Breaking the Limits of Subspace Inference

Claudia R. Solís-Lemus, Daniel L. Pimentel-Alarcón
Emory University, Georgia State University

***Abstract*—Inferring low-dimensional subspaces that describe high-dimensional, highly incomplete datasets has become a routinely procedure in modern data science. This paper is about a curious phenomenon related to the amount of information required to estimate a subspace. On one hand, it has been shown that information-theoretically, data in $\mathbb{R}^{\mathrm{d}}$ must be observed on at least $\ell = \mathrm{r}+1$ coordinates to uniquely identify an r-dimensional subspace that approximates it. On the other hand, it is well-known that the subspace containing a dataset can be estimated through its sample covariance matrix, which only requires observing 2 coordinates per datapoint (regardless of r!). At first glance, this may seem to contradict the information-theoretic bound. The key lies in the subtle difference between identifiability (uniqueness) and estimation (most probable). It is true that if we only observed $\ell \leq \mathrm{r}$ coordinates per datapoint, there will be infinitely many r-dimensional subspaces that perfectly agree with the observations. However, some subspaces may be more *likely* than others, which are revealed by the sample covariance. This raises several fundamental questions: what are the algebraic relationships hidden in 2 coordinates that allow estimating an r-dimensional subspace? Moreover, are $\ell = 2$ coordinates per datapoint necessary for estimation, or is it possible with only $\ell = 1$? In this paper we show that under certain assumptions, it is possible to estimate some subspaces up to finite choice with as few as $\ell = 1$ entry per column, and opens the question of whether there exist other subspace estimation methods that allow $\ell \leq \mathrm{r}$ coordinates per datapoint, and that are more efficient than the sample covariance, which converges slowly in the number of data points n.**

I. Introduction

This paper is about the following question. Consider a collection of arbitrarily many points $\{\mathbf{x}\}$ lying in an r-dimensional subspace $\mathbb{U} \subset \mathbb{R}^{\mathrm{d}}$. Suppose you only observe each point on $\ell \leq \mathrm{r}$ coordinates selected randomly. Can you infer $\mathbb{U}$? See Figure 1 to build some intuition.

It is easy to see that if $\ell = \mathrm{d}$ (full data), $\mathbb{U}$ can be identified as long as the set $\{\mathbf{x}\}$ contains r linearly independent vectors. If, on the contrary, $\mathrm{d} > \ell > \mathrm{r}$, it has been shown that $\mathbb{U}$ can be estimated uniquely as long as data is observed in the right places [8]. This is due to the fact that each incomplete vector will produce $\ell - \mathrm{r}$ polynomial constraints on $\mathbb{U}$. The pattern of observed data determines whether the constraints are redundant, and if such pattern satisfies certain combinatorial conditions (which happens with high probability under random samplings), then the constraints produced by all the data will determine $\mathbb{U}$ uniquely (see Figure 2 for more intuition).

However, if $\ell \leq \mathrm{r}$, then any r-dimensional subspace $\mathbb{V}$ in general position will agree with the observed data (even if $\mathrm{n} := |\{\mathbf{x}\}| \to \infty$). To understand this, let $\boldsymbol{\omega} \subset \{1, \dots, \mathrm{d}\}$ indicate a subset of $\ell \leq \mathrm{r}$ observed entries. Let $\mathbf{x}_{\boldsymbol{\omega}} \in \mathbb{R}^{\ell}$ and $\mathbb{V}_{\boldsymbol{\omega}} \subset \mathbb{R}^{\ell}$ indicate the restrictions of $\mathbf{x}$ and $\mathbb{V}$ to the coordinates in $\boldsymbol{\omega}$. Since $\mathbb{V}$ is r-dimensional and in general position, and

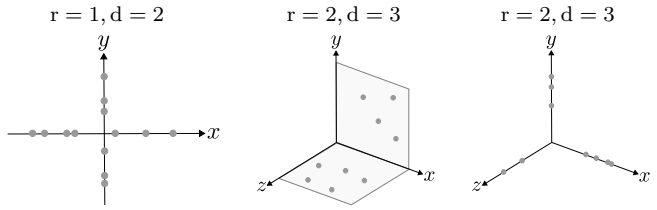

Fig. 1: **Can you find the r-dimensional subspace containing the full points if you only observe each point on $\ell \leq \mathrm{r}$ coordinates?** Each point in this figure represents an incomplete datum, or equivalently, the projection of the complete datum onto its observed coordinates. **Left:** Each point is only observed on $\ell = \mathrm{r} = 1$ coordinate. The points in the x axis are missing the y coordinate, and the points in the y axis are missing the x coordinate. Can you find the r-dimensional subspace (line) containing the full points? **Center:** Each point is only observed on $\ell = \mathrm{r} = 2$ coordinates. The points in the (x, y) plane (resp. (x, z) and (y, z)) are missing the z coordinate (resp. y and x). Can you find the r-dimensional subspace (plane) containing the full points? **Right:** Each point is only observed on $\ell = \mathrm{r} - 1 = 1$ coordinates. The points in the x axis (resp. y and z) are missing the $\{y, z\}$ coordinates (resp. $\{x, z\}$ and $\{x, y\}$). Can you find the r-dimensional subspace (plane) containing the full points?

$\ell \leq \mathrm{r}$, it follows that $\mathbb{V}_{\boldsymbol{\omega}} = \mathbb{R}^{\ell}$. In other words, $\mathbb{V}_{\boldsymbol{\omega}}$ covers the entire ℓ-dimensional space. Consequently, $\mathbf{x}_{\boldsymbol{\omega}} \in \mathbb{V}_{\boldsymbol{\omega}}$. This is true for every $\mathbf{x} \in \mathbb{R}^{\mathrm{d}}$ (not only in $\mathbb{U}$), for every $\boldsymbol{\omega}$ with $\ell \leq \mathrm{r}$ elements, and for every r-dimensional subspace $\mathbb{V}$ in general position. Equivalently, it is possible that each $\mathbf{x}_{\boldsymbol{\omega}}$ corresponds to a point in almost any $\mathbb{V}$, as illustrated in Figure 3.

This suggests that if $\ell \leq \mathrm{r}$, estimating $\mathbb{U}$ should be impossible. Surprisingly, it is not! Furthermore, it only takes elemental statistics to get convinced. Recall that if $\{\mathbf{x}\}$ are drawn independently according to an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{U}$, with finite fourth moment and zero mean (very standard and reasonable assumptions), then $\boldsymbol{\Sigma} := \mathsf{cov}(\mathbf{x}) = \mathbf{U}\mathbf{U}^{\mathsf{T}}$, where $\mathbf{U} \in \mathbb{R}^{\mathrm{d} \times \mathrm{r}}$ spans $\mathbb{U}$. Observing rows $\{\mathrm{i}, \mathrm{j}\}$ in $\mathbf{x}$ provides an estimate (through their outer product) of entries $\{\mathrm{i}, \mathrm{j}\} \times \{\mathrm{i}, \mathrm{j}\}$

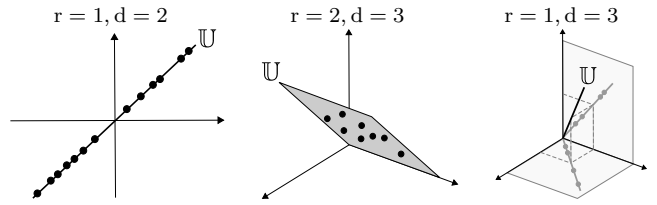

Fig. 2: It has been shown that $\mathbb{U}$ is identifiable if each point is observed on $\ell > \mathrm{r}$ entries. In the **left** and **center** this is the trivial case where $\ell = \mathrm{r}+1 = \mathrm{d}$, i.e., points are fully observed. In the **right**, $\mathbb{U}$ is a 1-dimensional subspace (line), and enough points are observed on $\ell = \mathrm{r}+1 = 2$ coordinates; consequently $\mathbb{U}$ is identifiable. This paper is about estimating $\mathbb{U}$ when each point is only observed on $\ell \leq \mathrm{r}$ coordinates (see Figure 1).

