Journal of Machine Learning Research 16 (2015) 1391-1436 Submitted 4/13; Revised 8/14; Published 8/15

The Algebraic Combinatorial Approach for Low-Rank Matrix Completion

Franz J. Király F.KIRALY@UCL.AC.UK
Department of Statistical Science
University College London
London WC1E 6BT, United Kingdom
and
Mathematisches Forschungsinstitut Oberwolfach
77709 Oberwolfach-Walke, Germany

Louis Theran LOUIS.THERAN@AALTO.FI
*Aalto Science Institute and Department of Information and Computer Science**
Aalto University
02150 Espoo, Finland

Ryota Tomioka TOMIOKA@TTIC.EDU
Toyota Technological Institute at Chicago
6045 S. Kenwood Ave.
Chicago, Illinois 60637, USA

Editor: Gábor Lugosi

Abstract

We present a novel algebraic combinatorial view on low-rank matrix completion based on studying relations between a few entries with tools from algebraic geometry and matroid theory. The intrinsic locality of the approach allows for the treatment of single entries in a closed theoretical and practical framework. More specifically, apart from introducing an algebraic combinatorial theory of low-rank matrix completion, we present probability-one algorithms to decide whether a particular entry of the matrix can be completed. We also describe methods to complete that entry from a few others, and to estimate the error which is incurred by any method completing that entry. Furthermore, we show how known results on matrix completion and their sampling assumptions can be related to our new perspective and interpreted in terms of a completability phase transition.[1]

Keywords: Low-rank matrix completion, entry-wise completion, matrix reconstruction, algebraic combinatorics

∗. Author's current address. Research was carried out while the author was at Freie Universität Berlin and supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 247029-SDModels

1. This paper is the much condensed, final version of (Király et al., 2013). For convenience, we have included references which have appeared since.

©2015 Franz J. Király, Louis Theran, and Ryota Tomioka.

1. Introduction

Matrix completion is the task to reconstruct (to "complete") matrices, given a subset of entries at known positions. It occurs naturally in many practically relevant problems, such as missing feature imputation, multi-task learning (Argyriou et al., 2008), transductive learning (Goldberg et al., 2010), or collaborative filtering and link prediction (Srebro et al., 2005; Acar et al., 2009; Menon and Elkan, 2011).

For example, in the "NetFlix problem", the rows of the matrix correspond to users, the columns correspond to movies, and the entries correspond to the rating of a movie by a user. Predicting how *one specific* user will rate *one specific* movie then reduces to completing a *single unobserved entry* from the observed ratings.

For arbitrarily chosen position (i, j), the primary questions are:

- *Is it possible to reconstruct the entry* (i, j)*?*
- *How many possible completions are there for the entry* (i, j)*?*
- *What is the value of the entry* (i, j)*?*
- *How accurately can one estimate the entry* (i, j)*?*

In this paper, we answer these questions *algorithmically* under the common *low-rank assumption* - that is, under the model assumption (or approximation) that there is an underlying complete matrix of some low rank r from which the partial observations arise. Our algorithms are the first in the low-rank regime that provide information about single entries. They adapt to the combinatorial structure of the observations in that, if it is possible, the reconstruction process can be carried out using much less than the full set of observations. We validate our algorithms on real data. We also identify *combinatorial* features of the low-rank completion problem. This then allows us to study low-rank matrix completion via tools from, e.g., graph theory.

1.1 Results

Here is a preview of the results and themes of this paper, including the answers to the main questions.

1.1.1 Is it possible to reconstruct the entry (i, j)?

We show that whether the entry (i, j) is completable depends, with probability one for *any continuous sampling regime*, only on the *positions* of the observations and the position (i, j) that we would like to reconstruct (Theorem 10). The proof is explicit and easily converted into an exact (probability one) algorithm for computing the set of completable positions (Algorithm 1).

1.1.2 How many possible completions are there for the entry (i, j)?

Whether the entry at position (i, j) is uniquely completable from the observations, or, more generally, how many completions there are also depends, with probability one, only on the positions of the observed entries and (i, j) (Theorem 17). We also give an efficient

(randomized probability one) algorithm (Algorithm 1) that verifies a sufficient condition for every unobserved entry to be uniquely completable.

1.1.3 What is the value of the entry (i, j)?

To reconstruct the missing entries, we introduce a general scheme based on finding polynomial relations between the observations and one unobserved one at position (i, j) (Algorithm 5). For rank one matrices (Algorithm 6), and, in any rank, observation patterns with a special structure (Algorithm 4) that allows "solving minor by minor", we instantiate the scheme completely and efficiently.

Since, for a specific (i, j), the polynomials needed can be very sparse, our approach has the property that it adapts to the combinatorial structure of the observed positions. To our knowledge, other algorithms for low-rank matrix completion do not have this property.

1.1.4 How accurately can one estimate entry (i, j)?

Our completion algorithms separate out *finding* the relevant polynomial relations from *solving* them. When there is more than one relation, we can use them as different estimates for the missing entry, allowing for estimation in the noisy setting (Algorithm 5). Because the polynomials are independent of specific observations, the same techniques yield *a priori estimates* of the variance of our estimators.

1.1.5 Combinatorics of matrix completion

Section 6 contains a detailed analysis of whether an entry (i, j) is completable in terms of a *bipartite graph* encoding the combinatorics of the observed positions. We obtain necessary (Theorem 38) and sufficient (Proposition 42) conditions for local completability, which are sharp in the sense that our local algorithms apply when they are met. We then relate the properties we find to standard graph-theoretic concepts such as edge-connectivity and cores. As an application, we determine a binomial sampling density that is sufficient for solving minor-by-minor nearly exactly via a random graph argument.

1.1.6 Experiments

Section 7 validates our algorithms on the Movie Lens data set and shows that the structural features identified by our theories predict completability and completability phase transitions in practice.

1.2 Tools and themes

Underlying our results are a new view of low-rank matrix completion based on algebraic geometry. Here are some of the key ideas.

1.2.1 Using the local-to-global principle

Our starting point is that the set of rank r, $(m \times n)$-matrices carries the additional structure of an *irreducible algebraic variety* (see Section 2.1). Additionally, the observation process is a polynomial map. The key feature of this setup is that it gives us access to fundamental

algebraic-geometric "local-to-global" results (see Appendix A) that assert the observation process will exhibit a *prototypical behavior*: the answers to the main questions will be the *same for almost all* low-rank matrices, so they are essentially properties of the rank and observation map. This lets us study the main questions in terms of observed and unobserved *positions* rather than specific *partial matrices*.

On the other hand, the same structural results show we can *certify* that properties like completability hold via *single examples*. We exploit this to replace very complex basis eliminations with fast algorithms based on numerical linear algebra.

1.2.2 Finding relations among entries using an ideal

Another fundamental aspect of algebraic sets are characterized exactly by the *vanishing ideal* of polynomials that evaluate to zero on them. For matrix completion, the meaning is: *every* polynomial relation between the observations and a specific position (i, j) is generated by a *finite* set of polynomials we can in principle identify (See Section 5).

