace per iteration. Since $|\omega_i^c|$ is close to and upper bounded by d, the computational complexity of the EM algorithm per iteration will be in the order of $NKdr$ and NKd^2r in the noiseless and noisy cases, respectively.

5. SIMULATIONS

The first experiment we present is a set of simulations of the EM setup above with $d = 100$, $K = 4$, $r = 5$. For each simulation we generated K subspaces and K initial estimates, each spanned by an orthonormal basis generated from r i.i.d. standard gaussian d-dimensional vectors and N_k columns from each subspace with $|\omega|$ observed entries each. We evaluated the performance of the EM algorithm derived before, batch k-GROUSE (BKG) [2] and the HRMC algorithm from [1]. We ran 450 independent trials of this experiment as a function of N_k, $|\omega|$ and σ^2. The results are summarized in Figure 1 (a)-(c).

For a second simulation, we consider an application in which unions of subspaces are indeed a good model for data. Distances in a network measured in number of hop counts between passive monitors and computers determine the network's topology. As measurements in such monitors are not controlled, not all distances can be observed. Fortunately, these distances lie in a union of K 2-dimensional subspaces with K being the number of subnets [9]. We simulated a network and measured hop counts based on shortest-path routing using a Heuristically Optimal Topology from with $d = 75$ passive monitors randomly located and $N_k = 225$ IP addresses from each of the $K = 12$ subnets. In Figure 1(d) we compare the results of the hop count matrix estimation from only 40% of the total hop counts using EM, BKG, HRMC and LRMC.

Finally, we tested our EM algorithm using real data from the Hopkins 155 Motion Segmentation Dataset. In each video of this dataset, a collection of points are identified over the



(a)

(b)

(c)

(d)

Fig. 1. (a) - (c): Proportion of misclassified points, (a) as a function of N_k with $|\omega| = \lceil r \log d \rceil = 24$ fixed; (b) as a function of $|\omega|$ with $N_k = 2.1d$ fixed; (c) as a function of σ^2 with $N_k = 3d$ and $|\omega| = \lceil r \log d \rceil$ fixed. (d) Cumulative Distribution of hop count estimation error for $K = 12$ subnets, $d = 75$ passive monitors and $N = 2700$ IP addresses from 40% of total observations.

frames. Each point belongs to an unknown cluster, e.g. a car, a person, background, etc., and the positions of these points are known to lie in a union of subspaces [10]. However, in real life it is unusual to be able to identify *every* point over *all* the frames of a video, due to occlusion, objects leaving the video window, objects rotation, miss detection, etc. Therefore missing data arises naturally. Table 1 shows a summary of EM's performance on this dataset, where we synthetically removed data uniformly at random.

Comparing to the related table in [10] we see that our EM algorithm performs about average of the algorithms in mean, but its performance is nearly as good in median, even with missing data, as the best algorithms with full data. We can interpret that to mean that there are a few datasets where EM does very poorly. In these datasets, there may be overlapping subspaces or ill conditioned data, which would be problematic for any algorithm.

Table 1. Classification Errors (in %) of EM on the Hopkins 155 Motion Segmentation Dataset

Two Motions

$\lvert\omega\rvert$(%)	Check.(78)		Traffic (31)		Articul. (11)	
	Mean	Median	Mean	Median	Mean	Median
100	4.3	0	0.1	0	0.5	0
70	3.6	0.5	0.9	0	4.6	0
50	3.2	0.3	1.3	0	2.4	0
30	5.8	0.9	3.4	0.4	2.4	0

Three Motions

$\lvert\omega\rvert$(%)	Check.(78)		Traffic (31)		Articul. (11)	
	Mean	Median	Mean	Median	Mean	Median
100	16.9	17.9	1.3	0.4	0	0
70	16.2	17.9	10.5	6.1	9.0	9.0
50	17.4	17.9	10.7	8.9	21.7	21.7
30	25.4	25.6	19.6	13.0	22.4	22.4

6. CONCLUSION

We showed that only $\mathscr{O}(Kd^{r+1})$ columns are sufficient to guarantee a unique solution for subspace clustering with missing data, as opposed to $\mathscr{O}(d^{\log d})$ from previous existing bounds. A powerful conclusion of our theory is that *if* there is an algorithm that finds a set of K low-dimensional subspaces that fit independent generic validation sets, then that solution is the true $\mathcal{S}$ a.s. Furthermore, we presented a novel EM-type algorithm that in practice performs very well even with fewer columns than theoretically derived, suggesting that our bound is over sufficient. The true sample complexity of this problem, conjectured to be $\mathscr{O}(Kd)$, and a practical algorithm that provably solves it under such conditions, remain important open questions that our immediate future work will aim to answer.

7. REFERENCES

[1] Brian Eriksson, Laura Balzano, and Robert Nowak, "High-Rank Matrix Completion and Subspace Clustering with Missing Data," in *Proceedings of the Conference on Artificial Intelligence and Statistics (AI Stats)*, 2012.

[2] Laura Balzano, Robert Nowak, Arthur Szlam, and Benjamin Recht, "*k*-Subspaces with missing data," in *Proceedings of the Statistical Signal Processing Workshop*, 2012.

[3] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *CoRR*, vol. abs/1203.1005, 2012.

[4] Daniel Pimentel, Laura Balzano, and Robert Nowak, ," 2014, `http://homepages.cae.wisc.edu/~pimentel/SSP14.pdf`.

[5] Benjamin Recht, "A simpler approach to matrix completion," *Jrnl. of Machine Learning Rsrch.*, vol. 12, pp. 3413–3430, 2011.

[6] Franz Király and Ryota Tomioka, "A combinatorial algebraic approach for the identifiability of low-rank matrix completion," in *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[7] Michael E. Tipping and Christopher M. Bishop, "Mixtures of probabilistic principal component analysers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.

[8] Zoubin Ghahramani and Michael I. Jordan, "Supervised learning from incomplete data via an em approach," in *Advances in Neural Information Processing Systems 6*. 1994, pp. 120–127, Morgan Kaufmann.

[9] Brian Eriksson, Paul Barford, Joel Sommers, and Robert Nowak, "DomainImpute: Inferring Unseen Components in the Internet," in *Proceedings of IEEE INFOCOM Mini-Conference*, April 2011, pp. 171–175.

[10] René Vidal, "Subspace clustering," *IEEE Signal Processing Magazine*, 2010.