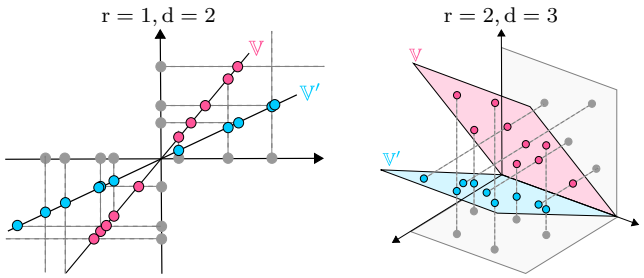

Fig. 3: If points are only observed on $\ell \leq r$ coordinates (depicted in gray), then each can be interpolated to almost any r-dimensional subspace $\mathbb{V}$ (here we only display two subspaces). In other words, it is possible that the (unknown) complete points lie in almost any $\mathbb{V}$. Hence, how could we possibly identify $\mathbb{U}$ when $\ell \leq r$?

of $\mathbf{\Sigma}$. If for every $i, j \in \{1, \ldots, d\}$, enough vectors are observed on rows $\{i, j\}$, then $\mathbf{\Sigma}$ can be estimated with arbitrary precision, which in turn implies that $\mathbb{U}$ can be inferred as the leading singular values even if vectors are only observed on $\ell = 2$ entries. This well-known estimation method is summarized in Algorithm 1. We point out that this method also applies if data is *near* a low-dimensional subspace, i.e., if instead of $\mathbf{x}$ we observe $\mathbf{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a noise vector with finite second moment and *reasonable* variance (i.e., comparable to the signal strength).

This apparent paradox can be understood in two different notions of identifiability. On one hand, $\mathbb{U}$ is unidentifiable in the sense that if $\ell \leq r$, then almost every r-dimensional subspace can potentially explain the observed data. On the other hand, the quadric outer products encoded in the covariance matrix reveal the subspace under which the full data would have minimum variance (see Figure 4), which coincides with $\mathbb{U}$ under standard regularity conditions.

This paper goes one step further to show an even more surprising result: in some cases it is indeed possible to estimate subspaces up to finite choice with as little as $\ell = 1$ coordinates. In fact, as we show in our experiments, one can infer $\mathbb{U}$ with as few as $\ell = 1$ samples per column and even under signal-to-noise ratio close to 1.

Algorithm 1: Subspace Estimation by Covariance SVD

Input: Partially observed vectors $\{\mathbf{x}_\omega\}$, subspace dimension r.

1. **Estimate** the incomplete covariance matrix:

$$\hat{\mathbf{\Sigma}}_{ij} = \frac{\sum_{\mathbf{x}_\omega} x_i x_j \mathbb{1}_{\{i,j \in \omega\}}}{\sum_{\omega} \mathbb{1}_{\{i,j \in \omega\}}}$$

2. **Singular value decomposition** of $\hat{\mathbf{\Sigma}}$:

$$\hat{\mathbf{U}} = \text{leading r left singular vectors of } \hat{\mathbf{\Sigma}}.$$

Output: Subspace $\hat{\mathbb{U}}$ spanned by $\hat{\mathbf{U}}$.

Organization of the paper

Section II describes the motivation behind this study. In Section III we give a formal setup of the problem and describe our main assumptions. Section IV presents our main theoretical result, and introduces our practical algorithm for subspace estimation. Section V discusses the paradox implied by our results. Section VI gives a discussion of the relation between subspace estimation and matrix completion. Section VII gives some insight into our assumptions. All experiments are in Section VIII, and Section IX discusses conclusions, future lines of work, and the general implications of our results.

II. Motivation and Prior Work

In many modern problems one aims to infer a linear subspace $\mathbb{U}$ that contains the columns of a highly incomplete data matrix $\mathbf{X}$. One typical example is the popular problem of low-rank matrix completion (LRMC) [1]–[8], where one aims to infer the missing entries in a low-rank matrix $\mathbf{X}$ (meaning that its columns lie in a low-dimensional subspace $\mathbb{U}$). Identifying $\mathbb{U}$ allows one to infer the missing entries in $\mathbf{X}$ (by projecting onto $\mathbb{U}$).

One classical motivation is recommender systems [9], [10], where each row of $\mathbf{X}$ represents an item, and each column represents a user. We only observe an entry in $\mathbf{X}$ whenever a user rates an item, and the goal is to predict unseen ratings in order to make good recommendations. Since the columns of $\mathbf{X}$ lie in a subspace, each column (user) can be represented as a linear combination of a few others. Equivalently, each column of $\mathbf{X}$ can be written as a linear combination of a basis of $\mathbb{U}$. Similar scenarios arise when monitoring large and complex networked systems such as the Internet, power grid, or wireless, social and biological networks. Hop counts over a network lie in a union of subspaces; by estimating the subspaces one can infer the topology of the network [11]. However, simultaneously measuring all individual components in these large systems is impossible with current technology, and even in smaller systems it is difficult or impractical to do so. In computer vision, the background and each moving object can be modeled using subspaces [12]–[15], but occlusions, shadows and other phenomenons naturally produce missing data. In classical settings of linear regression, like surveys, subsets of the data are simply unavailable (subjects do not know or do not want to provide information).