1.2.3 Connecting geometry to combinatorics using matroids

Our last major ingredient is the use of the Jacobian of the observation map, evaluated at a "generic point". The independence/dependence relation among its rows is invariant (with matrix-sampling probability one) over the set of rank r matrices that characterizes whether a position (i, j) is completable. Considering the subsets of independent rows as simply subsets of a finite set, we obtain a *linear matroid* characterizing completability. This perspective allows access to combinatorial tools of matroid theory, enabling the analysis in Section 6.

1.3 Context and novelty

Low-rank matrix completion has received a great deal of attention from the community. Broadly speaking, two main approaches have been developed: convex relaxations of the rank constraints (e.g., Candès and Recht, 2009; Candès and Tao, 2010; Negahban and Wainwright, 2011; Salakhutdinov and Srebro, 2010; Negahban and Wainwright, 2012; Foygel and Srebro, 2011; Srebro and Shraibman, 2005); and spectral methods (e.g., Keshavan et al., 2010; Meka et al., 2009). Both of these (see Candès and Tao, 2010; Keshavan et al., 2010) yield, in the noiseless case, optimal sample complexity bounds (in terms of the number of positions uniformly sampled) for exact reconstruction of an underlying matrix meeting certain analytic assumptions. All the prior work of which we are aware concentrates on: (A) sets of observed positions sampled from some known distribution; (B) completing all the unobserved entries. The results here, by contrast, apply specifically to *fixed* sets of observations and provide information about *any* unobserved position (i, j).

To point (A), there are three notable exceptions: Singer and Cucuringu (2010) discuss a mathematical analogy to combinatorial rigidity, studying which fixed observation patterns allow unique and stable completions; their work is to a large part conjectural but exposes the connection to graph combinatorics and anticipates some of our theoretical results. Lee and Shraibman (2013) study completion guarantees for fixed observation patterns with tools inspired by and related to the nuclear norm. Bhojanapalli and Jain (2014) showed a sufficient condition for exact recovery by nuclear norm minimization when the bipartite

graph corresponding to the observed positions has a large spectral gap under a strong incoherence assumption.

Regarding point (B) more specifically, all the prior work on low-rank matrix completion from noisy observations concentrates on: (i) estimating every missing entry; (ii) denoising every observed entry; and (iii) minimizing the MSE over the whole matrix. Our approach allows, for the first time, to construct *single-entry estimators* that minimize the variance of the entry under consideration; we have recently shown how to do this efficiently in rank 1 (Kiraly and Theran, 2013).

1.4 Organization

The sequel is structured as follows: Section 2 introduces the background material we need; Sections 3 and 4 develop our algebraic-combinatorial theory and derive algorithms for determining when an entry is completable; Section 5 formulates the reconstruction process itself algebraically; Section 6 contains a combinatorial analysis of the problem; finally Section 7 validates our approach on real data. The Appendix collects some technical results required in the proofs of the main theorems.

2. Background and Setup

In this section, we introduce two essential objects, the set of low-rank matrices $\mathcal{M}(m \times n, r)$ and the set of observed positions E. We also define the concept of *genericity*.

2.1 The determinantal variety

First, we set up basic notation. A matrix is denoted by upper-case bold character like $\mathbf{A}$. We denote by $[n]$ the set of integers $\{1, 2, \ldots, n\}$. $\mathbf{A}_{I,J}$ denotes the submatrix of an $m \times n$ matrix $\mathbf{A}$ specified by the sets of indices $I \subseteq [m]$ and $J \subseteq [n]$. The (i, j) element of a matrix $\mathbf{A}$ is denoted by A_{ij}. The cardinality of a set I is denoted by $|I|$.

Now we define the set of matrices of rank at most r.

Definition 1 *The set of all complex $(m \times n)$-matrices of rank r or less will be denoted by $\mathcal{M}(m \times n, r) = \{\mathbf{A} \in \mathbb{C}^{m \times n} : \operatorname{rank}(\mathbf{A}) \leq r\}$. We will always assume that $r \leq m \leq n$; by transposing the matrices, this is no loss of generality.*

Some basic properties of $\mathcal{M}(m \times n, r)$ are summarized in the following proposition.

Proposition 2 (Properties of the determinantal variety) *The following hold:*

(i) $\mathcal{M}(m \times n, r)$ is the image of the map $\Upsilon : (\mathbf{U}, \mathbf{V}) \mapsto \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{C}^{m \times r}$ and $\mathbf{V} \in \mathbb{C}^{n \times r}$, and is therefore irreducible.

(ii) $\mathcal{M}(m \times n, r)$ has dimension

$$d_r(m, n) := \dim \mathcal{M}(m \times n, r) = \begin{cases} r(m + n - r) & \text{if } m \geq r \text{ and } n \geq r \\ mn & \text{otherwise} \end{cases}$$

(iii) Every $(r+1)\times(r+1)$ *minor of a matrix in* $\mathcal{M}(m\times n, r)$ *is zero, namely,*

$$\det(\mathbf{A}_{I,J}) = 0, \quad \forall I \subseteq [m], J \subseteq [n],$$

where $|I| = r+1$, $|J| = r+1$, *and* $\mathbf{A} \in \mathcal{M}(m\times n, r)$.

(iv) The vanishing ideal of $\mathcal{M}(m\times n, r)$ *is generated by the vanishing of the minors from part 3.*

Proof (i) The existence of the singular-value decomposition imply that $\mathcal{M}(m\times n, r)$ is the surjective image of $\mathbb{C}^{r(m+n)}$ under the algebraic map Υ.

(ii) This follows from **(i)** and the uniqueness of the singular value decomposition, or Bruns and Vetter (1988, section 1.C, Proposition 1.1).

(iii) The rank of a matrix equals the order of the largest non-vanishing minor.

(iv) By Bruns and Vetter (1988, Theorem 2.10, Remark 2.12, and Corollary 5.17f), the ideal generated by the $r\times r$ minors is prime. Since it vanishes on the irreducible $\mathcal{M}(m\times n, r)$, it is the vanishing ideal. ■

The set of observed positions is denoted by E and can be viewed as a bipartite graph as follows.

Definition 3 *Let* $\mathcal{E} := [m]\times[n]$. *The set containing the positions of observed entries is denoted by* $E \subseteq \mathcal{E}$. *We define the bipartite graph* $G(E) = (V, W, E)$ *with vertices* $V = [m]$ *corresponding to rows and vertices* $W = [n]$ *corresponding to columns. We call the* $m\times n$ *adjacency matrix* $\mathbf{M}(E)$ *of the bipartite graph* $G(E)$ *a* mask. *The map*

$$\Omega : \mathbf{A} \mapsto (A_{ij})_{(i,j)\in E},$$

where $\mathbf{A} \in \mathcal{M}(m\times n, r)$, *is called a* masking *(in rank* r*).*

Note that the set of observed positions E, the adjacency matrix $\mathbf{M}$, and the map Ω can be used interchangeably. For example, we denote by $\mathbf{M}(\Omega)$ the adjacency matrix corresponding to the map Ω, and by $E(\mathbf{M})$ the set of positions specified by $\mathbf{M}$, and so on. Figure 1 shows two bipartite graphs G_1 and G_2 corresponding to the following two masks:

$$\mathbf{M}_1 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \mathbf{M}_2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

2.2 The Jacobian of the masking operator

Informally, the question we are going to address is:

Which entries of $\mathbf{A}$ *are (uniquely) reconstructable, given the masking* $\Omega(\mathbf{A})$*?*

The answer will depend on the interaction between the algebraic structure of $\mathcal{M}(m\times n, r)$ and the combinatorial structure of E. The main tool we use to study this is the Jacobian of the map Υ, since at smooth points, we can obtain information about the dimension of the pre-image $\Omega^{-1}(\mathbf{A})$ from its rank.

