
A Characterization of Deterministic Sampling Patterns for Low-Rank Matrix Completion

DANIEL L. PIMENTEL-ALARCÓN, NIGEL BOSTON, ROBERT D. NOWAK
University of Wisconsin - Madison

Abstract

Low-rank matrix completion (LRMC) problems arise in a wide variety of applications. Previous theory mainly provides conditions for completion under missing-at-random samplings. This paper studies deterministic conditions for completion. An incomplete $d \times N$ matrix is *finitely rank- r completable* if there are at most finitely many rank- r matrices that agree with all its observed entries. Finite completability is the tipping point in LRMC, as a few additional samples of a finitely completable matrix guarantee its *unique* completability. The main contribution of this paper is a deterministic sampling condition for finite completability. We use this to also derive deterministic sampling conditions for unique completability that can be efficiently verified. We also show that under uniform random sampling schemes, these conditions are satisfied with high probability if $\mathcal{O}(\max\{r, \log d\})$ entries per column are observed. These findings have several implications on LRMC regarding lower bounds, sample and computational complexity, the role of coherence, adaptive settings and the validation of any completion algorithm. We complement our theoretical results with experiments that support our findings and motivate future analysis of uncharted sampling regimes.

1 Introduction

Low-rank matrix completion (LRMC) has attracted a lot of attention in recent years because of its broad range of applications, e.g., recommender systems and collaborative filtering [1] and image processing [2].

The problem entails exactly recovering all the entries in a $d \times N$ rank- r matrix, given only a subset of its entries. LRMC is usually studied under a missing-at-random and bounded-coherence model. Under this model, necessary and sufficient conditions for perfect recovery are known [3–8]. Other approaches require additional coherence and spectral gap conditions [9], use rigidity theory [10], algebraic geometry and matroid theory [11] to derive necessary and sufficient conditions for completion of deterministic samplings, but a characterization of completable sampling patterns remains an important open question.

We say an incomplete matrix is *finitely rank- r completable* if there exist at most finitely many rank- r matrices that agree with all its observed entries. There exist sampling/observation patterns that guarantee finite completability, but if just a single one of the observed entries is instead missing, then there are *infinitely* many completions. Conversely, adding a few observations to such a pattern guarantees *unique* completability. Thus, finite completability is the tipping point in LRMC.

Whether a matrix is finitely completable depends on which entries are observed. Yet no characterization of the sets of observed entries that allow or prevent finite completability is known.

The main result of this paper is a sampling condition for finite completability, that is, a condition on the observed entries of a matrix to guarantee that it can be completed in at most finitely many ways. In addition, we provide deterministic sampling conditions for unique completability that can be efficiently verified. Finally, we show that uniform random samplings with $\mathcal{O}(\max\{r, \log d\})$ entries per column satisfy these conditions with high probability.

Our results have implications on LRMC regarding lower bounds, sample and computational complexity, the role of coherence, adaptive settings and validation conditions to verify the output of any completion algorithm. We complement our theoretical results with experiments that support our findings and motivate future analysis of uncharted sampling regimes.

Organization of the Paper

In Section 2 we formally state the problem and our main results. In Section 3 we discuss their implications in the context of previous work, and present our experiments. We present the proof of our main theorem in Section 4, and we leave the proofs of our other statements to Sections 5 and 6.

2 Model and Main Results

Let \mathbf{X}_Ω denote the incomplete version of a $d \times N$, rank- r data matrix \mathbf{X} , observed only in the nonzero locations of Ω , a $d \times N$ matrix with binary entries. The goal of LRMC is to recover \mathbf{X} from \mathbf{X}_Ω .

This problem is tantamount to identifying the r -dimensional subspace S^* spanned by the columns in \mathbf{X} , and this is how we will approach it. First observe that since \mathbf{X} is rank- r , a column with fewer than r samples cannot be completed. A column with exactly r observations can be uniquely completed once S^* is known, but it provides no information to identify S^* . We will thus assume that:

A1 Every column of \mathbf{X} is observed on exactly $r + 1$ entries.

The key insight of the paper is that observing $r + 1$ entries in a column of \mathbf{X} places one constraint on what S^* may be. For example, if we observe $r + 1$ entries of a particular column, then not all r -dimensional subspaces will be consistent with the entries. If we observe more columns with $r + 1$ entries, then even fewer subspaces will be consistent with them. In effect, each column with $r + 1$ observations places one constraint that an r -dimensional subspace must satisfy in order to be consistent with the observations. The observed entries in different columns may or may not produce redundant constraints. As we will see, the pattern of observed entries determines whether or not the constraints are redundant, thus indicating the number of subspaces that satisfy them. The main result of this paper is a simple condition on the pattern of observed entries that guarantees that only a finite number of subspaces satisfies all the constraints. This in turn provides a simple condition for exact matrix completion.

Remark 1. *We point out that any observation, in addition to the $r + 1$ per column that we assume, cannot increase the number of rank- r matrices that agree with the observations. So in general, if some columns of \mathbf{X} are observed on more than $r + 1$ entries, all we need is that the observed entries include a pattern with exactly $r + 1$ observations per column satisfying our sampling conditions.*

Also notice that completing \mathbf{X} is the same as completing \mathbf{X}^\top , so a row with fewer than r observations cannot be completed. While we do not assume that each row is observed on at least r entries, our sampling conditions guarantee that this is the case.

Let $\text{Gr}(r, \mathbb{R}^d)$ denote the Grassmannian manifold of r -dimensional subspaces in \mathbb{R}^d . Observe that each $d \times N$ rank- r matrix \mathbf{X} can be uniquely represented in terms of a subspace $S^* \in \text{Gr}(r, \mathbb{R}^d)$ (spanning the columns of \mathbf{X}) and an $r \times N$ coefficient matrix Θ^* . See Figure 1 to build some intuition. Let ν_{G} denote the uniform measure on $\text{Gr}(r, \mathbb{R}^d)$, and ν_{Θ} the Lebesgue measure on $\mathbb{R}^{r \times N}$. Our statements hold for *almost every* (a.e.) \mathbf{X} with respect to the product measure $\nu_{\text{G}} \times \nu_{\Theta}$.

The paper’s main result is the following theorem, which gives a deterministic sampling condition to guarantee that at most a finite number of r -dimensional subspaces are consistent with \mathbf{X}_{Ω} .

Given a matrix, let $n(\cdot)$ denote its number of columns and $m(\cdot)$ the number of its *nonzero* rows.

Theorem 1. *Let Ω be given, and suppose **A1** holds. For almost every \mathbf{X} , there exist at most finitely many rank- r completions of \mathbf{X}_{Ω} if and only if there exists a matrix $\tilde{\Omega}$ formed with $r(d - r)$ columns of Ω , such that*

- (i) *Every matrix Ω' formed with a subset of the columns in $\tilde{\Omega}$ satisfies*
- $$m(\Omega') \geq n(\Omega')/r + r. \tag{1}$$

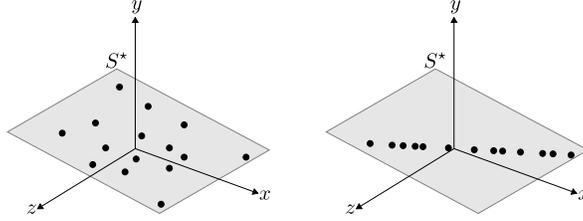


Figure 1: Each column in a rank- r matrix \mathbf{X} corresponds to a point in an r -dimensional subspace S^* . In these figures, S^* is a 2-dimensional subspace (plane) in general position. In the **left**, the columns of \mathbf{X} are in general position inside S^* , that is, drawn independently according to an absolutely continuous distribution with respect to the Lebesgue measure on S^* , for example, according to a gaussian distribution on S^* . In this case, the probability of observing a sample as in the **right**, where all columns lie in a line inside S^* , is zero. Our results hold for every rank- r matrix, except for a set of measure zero of pathological cases as in the right.

The proof of Theorem 1 is given in Section 4. In words, condition (i) asks that every subset of n columns of $\tilde{\Omega}$ has at least $n/r + r$ nonzero rows.

Example 1. Suppose Ω is given by:

$$\Omega = \underbrace{\left[\begin{array}{c|c|c|c} \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \hline \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \end{array} \right]}_{(r+1)(d-r)} \left. \vphantom{\left[\begin{array}{c|c|c|c} \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \\ \hline \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \end{array} \right]} \right\} \begin{array}{l} r \\ d-r \end{array}$$

such that Ω has exactly $r+1$ nonzero entries per column. This way, each column of Ω encodes exactly one constraint that candidate subspaces must satisfy in order to be consistent with the observed data. In this case we can simply take $\tilde{\Omega}$ to be the matrix formed with the first $r(d-r)$ columns of Ω . One may verify that $\tilde{\Omega}$ satisfies (i). Hence Ω satisfies the conditions of Theorem 1.

Unique Completability

Theorem 1 is easily extended to a condition on Ω that is sufficient to guarantee that one and only one subspace is consistent with \mathbf{X}_Ω , which in turn suffices for exact matrix completion.

Theorem 2. Let Ω be given, and suppose **A1** holds. Then almost every \mathbf{X} can be uniquely recovered from \mathbf{X}_Ω if Ω contains two disjoint submatrices: $\tilde{\Omega}$ of size $d \times r(d-r)$ and $\hat{\Omega}$ of size $d \times (d-r)$, such that $\tilde{\Omega}$ satisfies (i) and

(ii) Every matrix Ω' formed with a subset of the columns in $\hat{\Omega}$ satisfies

$$m(\Omega') \geq n(\Omega') + r. \quad (2)$$

The proof of Theorem 2 is given in Section 5. In words, condition (ii) asks that every subset of n columns of $\tilde{\Omega}$ has at least $n+r$ nonzero rows. Notice that (1) is a weaker condition than (2), but (1) is required to hold for all the subsets of $r(d-r)$ columns, while (2) is required to hold only for all the subsets of $d-r$ columns.

Example 2. Consider Ω as in Example 1. Take $\tilde{\Omega}$ to be the matrix formed with the first $r(d-r)$ columns of Ω and $\hat{\Omega}$ to be the matrix formed with the last $d-r$ columns of Ω . One may verify that $\tilde{\Omega}$ satisfies (i) and that $\hat{\Omega}$ satisfies (ii). Hence Ω satisfies the conditions of Theorem 2.

Theorem 1 implies that $r(d-r)$ columns with $r+1$ entries are necessary for finite completability (hence also for unique completability). There are cases when $r(d-r)$ columns are also sufficient for unique completability, e.g., if $r=1$, where finite completability is equivalent to unique completability (see Proposition 1).

In general, though, unique completability requires more columns than finite completability (see Example 5). Theorem 2 gives deterministic sufficient sampling conditions for unique completability that only require $(r+1)(d-r)$ columns. This shows that with just a few more observations, unique completability follows from finite completability.

We point out that when the conditions of Theorem 2 are met, S^* can be uniquely identified as

$$S^* = \text{span} \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix},$$

where \mathbf{V} is the unique solution to the polynomial system $\mathcal{F}(\mathbf{V}) = \mathbf{0}$, with \mathcal{F} as defined in Section 4.

Once S^* is known, \mathbf{X} can be perfectly recovered observing only r entries per column. To see this, let \mathbf{U}^* be a basis of S^* , and let \mathbf{v} be a subset of $\{1, \dots, d\}$ with exactly r elements. We will use the subscript \mathbf{v} to denote restriction to the rows in \mathbf{v} . Since the coefficients of column \mathbf{x} in the basis \mathbf{U}^* are given by $\boldsymbol{\theta}^* = (\mathbf{U}_{\mathbf{v}}^{*\top} \mathbf{U}_{\mathbf{v}}^*)^{-1} \mathbf{U}_{\mathbf{v}}^{*\top} \mathbf{x}_{\mathbf{v}}$, we can recover the entire column as $\mathbf{x} = \mathbf{U}^* \boldsymbol{\theta}^*$.

About Conditions (i) and (ii)

In general, verifying condition (i) in Theorems 1 and 2 may be computationally prohibitive, especially for large d . On the other hand, one can easily and efficiently verify whether (ii) is satisfied by checking the dimension of the null-space of a sparse matrix (Algorithm 1). Fortunately, there is a tight relation between conditions (i) and (ii), summarized in the following lemma. The proofs of the statements in this section are given in Section 6.