Motivated by these and other modern applications, recent developments in optimization, statistical signal processing, and information theory have resulted in theory and algorithms showing that subspaces can indeed be inferred from highly incomplete data. For instance, [1]–[4] show that with high probability, the r-dimensional subspace $\mathbb{U} \subset \mathbb{R}^d$ containing the columns of $\mathbf{X}$ can be uniquely identified if $\mathbf{X}$ is observed on $\ell = \mathcal{O}(r \log d)$ entries per column selected uniformly at random, and $\mathbb{U}$ has bounded coherence (parameter indicating how aligned a subspace is with the canonical axes). These results were later extended to include more general sampling schemes (non-uniform) [5], more coherent subspaces [6], and

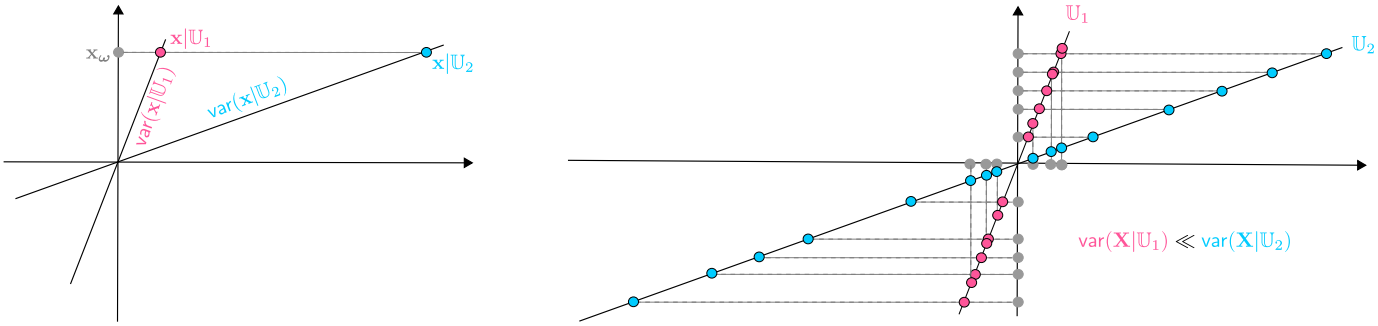

Fig. 4: **Left:** The incomplete column $\mathbf{x}_{\boldsymbol{\omega}}$ can be explained by infinitely many 1-dimensional subspaces, for example $\mathbb{U}_1$ and $\mathbf{U}_2$. However, the variance of the corresponding full vector $\mathbf{x}$ under $\mathbb{U}_1$ is much smaller than under $\mathbb{U}_2$. **Right:** The sample covariance reveals the subspace under which the full dataset $\mathbf{X}$ would have minimum variance, in this case $\mathbb{U}_1$.

deterministic information-theoretic sampling conditions [8], [16], [17]. Similarly, developing practical subspace estimation algorithms for incomplete data that can handle noise, outliers and other peculiarities, has been the focus of a wide variety of studies in the last years [1]–[6].

In practice, however, one generally ignores a priori the dimension of the underlying subspace that explains the dataset at hand. Moreover, the dimension r of the subspace often exceeds the number of observed coordinates ℓ per datapoint. Hence, it is of considerable interest to understand the limits of subspace inference from incomplete data when $\ell \leq \mathrm{r}$, and practical methods for this type of inference. For example, we know that the covariance method only requires $\ell = 2$ coordinates per datapoint; however, it requires a polynomial number of data points to produce an accurate estimator, which is highly suboptimal for $\ell > \mathrm{r}$ (see Figure 5). For instance, even if we observe $\ell = 50\%$ of the entries in $\mathrm{n} = 100$ data points in an $\mathrm{r} = 5$ dimensional subspace in ambient dimension $\mathrm{d} = 100$, selected uniformly at random, the subspace estimated through the covariance matrix achieves 25% error, while even the simplest LRMC methods, such as singular value thresholding (SVT) [3] achieve zero. This raises two important questions: (i) are $\ell = 2$ coordinates per datapoint the fundamental limit of subspace inference, or is it possible with only $\ell = 1$, and (ii) are there other algorithms, besides the covariance method, that can estimate subspaces in the over-incomplete regime ($\ell \leq \mathrm{r}$) but that do not require such huge number of data points, similar to SVT for $\ell < \mathrm{r}$? In this paper we show that in some cases it is indeed possible to identify $\mathbb{U}$ up to finite choice with as little as $\ell = 1$, thus addressing (i), and provide an estimation method, also addressing (ii).

III. Formal Setup and Assumptions

Let $\mathbb{U}$ be an r-dimensional subspace of $\mathbb{R}^{\mathrm{d}}$ and let $\{\mathbf{x}\}$ be a collection of n columns lying on $\mathbb{U}$. Let $\mathbf{x}_{\boldsymbol{\omega}}$ denote the incomplete version of $\mathbf{x}$, observed only on the entries indicated in $\boldsymbol{\omega} \subset \{1, \ldots, \mathrm{d}\}$. The goal is to estimate $\mathbb{U}$ from $\{\mathbf{x}_{\boldsymbol{\omega}}\}$.

Without any assumptions on $\{\mathbf{x}\}$ and $\{\boldsymbol{\omega}\}$, estimating $\mathbb{U}$ may be impossible. For instance, if $\{\mathbf{x}\}$ lies inside a proper subspace of $\mathbb{U}$ (for example, if $\mathbb{U}$ is a plane, and $\{\mathbf{x}\}$ lies in a line inside $\mathbb{U}$), then it would be impossible to estimate $\mathbb{U}$ (see Figure 6 to build some intuition). To avoid pathological cases like these, we will assume that $\{\mathbf{x}\}$ is *generically* spread over $\mathbb{U}$. More precisely:

(A1) The columns in $\{\mathbf{x}\}$ are drawn independently according to an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{U}$ with zero mean, and finite marginal variances $\sigma_{\mathrm{i}}^2 < \infty$ for $\mathrm{i} = 1, \ldots, \mathrm{d}$.

In words, **A1** requires that the data $\{\mathbf{x}\}$ are in general position over $\mathbb{U}$. This guarantees that $\{\mathbf{x}\}$ has at least a little bit of information in every direction of $\mathbb{U}$, and that observing more data will help estimating $\mathbb{U}$. Similar genericity assumptions are becoming increasingly common in matrix completion and related problems [7], [8], [15]–[20]. See Section VII for a further discussion about **A1** and its relation to other typical assumptions from the literature.

IV. Main Results

Assumption **A1** guarantees that $\{\mathbf{x}\}$ is well-spread over $\mathbb{U}$. However, this is not enough to identify $\mathbb{U}$. We also need to

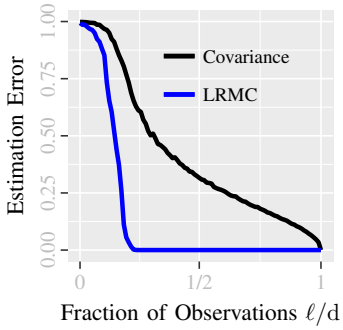

Fig. 5: Subspace estimation error (average over 100 trials) of the covariance method for subspace inference (Algorithm 1) and the LRMC method [3], as a function of the fraction of observed entries. This shows the sub-optimality of the covariance approach for $\ell > \mathrm{r}$ observations per column, and calls the attention for similar, more efficient methods for the over-incomplete regime $\ell \leq \mathrm{r}$.

guarantee that the samplings $\{\boldsymbol{\omega}\}$ have *enough* observations and in the right places. For example, it would be impossible to recover $\mathbb{U}$ unless each of its coordinates is sampled at least once. Let ℓ denote the maximum number of nonzero entries (observations) in a column of $\{\boldsymbol{\omega}\}$. Existing theory shows that $\ell > \mathrm{r}$ is necessary to *uniquely* identify $\mathbb{U}$ [8], [16], [17]. In fact, it is true that if $\ell \leq \mathrm{r}$, then *almost every* r-dimensional subspace will agree with $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ (like in Figure 3). However, under certain assumptions, some subspaces may be more *likely* than others.