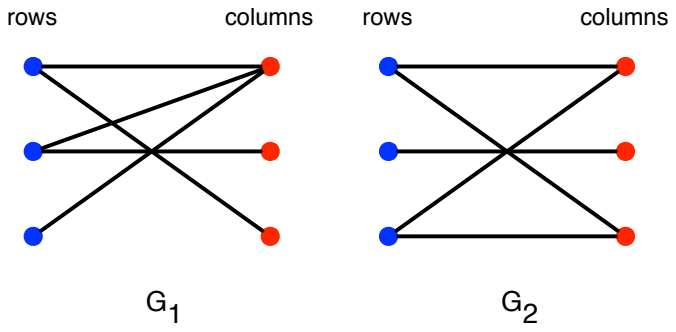


Figure 1: Two bipartite graphcs G_1 and G_2 corresponding to the masks $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively. Every non-edge corresponds to an unobserved entry.

Definition 4 *We denote by* $\mathbf{J}$ *the Jacobian of the map* $\Upsilon : \mathbf{U}, \mathbf{V} \mapsto \mathbf{A} = \mathbf{U}\mathbf{V}^\top$. *More specifically, the Jacobian of the map from* $\mathbf{U}$ *and* $\mathbf{V}$ *to* A_{ij} *can be written as follows:*

$$\left(\frac{\partial A_{ij}}{\partial \mathbf{u}_1^\top}, \ldots, \frac{\partial A_{ij}}{\partial \mathbf{u}_m^\top}, \frac{\partial A_{ij}}{\partial \mathbf{v}_1^\top}, \ldots, \frac{\partial A_{ij}}{\partial \mathbf{v}_n^\top}\right) = \underset{\substack{\uparrow \\ \text{Derivative wrt } \mathbf{u}_i}}{\left(0 \quad \cdots \quad \mathbf{v}_j^\top \quad \cdots 0 \quad 0 \cdots \right.} \underset{\substack{\uparrow \\ \text{Derivative wrt } \mathbf{v}_j}}{\left. \mathbf{u}_i^\top \quad \cdots \quad 0\right)} \tag{1}$$

where $\mathbf{u}_i^\top$ *is the* i*th row vector of* $\mathbf{U}$ *and* $\mathbf{v}_j^\top$ *is the* j*th row vector of* $\mathbf{V}$. *Stacking the above row vectors for* $(i, j) \in [m] \times [n]$, *we can write the Jacobian* $\mathbf{J}(\mathbf{U}, \mathbf{V})$ *as an* $mn \times r(m+n)$ *matrix as follows:*

$$\mathbf{J}(\mathbf{U}, \mathbf{V}) = \begin{pmatrix} \mathbf{I}_m \otimes \mathbf{v}_1^\top & \\ \mathbf{I}_m \otimes \mathbf{v}_2^\top & \mathbf{I}_n \otimes \mathbf{U} \\ \vdots & \\ \mathbf{I}_m \otimes \mathbf{v}_n^\top & \end{pmatrix}, \tag{2}$$

where $\otimes$ *denotes the Kronecker product. Here the rows of* $\mathbf{J}$ *correspond to the entries of* $\mathbf{A}$ *in the column major order.*

Lemma 5 *Every matrix* $\mathbf{S} \in \mathbb{C}^{m \times n}$ *whose vectorization* $\mathrm{vec}(\mathbf{S})$ *lies in the left null space of* $\mathbf{J}(\mathbf{U}, \mathbf{V})$ *satisfies*

$$\mathbf{U}^\top \mathbf{S} = 0, \qquad \mathbf{S}\mathbf{V} = 0,$$

and any $\mathbf{S}$ *satisfying the above lies in the left null space of* $\mathbf{J}(\mathbf{U}, \mathbf{V})$. *In addition, the dimension of the null space is* $(m-r)(n-r)$ *if* $\mathbf{U}$ *and* $\mathbf{V}$ *have full column rank* r.

Proof Let $\mathbf{P}$ be the $mn \times mn$ permutation matrix defined by

$$\mathbf{P}\,\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{X}^\top).$$

Note that $\mathbf{P}^\top \mathbf{P} = \mathbf{I}_{mn}$, and

$$\mathbf{P} \begin{pmatrix} \mathbf{I}_m \otimes \mathbf{v}_1^\top \\ \vdots \\ \mathbf{I}_m \otimes \mathbf{v}_n^\top \end{pmatrix} = \mathbf{I}_m \otimes \mathbf{V}.$$

Thus we have

$$\begin{aligned} \mathrm{vec}^\top(\mathbf{S})\mathbf{J} &= \left(\mathrm{vec}^\top(\mathbf{S})\mathbf{P}^\top (\mathbf{I}_m \otimes \mathbf{V}), \mathrm{vec}^\top(\mathbf{S}) (\mathbf{I}_n \otimes \mathbf{U})\right) \\ &= \left(\mathrm{vec}^\top(\mathbf{V}^\top \mathbf{S}^\top), \mathrm{vec}^\top(\mathbf{U}^\top \mathbf{S})\right), \end{aligned}$$

which is what we wanted. To show the last part of the lemma, let $\mathbf{U}_\perp \in \mathbb{C}^{m\times(m-r)}$ and $\mathbf{V}_\perp \in \mathbb{C}^{n\times(n-r)}$ be any basis of the orthogonal complement space of $\mathbf{U}$ and $\mathbf{V}$, respectively. Since the null space can be parametrized as $\mathbf{S} = \mathbf{U}_\perp \mathbf{S}' \mathbf{V}_\perp^\top$ by $\mathbf{S}' \in \mathbb{C}^{(m-r)\times(n-r)}$, and this parametrization is one-to-one, we see that the dimension of the null space is $(m-r)(n-r)$. ■

Now we define the Jacobian corresponding to the set of observed positions E.

Definition 6 *For a position (k,ℓ), we define $\mathbf{J}_{(k,\ell)}$ to be the single row of $\mathbf{J}$ corresponding to the position (k,ℓ). Similarly, we define $\mathbf{J}_E$ to be the submatrix of $\mathbf{J}$ consisting of rows corresponding to the set of observed positions E. Due to the chain rule, $\mathbf{J}_E$ is the Jacobian of the map $\Omega \circ \Upsilon$.*

2.3 Genericity

The pattern of zero and non-zero entries in (1) hints at a connection to purely combinatorial structure. To make the connection precise, we introduce *genericity*.