Lemma 1. Let $\tilde{\Omega}$ be a $d \times r(d-r)$ matrix formed with a subset of the columns in Ω . Suppose $\tilde{\Omega}$ can be partitioned into r matrices $\{\tilde{\Omega}_{\tau}\}_{\tau=1}^r$, each of size $d \times (d-r)$, such that (ii) holds for every $\tilde{\Omega}_{\tau}$. Then Ω satisfies (i).

As consequence of Lemma 1 we obtain an additional sufficient condition for completability that only involves (ii).

Corollary 1. *Let Ω be given, and suppose **A1** holds. Then almost every \mathbf{X} can be uniquely recovered from \mathbf{X}_Ω if Ω contains $r + 1$ disjoint matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^{r+1}$, each of size $d \times (d - r)$, such that (ii) holds for every $\hat{\Omega}_\tau$.*

Example 3. *Consider Ω as in Example 1. We can partition Ω into $[\hat{\Omega}_1 \mid \hat{\Omega}_2 \mid \dots \mid \hat{\Omega}_{r+1}]$, as depicted in Example 1. One may verify that $\hat{\Omega}_\tau$ satisfies (ii) for every $\tau = 1, \dots, r + 1$. Hence Ω satisfies the conditions of Corollary 1.*

With Corollary 1 we show that completable patterns appear with high probability under uniform random sampling schemes with as little as $\mathcal{O}(\max\{r, \log d\})$ samples per column.

Theorem 3. *Let $0 < \epsilon \leq 1$ be given. Suppose $r \leq \frac{d}{6}$ and that each column of \mathbf{X} is observed in at least ℓ entries, distributed uniformly at random and independently across columns, with*

$$\ell \geq \max\{12(\log(\frac{d}{\epsilon}) + 1), 2r\}. \quad (3)$$

Then with probability at least $1 - \epsilon$, \mathbf{X}_Ω will be finitely rank- r completable (if $N \geq r(d - r)$) and uniquely completable (if $N \geq (r + 1)(d - r)$).

In many situations, though, sampling is not uniform. For instance, in vision, occlusion of objects can produce missing data in very non-uniform random patterns. In cases like this, we can partition Ω (e.g., randomly) into matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^{r+1}$, each with $d - r$ columns. We can use Algorithm 1 below to determine whether each $\hat{\Omega}_\tau$ satisfies (ii). If this is the case, Ω is completable by Corollary 1. More about this is discussed in Section 3.

To present the algorithm, let us introduce the matrix \mathbf{A} that will allow us to determine efficiently whether a sampling $\check{\Omega}$ satisfies (ii). Let $\check{\Omega}$ be a matrix formed with $\check{N} \geq d - r$ columns of Ω , and let ω_i index the nonzero entries in the i^{th} column of $\check{\Omega}$. Let \mathbf{U} be a $d \times r$ matrix drawn according to $\nu_{\mathbf{U}}$, an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}^{d \times r}$, and let \mathbf{U}_{ω_i} denote the restriction of \mathbf{U} to the nonzero rows in ω_i . Let $\mathbf{a}_{\omega_i} \in \mathbb{R}^{r+1}$ be a nonzero vector in $\ker \mathbf{U}_{\omega_i}^T$, and \mathbf{a}_i be the vector in \mathbb{R}^d with the entries of \mathbf{a}_{ω_i} in the nonzero locations of ω_i and zeros elsewhere. Finally, let \mathbf{A} denote the $d \times \check{N}$ matrix with $\{\mathbf{a}_i\}_{i=1}^{\check{N}}$ as columns.

Algorithm 1 will verify whether $\dim \ker \mathbf{A}^T = r$, and this will determine whether $\check{\Omega}$ contains a $d \times (d - r)$ matrix $\hat{\Omega}$ satisfying (ii). The key insight behind Algorithm 1 is that \mathbf{A} encodes the information of the projections of $S = \text{span}\{\mathbf{U}\}$ onto the canonical coordinates indicated by $\check{\Omega}$. Theorem 1 in [21] shows that these projections will uniquely determine S if and only if $\dim \ker \mathbf{A}^T = r$, which will be the case if and only if $\check{\Omega}$ contains a $d \times (d - r)$ matrix $\hat{\Omega}$ satisfying (ii). We thus have the following corollary, which states that with probability 1, Algorithm 1 will determine whether $\check{\Omega}$ contains a matrix $\hat{\Omega}$ satisfying (ii).

Algorithm 1: Determine whether $\check{\Omega}$ contains a matrix $\hat{\Omega}$ satisfying (ii).

Input: Matrix $\check{\Omega}$ with $\check{N} \geq d - r$ columns of Ω .

- Draw $\mathbf{U} \in \mathbb{R}^{d \times r}$ according to $\nu_{\mathbf{U}}$.

- **for** $i = 1$ **to** \check{N} **do**

- ω_i = indices of the nonzero rows of the i^{th} column of $\check{\Omega}$.
 - \mathbf{a}_{ω_i} = nonzero vector in $\ker \mathbf{U}_{\omega_i}^{\text{T}}$.
 - \mathbf{a}_i = vector in \mathbb{R}^d with entries of \mathbf{a}_{ω_i} in the
 nonzero locations of ω_i and zeros elsewhere.

- \mathbf{A} = matrix formed with $\{\mathbf{a}_i\}_{i=1}^{\check{N}}$ as columns.

- **if** $\dim \ker \mathbf{A}^{\text{T}} = r$ **then**

- **Output:** $\check{\Omega}$ contains a $d \times (d - r)$ matrix $\hat{\Omega}$ satisfying (ii).

- **else**

- **Output:** $\check{\Omega}$ contains no $d \times (d - r)$ matrix $\hat{\Omega}$ satisfying (ii).

Corollary 2. Let $\check{\Omega}$ be a matrix formed with $\check{N} \geq d - r$ columns of Ω . Construct \mathbf{A} as in Algorithm 1. Then $\nu_{\mathbf{U}}$ -almost surely, $\check{\Omega}$ contains a $d \times (d - r)$ matrix $\hat{\Omega}$ satisfying (ii) if and only if $\dim \ker \mathbf{A}^{\text{T}} = r$.

Algorithm 1 can also be used to design completable samplings. As will be discussed in Section 3, this can be particularly useful for adaptive settings, where one may choose which entries to observe, yet it is undesirable or impossible to observe full columns or full rows.

Example 4. One may use Algorithm 1 to verify that each of the $r + 1$ blocks in the sampling matrix Ω below satisfies (ii). This implies that Ω satisfies the conditions of Corollary 1, and can thus be uniquely completed. In contrast with Example 1, this pattern does not sample full columns nor full rows.

$$\Omega = \underbrace{\left[\begin{array}{c|c|c|c} \overbrace{\begin{array}{c} d-r \\ \hline \mathbf{0} \end{array}} & \overbrace{\begin{array}{c} d-r \\ \hline \mathbf{0} \end{array}} & \cdots & \overbrace{\begin{array}{c} d-r \\ \hline \mathbf{0} \end{array}} \\ \hline \overbrace{\begin{array}{c} \mathbf{1} \\ \hline \mathbf{0} \end{array}} & \overbrace{\begin{array}{c} \mathbf{1} \\ \hline \mathbf{0} \end{array}} & \cdots & \overbrace{\begin{array}{c} \mathbf{1} \\ \hline \mathbf{0} \end{array}} \end{array} \right]}_{(r+1)(d-r)} \left. \begin{array}{l} \right\} d - r - 1 \\ \left. \right\} r + 1. \end{array}$$

3 Experiments and Implications

In this section we discuss implications of the results stated above and explore how well they predict performance in a series of simulation experiments. In all our experiments we use the so-called iterative hard-thresholded SVD (IHTSVD)

algorithm [12]. This algorithm iterates between truncating the SVD of the current estimate to a user-specified rank r , and then replacing the values in the observed entries with their original (observed) values. This algorithm is also quite similar to the Singular Value Thresholding algorithm [13], OptSpace [14] and FPCA [15]. In the very low sampling regimes of interest in our studies, we found the IHTSVD algorithm typically performed as well or better than several other completion algorithms (e.g., SVT [13], GROUSE [16], alternating minimization [17] and EM [18]).

Lower bound

It is easy to see that $\ell = \mathcal{O}(\max\{r, \log d\})$ uniformly randomly sampled entries per column are necessary to complete an $\mathcal{O}(d) \times \mathcal{O}(d)$ matrix. This is because a column with fewer than r observed entries cannot be completed, and if fewer than $\mathcal{O}(\log d)$ uniformly random samples per column are observed, then a row may be completely unobserved with large probability, making it impossible to complete a matrix. Thus, $\ell = \mathcal{O}(\max\{r, \log d\})$ is a lower bound for LRMC.

It was further shown [4] that there exist matrices that cannot be completed unless $\ell = \mathcal{O}(\mu r \log d)$ uniformly randomly sampled entries per column are observed, where $\mu \in [1, \frac{d}{r}]$ is the standard coherence parameter defined as

$$\mu := \frac{d}{r} \max_{1 \leq j \leq d} \|\mathbf{P}^* \mathbf{e}_j\|_2^2,$$

where \mathbf{P}^* denotes the projection operator onto S^* , and \mathbf{e}_j the j^{th} canonical vector in \mathbb{R}^d . Our results imply that this is only the case for a set of matrices with measure zero, and that a.e. matrix can be uniquely completed with as little as $\ell = \mathcal{O}(\max\{r, \log d\})$ uniformly randomly sampled samples per column, regardless of μ .

To better understand this, and see that our results do not contradict previous theory, let us revisit the proof of Theorem 1.7 in [4]. The proof is based on the construction of block-diagonal matrices with blocks of size $\frac{d}{r\mu}$ and coherence $\leq \mu$ that cannot be recovered with fewer than $\ell = \mathcal{O}(\mu r \log d)$ uniformly random samples per column, e.g.,

$$\mathbf{X} = \left[\begin{array}{c|c} \overbrace{\mathbf{B}_1}^{\frac{d}{r\mu}} & \mathbf{0} \\ \hline \mathbf{0} & \underbrace{\mathbf{B}_{r\mu}}_{\frac{d}{r\mu}} \end{array} \right] \left. \vphantom{\begin{array}{c|c} \mathbf{B}_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B}_{r\mu} \end{array}} \right\} d. \quad (4)$$

This is so because zero valued entries provide no information for the reconstruction process. It follows that the larger μ , the smaller the blocks will be, and more intensive random sampling would be required to guarantee that entries in

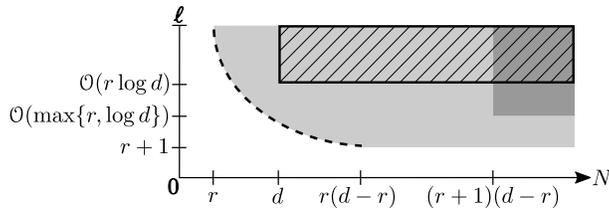


Figure 2: Theoretical sampling regimes of LRMC. In the white region, where the dashed line is given by $\ell = \frac{r(d-r)}{N} + r$, it is easy to see that LRMC is impossible by a simple count of the degrees of freedom in a subspace (see Section 4). In the light-gray region, LRMC is possible provided the entries are observed in the right places, e.g., satisfying the conditions of Theorem 2. By Theorem 3, uniform random samplings will satisfy these conditions with high probability as long as $N \geq (r+1)(d-r)$ and $\ell \geq \max\{12(\log(\frac{d}{\epsilon}) + 1), 2r\}$, hence with high probability LRMC is possible in the dark-grey region. Previous analyses showed that LRMC is possible from uniform random sampling in the striped region [3], but the rest remained unclear until now.

the diagonal blocks are observed. This is why more samples ($\mathcal{O}(r\mu \log d)$ per column) are required to reconstruct more coherent matrices *like this one*, and hence the dependency on $r\mu$ in the bound of Theorem 1.7 in [4].