The main result of this paper is Theorem 1 below. It shows that even if all columns of $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ only have $\ell = 1$ observation, it is still possible to estimate $\mathbb{U}$ up to finite choice with arbitrary accuracy (even though all r-dimensional subspaces would agree with $\{\mathbf{x}_{\boldsymbol{\omega}}\}$) when $\mathrm{r} = 1$. The key idea behind this result is to think of $\{\mathbf{x}\}$ as a collection of random vectors. Since these vectors lie in $\mathbb{U}$, they will have a covariance matrix $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{U}^{\mathsf{T}}$, where $\mathbf{U} \in \mathbb{R}^{\mathrm{d}\times\mathrm{r}}$ is a basis of $\mathbb{U}$. Thus, the variance of the observed entries in the canonical axes provide information of the one-dimensional subspace under which the full data $\{\mathbf{x}\}$ would have the least variance (see Figures 7 and 4 for some intuition), which under the standard regularity conditions in **A1**, coincides with the true subspace $\mathbb{U}$.

Theorem 1. *Let* **A1** *hold. Suppose that* $\mathbb{U}$ *is a one-dimensional subspace* ($\mathrm{r} = 1$)*, and that* $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ *has* $\ell = 1$ *nonzero entry per column, drawn uniformly at random and independently across columns. Let* $\{\hat{\mathbb{U}}\}$ *be the set of* $2^{\mathrm{d}} - 2$ *subspaces spanned by* $\{[\pm\hat{\sigma}_1 \;\; \pm\hat{\sigma}_2 \;\; \cdots \;\; \pm\hat{\sigma}_{\mathrm{d}}]^{\mathsf{T}}\}$*, where* $\hat{\sigma}_{\mathrm{i}}$ *represents the estimated marginal standard deviation of the* i^{th} *coordinate. Then* $\mathsf{P}(\mathbb{U} \in \{\hat{\mathbb{U}}\}) \to 1$*, as* $\mathrm{n} \to \infty$*.*

Theorem 1 shows that we can estimate the underlying subspace (up to finite choice) by observing only one sample per column in $\{\mathbf{x}\}$. The subspace estimate is not unique because we never observe two entries of each column simultaneously (which would provide us with the covariance estimate), and so we cannot know the direction of the one-dimensional subspace. With only one entry per column, we can only estimate the angle to the canonical axes (see Figure 8).

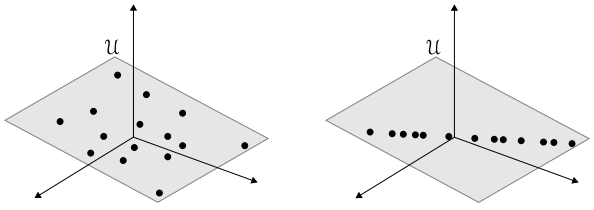

Fig. 6: Each column in $\{\mathbf{x}\}$ corresponds to a point in an r-dimensional subspace $\mathbb{U}$. In these figures, $\mathbb{U}$ is a 2-dimensional subspace (plane) of $\mathbb{R}^3$. In the **left**, $\{\mathbf{x}\}$ are drawn *generically* from $\mathbb{U}$, that is, independently according to an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{U}$, for example, according to a gaussian distribution on $\mathbb{U}$. In this case, the probability of observing a sample as in the **right**, where all columns lie in a line inside $\mathbb{U}$, is zero.

Proof. Let $\mathbb{U}$ be a one-dimensional subspace ($\mathrm{r} = 1$) in $\mathbb{R}^{\mathrm{d}}$ that independently generates the columns $\{\mathbf{x}\}$. Let $\mathbf{x}_{\boldsymbol{\omega}}$ denote the incomplete version of $\mathbf{x}$ where we observe $\ell = 1$ entry per column, chosen randomly. Let $\{\hat{\mathbb{U}}\}$ be the set of $2^{\mathrm{d}} - 2$ subspaces spanned by $\{[\pm\hat{\sigma}_1 \;\; \pm\hat{\sigma}_2 \;\; \cdots \;\; \pm\hat{\sigma}_{\mathrm{d}}]^{\mathsf{T}}\}$, where $\hat{\sigma}_{\mathrm{i}}$ represents the estimated marginal standard deviation of the entries in the i^{th} coordinate.

To prove the theorem, we need to show two statements:

1) $\mathbb{U}$ is spanned by one of the vectors in $\{[\pm\sigma_1 \;\; \pm\sigma_2 \;\; \cdots \;\; \pm\sigma_{\mathrm{d}}]^{\mathsf{T}}\}$ where σ_{i} represents the true marginal standard deviation of the entries in the i^{th} coordinate.
2) The estimated standard deviations $\hat{\sigma}_{\mathrm{i}}$ converge in probability to the true standard deviations σ_{i}.

Statement (2) is trivially true by the Law of Large Numbers (LLN). To see this, let $\mathrm{x}_1, \ldots, \mathrm{x}_{\mathrm{n}_{\mathrm{i}}}$ denote all the entries observed on the i^{th} coordinate/row. By **A1**, $\mathrm{x}_1, \ldots, \mathrm{x}_{\mathrm{n}_{\mathrm{i}}}$ are an i.i.d. random sample with zero mean and finite marginal variance. Notice that $\mathrm{n}_{\mathrm{i}} < \mathrm{n} := |\{\mathbf{x}\}|$, but $\mathrm{n}_{\mathrm{i}} \to \infty$ as $\mathrm{n} \to \infty$ due to the random sampling assumption of Theorem 1. Since $\sigma_{\mathrm{i}}^2 < \infty$, by LLN, $\hat{\sigma}_{\mathrm{i}}^2 = \frac{1}{\mathrm{n}_{\mathrm{i}}} \sum_{\mathrm{j}=1}^{\mathrm{n}_{\mathrm{i}}} \mathrm{x}_{\mathrm{j}}^2 \xrightarrow{\mathsf{P}} \sigma_{\mathrm{i}}^2$.

To show (1), we will first show that if all data is positive, then $\mathbb{U}$ is spanned by $[\sigma_1 \;\; \sigma_2 \;\; \cdots \;\; \sigma_{\mathrm{d}}]^{\mathsf{T}}$. Then we will argue that since we only observe one entry per column, we cannot estimate the sign of the correlation between the i^{th} and j^{th} coordinates, and hence if data contains negative numbers, then $\mathbb{U}$ must be spanned by one of the vectors in $\{[\pm\sigma_1 \;\; \pm\sigma_2 \;\; \cdots \;\; \pm\sigma_{\mathrm{d}}]^{\mathsf{T}}\}$.