Definition 7 *We say a boolean statement $P(\mathbf{X})$ holds* for a generic $\mathbf{X}$ *in irreducible algebraic variety $\mathcal{X}$, if for any Hausdorff continuous measure μ on $\mathcal{S}$, $P(\mathbf{X})$ holds with probability* 1.

These kinds of statements are sometimes called "generic properties," and they are properties of $\mathcal{X}$, rather than any specific μ. The prototypical example of a generic property is where $\mathcal{X} = \mathbb{C}^n$, $p \neq 0$ is a polynomial, and the statement P is "$p(\mathbf{X}) \neq 0$."

Here, we are usually concerned with the case $\mathcal{X} = \mathcal{M}(m \times n, r)$. Proposition 2 tells us that m, n and r define $\mathcal{M}(m \times n, r)$ completely. Assertions of the form "For generic $\mathbf{X} \in \mathcal{M}(m \times n, r)$, $P(\mathbf{X})$ depends only on $(t_1, t_2, \ldots)$" mean $P(\mathbf{X})$ is a generic statement for *all* $\mathcal{M}(m \times n, r)$ with the parameters t_i fixed.

Although showing whether some statement P holds generically might seem hard, we are interested in P defined by polynomials. In this case, results in Appendix A imply that it is enough to show that either P holds: (a) on an open subset of $\mathcal{X}$ in the metric topology; or (b) almost surely, with respect to a Hausdorff continuous measure.

As a first step, and to illustrate the "generic philosophy" we show that the generic behavior of the Jacobian $\mathbf{J}_E(\mathbf{U}, \mathbf{V})$ is a property of E. We first start by justifying the definition via $(\mathbf{U}, \mathbf{V})$ (as opposed to $\mathbf{A}$).

Lemma 8 *For all $E \subset \mathcal{E}$ and $\mathbf{A} \in \mathcal{M}(m \times n, r)$ generic, with $\mathbf{A} = \Upsilon(\mathbf{U}, \mathbf{V})$, and $\mathbf{U}$ and $\mathbf{V}$ generic, the rank of $\mathbf{J}_E$ is independent of $\mathbf{A}$, $\mathbf{U}$, and $\mathbf{V}$.*

Proof We first consider the composed map $\Omega \circ \Upsilon$. This is a polynomial map in the entries of $\mathbf{U}$ and $\mathbf{V}$, so its critical points (at which the differential $\mathbf{J}_E$ attains less than its maximum rank) is an algebraic subset of $\mathbb{C}^{r(m+n)}$. The "Semialgebraic Sard Theorem" (Kurdyka et al., 2000, Theorems 3.1, 4.1) then implies that the set of critical points is, in fact, a *proper* algebraic subset of $\mathbb{C}^{r(m+n)}$.

So far, we have proved that the rank of $\mathbf{J}_E$ is independent of $\mathbf{U}$ and $\mathbf{V}$. However, $\mathbf{U}$ and $\mathbf{V}$ are not uniquely determined by $\mathbf{A}$. To reach the stronger conclusion, we first observe that a generic $\mathbf{A} \in \mathcal{M}(m \times n, r)$ is a regular value of Υ, again by Semialgebraic Sard. Thus, the set of $(\mathbf{U}, \mathbf{V})$ such that $\Upsilon(\mathbf{U}, \mathbf{V})$ and $\Omega \circ \Upsilon(\mathbf{U}, \mathbf{V})$ are both regular values is the intersection of two dense sets in $\mathbb{C}^{r(m+n)}$. ■

3. Finite Completability

This section is devoted to the question *"Is it possible to reconstruct the entry (i, j)?"*. We will show that under mild assumptions, the answer depends only on the *position* (i, j), the observed positions, and the rank, but not the observed entries. The main idea behind this result is relating reconstructability to the rows of the Jacobian $\mathbf{J}$, and their rank, which can be shown to be independent of the actual entries for almost all low-rank matrices. Therefore, we can later separate the question of reconstructability from the actual reconstruction process.

3.1 Finite completability as a property of the positions

We show how to predict whether the entry at a specific *position* (k, ℓ) will be reconstructable from a specific set of positions $E \subset \mathcal{E}$. For the rest of this section, we fix the parameters r, m and n, and denote by E a set of observed positions. The symbol $\mathbb{K}$ will denote either of the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

We start by precisely defining what it means for one set of entries to imply the imputability of another entry.

Definition 9 *Let $E \subset \mathcal{E}$ be a set of observed positions and $\mathbf{A}$ be a rank r true matrix. The entry $A_{k\ell}$ is* finitely completable in rank r *from the observed set of entries $\{A_{ij} : (i, j) \in E\}$ if the entry $A_{k\ell}$ can take only finitely many values when fixing $\Omega(\mathbf{A})$.*

There are two subtleties here: the first is that, even if there is an infinity of possible completions for the whole matrix $\mathbf{A}$, it is possible that some specific $A_{k\ell}$ takes on only finitely many values; the question of whether the entry $A_{k\ell}$ at position (k, ℓ) is finitely completable may have different answers for different $\mathbf{A}$. The theoretical results in this section take care of both issues.

Theorem 10 *Let $E \subset \mathcal{E}$ be a set of positions, $(k, \ell) \in \mathcal{E} \setminus E$ be arbitrary, and let $\mathbf{A} \in \mathbb{K}^{m \times n}$ be a generic, $(m \times n)$-matrix of rank r. Whether the entry $A_{k\ell}$ at position (k, ℓ) is finitely completable depends only on the position (k, ℓ), the true rank r, and the observed positions E (and not on $\mathbf{A}$, m, n, or $\mathbb{K}$).*

This lets us talk about the finite completability of positions instead of entries.

Definition 11 *Let $E \subset \mathcal{E}$ be a set of observed positions, and $(k,\ell) \in \mathcal{E} \setminus E$. We say that the* position (k,ℓ) is finitely completable from E in rank r *if, for generic $\mathbf{A}$, the entry $A_{k\ell}$ is finitely completable from $\Omega(\mathbf{A})$. The* rank r finitely completable closure $\mathrm{cl}_r(E)$ *is the set of positions generically finitely completable from E.*

The main tool we use to prove Theorem 10 is the Jacobian matrix $\mathbf{J}_E$. For it, we obtain

Theorem 12 *Let $E \subset \mathcal{E}$ and let $\mathbf{A}$ be a generic, rank r matrix. Then*

$$\mathrm{cl}_r(E) = \{(k,\ell) \in \mathcal{E} : \mathbf{J}_{\{(k,\ell)\}} \in \operatorname{rowspan} \mathbf{J}_E\}.$$

One implication of Theorem 12 is that linear independence of subsets of rows of $\mathbf{J}_E$ is also a generic property. (In fact, the proof in Section 3.3 goes in the other direction.) The combinatorial object that captures this independence is a matroid.

Definition 13 *Let $\mathbf{A}$ be a generic rank r matrix. The* rank r determinantal matroid *is the linear matroid $(\mathcal{E}, \mathrm{rank}_r)$, with rank function $\mathrm{rank}_r(E) = \operatorname{rank} \mathbf{J}_E$.*

Note that due to Lemma 8, $\operatorname{rank} \mathbf{J}_E$ is independent of $\mathbf{A}$ as long as we are concerned with generic matrices and the rank function is well defined.