However, matrices with this block structure have measure zero (with respect to the measure defined above). Our results show that for a.e. matrix, an incomplete column contains the same exploitable information regardless of the coherence parameter, and $\mathcal{O}(\max\{r, \log d\})$ uniform random entries per column are sufficient for completion. This means that while there are some matrices that require $\mathcal{O}(r\mu \log d)$ uniform random samples per column for reconstruction, a.e. matrix only requires $\mathcal{O}(\max\{r, \log d\})$, regardless of μ .

Sample Complexity

Coherence aside, it is also known that $N = d$ columns, and $\ell = \mathcal{O}(r \log d)$ uniform random samples per column are sufficient for completion [3]. Theorem 3 extends this result, showing that $N = (r+1)(d-r)$ columns and $\ell = \mathcal{O}(\max\{r, \log d\})$ uniform random samples per column are sufficient to uniquely complete a.e. matrix. This exposes an interesting tradeoff between the required number of columns and observed entries per column for completion, defining new unstudied sampling regimes where completion is now known to be possible (Figure 2).

The purpose of our first experiment is to support that $\ell = \mathcal{O}(\max\{r, \log d\})$ random samples per column are truly sufficient for LRMC, as opposed to $\mathcal{O}(r \log d)$. To this end, we will study the behavior of the IHTSVD algorithm as a function of the ambient dimension d and the rank r (see the beginning of Section 3 for a discussion of this algorithmic choice).

To obtain low-rank matrices, we first generated a $d \times r$ random matrix \mathbf{U}^* with $\mathcal{N}(0, 1)$ i.i.d. entries to use as basis of S^* . We then generated an $r \times (r+1)(d-r)$ random matrix Θ^* , also with $\mathcal{N}(0, 1)$ i.i.d. entries, to use as coefficient vectors, to construct $\mathbf{X} = \mathbf{U}^* \Theta^*$. Matrices generated this way are known to have low coherence.

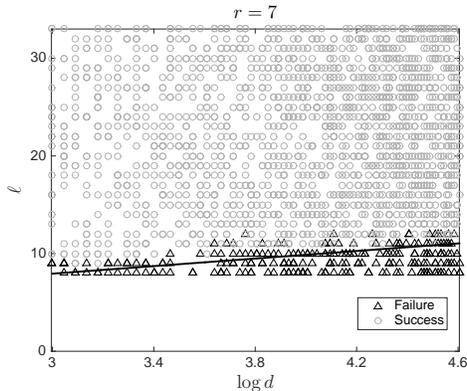


Figure 3: Results of the IHTSVD algorithm as a function of the ambient dimension d and the number of uniform random samples per column ℓ , for rank $r = 7$. In each of the 2,000 trials we declared a success if the normalized completion error was below 10^{-12} (using normalized Frobenius norm). The black line represents the linear discriminant between success and failure trials.

Next, for different values of the rank r , we tested whether a matrix could be completed as a function of its ambient dimension d and the number of uniform random samples per column ℓ . For example, the results of this experiment for $r = 7$ can be seen in Figure 3.

We then computed the linear discriminant between successful and failure trials for each value of r . If $\ell = \mathcal{O}(r \log d)$ samples were necessary, we would expect the slope between these lines to grow proportionally to r . However, the results, depicted in Figure 4, show that the slope of these lines remain fairly constant, and the offset grows with r , supporting that $\ell = \mathcal{O}(\max\{r, \log d\})$ samples are sufficient.

Computational Complexity

Our results show that completion is theoretically possible with as little as with $\ell \geq \mathcal{O}(\max\{r, \log d\})$ uniform random samples per column, or even with as little as $\ell = r + 1$ (provided they are located in the right places). Nevertheless, this may involve solving the system of polynomial equations $\mathcal{F} = \mathbf{0}$ (see Section 4), which is computationally impractical. It is thus currently unknown whether there exist practical completion algorithms for these uncharted sampling regimes.

We now present a series of experiments that suggest three things: first, that even in cases where LRMC is theoretically possible, missingness seems to come at a price: the more missing data the more computationally expensive completion seems to be. This further suggests that there is a minimal sampling regime where, though theoretically possible, LRMC might be computationally prohibitive in practice. Second, that even though theoretically, whether a.e. matrix can be completed does not depend on its coherence, in practice, extremely coherent matrices may be computationally more expensive to complete. Sim-

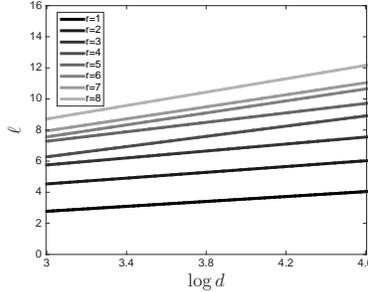
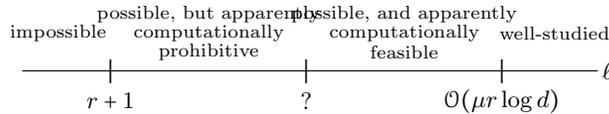


Figure 4: Linear discriminants for different values of the rank r , between successful (above line) and unsuccessful (below line) completions for the experiment in Figure 3. That is, for a given r , any pair $(\log d, \ell)$ above the linear discriminant typically succeeds at completion, and below the linear discriminant typically fails. Theorem 3 shows that $\ell = \mathcal{O}(\max\{r, \log d\})$ uniform random observations per column are sufficient for completion. The slope of these lines remain fairly constant, and the offset grows with r , supporting this result.

ilarly, this suggests that there is a maximal coherence regime where, though theoretically possible, LRMC might be computationally prohibitive in practice. And third, there seems to be an additional uncharted sampling regime with $\ell < \mathcal{O}(\mu r \log d)$ samples per column where completion is computationally feasible.

To summarize, we have the following sampling regimes, where LRMC is:



We first study the computational cost of missing data. To this end we computed the minimum number of iterations required to complete a matrix, as a function of the number of uniform random samples per column ℓ . The results are summarized in Figure 5. Unsurprisingly, the more missing data, the more iterations are required to complete the matrix.

In addition, we constructed samplings Ω with only $\ell = r + 1$ samples per column selected uniformly at random, and kept only those samplings satisfying the conditions of Corollary 1, to guarantee that \mathbf{X}_Ω were uniquely completable (we used Algorithm 1 to determine whether each sampling satisfied these conditions). Unfortunately, even though \mathbf{X}_Ω was uniquely completable, the matrix was incorrectly completed in every single trial. This suggests that completion in this regime, now known to be theoretically possible (through the solution of the polynomial system $\mathcal{F} = \mathbf{0}$; see Section 4), might be computationally prohibitive in practice.

We thus tested how much missing data can practical algorithms handle while remaining computationally efficient. To this end, we sampled $\ell < \mu r \log d$ entries per column, drawn uniformly at random, and ran the IHTSVD algorithm for

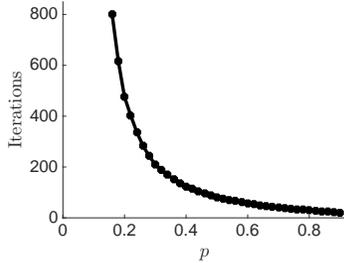


Figure 5: Average number of iterations (over 500 trials) required by IHTSVD to complete a matrix with low coherence ($\mu < 3$) with an accuracy of 10^{-12} (using normalized Frobenius norm), as a function of $p := \ell/d$, the proportion of uniform random samples per column, with ambient dimension $d = 500$ and rank $r = 10$.

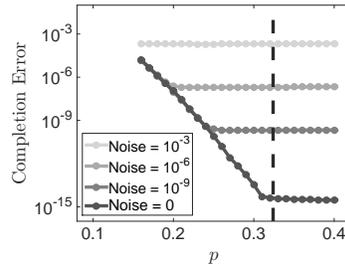


Figure 6: Average completion error of IHTSVD (over 500 trials) for different levels of additive i.i.d. zero-mean Gaussian noise (noise variance as indicated in the legend), after at most 250 iterations, as a function of $p := \ell/d$, the proportion of uniform random samples per column, with ambient dimension $d = 500$ and rank $r = 10$. Previous guarantees would require all entries to be observed, and so in practice, existing theory would not allow one to confirm the correctness of a completion. Our results do. The dashed line represents $p = \max\{12(\log(\frac{d}{\epsilon}) + 1), 2r\}/d$, with $\epsilon = \frac{1}{d}$, the sufficient condition of Theorem 3, which implies that with probability at least $1 - \epsilon$, for any p above this threshold, a rank- r completion is guaranteed to be correct, regardless of the completion method.

at most $T = d$ iterations (see the beginning of Section 3 for a discussion of this algorithmic choice).

To truly test this regime, we considered a setup where previous theory would require all entries to be observed to guarantee a correct completion with probability at least $1 - \epsilon$. There are plenty of such scenarios. We arbitrarily selected $d = 500$ and $r = 10$, and $\epsilon = 1/d$.

Our simulations, summarized in Figure 6, show that practical algorithms tend to work consistently well with $\ell < \mu r \log(\frac{d}{\epsilon})$. This suggests that there is a regime with $\ell < \mu r \log(\frac{d}{\epsilon})$ samples per column where completion is computationally feasible, even in the presence of noise.

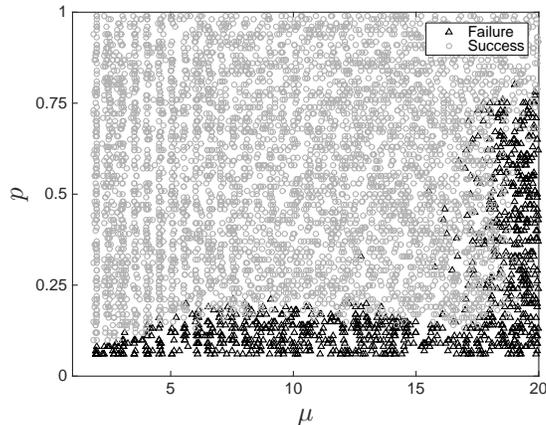


Figure 7: Results of the IHTSVD algorithm as a function of the coherence parameter $\mu \in [1, \frac{d}{r}]$ and the proportion of uniform random samples per column $p := \ell/d$, with ambient dimension $d = 100$ and rank $r = 5$. Similar results were observed for other algorithms, including alternating minimization [17] and EM [18]. In each of the 5,000 trials we declared a success if the normalized completion error was below 10^{-12} (using normalized Frobenius norm). Our theoretical results show that whether a.e. matrix can be uniquely completed does not depend on its coherence. This experiment suggests that in practice, this is also the case for most of the range of μ . For instance, given p , the success rate of this algorithm is about the same for most of the range of μ (about $1 \leq \mu \leq 17$). Nevertheless, the success rate quickly decays if the coherence is extremely high (μ close to the maximum possible, $\frac{d}{r} = 20$).

Dependence on Coherence Parameter

In our next experiment, we study the practical role of coherence in LRMC. More precisely, we tested whether a matrix could be computationally efficiently completed as a function of its coherence parameter μ , and the number of uniform random samples per column ℓ (to generate matrices with a specific coherence parameter, we simply increased the magnitude of a few entries in \mathbf{U}^* , until it had the desired coherence). The results, summarized in Figure 7, suggest that for most of the coherence range, whether this algorithm can correctly complete the matrix mainly depends on the number of samples rather than on the coherence parameter. For instance, see in Figure 7 that given the number of samples, the success rate of this algorithm is about the same for most of the range of μ . Nonetheless, there are some cases with extremely large coherence (μ close to the maximum possible, $\frac{d}{r}$, corresponding to subspaces almost perfectly aligned with the canonical axes), where this algorithm tends to fail more often at reconstructing the matrix (these are cases where most of the information is concentrated in only a few entries, which brings computational and numerical accuracy problems).