To this end, let $\mathbf{x}$ be a positive random vector lying in $\mathbb{U}$, and let $\mathbf{u} \in \mathbb{R}^{\mathrm{d}}$ be a vector spanning $\mathbb{U}$. Since $\mathbf{x} \in \mathbb{U}$, we know that $\mathbf{x} = \mathbf{u}\theta$ for some $\theta \in \mathbb{R}$, which further implies that $\boldsymbol{\Sigma} := \mathsf{cov}(\mathbf{x}) = \alpha\mathbf{u}\mathbf{u}^{\mathsf{T}}$, where α is the variance of θ. This implies that the leading eigenvector of $\boldsymbol{\Sigma}$ provides a basis for $\mathbb{U}$. We will show that this eigenvector is $[\sigma_1 \;\; \sigma_2 \;\; \cdots \;\; \sigma_{\mathrm{d}}]^{\mathsf{T}}$. To see this, let $\mathbf{x}' = [\mathrm{x}_{\mathrm{i}} \;\; \mathrm{x}_{\mathrm{j}}]^{\mathsf{T}}$ and $\mathbf{u}' = [\mathrm{u}_{\mathrm{i}} \;\; \mathrm{u}_{\mathrm{j}}]$ be vectors containing the i^{th} and j^{th} entries of $\mathbf{x}$ and $\mathbf{u}$. Since $\mathrm{x}_{\mathrm{i}}, \mathrm{x}_{\mathrm{j}} > 0$ (implying that $\mathbf{x}'$ lies in the first quadrant $[0, \infty)^2$) and $\mathbb{U}' := \mathrm{span}[\mathbf{u}']$ is a line, it follows that $\mathsf{corr}(\mathrm{x}_{\mathrm{i}}, \mathrm{x}_{\mathrm{j}}) = 1$, which further implies that $\mathsf{cov}(\mathrm{x}_{\mathrm{i}}, \mathrm{x}_{\mathrm{j}}) = \sigma_{\mathrm{i}}\sigma_{\mathrm{j}}$. This means that we can compute the covariance matrix of $\mathbf{x}'$ by only observing the variances:

$$\boldsymbol{\Sigma}' = \begin{bmatrix} \sigma_{\mathrm{i}}^2 & \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 \end{bmatrix}.$$

Next we can use elemental linear algebra to see that the leading eigenvector of $\boldsymbol{\Sigma}'$ is given by $[\sigma_{\mathrm{i}} \;\; \sigma_{\mathrm{j}}]^{\mathsf{T}}$:

$$\begin{bmatrix} \sigma_{\mathrm{i}}^2 & \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} a \\ b \end{bmatrix},$$

or equivalently:

$$\begin{bmatrix} \sigma_{\mathrm{i}}^2 - \lambda & \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 - \lambda \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{1}$$

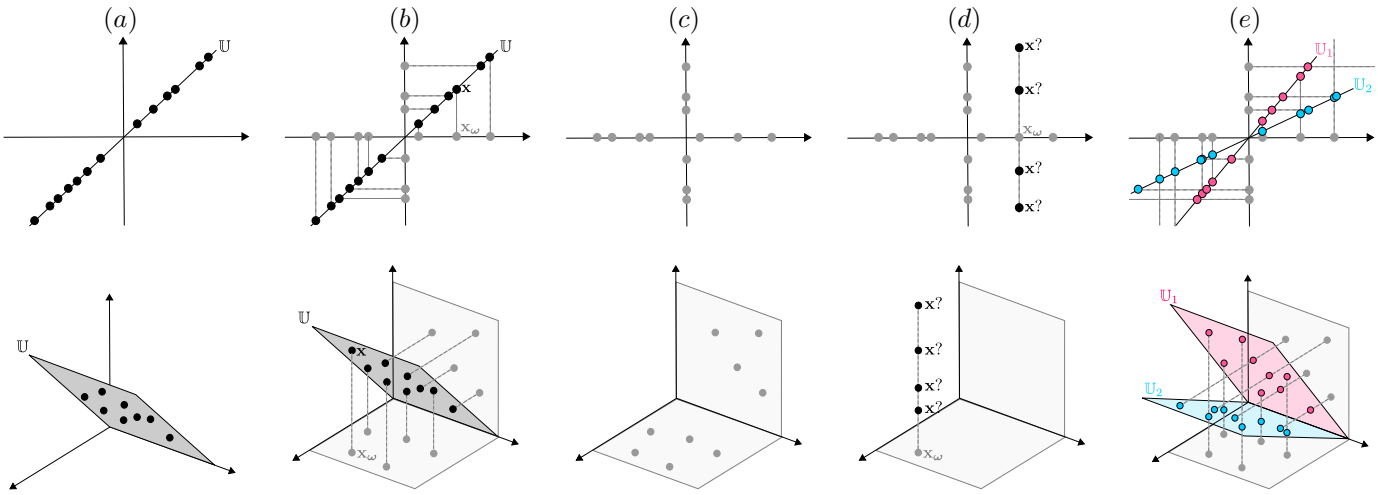

Fig. 7: (a) Consider a set of points $\{\mathbf{x}\}$ in an r-dimensional subspace $\mathbb{U} \subset \mathbb{R}^{\mathrm{d}}$: $\mathrm{r} = 1, \mathrm{d} = 2$. (b) Suppose we only see each point on r coordinates; equivalently, we only see the projection $\mathbf{x}_{\boldsymbol{\omega}}$ of each point $\mathbf{x}$ onto its observed coordinates. (c) Can we identify $\mathbb{U}$ from the collection of incomplete points $\{\mathbf{x}_{\boldsymbol{\omega}}\}$? (d) The challenge is that given $\mathbf{x}_{\boldsymbol{\omega}}$, there are infinitely many possibilities of where its complete point $\mathbf{x}$ could be, and so (e) each point could be on any of infinitely many subspaces (here we only depict two).

Then

$$\det \begin{bmatrix} \sigma_{\mathrm{i}}^2 - \lambda & \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 - \lambda \end{bmatrix} = 0,$$

and solving for λ:

$$(\sigma_{\mathrm{i}}^2 - \lambda)(\sigma_{\mathrm{j}}^2 - \lambda) - \sigma_{\mathrm{i}}^2\sigma_{\mathrm{j}}^2 = 0,$$

we conclude that $\lambda = 0$ or $\lambda = \sigma_{\mathrm{i}}^2 + \sigma_{\mathrm{j}}^2$. Setting $\lambda = \sigma_{\mathrm{i}}^2 + \sigma_{\mathrm{j}}^2$, we can rewrite (1) as:

$$\begin{aligned} (\sigma_{\mathrm{i}}^2 - \lambda)a + \sigma_{\mathrm{i}}\sigma_{\mathrm{j}}b &= 0 \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}}a + (\sigma_{\mathrm{j}}^2 - \lambda)b &= 0, \end{aligned}$$

or equivalently,

$$\begin{aligned} -\sigma_{\mathrm{j}}^2 a + \sigma_{\mathrm{i}}\sigma_{\mathrm{j}}b &= 0 \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}}a - \sigma_{\mathrm{i}}^2 b &= 0. \end{aligned}$$

Given that $\sigma_{\mathrm{i}} > 0$ and $\sigma_{\mathrm{j}} > 0$, we get $a = \frac{\sigma_1}{\sigma_2}$ and $b = 1$.