In the language of matroids, Theorem 12 says that, generically, the finitely completable closure is equal to the matroid closure in the rank r determinantal matroid. This perspective will prove profitable when we consider entry-by-entry algorithms for completion in Section 5 and combinatorial conditions related to finite completability in Section 6.

3.2 Computing the finite closure

We describe, in pseudo-code, Algorithm 1 which computes the finite closure of E. An algorithm for testing whether a single entry (k,ℓ) is finitely completable is easily obtained by only testing the entry (k,ℓ) in step 4. The correctness of Algorithm 1 follows from Theorem 12 and the fact that, if we sample $\mathbf{U}$ and $\mathbf{V}$ from any continuous density, with probability one, we obtain generic $\mathbf{U}$ and $\mathbf{V}$[2].

Remark 14 *For clarity and practicality, we have presented Algorithm 1 as a numerical routine based on SVD. To analyze it in the RAM model, instead of sampling $\mathbf{U}$ and $\mathbf{V}$ from a continuous density, we sample the entries uniformly from a finite field $\mathbb{Z}_p$ of prime order $p \approx (n+m)^2$. With this modification Algorithm 1 becomes strongly polynomial time via, e.g., Gaussian elimination. Using the main results of Schwartz (1980), one can show that this finite field variant computes the generic rank with probability $1 - O(1/(n+m))$.*

3.3 Proofs

3.3.1 Proof of Theorem 12

Let $(i,j) \in \mathcal{E} \setminus E$. Factor the map $\Omega \circ \Upsilon$ into

$$\mathbb{C}^{r(m+n)} \xrightarrow{\Upsilon} \mathcal{M}(m \times n, r) \xrightarrow{f} \mathbb{C}^{|E|+1} \xrightarrow{g} \mathbb{C}^{|E|}$$

2. If we discretize the continuous density, then "probability one" becomes "with high probability".

Algorithm 1 Completable closure.
Input: A set $E \subset \mathcal{E}$ of observed positions.
Output: The rank r completable closure $\mathrm{cl}_r(E)$.

1: Sample $\mathbf{U} \in \mathbb{R}^{m\times r}, \mathbf{V} \in \mathbb{R}^{n\times r}$ from a continuous density.
2: Compute the Jacobian matrix $\mathbf{J}_E(\mathbf{U}, \mathbf{V})$.
3: Compute the singular value decomposition of $\mathbf{J}_E(\mathbf{U}, \mathbf{V})$. Let $\mathbf{V}_E$ be the right singular vectors corresponding to singular values greater than 10^{-12}.
4: For each $e \in \mathcal{E}\backslash E$, compute the projection of $\mathbf{J}_{\{e\}}(\mathbf{U}, \mathbf{V}) \in \mathbb{R}^{r(m+n)}$ on the subspace spanned by $\mathbf{V}_E$. Let the Euclidean norm of the residual of the projection be r_e; let $r_e = 0$ for $e \in E$.
5: Return $\mathrm{cl}_r(E) := \{(i,j) \in \mathcal{E} \,;\, r_e \leq 10^{-8}\}$.

so that f is the projection of $\Upsilon(\mathbf{U}, \mathbf{V})$ onto the set of entries at positions $E \cup \{(i,j)\}$ and g then projects out the coordinate corresponding to (i,j). Lemma 8 implies that, since $(\mathbf{U}, \mathbf{V})$ is generic, all the intermediate image points are smooth. The constant rank theorem then implies:

1. We can find open neighborhoods $f(\mathcal{M}(m\times n, r)) \supset M \ni f(\Upsilon(\mathbf{U}, \mathbf{V}))$ and $g\circ f(\mathcal{M}(m\times n, r)) \supset N \ni g(f(\Upsilon(\mathbf{U}, \mathbf{V})))$ such that the restriction of g to M is smooth and $g^{-1}(N) \subset M$.

2. We have
$$\dim\left(g^{-1}(N)\right) + \dim N = \dim M.$$

Since by using smoothness again

$$\dim N = \dim\left(g(f(\Upsilon(\mathbf{U}, \mathbf{V})))\right) = \operatorname{rank}\left(\mathbf{J}_E(\mathbf{U}, \mathbf{V})\right),$$

and

$$\dim M = \dim\left(f(\Upsilon(\mathbf{U}, \mathbf{V}))\right) = \operatorname{rank}\left(\mathbf{J}_{E\cup\{i,j\}}(\mathbf{U}, \mathbf{V})\right),$$

$\dim\left(g^{-1}(N)\right) = 0$, that is, the position (i,j) is finitely completable from E and $\Upsilon(\mathbf{U}, \mathbf{V})$, if and only if

$$\operatorname{rank}\left(\mathbf{J}_E(\mathbf{U}, \mathbf{V})\right) = \operatorname{rank}\left(\mathbf{J}_{E\cup\{i,j\}}(\mathbf{U}, \mathbf{V})\right). \tag{3}$$

Equation (3) is just the assertion that $\mathbf{J}_{\{(i,j)\}} \in \operatorname{rowspan} \mathbf{J}_E$.

By Lemma 8, Equation (3) is a generic statement, independent of $\mathbf{A}$, $\mathbf{U}$ and $\mathbf{V}$. Because the rows of $\mathbf{J}_E$ and $\mathbf{J}_{\{(i,j\})}$ have non-zero columns only at positions depending on E and (i,j), whether (3) holds does not depend on m and n (which are, by hypothesis, large enough).

Finally, statement that finite completability is the same for $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ follows from Theorem 68 in the appendix.

3.3.2 Proof of Theorem 10

The theorem follows directly from Theorem 12 and the definition of closure.

3.4 Discussion

The kernel of $\mathbf{J}_E$ spans the space of infinitesimal deformations of $(\mathbf{U}, \mathbf{V})$ that preserve $\Omega \circ \Upsilon(\mathbf{U}, \mathbf{V})$. Because generic points are smooth, Milnor (1968, Curve Selection Lemma) implies that every infinitesimal deformation can be integrated to a finite deformation. Conversely (this is the harder direction) every curve in $(\Omega \circ \Upsilon)^{-1}(\mathbf{A})$ through $(\mathbf{U}, \mathbf{V})$ has, as its tangent vector a non-zero infinitesimal deformation. At non-generic points, this equivalence does not hold, so the arguments here require genericity and smoothness in an essential way.

The finite identifiability statements in this section are instances of a more general phenomenon, which is explored in Király et al. (2013). The results there imply similar identifiability results, such as Bamber (1985); Allman et al. (2009); Hsu et al. (2012); Mahdi et al. (2014); Meshkat et al. (2014), that use criteria based on a Jacobian, and also show that our use of the "Υ" parameterization of $\mathcal{M}(m \times n, r)$ is not essential.

Another connection is that, since permuting the rows and columns of a matrix preserves its rank, we get:

Corollary 15 *The rank function* $\mathrm{rank}_r(\cdot)$ *of the determinantal matroid depends only on the graph isomorphism type of the graph associated with* E*.*

In Section 6, we consider completability as a property of *graphs*. This relies on Corollary 15.