To further study the role of coherence in practice, we recorded the number of iterations that were required to complete each matrix (in the success cases of the previous experiment). The results, summarized in Figure 8, suggest

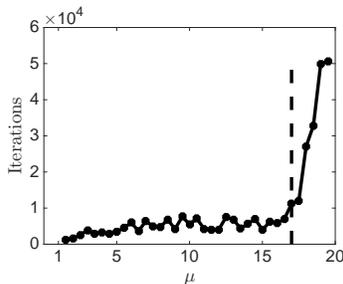


Figure 8: Average number of iterations (of the success trials from Figure 7) required by IHTSVD to complete a matrix with an accuracy of 10^{-12} (using normalized Frobenius norm), as a function of its coherence parameter μ . This suggests the existence of a maximal coherence regime (e.g., after the dashed line) where, though theoretically possible, completion may become computationally impractical.

that while coherent matrices may be theoretically as completable as incoherent ones, in practice, the more coherent a matrix is, the more computationally expensive it may be to complete it. Furthermore, the number of iterations seems to increase steadily for most of the coherence range, but after a transition point it suddenly seems to grow exponentially, suggesting the existence of a maximal coherence regime where, though theoretically possible, completion may be computationally impractical (similar to the minimal sampling regime from our previous experiment).

New Guarantees

It is known that $\mathcal{O}(\mu r \log d)$ uniform random samples per column (with constants greater than 1) are sufficient for completion [3]. There are non-pathological regimes (e.g., $d = 500$ and $r = 10$, or $d = 100$ and $r = 5$, and ideal coherence, as in our experiments) where these conditions end up requiring that all entries are observed. Experiments show that the IHTSVD algorithm can exactly complete such matrices when even fewer than half of the entries are observed, but prior theory gives no guarantees in these regimes, and so in practice, one would be unable to confirm the correctness of a completion.

Furthermore, typical conditions for LRMC usually apply to matrices with bounded coherence, and require uniform random sampling with rates that depend on the coherence parameter μ . In many practical applications, sampling is hardly uniform (e.g., vision, where occlusion of objects produce missing data in very non-uniform random patterns), and μ is typically unknown, so the existing theory does not allow one to confirm the correctness of a completion.

Our results shed new light on these issues. Theorem 2 states that regardless of coherence and the sampling model, if the observation pattern satisfies the conditions of the theorem, a rank- r completion, obtained by any method whatsoever, is guaranteed to be *the* correct completion. In particular, Theorem 3 states that this will be the case with high probability under uniform random

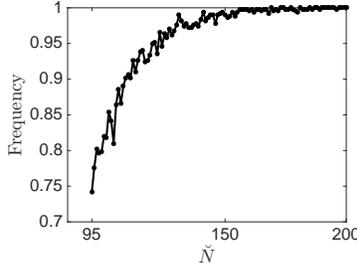


Figure 9: We generated $d \times \check{N}$ matrices $\check{\Omega}$ with only $r+1$ samples per column, selected uniformly at random, with $d = 100$, $r = 5$. This figure shows the proportion of times (over 500 trials) that $\check{\Omega}$ contains a $d \times (d-r)$ matrix $\hat{\Omega}$ satisfying (ii), as a function of \check{N} . We used Algorithm 1 to determine whether this was the case. Notice that as \check{N} grows, the probability that each $\check{\Omega}$ contains an $\hat{\Omega}_\tau$ satisfying (ii) quickly approaches 1.

sampling models.

In some cases one can use Corollary 1 together with Algorithm 1 to verify efficiently and deterministically whether these conditions are satisfied. Recall that Corollary 1 states that unique completability is possible if Ω contains $r+1$ disjoint matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^{r+1}$, each of size $d \times (d-r)$ satisfying (ii). Given a matrix $\hat{\Omega}_\tau$, Algorithm 1 allows to verify whether it satisfies (ii). However, it provides no means to select the $\hat{\Omega}_\tau$'s.

In general, one can construct samplings for which finding the right $\hat{\Omega}_\tau$'s would require exponential time. However, if the samples are well spread across the rows, then one may validate a completion deterministically by selecting the $\hat{\Omega}_\tau$'s randomly.

To see this, suppose Ω has $(r+1)\check{N}$ columns, with $\check{N} \geq d-r$ and exactly $r+1$ observations per column (see Remark 1). We can randomly partition Ω into $r+1$ disjoint submatrices $\{\check{\Omega}_\tau\}_{\tau=1}^{r+1}$, each of size \check{N} . One can then use Algorithm 1 to verify whether each $\check{\Omega}_\tau$ contains an $d \times (d-r)$ submatrix $\hat{\Omega}_\tau$ satisfying (ii). If this is the case, then we know deterministically that the completion is correct.

Figure 9 shows that as \check{N} grows, the probability that each $\check{\Omega}_\tau$ contains an $\hat{\Omega}_\tau$ satisfying (ii) quickly approaches 1. For example, with \check{N} as small as $2(d-r)$, i.e., with only twice as many columns as strictly necessary, each $\check{\Omega}_\tau$ will contain an $\hat{\Omega}_\tau$ satisfying (ii) with probability larger than .999. This suggests that if we find a low-rank completion of a matrix, we can expect that a random partition will certify it through Corollary 1 and Algorithm 1.

This way, our results can be used to certify the correctness of a completion, thus bringing guarantees applicable to any algorithm, under any sampling model, in lieu of coherence assumptions.

Adaptive Sampling

If one could select which entries of \mathbf{X} to observe, perhaps the easiest way to recover \mathbf{X} is to sample r linearly independent columns to obtain a basis of

the subspace, and then r rows to obtain the coefficients of each column in this basis. However, in many LRMC applications, the entries one may observe can be limited. Take for example recommender systems, where obtaining a complete column equates to asking a single user (column) to evaluate every item (row). In these problems the number of rows can be very large, hence this can be an unreasonable thing to ask. Moreover, the combinations of rows that one may sample could be restricted. An other example arises in distributed settings, where at each location one may only sample certain subsets of all the information.

Our results tell us exactly which entries to look for. Furthermore, it is fairly simple to construct sampling patterns that satisfy the conditions of Theorems 1 and 2 and Corollary 1 that do not require to sample full columns or rows. For instance, we can generate random samplings, use Algorithm 1 to verify whether they satisfy condition (ii) (most of them will; see Figure 9), and keep them or discard them depending on this.

Deterministic constructions are also possible. For instance, it is easy to verify that each of the blocks in Example 4 satisfies (ii), which implies Ω satisfies the conditions of Corollary 1, and can thus be uniquely completed. This example corresponds to asking the i^{th} user of each block to rate items i through $i+r$. We conclude that if the entries one may choose to observe are limited (as is the case in many LRMC applications), one can directly apply our results to adaptively design observation patterns that guarantee completability.

4 Proof of Theorem 1

For any subspace, matrix or vector that is compatible with a set of indices ω , we will use the subscript ω to denote its restriction to the coordinates/rows in ω . For example, letting ω_i denote the indices of the nonzero rows of the i^{th} column of Ω , then $\mathbf{x}_{\omega_i} \in \mathbb{R}^{r+1}$ and $S_{\omega_i}^* \subset \mathbb{R}^{r+1}$ denote the restrictions of the i^{th} column in \mathbf{X} and S^* , to the indices in ω_i . We say that an r -dimensional subspace S fits \mathbf{X}_Ω if $\mathbf{x}_{\omega_i} \in S_{\omega_i} \forall i$.

The Variety \mathcal{S}

Let us start by studying the variety of all r -dimensional subspaces that fit \mathbf{X}_Ω . First observe that in general, the restriction of an r -dimensional subspace to $\ell \leq r$ coordinates is \mathbb{R}^ℓ . We formalize this in the following definition, which essentially states that a subspace is non-degenerate if its restrictions to $\ell \leq r$ coordinates are \mathbb{R}^ℓ .

Definition 1 (Degenerate subspace). *We say $S \in \text{Gr}(r, \mathbb{R}^d)$ is degenerate if and only if there exists a set $\omega \subset \{1, \dots, d\}$ with $|\omega| \leq r$, such that $\dim S_\omega < |\omega|$.*

Let ν_G denote the uniform measure on $\text{Gr}(r, \mathbb{R}^d)$. A subspace is degenerate if and only if an $r \times r$ submatrix of one of its bases is rank-deficient. This is

equates to having a zero determinant. Since the determinant is a polynomial in the entries of a matrix, this is a condition of ν_G -measure zero.

Since ν_G -almost every subspace is non-degenerate, let us consider only the subspaces in $\text{Gr}_*(r, \mathbb{R}^d) \subset \text{Gr}(r, \mathbb{R}^d)$, the set of all non-degenerate r -dimensional subspaces of \mathbb{R}^d .

Define $\mathcal{S}(\mathbf{X}_\Omega) \subset \text{Gr}_*(r, \mathbb{R}^d)$ such that every $S \in \mathcal{S}(\mathbf{X}_\Omega)$ fits \mathbf{X}_Ω , i.e.,

$$\mathcal{S}(\mathbf{X}_\Omega) := \left\{ S \in \text{Gr}_*(r, \mathbb{R}^d) : \{\mathbf{x}_{\omega_i} \in S\}_{i=1}^N \right\}.$$

Let $\mathbf{U} \in \mathbb{R}^{d \times r}$ be a basis of $S \in \mathcal{S}(\mathbf{X}_\Omega)$. The condition $\mathbf{x}_{\omega_i} \in S_{\omega_i}$ is equivalent to saying that there exists a vector $\boldsymbol{\theta}_i \in \mathbb{R}^r$ such that

$$\mathbf{x}_{\omega_i} = \mathbf{U}_{\omega_i} \boldsymbol{\theta}_i. \quad (5)$$

We can see that if \mathbf{x}_{ω_i} has fewer than r observations, (5) will be an underdetermined system with infinitely many solutions, and hence \mathbf{x}_{ω_i} can be completed in infinitely many ways.

If \mathbf{x}_{ω_i} has exactly r observations, (5) becomes a system with r equations and r unknowns (the elements of $\boldsymbol{\theta}_i$). This will be the case for every $S \in \text{Gr}_*(r, \mathbb{R}^d)$. Hence a column with exactly r observations can be uniquely completed once S^* is known, but it provides no information to identify S^* .

On the other hand, if \mathbf{x}_{ω_i} has exactly $r + 1$ observations, then (5) becomes an overdetermined system with $r + 1$ equations and r unknowns. This imposes one constraint on the elements of \mathbf{U}_{ω_i} , thus restricting the set of subspaces that fit \mathbf{x}_{ω_i} .

In general, each column with $r + 1$ observations will impose one constraint that may reduce one of the $r(d - r)$ degrees of freedom in $\text{Gr}_*(r, \mathbb{R}^d)$. Therefore, one necessary condition for completion is that \mathbf{X}_Ω imposes at least $r(d - r)$ constraints.

We will now study these constraints and characterize when exactly will they reduce all the $r(d - r)$ degrees of freedom in $\text{Gr}_*(r, \mathbb{R}^d)$, thus restricting $\mathcal{S}(\mathbf{X}_\Omega)$ to a set with at most finitely many elements.

Let $\{\Delta_i, \nabla_i\}$ be a partition of the $r + 1$ elements of ω_i , such that Δ_i has exactly r elements, and ∇_i has only one element. We can then expand (5) as

$$\begin{matrix} r \\ 1 \end{matrix} \left\{ \begin{matrix} \left[\mathbf{x}_{\Delta_i} \right] \\ \left[\mathbf{x}_{\nabla_i} \right] \end{matrix} \right\} = \left[\begin{matrix} \mathbf{U}_{\Delta_i} \\ \mathbf{U}_{\nabla_i} \end{matrix} \right] \boldsymbol{\theta}_i.$$

Since S is non-degenerate, \mathbf{U}_{Δ_i} is full-rank, so we may solve for $\boldsymbol{\theta}_i$ using the top block to obtain $\boldsymbol{\theta}_i = \mathbf{U}_{\Delta_i}^{-1} \mathbf{x}_{\Delta_i}$. Plugging this on the last row, we have that (5) is equivalent to:

$$\mathbf{x}_{\nabla_i} = \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^{-1} \mathbf{x}_{\Delta_i}. \quad (6)$$

On the other hand, \mathbf{x}_{ω_i} lies in $S_{\omega_i}^*$ by assumption. This implies that there exists a unique $\boldsymbol{\theta}_i^* \in \mathbb{R}^r$ such that

$$\mathbf{x}_{\omega_i} = \mathbf{U}_{\omega_i}^* \boldsymbol{\theta}_i^*, \quad (7)$$

where \mathbf{U}^* is a basis of S^* . Substituting (7) in (6) we obtain

$$\mathbf{U}_{\nabla_i}^* \boldsymbol{\theta}_i^* = \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^{-1} \mathbf{U}_{\Delta_i}^* \boldsymbol{\theta}_i^*. \quad (8)$$

Recall that $\mathbf{U}_{\Delta_i}^{-1} = \mathbf{U}_{\Delta_i}^\dagger / |\mathbf{U}_{\Delta_i}|$, where $\mathbf{U}_{\Delta_i}^\dagger$ and $|\mathbf{U}_{\Delta_i}|$ denote the adjugate and the determinant of \mathbf{U}_{Δ_i} . Therefore, we may rewrite (8) as the following polynomial equation:

$$\left(|\mathbf{U}_{\Delta_i}| \mathbf{U}_{\nabla_i}^* - \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^\dagger \mathbf{U}_{\Delta_i}^* \right) \boldsymbol{\theta}_i^* = 0. \quad (9)$$

We conclude that a subspace S with basis \mathbf{U} fits \mathbf{X}_Ω if and only if \mathbf{U} satisfies (9) for every $i = 1, \dots, N$.