We thus conclude that the leading eigenvector of $\boldsymbol{\Sigma}'$ is given by $[\sigma_{\mathrm{i}} \;\; \sigma_{\mathrm{j}}]^{\mathsf{T}}$. Notice that ignoring a scaling factor, the leading eigenvector of $\boldsymbol{\Sigma}'$ contains the i^{th} and j^{th} entries of the leading eigenvector of $\boldsymbol{\Sigma}$ (because $\boldsymbol{\Sigma}' = \alpha \mathbf{u}'\mathbf{u}'^{\mathsf{T}}$ is the minor of $\boldsymbol{\Sigma} = \alpha \mathbf{u}\mathbf{u}^{\mathsf{T}}$ containing rows and columns i and j). Since this is true for every i and j, we conclude that $[\sigma_1 \;\; \sigma_2 \;\; \cdots \;\; \sigma_{\mathrm{d}}]^{\mathsf{T}}$ is the eigenvector of $\boldsymbol{\Sigma}$, and hence spans $\mathbb{U}$.

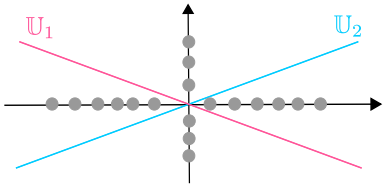

Fig. 8: With only one entry per column, we can only estimate the angle to the canonical axes, but not the direction. Consequently, we can only identify $\mathbb{U}$ up to finite choice. In this figure there are two subspaces, $\mathbb{U}_1$ and $\mathbb{U}_2$ that are equally likely to have produced the incomplete data $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ (gray points).

At this point we know that if $\mathbf{x}$ is positive, then $\mathbb{U} = \mathrm{span}[\sigma_1 \;\; \sigma_2 \;\; \cdots \;\; \sigma_{\mathrm{d}}]^{\mathsf{T}}$. If $\mathbf{x}$ is not necessarily positive, then we cannot know the sign of $\mathrm{corr}(\mathrm{x_i}, \mathrm{x_j})$. In other words, we cannot know whether

$$\boldsymbol{\Sigma}' = \begin{bmatrix} \sigma_{\mathrm{i}}^2 & \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ \sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 \end{bmatrix} \quad \text{or} \quad \boldsymbol{\Sigma}' = \begin{bmatrix} \sigma_{\mathrm{i}}^2 & -\sigma_{\mathrm{i}}\sigma_{\mathrm{j}} \\ -\sigma_{\mathrm{i}}\sigma_{\mathrm{j}} & \sigma_{\mathrm{j}}^2 \end{bmatrix}.$$

Consequently, the leading eigenvector of $\boldsymbol{\Sigma}'$ may be either $[\sigma_{\mathrm{i}} \;\; \sigma_{\mathrm{j}}]^{\mathsf{T}}$ or $[-\sigma_{\mathrm{i}} \;\; \sigma_{\mathrm{j}}]^{\mathsf{T}}$, and by extension, the leading eigenvector of $\boldsymbol{\Sigma}$ may be any of the vectors in $\{[\pm\sigma_1 \;\; \pm \sigma_2 \;\; \cdots \;\; \pm\sigma_{\mathrm{d}}]^{\mathsf{T}}\}$, as claimed.

□

The subspace estimation method described by Theorem 1 is summarized in Algorithm 2.

Algorithm 2: Subspace Estimation by Variance

Input: Partially observed data $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ with $\ell = 1$ entries per column.

1. **Estimate** variances:

$$\hat{\sigma}_{\mathrm{i}}^2 = \frac{\sum_{\mathbf{x}_{\boldsymbol{\omega}}} \mathrm{x_i}^2 \mathbb{1}_{\{\mathrm{i} \in \boldsymbol{\omega}\}}}{\sum_{\boldsymbol{\omega}} \mathbb{1}_{\{\mathrm{i} \in \boldsymbol{\omega}\}}}$$

2. **Construct** standard deviations vectors:

$$\{\hat{\mathbf{u}}\} = \{[\pm\sigma_1 \;\; \pm\sigma_2 \;\; \cdots \;\; \pm\sigma_{\mathrm{d}}]^{\mathsf{T}}\}$$

Output: Collection of subspaces $\{\hat{\mathbb{U}}\}$ spanned by all possible directions of $\{\hat{\mathbf{u}}\}$.

V. The Apparent Paradox

The importance of Theorem 1 is that it shows that it is possible to estimate $\mathbb{U}$ with as few as $\ell = 1$ samples per column in $\{\mathbf{x}\}$ (for $\mathrm{r} = 1$). This is somewhat paradoxic,

because if $\ell \leq \mathrm{r}$, then almost every r-dimensional subspace will agree with $\{\mathbf{x}_{\boldsymbol{\omega}}\}$.

To better understand this, recall that $\mathbf{x}_{\boldsymbol{\omega}} \in \mathbb{R}^{|\boldsymbol{\omega}|}$ represents an incomplete column, observed only in the coordinates indicated in $\boldsymbol{\omega} \subset \{1, \ldots, \mathrm{d}\}$. Let $\mathbf{V} \in \mathbb{R}^{\mathrm{d} \times \mathrm{r}}$ be a basis of an r-dimensional subspace $\mathbb{V}$, and let $\mathbf{V}_{\boldsymbol{\omega}} \in \mathbb{R}^{|\boldsymbol{\omega}| \times \mathrm{r}}$ denote the restriction of $\mathbf{V}$ to the rows in $\boldsymbol{\omega}$. Then $\mathbb{V}$ will agree with $\mathbf{x}_{\boldsymbol{\omega}}$ if and only if there exists a coefficient $\boldsymbol{\theta} \in \mathbb{R}^{\mathrm{r}}$ such that

$$\mathbf{x}_{\boldsymbol{\omega}} = \mathbf{V}_{\boldsymbol{\omega}} \boldsymbol{\theta}. \tag{2}$$

Notice that (2) defines a system of $\ell = |\boldsymbol{\omega}|$ equations and r variables (the r entries in $\boldsymbol{\theta}$).

If $\ell > \mathrm{r}$, (2) becomes a polynomial constraint on $\mathbf{V}$ (see Section 4 in [8] for the details), and so not all subspaces $\mathbb{V}$ will agree with $\mathbf{x}_{\boldsymbol{\omega}}$. If there are enough columns in $\mathbf{X}_{\boldsymbol{\Omega}}$ with $\ell > \mathrm{r}$ samples *in the right places*, the polynomials produced by $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ will be independent, and $\mathbb{V} = \mathbb{U}$ will be the only subspace that agrees with $\{\mathbf{x}_{\boldsymbol{\omega}}\}$.

However, if $\ell \leq \mathrm{r}$, (2) will always have a solution (as long as $\mathbb{V}$ is in general position), and so *any* subspace $\mathbb{V}$ in general position will agree with $\mathbf{x}_{\boldsymbol{\omega}}$. Similarly, if *all* columns of $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ are observed on $\ell \leq \mathrm{r}$ entries, *any* subspace $\mathbb{V}$ in general position will agree with $\{\mathbf{x}_{\boldsymbol{\omega}}\}$.