4. Unique Completability

In this section, we will address the question *"How many possible completions are there for the entry* (i, j)*?"*. In Section 3.1, it was shown that whether the *entry* (i, j) is completable depends (under mild assumptions) only on the *position* (i, j), the observed entries, and the rank. In this section, we show an analogue result that the same holds for the number of possible completions as well. Whether there is exactly one solution is of the most practical relevance, and we give a sufficient condition for unique completability.

4.1 Unique completability as a property of the positions

We start by defining what it means for one entry to be uniquely completable:

Definition 16 *Let* $E \subset \mathcal{E}$ *be a set of observed positions and* $\mathbf{A}$ *be a rank* r *true matrix. The entry* $A_{k\ell}$ *at position* $(k, \ell) \in \mathcal{E} \setminus E$ *is called* uniquely completable *from the entries* $A_{ij}, (i, j) \in E$*, if* $A_{k\ell}$ *is uniquely determined by the* $A_{ij}, (i, j) \in E$*.*

The main theoretical statement for unique completability is an analogue to the main theorem for finite completability; again, whether an entry is uniquely completable, depends only on the positions of the observations, assuming the true matrix is generic.

Theorem 17 *Let* $\mathbf{A} \in \mathbb{C}^{m \times n}$ *be a generic* $(m \times n)$*-matrix of rank* r*, and consider a masking where the entries* A_{ij} *with* $(i, j) \in E \subseteq [m] \times [n]$ *are observed. Let* $(k, \ell) \in [m] \times [n]$ *be arbitrary. Then, whether* $A_{k\ell}$ *is uniquely completable from the* $A_{ij}, (i, j) \in E$ *depends only on the position* (k, ℓ)*, the true rank* r *and the observed positions* E *(and not on* $\mathbf{A}$*,* m *or* n*).*

The proof of Theorem 17 is a bit more technical than its finite completability analogue, Theorem 10. The main problem is that the constant rank theorem cannot be applied since the latter is a local statement only and does not make say anything about the global number of solutions. The proper tools to overcome that are found in algebraic geometry; a complete proof is deferred to Section 4.4. The proof we give also shows that there is an analog statement for the total number of possible completions, even if there is more than one. Since the number of completions over the reals can potentially change even with generic $\mathbf{A}$, the result is stated only over the complex numbers.

Theorem 17 shows that it makes sense to talk about positions instead of entries that are uniquely completable, in analogy to the finite case; moreover, it shows that there is a biggest such set:

Definition 18 *Let $E \subseteq [m] \times [n]$ be the set of observed positions, and let $(k, \ell) \in [m] \times [n]$ be a position. We will call (k, ℓ) uniquely completable if $A_{k\ell}$ is uniquely completable from $A_{ij}, (i, j) \in E$ for a generic matrix $\mathbf{A} \in \mathbb{K}^{m \times n}$ of rank r.*

Furthermore, we will denote by $\operatorname{ucl}_r(E)$ the inclusion-wise maximal set of positions such that every index $(k, \ell) \in \operatorname{ucl}_r(E)$ is uniquely completable from E. We will call the $\operatorname{ucl}_r(E)$ unique closure *of E in rank r.*

As for finite completability, we can check generic unique completability of a position by testing a random $\mathbf{A}$. However, we don't have an analogue for the Jacobian $\mathbf{J}_E$ that exactly characterizes unique completability. One could, of course, use general Gröbner basis methods, but these are computationally impractical. In the next section, we describe an easy-to-check sufficient condition for unique completability in terms of the Jacobian.

4.2 Characterization by Jacobian stresses

As for the case of finite completability, the Jacobian of the masking can be used to provide algorithmic criteria to determine whether an entry is uniquely completable. The characterizing objects will be the so-called *stresses*, dual objects to the column space of the Jacobian. Intuitively, they correspond to infinitesimal dual deformations. Singer and Cucuringu (2010, Equation 3.7) have defined a similar concept which is closely related to the equilibrium stresses of Connelly (2005, Section 1.3).

Mathematically, stresses are left kernels of the Jacobian:

Definition 19 *A rank-r* stress *of the matrix $\mathbf{A} = \mathbf{U}\mathbf{V}^\top$ is a matrix $\mathbf{S} \in \mathbb{C}^{m \times n}$ whose vectorization is in the left kernel of the Jacobian $\mathbf{J}(\mathbf{U}, \mathbf{V})$; that is,*

$$\operatorname{vec} \mathbf{S} \cdot \mathbf{J}(\mathbf{U}, \mathbf{V}) = 0.$$

*Let $E \subseteq [m] \times [n]$ be a set of observed entries. A stress $\mathbf{S}$ such that $\mathbf{S}_{ij} = 0$ for all $(ij) \notin E$ is called E-*stress *of $\mathbf{A}$.*

The $\mathbb{C}$-vector space of E-stresses of $\mathbf{A}$ will be denoted by $\Psi_{\mathbf{A}}(E)$, noting that it does not depend on the choice of $\mathbf{U}, \mathbf{V}$.

Note that E-stresses are, after vectorization and removing zeroes, in the left kernel of the partial Jacobian $\mathbf{J}_E$.

The central property of the stress which allows to test for unique completability is its rank (as a matrix in $\mathbb{C}^{m\times n}$):

Definition 20 *Let $E \subseteq [m] \times [n]$ be as set of observed entries. We define the* maximal E-stress rank *of $\mathbf{A}$ in rank r to be*

$$\rho_{\mathbf{A}}(E) = \max_{\mathbf{S}\in\Psi_{\mathbf{A}}(E)} \operatorname{rank} \mathbf{S}.$$

As for the rank of the Jacobian, the dependence on $\mathbf{A}$ can be removed for generic matrices:

Proposition 21 *Let $\mathbf{A}$ be a generic $(m \times n)$-matrix of rank r. The maximal stress rank $\rho_{\mathbf{A}}(E)$ depends only on E and r. In particular, $\rho_{\mathbf{A}}(E)$ does not depend on the entries of $\mathbf{A}$.*

Proof Let $\mathbf{A} = \Upsilon(\mathbf{U}, \mathbf{V})$. By Cramer's rule, if $\mathbf{S} \in \Psi_{\mathbf{A}}(E)$, the entries of $\mathbf{S}$ are rational functions of the entries of $\mathbf{U}$ and $\mathbf{V}$. After clearing denominators, the proof is similar to that of Lemma 8. ■

We can therefore just talk about the generic E-stress rank, omitting again the dependence on the entries $\mathbf{A}$:

Definition 22 *Let $E \subseteq [m] \times [n]$ be as set of observed entries. We define the* generic E-stress rank $\rho(E)$ *to be equal to $\rho_{\mathbf{A}}(E)$ for generic $\mathbf{A}$ or rank r.*

Our main theorem states that if the generic E-stress rank is maximal for finitely completable E, then E is also uniquely completable:

Theorem 23 *Let $E \subseteq \mathcal{E}$. If the generic E-stress rank in rank r is $\rho(E) \geq \min(m, n) - r$, then $\operatorname{cl}_r(E) = \operatorname{ucl}_r(E)$.*

We defer the somewhat technical proof to Section 4.4.