Since every nontrivial subspace has infinitely many bases, even if there is only one r -dimensional subspace in $\mathcal{S}(\mathbf{X}_\Omega)$, the variety

$$\left\{ \mathbf{U} \in \mathbb{R}^{d \times r} : \left(|\mathbf{U}_{\Delta_i}| \mathbf{U}_{\nabla_i}^* - \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^\dagger \mathbf{U}_{\Delta_i}^* \right) \boldsymbol{\theta}_i^* = 0 \quad \forall i \right\}$$

has infinitely many solutions. Therefore, we will associate a unique \mathbf{U} with each subspace as follows. Observe that for every $S \in \text{Gr}_*(r, \mathbb{R}^d)$, we can write $S = \text{span}\{\mathbf{U}\}$ for a unique \mathbf{U} in the following column echelon form:

$$\mathbf{U} = \left[\begin{array}{c} \mathbf{I} \\ \mathbf{V} \end{array} \right] \left. \begin{array}{l} \left. \vphantom{\begin{array}{c} \mathbf{I} \\ \mathbf{V} \end{array}} \right\} r \\ \left. \vphantom{\begin{array}{c} \mathbf{I} \\ \mathbf{V} \end{array}} \right\} d-r. \end{array} \quad (10)$$

On the other hand, every $\mathbf{V} \in \mathbb{R}^{(d-r) \times r}$ defines a unique r -dimensional subspace of \mathbb{R}^d , via $\text{span}\{\mathbf{U}\}$. Moreover, $\text{span}\{\mathbf{U}\}$ will be non-degenerate for almost every \mathbf{V} , with respect to ν_V : the Lebesgue measure on $\mathbb{R}^{(d-r) \times r}$. Let $\mathbb{R}_*^{(d-r) \times r} \subset \mathbb{R}^{(d-r) \times r}$ denote the set of all $(d-r) \times r$ matrices \mathbf{V} whose $\text{span}\{\mathbf{U}\}$ is non-degenerate, or equivalently, whose $r \times r$ submatrices of \mathbf{U} are full-rank. Then we have a bijection between $\text{Gr}_*(r, \mathbb{R}^d)$ and $\mathbb{R}_*^{(d-r) \times r}$ via $S^* = \text{span}\{\mathbf{U}\}$. It follows that a statement holds for $(\nu_G \times \nu_\Theta)$ -almost every pair $\{S^*, \boldsymbol{\Theta}^*\}$ if and only if it holds for $(\nu_V \times \nu_\Theta)$ -almost every pair $\{\mathbf{V}^*, \boldsymbol{\Theta}^*\}$. We will use these measures interchangeably.

The Set \mathcal{F}

Continuing with our analysis, recall that a subspace S with basis \mathbf{U} will fit \mathbf{X}_Ω if and only if \mathbf{U} satisfies (9) for every i . With this in mind, define

$$f_i(\mathbf{V}|\mathbf{V}^*, \boldsymbol{\theta}_i^*) := \left(|\mathbf{U}_{\Delta_i}| \mathbf{U}_{\nabla_i}^* - \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^\dagger \mathbf{U}_{\Delta_i}^* \right) \boldsymbol{\theta}_i^*,$$

with \mathbf{U} and \mathbf{U}^* in the column echelon form in (10). We will use f_i as shorthand, with the understanding that f_i is a polynomial in the elements of \mathbf{V} , and that the elements of \mathbf{V}^* and $\boldsymbol{\theta}_i^*$ play the role of coefficients.

Furthermore, let

$$\mathcal{F}(\mathbf{V}|\mathbf{V}^*, \boldsymbol{\Theta}^*) := \{f_i\}_{i=1}^N,$$

and use $\mathcal{F}(\mathbf{V})$, or simply \mathcal{F} as shorthand, with the understanding that \mathcal{F} is a set of polynomials in the elements of \mathbf{V} , and that the elements of \mathbf{V}^* and $\boldsymbol{\Theta}^*$ play the role of coefficients. We will also use $\mathcal{F} = \mathbf{0}$ as shorthand for $\{f_i = 0\}_{i=1}^N$.

This way, we may rewrite:

$$\mathcal{S}(\mathbf{X}_\Omega) = \left\{ \text{span} \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix} \in \text{Gr}_*(r, \mathbb{R}^d) : \mathcal{F}(\mathbf{V}) = \mathbf{0} \right\}.$$

In general, the affine variety

$$\mathcal{V}(\mathcal{F}) := \left\{ \mathbf{V} \in \mathbb{R}_*^{(d-r) \times r} : \mathcal{F}(\mathbf{V}) = \mathbf{0} \right\}$$

could contain an infinite number of elements. We are interested in conditions that guarantee there is only one or (slightly less demanding) only a finite number. The following lemma states that this will be the case if and only if $r(d-r)$ polynomials in \mathcal{F} are algebraically independent.

Lemma 2. *Let A1 hold. For a.e. \mathbf{X} , $\mathcal{S}(\mathbf{X}_\Omega)$ contains at most finitely many subspaces if and only if $r(d-r)$ polynomials in \mathcal{F} are algebraically independent.*

Proof. By our previous discussion, for a.e. \mathbf{X} there are at most finitely many subspaces in $\mathcal{S}(\mathbf{X}_\Omega)$ if and only if there are at most finitely many points in $\mathcal{V}(\mathcal{F})$. We know from algebraic geometry that this will be the case if and only if $\dim \mathcal{V}(\mathcal{F}) = 0$ (see, e.g., Proposition 6 in Chapter 9, Section 4 of [19]).

Since $\mathcal{V}(\mathcal{F}) \subset \mathbb{R}_*^{(d-r) \times r}$, we know that if $\dim \mathcal{V}(\mathcal{F}) = 0$, then \mathcal{F} must contain $r(d-r)$ algebraically independent polynomials (see, e.g., Exercise 16 in Chapter 9, Section 6 of [19]).

On the other hand, we know that $\dim \mathcal{V}(\mathcal{F}) = 0$ if $r(d-r)$ polynomials in \mathcal{F} are a regular sequence (see, e.g., Exercise 8 in Chapter 9, Section 4 of [19]).

Finally, since being a regular sequence is an open condition, it follows that for $(\nu_{\mathbf{V}} \times \nu_{\boldsymbol{\Theta}})$ -almost every $\{\mathbf{V}^*, \boldsymbol{\Theta}^*\}$, polynomials in \mathcal{F} are algebraically independent if and only if they are a regular sequence (see, e.g., Remark 3.4 in [20]). \square

Remark 2. *The next part of our analysis studies conditions to guarantee that the polynomials in \mathcal{F} are algebraically independent. Following up on Remark 1, any observation, in addition to the $r+1$ per column that we assume, cannot increase the number of subspaces that agree with the observations. In effect, each observed entry, in addition to the first $r+1$ observations, places one additional polynomial constraint analogous to f_i . However, the polynomials produced by the same column share the same coefficient $\boldsymbol{\theta}_i^*$. Intuitively, this means that the polynomials are no longer generic. While these polynomials might or might not*

be algebraically dependent, in general it is difficult to determine which is the case.

For this reason we assume **A1**: that each column is observed on exactly $r+1$ entries. This way the $r+1$ entries in each column produce only one polynomial constraint. This guarantees that we only use one polynomial per column, so that all the coefficients of the polynomials in \mathcal{F} are generic, and easier to study. In general, if some columns are observed on more than $r+1$ entries, all we need is that the observed entries contain a pattern with exactly $r+1$ observations per column satisfying our sampling conditions.

Algebraic Independence

By the previous discussion, there are at most finitely many r -dimensional subspaces that fit \mathbf{X}_Ω if and only if there is a subset $\tilde{\mathcal{F}}$ of $r(d-r)$ polynomials in \mathcal{F} that is algebraically independent.

Whether this is the case depends on the supports of the polynomials in $\tilde{\mathcal{F}}$, i.e., on $\tilde{\Omega}$: the subset of columns in Ω corresponding to such polynomials. Lemma 3 shows that the polynomials in $\tilde{\mathcal{F}}$ will be algebraically independent if and only if $\tilde{\Omega}$ satisfies the conditions in Theorem 1.

Lemma 3. *Let **A1** hold. For a.e. \mathbf{X} , the polynomials in $\tilde{\mathcal{F}}$ are algebraically dependent if and only if $n(\Omega') > r(m(\Omega') - r)$ for some matrix Ω' formed with a subset of the columns in $\tilde{\Omega}$.*

To show this statement we will use Lemmas 4 and 5 below.

Let Ω' be a subset of the columns in $\tilde{\Omega}$, and let \mathcal{F}' be the subset of the $n(\Omega')$ polynomials in $\tilde{\mathcal{F}}$ corresponding to such columns. Notice that \mathcal{F}' only involves the variables in \mathbf{U} corresponding to the $m(\Omega')$ nonzero rows of Ω' .

Let $\varkappa(\Omega')$ be the largest number of algebraically independent polynomials in \mathcal{F}' .

Lemma 4. *For a.e. \mathbf{X} , $\varkappa(\Omega') \leq r(m(\Omega') - r)$.*

Proof. Observe that the column echelon form in (10) was chosen arbitrarily. As a matter of fact, for every permutation of rows Π and every $S \in \text{Gr}_*(r, \mathbb{R}^d)$, we may write $S = \text{span}\{\mathbf{U}\}$, for a unique \mathbf{U} in the following permuted column echelon form:

$$\mathbf{U} = \Pi \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix}.$$

For example, we could take Π to swap the top and bottom blocks in (10), and take \mathbf{U} in the following form:

$$\mathbf{U} = \Pi \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{V} \\ \mathbf{I} \end{bmatrix}.$$

Observe that in general, \mathbf{U} , \mathbf{V} and \mathcal{F} will be different for each choice of $\mathbf{\Pi}$. Nevertheless, the condition $\mathbf{x}_{\omega_i} \in S_{\omega_i}$ is invariant to the choice of basis of S . This implies that while different choices of $\mathbf{\Pi}$ produce different \mathcal{F} 's, the variety

$$\mathcal{S}(\mathbf{X}_{\Omega}) = \left\{ \text{span } \mathbf{\Pi} \begin{bmatrix} \mathbf{I} \\ \mathbf{V} \end{bmatrix} \in \text{Gr}_*(r, \mathbb{R}^d) : \mathcal{F}(\mathbf{V}) = \mathbf{0} \right\}$$

is the same for every $\mathbf{\Pi}$.

This implies that the number of algebraically independent polynomials in \mathcal{F}' is invariant to the choice of $\mathbf{\Pi}$. Therefore, showing that Lemma 4 holds for one particular $\mathbf{\Pi}$ suffices to show that it holds for every $\mathbf{\Pi}$.

With this in mind, take $\mathbf{\Pi}$ such that \mathbf{U} is written with the identity block in the position of r nonzero rows of Ω' .