The paradox shown by Theorem 1 is that even though *any* subspace $\mathbb{V}$ would agree with a matrix $\{\mathbf{x}_{\boldsymbol{\omega}}\}$ with $\ell \leq \mathrm{r}$ samples per column, we can still have information on the *true* subspace $\mathbb{U}$, because it will be more likely than the rest.

VI. Subspace Estimation vs. Matrix Completion

As we mentioned in Section I, subspace estimation is tightly related to matrix completion, where one aims to recover a $\mathrm{d} \times \mathrm{n}$ rank-r matrix $\mathbf{X}$ from a subset of its entries, indicated by $\boldsymbol{\Omega} \in \{0,1\}^{\mathrm{d} \times \mathrm{n}}$. Once $\mathbf{X}$ is recovered, one can compute the r leading singular vectors of $\mathbf{X}$ to obtain a basis for $\mathbb{U}$. On the other hand, once $\mathbb{U}$ is identified, one can potentially complete $\mathbf{X}_{\boldsymbol{\Omega}}$ (the observed entries of $\mathbf{X}$) by projecting each column onto $\mathbb{U}$. More precisely, suppose we have already identified the subspace $\mathbb{U}$ (with basis $\mathbf{U}$) containing the columns of $\mathbf{X}_{\boldsymbol{\Omega}}$. Let $\mathbf{x}_{\boldsymbol{\omega}} \in \mathbb{R}^{|\boldsymbol{\omega}|}$ denote a column of $\mathbf{X}_{\boldsymbol{\Omega}}$ with $\ell = |\boldsymbol{\omega}|$ entries, and $\mathbf{U}_{\boldsymbol{\omega}} \in \mathbb{R}^{|\boldsymbol{\omega}| \times \mathrm{r}}$ denotes the restriction of $\mathbf{U}$ to the rows in $\boldsymbol{\omega}$. Since $\mathbf{x}_{\boldsymbol{\omega}}$ agrees with $\mathbb{U}$, there exists a coefficient $\boldsymbol{\theta} \in \mathbb{R}^{\mathrm{r}}$ such that

$$\mathbf{x}_{\boldsymbol{\omega}} = \mathbf{U}_{\boldsymbol{\omega}} \boldsymbol{\theta}. \tag{3}$$

Notice that (3) defines a system of $\ell = |\boldsymbol{\omega}|$ equations and r variables (the r entries in $\boldsymbol{\theta}$).

If $\ell \geq \mathrm{r}$ and $\mathbb{U}$ is in general position, then (3) will always have a unique solution. In fact, we can solve for $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} \;=\; (\mathbf{U}_{\boldsymbol{\omega}}^{\mathsf{T}} \mathbf{U}_{\boldsymbol{\omega}})^{-1} \mathbf{U}_{\boldsymbol{\omega}}^{\mathsf{T}} \mathbf{x}_{\boldsymbol{\omega}},$$

and complete $\mathbf{x}_{\boldsymbol{\omega}}$ as $\mathbf{x} = \mathbf{U}\boldsymbol{\theta}$. However, if $\ell < \mathrm{r}$, then (3) will have infinitely many solutions $\boldsymbol{\theta}$, and $\mathbf{x}_{\boldsymbol{\omega}}$ will have infinitely many completions. Consequently, if *any* column of $\mathbf{X}_{\boldsymbol{\Omega}}$ is observed on $\ell < \mathrm{r}$ samples, there will be infinitely many rank-r matrices that agree with $\mathbf{X}_{\boldsymbol{\Omega}}$ (and lie in $\mathbb{U}$).

In this paper we show that it is possible to estimate the *true* subspace $\mathbb{U}$, even if $\ell \leq \mathrm{r}$ (implying that there are infinitely many r-dimensional subspaces $\mathbb{V}$ that agree with $\mathbf{X}_{\boldsymbol{\Omega}}$; see Section V). It remains an open question to determine whether it is possible to estimate the *true* completion $\mathbf{X}$ even if $\ell < \mathrm{r}$ for some columns of $\mathbf{X}_{\boldsymbol{\Omega}}$ (implying that there are infinitely many rank-r matrices that agree with $\mathbf{X}_{\boldsymbol{\Omega}}$).

VII. More About Our Assumptions

Essentially, **A1** requires that $\mathbf{X}$ is a generic low-rank matrix. This discards pathological cases, like matrices with identical columns or exact-zero entries. Examples of these cases could arise in unnatural, cartoon-like images.

However, **A1** allows realistic cases, like natural images. For instance, backgrounds in natural images can be highly structured but are not perfectly constant, as there is always some degree of natural variation that is reasonably modeled by an absolutely continuous (but possibly highly inhomogeneous) distribution. For example, the sky in a natural image might be strongly biased towards blue values, but each sky pixel will have at least small variations that will make the sky not perfectly constant blue. So while these are structured images, these variations make them generic enough so that our theoretical results are applicable.

Furthermore, because absolutely continuous distributions may be strongly inhomogeneous, they can be used to represent highly coherent matrices (that is, matrices whose underlying subspace is highly aligned with the canonical axes). We point out that **A1** does not imply coherence nor vice-versa. For example, typical coherence assumptions indeed allow some identical columns, or exact-zero entries [1]–[6]. However, they rule-out cases that our theory allows. For example, consider a case where a few rows of $\mathbf{U}$ are drawn i.i.d. $\mathcal{N}(0, \sigma_1^2)$ and many rows of $\mathbf{U}$ are drawn i.i.d. $\mathcal{N}(0, \sigma_2^2)$, with $\sigma_1 \gg \sigma_2$. This is a good model for some microscopy and astronomical applications that have a few high-intensity pixels, and many low-intensity pixels. Such $\mathbf{U}$ would yield a highly coherent matrix, which existing theory and algorithms cannot handle, while our results can.

To sum up, our assumptions are different, not stronger nor weaker than the usual coherence assumptions [1]–[6], and we believe they are also more reasonable in many practical applications.

VIII. Experiments

We now present a series of experiments to support our theoretical findings, and analyze the performance of our method (Algorithm 2). Since there are no other methods that can handle the case of $\ell = 1$, we cannot compare our method to existing methods. Thus, we will provide a study on the convergence rate of our method.