4.3 Computing the generic stress rank

Theorem 23 implies that the generic stress rank $\rho(E)$ can be used to certify unique completability of an observation pattern E. We explicitly describe the necessary computational steps in Algorithm 2.

As the algorithm for finite completion, it uses a randomized strategy which allows to compute over the real numbers instead of a field of rational functions by substituting a generic entry. Steps 1 and the beginning of step 2 are thus analogous as in Algorithm 1. In step 2, the completion matrix $\mathbf{J}_E$ is computed, evaluated at the matrices $(\mathbf{U}, \mathbf{V})$. In 3, an evaluated stress $\mathbf{S}$ is obtained in the left kernel of $\mathbf{J}_E$. Its rank, which is computed in step 5, will be the generic stress rank. Correctness (with probability one) is implied by Proposition 21. Also, similar to Algorithm 1, Algorithm 2 is a randomized algorithm for which considerations analogue to those in Remark 14 hold.

Algorithm 2 Generic stress rank.
Input: Observed positions $E \subseteq \mathcal{E}$. *Output:* The generic stress rank $\rho(E)$ of E in rank r.

1: Randomly sample $\mathbf{U} \in \mathbb{R}^{m\times r}, \mathbf{V} \in \mathbb{R}^{n\times r}$.
2: Compute $\mathbf{J}_E(\mathbf{U}, \mathbf{V})$ with rows $\mathbf{J}_{(i,j)} := (\mathbf{e}_i \otimes \mathbf{v}_j^\top, \mathbf{e}_j \otimes \mathbf{u}_i^\top)$,
3: where $\mathbf{u}_i$ is the i-th row of $\mathbf{U}$, and $\mathbf{v}_j$ the j-th row of $\mathbf{V}$.
4: Compute a random vector $\mathbf{S} \in \mathbb{R}^{|E|}$ in the left kernel of $\mathbf{J}_E$. Reformat $\mathbf{S}$ as $(m \times n)$ matrix, where entries with index not in E are zero, and the remaining indices correspond to the row positions in $\mathbf{J}_E$.
5: Output $\rho(E) = \operatorname{rank}(\mathbf{S})$.

4.4 Proofs

4.4.1 Proof of Theorem 17

Proof Consider the algebraic map

$$g : (A_{k\ell}; A_{ij}, (i,j) \in E) \mapsto (A_{ij}, (i,j) \in E)$$

By Proposition 58 in the appendix, Ω is a surjective algebraic map of irreducible varieties. Therefore, the generic fiber cardinality $\left|g^{-1} \circ g(x)\right|$ for generic $x \in \mathcal{X}$ does not depend on x by Corollary 62. In particular, whether $1 = \left|g^{-1} \circ g(x)\right|$ or not.

■

4.4.2 Proof of Theorem 23

This sections contains the proof for Theorem 23 and some related results.

Lemma 24 *Let $\mathbf{S} \in \mathbb{C}^{m\times n}$ be a stress w.r.t. $m, n, r, \mathbf{A} = \mathbf{U}\mathbf{V}^\top$. Then,*

$$\mathbf{U}^\top \cdot \mathbf{S} = 0 \quad \textit{and} \quad \mathbf{S} \cdot \mathbf{V} = 0$$

(where 0 denotes the zero matrix of the correct size).

Proof Since $\mathbf{S}$ is an stress, it holds by definition that $\operatorname{vec}\mathbf{S} \cdot \mathbf{J}(\mathbf{U}, \mathbf{V}) = 0$. The statement then follows from Lemma 5. ■

Lemma 24 immediately implies a rank inequality:

Corollary 25 *Let $E \subseteq [m] \times [n]$, assume the true matrix has full rank r. Then, it holds that $\rho(E) \leq \min(m, n) - r$*

Proof Keep the notations of Lemma 24. The statement Lemma 24 implies that for arbitrary $\mathbf{S}$, one has $\mathbf{S} \cdot \mathbf{V} = 0$. Since $\mathbf{V}$ is a matrix of full rank r, this implies that the null space dimension of $\mathbf{S}$ is at least r, which is equivalent to the statement by the rank-nullity theorem.
■

In keeping with our development of finite completability in terms of $\mathbf{J}_E$, we have defined

stresses in a way that might depend on the coordinates $(\mathbf{U}, \mathbf{V})$. In the proof of Theorem 23, we will check that this can be removed when necessary. An alternative but probably less concise approach would be to express the matrix $\mathbf{J}_E$ directly in terms of the entries $\mathbf{A}$.

4.4.3 Proof of Theorem 23

We start with a general statement that stresses are invariant over the pre-image $(\Omega \circ \Upsilon)^{-1}(\Omega(\mathbf{A}))$, loosely inspired by the work of Connelly (2005).

Lemma 26 *Let* $\mathbf{A}$ *be generic, with* $\Upsilon(\mathbf{U}, \mathbf{V}) = \mathbf{A}$*,* $E \subset \mathcal{E}$*, and* $\mathbf{S}$ *an* E*-stress. Then* $\mathbf{S}$ *is also an* E*-stress for any* $(\mathbf{U}', \mathbf{V}')$ *with* $\Omega \circ \Upsilon(\mathbf{U}', \mathbf{V}') = \Omega(\mathbf{A})$

Proof Let $(\mathbf{U}', \mathbf{V}') \in (\Omega \circ \Upsilon)^{-1}(\Omega \circ \Upsilon(\mathbf{U}, \mathbf{V}))$ be a point different from $(\mathbf{U}, \mathbf{V})$. Because $\Omega(\mathbf{A})$ is a regular value of the composed map $\Omega \circ \Upsilon$, the Inverse Function Theorem provides diffeomorphic neighborhoods $M \ni (\mathbf{U}, \mathbf{V})$ and $N \ni (\mathbf{U}', \mathbf{V}')$; let $f : M \to N$ be the diffeomorphism.

By construction, df is non-singular. The chain rule then implies that $(\mathbf{J}_E)_{(\mathbf{U}', \mathbf{V}')} = (\mathbf{J}_E)_{(\mathbf{U}, \mathbf{V})} \cdot df^{-1}$, so the left kernels of both Jacobians are the same. The definition of stress as a vector in the left kernel then proves the lemma. ■

Proof [of Theorem 23] It is clear that $\mathrm{cl}_r(E) \supseteq \mathrm{ucl}_r(E)$. Thus we show that $\mathrm{cl}_r(E) \subseteq \mathrm{ucl}_r(E)$. By Lemma 26, $\mathbf{S}$ is a stress for any $(\mathbf{U}, \mathbf{V})$ that agrees with the observed entries $\Omega(\mathbf{A})$ on the observed positions E. Then by Lemma 24, any such pair $(\mathbf{U}, \mathbf{V})$ must satisfy $\mathbf{U}^\top \cdot \mathbf{S} = 0$ and $\mathbf{S} \cdot \mathbf{V}$. Since generically the stress has rank $\min(m, n) - r$, these equations determine the row and column spans of $\mathbf{A}$. Once the row and column spans are fixed, any row or column with at least r observed positions can be uniquely determined. On the other hand, any row or column with fewer than r observed positions cannot be recovered (even if the row or column span is known). Therefore we have $\mathrm{cl}_r(E) \subseteq \mathrm{ucl}_r(E)$. ■

5. Local Completion

In this section, we connect our theoretical results to the process of reconstructing the missing entries. In a nutshell, the idea is that a completable missing entry $(i, j) \in \mathcal{E} \setminus E$ is covered by at least one so-called *circuit* in $E \cup \{(i, j)\}$, to which we can associate *circuit polynomials* which can be used to solve for A_{ij} in terms of the observations, addressing the question *"What is the value of the entry* (i, j)*?"*. Just as in theory where we could separate the reconstructability from the reconstruction, we can obtain a quantitative version of this separation by estimating the entry-wise reconstruction error without actually performing the reconstruction, allowing to give an answer to *"How accurately can one estimate the entry* (i, j)*?"*. We give general algorithms for arbitrary rank, and a closed-form solution for rank one.