Since the polynomials in \mathcal{F}' only involve the elements of the $m(\Omega')$ rows of \mathbf{U} corresponding to the nonzero rows of Ω' , and \mathbf{U} has the identity block in the position of r nonzero rows of Ω' , it follows that the polynomials in \mathcal{F}' only involve the $r(m(\Omega') - r)$ variables in the $m(\Omega') - r$ corresponding rows of \mathbf{V} . Furthermore, $\mathcal{F}' = \mathbf{0}$ has at least one solution. This implies $\aleph(\Omega') \leq r(m(\Omega') - r)$, as desired. \square

We say \mathcal{F}' is *minimally algebraically dependent* if the polynomials in \mathcal{F}' are algebraically dependent, but every proper subset of the polynomials in \mathcal{F}' is algebraically independent.

Lemma 5. *For a.e. \mathbf{X} , if \mathcal{F}' is minimally algebraically dependent, then $n(\Omega') = r(m(\Omega') - r) + 1$.*

In order to prove Lemma 5 we will need the next two lemmas.

Lemma 6. *Take $\mathbf{\Pi}$ such that $\mathbf{U}_{\Delta_i} = \mathbf{U}_{\Delta_i}^* = \mathbf{I}$. For a.e. \mathbf{X} , if $\mathcal{F}' = \{\mathcal{F}'', f_i\}$ is minimally algebraically dependent, then all solutions to $\mathcal{F}' = \mathbf{0}$ satisfy $\mathbf{U}_{\nabla_i} = \mathbf{U}_{\nabla_i}^*$.*

The intuition behind this lemma is as follows: suppose for contrapositive that there are infinitely many solutions \mathbf{U}_{ω_i} to $\mathcal{F}'' = \mathbf{0}$ with $\mathbf{U}_{\Delta_i} = \mathbf{I}$. Each of these solutions defines a different subspace. Since $\{\mathcal{F}'', f_i\}$ is minimally algebraically dependent, a.e. solution to \mathcal{F}'' must fit \mathbf{x}_{ω_i} . This will only happen if \mathbf{x}_{ω_i} lies in the intersection of infinitely many r -dimensional subspaces, which is at most $(r-1)$ -dimensional. But since \mathbf{x}_{ω_i} is drawn from S^* (an r -dimensional subspace), we know that almost surely \mathbf{x}_{ω_i} will not lie in such $(r-1)$ -dimensional subspace.

Proof. Suppose that $\mathcal{F}' = \{\mathcal{F}'', f_i\}$ is minimally algebraically dependent, and let \mathbf{v}_i denote the row of \mathbf{V} corresponding to \mathbf{U}_{∇_i} , such that f_i simplifies into

$$\begin{aligned} f_i(\mathbf{v}_i, \mathbf{U}_{\Delta_i} | \mathbf{V}^*, \boldsymbol{\theta}_i^*) &= \left(|\mathbf{U}_{\Delta_i} | \mathbf{v}_i^* - \mathbf{v}_i \mathbf{U}_{\Delta_i}^\dagger \mathbf{U}_{\Delta_i}^* \right) \boldsymbol{\theta}_i^* \\ &= (\mathbf{v}_i^* - \mathbf{v}_i) \boldsymbol{\theta}_i^*. \end{aligned}$$

Since f_i involves \mathbf{v}_i , \mathcal{F}'' must contain at least one polynomial in \mathbf{v}_i (otherwise \mathcal{F}' cannot be minimally algebraically dependent). This means that \mathcal{F}'' contains at least one polynomial f_j involving \mathbf{v}_i :

$$f_j(\mathbf{v}_i, \mathbf{U}_{\Delta_j} | \mathbf{V}^*, \boldsymbol{\theta}_j^*) = \left(|\mathbf{U}_{\Delta_j} | \mathbf{v}_i^* - \mathbf{v}_i \mathbf{U}_{\Delta_j}^\dagger | \mathbf{U}_{\Delta_j}^* \right) \boldsymbol{\theta}_j^*.$$

For a.e. \mathbf{X} , $\boldsymbol{\theta}_j^*$ is independent of $\boldsymbol{\theta}_i^*$, so $(\nu_{\mathbf{V}} \times \nu_{\Theta})$ -almost surely, $f_i \neq f_j$.

We want to show that if \mathcal{F}' is minimally algebraically dependent, then $\mathbf{v}_i = \mathbf{v}_i^*$ is the only solution to $\mathcal{F}' = \mathbf{0}$. So define $\mathbf{v}_i =: [v_{i1} \ v_{i2}]$, and assume for contradiction that there exists a solution to $\mathcal{F}'' = \mathbf{0}$ with $\mathbf{v}_{i2} = \gamma \neq \mathbf{v}_{i2}^*$ and $\mathbf{U}_{\Delta_j} = \boldsymbol{\Gamma}_j$, that is also a solution to $\mathcal{F}' = \mathbf{0}$.

Next consider the univariate polynomials in v_{i1} evaluated at this solution:

$$\begin{aligned} g_i(v_{i1} | \mathbf{V}^*, \boldsymbol{\theta}_i^*) &:= f_i(v_{i1}, \mathbf{v}_{i2}, \mathbf{U}_{\Delta_i} | \mathbf{V}^*, \boldsymbol{\theta}_i^*) \Big|_{\mathbf{v}_{i2}=\gamma, \mathbf{U}_{\Delta_i}=\boldsymbol{\Gamma}_i}, \\ g_j(v_{i1} | \mathbf{V}^*, \boldsymbol{\theta}_j^*) &:= f_j(v_{i1}, \mathbf{v}_{i2}, \mathbf{U}_{\Delta_j} | \mathbf{V}^*, \boldsymbol{\theta}_j^*) \Big|_{\mathbf{v}_{i2}=\gamma, \mathbf{U}_{\Delta_j}=\boldsymbol{\Gamma}_j}, \end{aligned}$$

and observe that since $\{\gamma, \boldsymbol{\Gamma}_j\}$ are a solution to \mathcal{F}' , then g_i and g_j must have a common root.

We know from elimination theory that two distinct polynomials g_i, g_j have a common root if and only if their resultant $\mathbf{Res}(g_i, g_j)$ is zero (see, for example, Proposition 8 in Chapter 3, Section 5 of [19]).

But $\mathbf{Res}(g_i, g_j)$ is a polynomial in the coefficients of g_i and g_j . In other words, $\mathbf{Res}(g_i, g_j) = h(\mathbf{V}^*, \boldsymbol{\theta}_i^*, \boldsymbol{\theta}_j^*)$ for some nonzero polynomial h in \mathbf{V}^* , $\boldsymbol{\theta}_i^*$ and $\boldsymbol{\theta}_j^*$. Therefore, $h \neq 0$ for $(\nu_{\mathbf{V}} \times \nu_{\Theta})$ -almost every $\{\mathbf{V}^*, \boldsymbol{\Theta}^*\}$ (since the variety defined by $h = 0$ has measure zero). Equivalently, $h \neq 0$ for a.e. \mathbf{X} . Since $\mathbf{Res}(g_i, g_j) \neq 0$, it follows that g_i and g_j do not have a common root v_{i1} , which is the desired contradiction.

This will be true for either almost every γ in an infinite collection, or for every γ in a finite collection. In the first case, we would conclude that $\mathcal{F}' = \mathbf{0}$ has infinitely fewer solutions than $\mathcal{F}'' = \mathbf{0}$, in contradiction to the minimally algebraically dependent assumption. In the second case, we conclude that \mathbf{v}_{i2}^* is the only solution to $\mathcal{F}' = \mathbf{0}$.

Since v_{i1} was an arbitrary entry of \mathbf{U}_{ω_i} , we conclude that for a.e. \mathbf{X} , if \mathcal{F}' is minimally algebraically dependent, then $\mathbf{U}_{\nabla_i} = \mathbf{U}_{\nabla_i}^*$ is the only solution to $\mathcal{F}' = \mathbf{0}$, as desired. \square

Define $\{\mathbf{V}_t, \mathbf{V}_t^c\}$ as the partition of the variables involved in the polynomials in $\mathcal{F}'_t \subset \mathcal{F}'$, such that all the variables in \mathbf{V}_t are uniquely determined by $\mathcal{F}' = \mathbf{0}$.

Lemma 7. *Suppose $\mathbf{V}_t \neq \emptyset$ and that every $f_i \in \mathcal{F}'_t$ is a polynomial in at least one of the variables in \mathbf{V}_t . Then for a.e. \mathbf{X} , all the variables involved in \mathcal{F}'_t are uniquely determined by $\mathcal{F}' = \mathbf{0}$.*

Proof. Let v^c be one of the variables in \mathbf{V}_t^c and let f_i be a polynomial in \mathcal{F}'_t involving v^c . By assumption on \mathcal{F}'_t , f_i also involves at least one of the variables in \mathbf{V}_t , say v .

Let \mathbf{w} denote the set of all variables involved in f_i except v . Observe that $v^c \in \mathbf{w}$. This way, f_i is shorthand for $f_i(v, \mathbf{w} | \mathbf{V}^*, \boldsymbol{\theta}_i^*)$.

We will show that for a.e. \mathbf{X} , all the variables in \mathbf{w} are also uniquely determined by $\mathcal{F}' = \mathbf{0}$.

Suppose there exists a solution to $\mathcal{F}' = \mathbf{0}$ with $\mathbf{w} = \boldsymbol{\gamma}$, and define the univariate polynomial

$$g(v | \mathbf{V}^*, \boldsymbol{\theta}_i^*) := f_i(v, \mathbf{w} | \mathbf{V}^*, \boldsymbol{\theta}_i^*) \Big|_{\mathbf{w}=\boldsymbol{\gamma}} .$$

Now assume for contradiction that there exists another solution to $\mathcal{F}' = \mathbf{0}$ with $\mathbf{w} \neq \boldsymbol{\gamma}$. Let $\mathbf{w} = \boldsymbol{\gamma}'$ be an other solution to $\mathcal{F}' = \mathbf{0}$, and define

$$g'(v | \mathbf{V}^*, \boldsymbol{\theta}_i^*) := f_i(v, \mathbf{w} | \mathbf{V}^*, \boldsymbol{\theta}_i^*) \Big|_{\mathbf{w}=\boldsymbol{\gamma}'} .$$

We will first show that $g \neq g'$. To see this, recall the definition of f_i , and observe that it depends on the choice of ∇_i . Nevertheless, it is easy to see that $f_i = 0$ describes the same variety regardless of the choice of ∇_i . Intuitively, this means that even though f_i might look different for each choice of ∇_i , it really *is* the same.

Therefore, we may select ∇_i to be the element of $\boldsymbol{\omega}_i$ corresponding to the position of a variable of \mathbf{w} that takes different values in $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. This way, a variable with multiple solutions is located in the location of \mathbf{U}_{∇_i} . Since f_i is linear in \mathbf{U}_{∇_i} , it follows that $g \neq g'$ for $(\nu_V \times \nu_\Theta)$ -almost every $\{\mathbf{V}^*, \boldsymbol{\Theta}^*\}$.

Now observe that since v is uniquely determined by $\mathcal{F}' = \mathbf{0}$, g and g' have a common root, which immediately implies that there are at most finitely many distinct g' . Otherwise, v would be a common root to infinitely many distinct polynomials, which $(\nu_V \times \nu_\Theta)$ -almost surely cannot be the case.

We know from elimination theory that two distinct polynomials g, g' have a common root if and only if their resultant $\mathbf{Res}(g, g')$ is zero (see, for example, Proposition 8 in Chapter 3, Section 5 of [19]).

But $\mathbf{Res}(g, g')$ is a polynomial in the coefficients of g and g' . In other words, $\mathbf{Res}(g, g') = h(\mathbf{V}^*, \boldsymbol{\theta}_i^*)$ for some nonzero polynomial h in \mathbf{V}^* and $\boldsymbol{\theta}_i^*$. Therefore, $h \neq 0$ for $(\nu_V \times \nu_\Theta)$ -almost every $\{\mathbf{V}^*, \boldsymbol{\Theta}^*\}$ (since the variety defined by $h = 0$ has measure zero). Equivalently, $h \neq 0$ for a.e. \mathbf{X} .

Since $\mathbf{Res}(g, g') \neq 0$, it follows that g and g' do not have a common root v , which is the desired contradiction. This is true for all of the finitely many g' . This shows that for a.e. \mathbf{X} , all the variables in \mathbf{w} (including v^c) are uniquely determined by $\mathcal{F}' = \mathbf{0}$.