In our experiment we study how accurately can our method estimate $\mathbb{U}$ as a function of the ambient dimension d, the number of columns n, and the level of noise σ. To this end, we first generate a vector $\mathbf{u} \in \mathbb{R}^{\mathrm{d}}$ with i.i.d. $\mathcal{N}(0,1)$ entries, to use as a basis of $\mathbb{U}$. Then we create a coefficient matrix $\boldsymbol{\Theta} \in \mathbb{R}^{1 \times \mathrm{n}}$

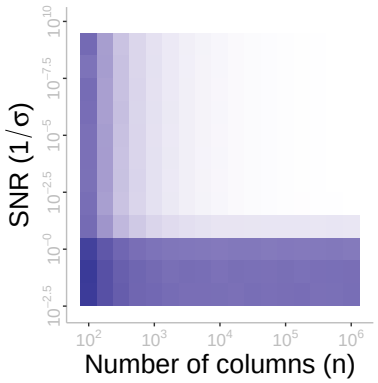

Fig. 9: Subspace estimation error (average over 100 trials) of our method (Algorithm 2) as a function of the number of columns n and the signal-to-noise ratio $1/\sigma$, with $\ell = 1$ sample per column (extreme low sampling setting). The darkest color represents error = 1 (maximum), and the lightest represents 0 (the lighter the better).

with i.i.d. $\mathcal{N}(0,1)$ entries, and construct our data matrix $\mathbf{X} = \mathbf{u}\boldsymbol{\Theta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \in \mathbb{R}^{\mathrm{d}\times\mathrm{n}}$ has i.i.d. $\mathcal{N}(0,\sigma^2)$ entries. Next we generate a sampling matrix $\boldsymbol{\Omega} \in \{0,1\}^{\mathrm{d}\times\mathrm{n}}$ with exactly $\ell = 1$ nonzero entry per column, selected uniformly at random. Finally, we use our method to obtain an estimate of the set $\{\hat{\mathbb{U}}\}$. We repeated this procedure 100 trials, and recorded the error, measured as the minimum Frobenius norm of the difference between the projector operators of $\mathbb{U}$ and all those in $\{\hat{\mathbb{U}}\}$. The results are summarized in Figure 9. Consistent with Theorem 1, we can see that $P(\mathbb{U} \in \{\hat{\mathbb{U}}\}) \to 1$ as $\mathrm{n} \to \infty$.

IX. Discussion

Theorem 1 shows that it is possible to estimate $\mathbb{U}$ (up to finite choice) with as few as $\ell = 1$ observations per column. Using less than r entries per column to reconstruct $\mathbb{U}$ is not new. Standard subspace estimation methods involving the eigenvectors of the covariance matrix had already successfully reconstructed $\mathbb{U}$ with as few as $\ell = 2$ entries per column. These covariance approaches, combined with our proposed new method with $\ell = 1$, draw attention to the question of where does the limit of subspace estimation *truly* lies. Conventional theory assumes a limiting bound of $\ell = \mathrm{r} + 1$ entries per column for the unique estimation of the subspace. However, the covariance matrix approach manages to uniquely estimate the subspace with $\ell = 2$, while here we estimate the subspace up to finite choice with as few as $\ell = 1$ entry per column. Just as the covariance method and our proposed method profit from the estimation of the *most likely* subspace (subspace under which the data would have least variance), as opposed to the best fit, this work leaves the question open as to whether there exist other estimation criteria (new column in table I) that permits the unique estimation of $\mathbb{U}$ with $\ell = 1$ entry per column.

Finally, the proposed method reconstructs $\mathbb{U}$ under the assumption of uniform samplings. In practice, however, the sampling in $\boldsymbol{\Omega}$ may not be uniformly distributed. For example, in recommender systems, the items that each user rates tend to be highly correlated (a child is more likely to rate children movies; some movies are more popular than others, etc.). In computer vision, the location of objects producing occlusions will be highly correlated over time. In surveys, people often cannot or do not want to answer similar questions, and so on. Our future work will focus on these more general sampling schemes, and on more efficient estimation techniques that require fewer data points.

Entries per column	Best fit subspace	Most likely subspace
$\ell = \mathrm{r} + 1$	unique solution	unique solution
$2 \leq \ell < r$	∞ solutions	unique solution
$\ell = 1$	∞ solutions	finite solutions

TABLE I: Summary of current subspace inference capabilities.

References

[1] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 2009.

[2] E. Candès and T. Tao, *The power of convex relaxation: near-optimal matrix completion*, IEEE Transactions on Information Theory, 2010.

[3] J. Cai, E. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization, 2010.

[4] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 2011.

[5] Y. Chen, *Incoherent-optimal matrix completion*, IEEE Transactions on Information Theory, 2015.

[6] Y. Chen, S. Bhojanapalli, S. Sanghavi and R. Ward, *Coherent matrix completion*, International Conference on Machine Learning, 2014.

[7] F. Király, L. Theran and R. Tomioka, *The algebraic combinatorial approach for low-rank matrix completion*, Journal of Machine Learning Research, 2015.

[8] D. Pimentel-Alarcón, N. Boston and R. Nowak, *A characterization of deterministic sampling patterns for low-rank matrix completion*, IEEE Journal of Selected Topics in Signal Processing, 2016.

[9] J. Rennie and N. Srebro, *Fast maximum margin matrix factorization for collaborative prediction*, International Conference on Machine Learning, 2005.

[10] A. Zhang, N. Fawaz, S. Ioannidis and A. Montanari, *Guess who rated this movie: identifying users through subspace clustering*, Conference on Uncertainty in Artificial Intelligence, 2012.

[11] B. Eriksson, P. Barford, J. Sommers and R. Nowak, *DomainImpute: inferring unseen components in the Internet*, Proceedings of the IEEE INFO-COM, 2011.

[12] J. Costeira and T. Kanade, *Body factorization method for independently moving objects*, International Journal of Computer Vision, 1998.

[13] K. Kanatani, *Motion segmentation by subspace separation and model selection*, IEEE International Conference on Computer Vision, 2001.

[14] J. Wright, A. Ganesh, S. Rao, Y. Peng and Y. Ma, *Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization*, Advances in Neural Information Processing Systems, 2009.

[15] D. Pimentel-Alarcón and R. Nowak, *Random consensus robust PCA*, Artificial Intelligence and Statistics, 2017.

[16] D. Pimentel-Alarcón, N. Boston and R. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, IEEE International Symposium on Information Theory, 2015.

[17] D. Pimentel-Alarcón and R. Nowak, *The information-theoretic requirements of subspace clustering with missing data*, International Conference on Machine Learning, 2016.

[18] D. Pimentel-Alarcón and R. Nowak, *A converse to low-rank matrix completion*, IEEE International Symposium on Information Theory, 2016.

[19] M. Ashraphijuo, X. Wang and V. Aggarwal, *A characterization of sampling patterns for low-rank multiview data completion problem*, IEEE International Symposium on Information Theory, 2017.

[20] M. Ashraphijuo, V. Aggarwal and X. Wang, *A characterization of sampling patterns for low-tucker-rank tensor completion problem*, IEEE International Symposium on Information Theory, 2017.

[21] Z. Wen, W. Yin and Y. Zhang, *Solving a low-rank factorization model for matrix completion by a non-linear successive over-relaxation algorithm*, Mathematical Programming Computation, 2012.
[22] Y. Shen, Z. Wen and Y. Zhang, *Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization*, International Conference on Numerical Optimization and Numerical Linear Algebra, 2014.
[23] J. Cai, E. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization, 2010.
[24] R. Keshavan, A. Montanari and S. Oh, *Matrix completion from a few entries*, IEEE Transactions on Information Theory, 2010.
[25] S. Ma, D. Goldfarb, L. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization*, Mathematical Programming, 2011.
[26] L. Balzano, R. Nowak, and B. Recht, *Online identification and tracking of subspaces from highly incomplete information*, Allerton Conference on Communication, Control and Computing, 2010.
[27] P. Jain, P. Netrapalli and S. Sanghavi, *Low-rank matrix completion using alternating minimization*, ACM Symposium on Theory of Computing, 2013.