Since v^c was an arbitrary element in \mathbf{V}_i^c , we conclude that all the variables in \mathbf{V}_i^c are uniquely determined by $\mathcal{F}' = \mathbf{0}$. \square

With this, we are now ready to present the proofs of Lemma 5, Lemma 3 and Theorem 1.

Proof. (Lemma 5) By the same arguments as in Lemma 4, whether \mathcal{F}' is minimally algebraically dependent is invariant to any permutation $\boldsymbol{\Pi}$ of the rows of

the column echelon form in (10). Therefore, showing that Lemma 5 holds for one particular choice of $\mathbf{\Pi}$ suffices to show it holds for every $\mathbf{\Pi}$.

With this in mind, suppose $\mathcal{F}' = \{\mathcal{F}'', f_i\}$ is minimally algebraically dependent. Take $\mathbf{\Pi}$ such that \mathbf{U} and \mathbf{U}^* are written in the column echelon form in (10) with the identity block in the rows indexed by Δ_i , and let \mathbf{v}_i denote the row of \mathbf{V} corresponding to \mathbf{U}_{∇_i} , such that

$$\mathbf{U}_{\omega_i} = \left[\begin{array}{c} \mathbf{I} \\ \hline \mathbf{v}_i \end{array} \right] \left. \vphantom{\begin{array}{c} \mathbf{I} \\ \hline \mathbf{v}_i \end{array}} \right\} r \\ \left. \vphantom{\begin{array}{c} \mathbf{I} \\ \hline \mathbf{v}_i \end{array}} \right\} 1.$$

We know by Lemma 6 that \mathbf{v}_i is uniquely determined by $\mathcal{F}' = \mathbf{0}$. We will now iteratively use Lemma 7 to show that all the variables in \mathcal{F}' (which are the same as the variables in \mathcal{F}'') are also uniquely determined by $\mathcal{F}' = \mathbf{0}$. This will imply that all the variables in \mathcal{F}'' are finitely determined by $\mathcal{F}'' = \mathbf{0}$, and that \mathcal{F}'' contains the same number of polynomials, $n(\mathbf{\Omega}'')$, as variables, $r(m(\mathbf{\Omega}'') - r)$, which is the desired conclusion.

First observe that since \mathbf{v}_i is finitely determined by $\mathcal{F}'' = \mathbf{0}$, \mathcal{F}'' must contain at least r polynomials in \mathbf{v}_i . Denote these polynomials by $\mathcal{F}'_1 \subset \mathcal{F}''$.

We will proceed inductively, indexed by $t \geq 1$. First, set $t = 1$ and define $\mathbf{V}_1 = \{\mathbf{v}_i\}$. We showed above that the variables in \mathbf{V}_1 are uniquely determined by $\mathcal{F}' = \mathbf{0}$. Suppose that \mathcal{F}'_1 involves some variables other than those in \mathbf{V}_1 . Note that every polynomial in \mathcal{F}'_1 involves at least one of the variables in \mathbf{V}_1 . Let \mathbf{V}_2 be the set of all variables involved in \mathcal{F}'_1 . By Lemma 7, all the variables in \mathbf{V}_2 are uniquely determined by $\mathcal{F}' = \mathbf{0}$.

We will now proceed inductively. For any $t \geq 2$, let \mathbf{V}_t be a subset of n_t variables in \mathbf{V} . Assume that all the variables in \mathbf{V}_t are uniquely determined by $\mathcal{F}' = \mathbf{0}$. Since $\dim \mathcal{V}(\mathcal{F}'') = \dim \mathcal{V}(\mathcal{F}')$, it follows that all the variables in \mathbf{V}_t are finitely determined by $\mathcal{F}'' = \mathbf{0}$. It follows that \mathcal{F}'' must contain at least n_t algebraically independent polynomials, each involving at least one of the variables in \mathbf{V}_t . Let \mathcal{F}'_t be this set of polynomials. Suppose \mathcal{F}'_t involves some variables other than \mathbf{V}_t . Define \mathbf{V}_{t+1} to be the set of all variables involved in \mathcal{F}'_t . By Lemma 7, all the variables in \mathbf{V}_{t+1} are uniquely determined by $\mathcal{F}' = \mathbf{0}$.

Since this is true for every t , and there are finitely many variables, this process must terminate at some finite step T , at which point \mathcal{F}'_T is a set of n_T algebraically independent polynomials in n_T variables.

This means that all the variables in \mathcal{F}'_T are finitely determined by $\mathcal{F}'_T = \mathbf{0}$, and since f_i only involves a subset of the variables in \mathcal{F}'_T , it follows that the polynomials in $\{\mathcal{F}'_T, f_i\} \subset \mathcal{F}'$ are algebraically dependent. Furthermore, since \mathcal{F}' is minimally algebraically dependent by assumption, we have that $\mathcal{F}'_T = \mathcal{F}''$.

Finally, observe that \mathcal{F}'' contains $n(\mathbf{\Omega}'')$ polynomials in $r(m(\mathbf{\Omega}'') - r)$ variables. Since $\mathcal{F}'' = \mathcal{F}'_T$, and \mathcal{F}'_T has n_T polynomials in n_T variables, it follows that $n(\mathbf{\Omega}'') = r(m(\mathbf{\Omega}'') - r)$, as desired. \square

Proof. (Lemma 3)

- (\Rightarrow) Suppose \mathcal{F}' is minimally algebraically dependent. By Lemma 5, $n(\Omega') = r(m(\Omega') - r) + 1 > r(m(\Omega') - r)$, and we have the first implication.
- (\Leftarrow) Suppose there exists an Ω' with $n(\Omega') > r(m(\Omega') - r)$. By Lemma 4, $n(\Omega') > \kappa(\Omega')$, which implies the polynomials in \mathcal{F}' , and hence $\tilde{\mathcal{F}}$, are algebraically dependent. □

Proof. (Theorem 1)

- (\Rightarrow) Suppose for contrapositive that for every $\tilde{\Omega}$ formed with $r(d-r)$ columns of Ω , there exists an Ω' formed with a subset of its columns such that $m(\Omega') < n(\Omega')/r + r$. Lemma 3 implies that the polynomials in \mathcal{F}' , and hence $\tilde{\mathcal{F}}$, are algebraically dependent. It follows by Lemma 2 that there are infinitely many subspaces in $\mathcal{S}(\mathbf{X}_\Omega)$.
- (\Leftarrow) Suppose that for some $\tilde{\Omega}$ formed with $r(d-r)$ columns of Ω , every Ω' formed with a subset of the columns in $\tilde{\Omega}$ satisfies $m(\Omega') \geq n(\Omega')/r + r$, including $\tilde{\Omega}$. By Lemma 3, the $r(d-r)$ polynomials in $\tilde{\mathcal{F}}$ are algebraically independent. It follows by Lemma 2 that there are at most finitely many subspaces in $\mathcal{S}(\mathbf{X}_\Omega)$, hence at most finitely many rank- r completions of \mathbf{X}_Ω . □

5 Proof of Theorem 2

In this section we give the proof of Theorem 2. We will use $\tilde{\mathbf{X}}_{\tilde{\Omega}}$ and $\hat{\mathbf{X}}_{\hat{\Omega}}$ to denote the $d \times r(d-r)$ and $d \times (d-r)$ submatrices of \mathbf{X}_Ω corresponding to $\tilde{\Omega}$ and $\hat{\Omega}$. In addition, let $\hat{\omega}_i$ and $\hat{\mathbf{x}}_{\hat{\omega}_i}$ denote the i^{th} columns of $\hat{\Omega}$ and $\hat{\mathbf{X}}_{\hat{\Omega}}$.

In order to prove Theorem 2, we will require Theorem 1 in [21], which we state here as the following lemma, with some minor adaptations to our context.

Lemma 8. *Suppose $\hat{\Omega}$ is a $d \times (d-r)$ matrix with binary entries for which (i) holds and let $S \in \text{Gr}(r, \mathbb{R}^d)$. Then for ν_G -almost every S^* , $\{S_{\hat{\omega}_i} = S_{\hat{\omega}_i}^*\}_{i=1}^{d-r}$ if and only if $S = S^*$.*

With this, we are ready to give the proof of Theorem 2.

Proof. (Theorem 2) Suppose Ω contains two disjoint matrices $\tilde{\Omega}$ and $\hat{\Omega}$ satisfying the conditions of Theorem 2.

Since $\tilde{\Omega}$ satisfies (i), by Theorem 1 there are at most finitely many r -dimensional subspaces that fit $\tilde{\mathbf{X}}_{\tilde{\Omega}}$. Equivalently, the set $\tilde{\mathcal{F}}$, containing the $r(d-r)$ polynomials defined by the columns in $\tilde{\mathbf{X}}_{\tilde{\Omega}}$, is algebraically independent. Let \hat{f}_i be the polynomial defined by $\hat{\mathbf{x}}_{\hat{\omega}_i}$. It follows that the set $\{\tilde{\mathcal{F}}, \hat{f}_i\}$ is

algebraically dependent. Let \mathcal{F}'' be a subset of the polynomials in $\tilde{\mathcal{F}}$, such that $\mathcal{F}' = \{\mathcal{F}'', \hat{f}_i\}$ is minimally algebraically dependent. Then any subspace S with basis \mathbf{U} that fits $\hat{\mathbf{x}}_{\hat{\omega}_i}$ must satisfy $\mathcal{F}' = \mathbf{0}$, implying by Lemma 6 that $\mathbf{U}_{\hat{\omega}_i} = \mathbf{U}_{\hat{\omega}_i}^*$.

Therefore, every S that fits both $\tilde{\mathbf{X}}_{\tilde{\Omega}}$ and $\hat{\mathbf{X}}_{\hat{\Omega}}$ must satisfy $\{S_{\hat{\omega}_i} = S_{\hat{\omega}_i}^*\}_{i=1}^{d-r}$. Since $\hat{\Omega}$ satisfies (ii), it follows by Lemma 8 that $S = S^*$. \square

In Section 2 we mentioned that there are cases where $r(d-r)$ columns with only $r+1$ samples are sufficient for unique completability. The next result states that this is indeed the case if $r = 1$.

Proposition 1. *If $r = 1$, finite completability is equivalent to unique completability.*

Proof. Assume $r = 1$. Then \mathbf{U}_{Δ_i} and \mathbf{U}_{∇_i} are scalars, so f_i simplifies into:

$$f_i = (\mathbf{U}_{\Delta_i} \mathbf{U}_{\nabla_i}^* - \mathbf{U}_{\nabla_i} \mathbf{U}_{\Delta_i}^*) \theta_i^*.$$

This implies that $\mathcal{F} = \mathbf{0}$ is a system of linear equations, hence if it has finitely many solutions, it has only one. \square

In Section 2 we also mentioned that in general, strictly more than $r(d-r)$ columns with only $r+1$ samples are necessary for unique completability. We would like to close this section with an example where this is the case.

Example 5. *Consider $d = 4$ and $r = 2$, such that $N = r(d-r) = 4$. Let*

$$\mathbf{\Omega} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

It is easy to see that that $\tilde{\mathbf{\Omega}} = \mathbf{\Omega}$ satisfies the conditions of Theorem 1. One may also verify (for example, solving explicitly $\mathcal{F}(\mathbf{V}) = \mathbf{0}$) that for a.e. \mathbf{X} there exist two subspaces that fit $\mathbf{X}_{\mathbf{\Omega}}$.

As a matter of fact, this will also be the case for any permutation of the rows and columns of this matrix. One may construct similar samplings with the same property for larger d and r . All this to say that this is not a singular pathological example; there are many samplings that cannot be uniquely recovered with only $r(d-r)$ columns with $r+1$ samples.

6 Additional Proofs

In this section we present the proofs of Lemma 1 and Theorem 3. The proof of Corollary 1 follows directly from Lemma 1 and Theorem 2, and the proof of Corollary 2 follows directly from Theorem 1 in [21].

Proof. (Lemma 1) Suppose $\tilde{\Omega}$ contains disjoint matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^r$ satisfying the conditions of Lemma 1. Let Ω' be a matrix formed with a subset of the columns in $\tilde{\Omega}$. Then $\Omega' = [\Omega'_1 \cdots \Omega'_r]$ for some matrices $\{\Omega'_\tau\}_{\tau=1}^r$ formed with subsets of the columns in $\{\hat{\Omega}_\tau\}_{\tau=1}^r$.

It follows that

$$n(\Omega') = \sum_{\tau=1}^r n(\Omega'_\tau) \leq \sum_{\tau=1}^r \max_{\tau} n(\Omega'_\tau).$$

Assume without loss of generality that this maximum is achieved when $\tau = 1$. Then

$$n(\Omega') \leq rn(\Omega'_1) \leq r(m(\Omega'_1) - r) \leq r(m(\Omega') - r),$$

where the last two inequalities follow because (2) holds for every Ω'_τ by assumption, and because $m(\Omega') \geq m(\Omega'_\tau)$ for every τ .

Since Ω' was arbitrary, we conclude that (1) holds for every matrix Ω' formed with a subset of the columns in $\tilde{\Omega}$. \square

The following lemma shows that (ii) is satisfied with high probability under uniform random sampling schemes with only $\mathcal{O}(\max\{r, \log d\})$ samples per column.

Lemma 9. *Let the assumptions of Theorem 3 hold, and let $\hat{\Omega}$ be a matrix formed with $d - r$ columns of Ω . With probability at least $1 - \frac{\epsilon}{d}$, $\hat{\Omega}$ will satisfy (ii).*

Proof. Let \mathcal{E} be the event that $m(\Omega') < n(\Omega') + r$ for some matrix Ω' formed with a subset of the columns in $\hat{\Omega}$. It is easy to see that this will only occur if there is a matrix Ω' formed with n columns of $\hat{\Omega}$ that has all its nonzero entries in the same $n + r - 1$ rows. Let \mathcal{E}_n denote the event that the matrix formed with the first n columns from $\hat{\Omega}$ has all its nonzero entries in the first $n + r - 1$ rows. Then

$$\mathbb{P}(\mathcal{E}) \leq \sum_{n=1}^{d-r} \binom{d-r}{n} \binom{d}{n+r-1} \mathbb{P}(\mathcal{E}_n) \quad (11)$$

If each column of $\hat{\Omega}$ contains at least ℓ nonzero entries, distributed uniformly and independently at random with ℓ as in (3), it is easy to see that $\mathbb{P}(\mathcal{E}_n) = 0$ for $n \leq \ell - r$, and for $\ell - r < n \leq d - r$,

$$\mathbb{P}(\mathcal{E}_n) \leq \left(\frac{\binom{n+r-1}{\ell}}{\binom{d}{\ell}} \right)^n < \left(\frac{n+r-1}{d} \right)^{\ell n}.$$

Since $\binom{d-r}{n} < \binom{d}{n+r-1}$, continuing with (11) we obtain:

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &< \sum_{n=\ell-r+1}^{d-r} \binom{d}{n+r-1}^2 \left(\frac{n+r-1}{d}\right)^{\ell n} \\ &< \sum_{n=\ell}^{d/2} \binom{d}{n}^2 \left(\frac{n}{d}\right)^{\ell(n-r+1)} \end{aligned} \quad (12)$$

$$+ \sum_{n=1}^{d/2} \binom{d}{d-n}^2 \left(\frac{d-n}{d}\right)^{\ell(d-n-r+1)} \quad (13)$$

For the terms in (12), write

$$\binom{d}{n}^2 \left(\frac{n}{d}\right)^{\ell(n-r+1)} \leq \left(\frac{de}{n}\right)^{2n} \left(\frac{n}{d}\right)^{\ell(n-r+1)}. \quad (14)$$

Since $n \geq \ell \geq 2r$,

$$(14) < \left(\frac{de}{n}\right)^{2n} \left(\frac{n}{d}\right)^{\frac{\ell n}{2}} = e^{2n} \left(\frac{n}{d}\right)^{\left(\frac{\ell}{2}-2\right)n}, \quad (15)$$

and since $n \leq \frac{d}{2}$,

$$(15) \leq e^{2n} \left(\frac{1}{2}\right)^{\left(\frac{\ell}{2}-2\right)n} = \left(e^2 \cdot 2^{-\frac{\ell}{2}+2}\right)^n < \frac{\epsilon}{d^2}, \quad (16)$$

where the last step follows because $\ell > 2 \log_2 \left(\frac{(de)^2}{\epsilon}\right) + 4$.

For the terms in (13), write

$$\binom{d}{d-n}^2 \left(\frac{d-n}{d}\right)^{\ell(d-n-r+1)} \leq \left(\frac{de}{n}\right)^{2n} \left(\frac{d-n}{d}\right)^{\ell(d-n-r+1)}. \quad (17)$$

In this case, since $1 \leq n \leq \frac{d}{2}$ and $r \leq \frac{d}{6}$, we have

$$\begin{aligned} (17) &< (de)^{2n} \left(\frac{d-n}{d}\right)^{\frac{\ell d}{3}} = (de)^{2n} \left[\left(1 - \frac{n}{d}\right)^d\right]^{\frac{\ell}{3}} \\ &\leq (de)^{2n} [e^{-n}]^{\frac{\ell}{3}}, \end{aligned}$$

which we may rewrite as

$$\left(e^{2 \log d}\right)^n \left(e^2\right)^n \left(e^{-\frac{\ell}{3}}\right)^n = \left(e^{2 \log d + 2 - \frac{\ell}{3}}\right)^n < \frac{\epsilon}{d^2}, \quad (18)$$

where the last step follows because $\ell > 3 \log \left(\frac{d^2}{\epsilon}\right) + 6 \log d + 6$.

Substituting (16) and (18) in (12) and (13), we have that $\mathbb{P}(\mathcal{E}) < \frac{\epsilon}{d}$. \square

We are now ready to give the proof of Theorem 3.

Proof. (Theorem 3) If $N \geq r(d-r)$, randomly select disjoint matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^r$, each formed with $d-r$ columns of Ω .

Union bounding over τ , we may upper bound the probability that Ω fails to satisfy the conditions of Lemma 1 by

$$\sum_{\tau=1}^r \mathbb{P}(\mathcal{E}) < \sum_{\tau=1}^r \frac{\epsilon}{d} < \sum_{\tau=1}^r \frac{\epsilon}{r} = \epsilon.$$

The first part of the statement follows because the conditions in Lemma 1 imply the conditions in Theorem 1.

If $N \geq (r+1)(d-r)$, randomly select disjoint matrices $\{\hat{\Omega}_\tau\}_{\tau=1}^{r+1}$, each formed with $d-r$ columns of Ω . By the same arguments, the probability that Ω fails to satisfy the conditions of Theorem 2 is upper bounded by:

$$\sum_{\tau=1}^{r+1} \mathbb{P}(\mathcal{E}) < \sum_{\tau=1}^{r+1} \frac{\epsilon}{d} < \sum_{\tau=1}^{r+1} \frac{\epsilon}{r+1} = \epsilon.$$

□

7 Conclusions

In this paper we give sampling conditions for finite rank- r completability, that is, conditions on the set of observed entries to guarantee that a matrix can be completed in at most finitely many ways. We also provide deterministic sampling conditions for unique completability that can be efficiently verified. In addition, we show that uniform random samplings with only $\mathcal{O}(\max\{r, \log d\})$ observed entries per column satisfy these conditions with high probability. These findings have several implications on LRMC regarding lower bounds, sample and computational complexity, the role of coherence, adaptive settings and the validation of any completion algorithm.

8 Acknowledgements

We would like to thank Louis Theran for pointing out a mistake in a previous version of the paper. In that earlier version we erroneously assumed that columns with more than $r+1$ observed entries would yield multiple independent constraints. However, as Theran pointed out through the following example in [11], these constraints may be algebraically dependent. For this reason, in our current analysis we use only one constraint per column.

Example 6. Suppose \mathbf{X} is a rank-2 matrix observed on the entries indicated by 1's

$$\Omega = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Here \mathbf{x}_3 is observed on $r + 2$ entries. Using Definition 1 in our earlier version of this paper, \mathbf{x}_3 yields the two central columns of $\check{\mathbf{\Omega}}$

$$\check{\mathbf{\Omega}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

The two central columns in $\check{\mathbf{\Omega}}$ correspond to $\mathbf{\Omega}_3$, and encode the two polynomial constraints obtained from the observed entries in \mathbf{x}_3 .

While the sampling $\check{\mathbf{\Omega}}$ satisfies the completability conditions of Theorem 1 in our earlier version of this paper, \mathbf{X} cannot be completed. This is because the two polynomials defined by \mathbf{x}_3 share the same coefficients, which makes them algebraically dependent. If instead of \mathbf{x}_3 we observed two columns on the entries indicated by $\mathbf{\Omega}_3$, then we would also obtain two polynomial constraints. Only these polynomials would have generic coefficients, and would be algebraically independent with probability 1. In fact, a matrix observed on the entries indicated in $\check{\mathbf{\Omega}}$ (or more) can indeed be completed.

References

- [1] J. Rennie and N. Srebro, *Fast maximum margin matrix factorization for collaborative prediction*, International Conference on Machine Learning, pp. 713-719, 2005.
- [2] K. Weinberger and L. Saul, *Unsupervised learning of image manifolds by semidefinite programming*, International Journal of Computer Vision, 70(1), pp. 77-90, 2006.
- [3] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9(6), pp. 717-772, 2009.
- [4] E. Candès and T. Tao, *The power of convex relaxation: near-optimal matrix completion*, IEEE Transactions on Information Theory, 56(5), pp. 2053-2080, 2010.
- [5] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 12, pp. 3413-3430, 2011.
- [6] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Transactions on Information Theory, 57(3), pp. 1548-1566, 2011.
- [7] Y. Chen, *Incoherence-optimal matrix completion*, IEEE Transactions on Information Theory, 61(5), pp. 2909-2923, 2013.
- [8] Y. Chen, S. Bhojanapalli, S. Sanghavi and R. Ward, *Coherent matrix completion*, International Conference on Machine Learning, pp. 674-682, 2014.

- [9] S. Bhojanapalli and P. Jain, *Universal matrix completion*, International Conference on Machine Learning, pp. 1881-1889, 2014.
- [10] A. Singer and M. Cucuringu, *Uniqueness of low-rank matrix completion by rigidity theory*, SIAM Journal on Matrix Analysis and Applications, 31(4), pp. 1621-1641, 2010.
- [11] F. Király, L. Theran and R. Tomioka, *The algebraic combinatorial approach for low-rank matrix completion*, Journal of Machine Learning Research, pp. 1391-1436, 2015.
- [12] E. Chunikhina, R. Raich and T. Nguyen, *Performance analysis for matrix completion via iterative hard-thresholded SVD*, IEEE Statistical Signal Processing Workshop, pp. 392-395, 2014.
- [13] J. Cai, E. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization, 20(4), pp. 1956-1982, 2010.
- [14] R. Keshavan, A. Montanari and S. Oh, *Matrix completion from a few entries*, IEEE Transactions on Information Theory, pp. 324-328, 2010.
- [15] S. Ma, D. Goldfarb, L. Chen, *Fixed point and Bregman iterative methods for matrix rank minimization*, Mathematical Programming, 128(1-2), pp. 321-353, 2011.
- [16] L. Balzano, R. Nowak, and B. Recht, *Online identification and tracking of subspaces from highly incomplete information*, Allerton Conference on Communication, Control and Computing, pp. 704-711, 2010.
- [17] P. Jain, P. Netrapalli and S. Sanghavi, *Low-rank matrix completion using alternating minimization*, ACM Symposium on Theory Of Computing, pp. 665-674, 2013.
- [18] D. Pimentel-Alarcón, L. Balzano and R. Nowak, *On the sample complexity of subspace clustering with missing data*, IEEE Statistical Signal Processing Workshop, pp. 280-283, 2014.
- [19] D. Cox, J. Little and D. O'shea, *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*, Third Edition, Springer, 2007.
- [20] A. Aramova, L. Avramov and J. Herzog, *Resolutions of monomial ideals and cohomology over exterior algebras*, Transactions of the American Mathematical Society, 352(2), pp. 579-594, 2000.
- [21] D. Pimentel-Alarcón, N. Boston and R. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, IEEE International Symposium on Information Theory, pp. 2191-2195, 2015.