5.1 Circuits as rank certificates

We start with some concepts from matroid theory.

Definition 27 *A set of observed positions $C \subseteq \mathcal{E}$ is called a* circuit *of rank r if* $\operatorname{rank}_r(C) = |C| - 1$ *and* $\operatorname{rank}_r(S) = |S|$ *for all proper subsets $S \subsetneq C$. The graph $G(C)$ is called* circuit graph *of rank r.*

A reformulation of Theorem 10, in terms of circuits is the following.

Theorem 28 *The position (i, j) is finitely completable if and only if there is a circuit $C \subset E \cup \{(i, j)\}$ with $(i, j) \in C$.*

Proof See (Oxley, 2011, Lemma 1.4.3) ■

The connection to reconstructing missing entries is that every circuit comes with a unique polynomial:

Theorem 29 *Let $C \subseteq \mathcal{E}$ be a circuit in rank r, Ω_C be the mask corresponding to C, and $\mathbf{A} \in \mathbb{C}^{m\times n}$. There is a unique, up to scalar multiplication, square-free polynomial θ_C such that: $\theta_C(\Omega_C(\mathbf{A})) = 0$ if and only if there is $\mathbf{A}' \in \mathcal{M}(m \times n, r)$ and $\Omega_C(\mathbf{A}) = \Omega_C(\mathbf{A}')$.*

Proof This follows indirectly from Theorem 1.1 in Dress and Lovász (1987), or from the discussion in Section 5.2 of Király et al. (2013) ■

In other words, circuit polynomials minimally certify for the rank r condition being fulfilled on the entries in C. The simplest example of a circuit is an $(r + 1) \times (r + 1)$ rectangle in $\mathcal{E}$. The associated polynomial is the determinant of an $(r + 1) \times (r + 1)$ minor of $\mathbf{A}$. Thus, Theorem 29 is a generalization of the linear algebra fact that a matrix is rank r if and only if all $(r + 1)$-minors vanish.

Definition 30 *We will call the polynomial θ_C from Theorem 29 a* circuit polynomial *associated with the circuit C. Understanding that there are an infinity up to multiplication with a scalar multiple, we will also talk about* the *circuit polynomial when that does not make a difference.*

Remark 31 *The circuit polynomial can be interpreted algorithmically as follows: let $C \subseteq \mathcal{E}$ be a circuit, assume all entries but one in C are observed, e.g., $(k, \ell) \in C$ is not observed and $E = C \setminus (k, \ell)$ is observed. Then, $\theta_C(\Omega_C(\mathbf{A})) = \theta_C(A_{k\ell}, \Omega_E(\mathbf{A}))$ can be interpreted as a polynomial in the one unknown $A_{k\ell}$. That is, the circuit polynomial allows to solve entry-wise for single missing entries.*

Definition 32 *Fix some set of observed entries $E \subseteq \mathcal{E}$. A circuit $C \subseteq \mathcal{E}$ is called* completing *for the observations in E, or with respect to E, if $|C \cap E| \geq |C| - 1$.*

5.2 Completion with circuit polynomials

The circuit properties inspire a general solution strategy. In general, Algorithm 3 is ineffective, in the sense that Step 4 is unlikely to have a sub-exponential time algorithm in the general case. However, there is a specific instance in which it is effective: when the circuit C is always an $(r + 1) \times (r + 1)$ rectangle. In this case, the circuit polynomial is the corresponding $(r+1) \times (r+1)$ minor. This means that enumerating all the circuits through (i, j) is not necessary, because a minor is linear in the unknown entry $A_{k\ell}$.

Algorithm 3 Completion with circuits.
Input: A set $E \subset \mathcal{E}$ of observed positions.
Output: Estimates for the entries $\mathrm{cl}_r(E) \setminus E$

repeat
  Find an unobserved entry $(k, \ell) \in \mathrm{cl}_r(E) \setminus E$,
  Find the set $\mathcal{C} = \{C_1, \ldots, C_t\}$ of all circuits (w.r.t. E) containing (k, ℓ).
  Compute the circuit polynomials θ_{C_i}.
  Substitute the entries $\{A_{ij} : (i, j) \in E\}$ into the θ_{C_i} to get a family of polynomials in the variable $A_{k\ell}$ and find a solution $A_{k\ell}$ common to all of them.
  $E \leftarrow E \cup (k, \ell)$
until $E = \mathrm{cl}_r(E)$.

A practical algorithm for computing the closure of a mask E and recovering the corresponding entries based on $(r+1) \times (r+1)$ minors is given in Algorithm 4. In Step 5, $N(j)$ and $N(i)$ denote the set of neighbors of vertices $j \in W$ and $i \in V$, respectively. In Step 10, $A^+_{I',J'}$ denotes the Moore-Penrose pseudoinverse of $A_{I',J'}$. Intuitively, the algorithm iterates over missing edges and look if there is a $(r+1) \times (r+1)$ biclique in the union of current set of edges E_k and (i, j). If such a biclique exists, then the edge (i, j) is added to E_{k+1} so that the edge is used in the next round. The iteration terminates when there is no more edge to add.

Algorithm 4 `MinorClosure`$((V, W, E), r)$
Inputs: bipartite graph (V, W, E), rank r.
Outputs: completed matrix A and minor closure of E.

1: Let $E_0 \leftarrow E$ and $k \leftarrow 0$.
2: **repeat**
3: $E_{k+1} \leftarrow E_k$
4: **for** each missing edge (i, j) in $\mathcal{E} \setminus E_k$ **do**
5: Let $I \leftarrow N(j) \subseteq V$, $J \leftarrow N(i) \subseteq W$, where the neighbors are defined with respect to graph (V, W, E_k).
6: $E'_k \leftarrow I \times J \cap E_k$.
7: $(I', J') \leftarrow$ `FindAClique`$((I, J, E'_k), r, r)$.
8: **if** $|I'| > 0$ and $|J'| > 0$ **then**
9: $E_{k+1} \leftarrow E_{k+1} \cup (i, j)$.
10: $A_{ij} \leftarrow A_{i,J'} A^+_{I',J'} A_{I',j}$.
11: **end if**
12: **end for**
13: $k \leftarrow k + 1$.
14: **until** $E_k = E_{k-1}$ or $E_k = \mathcal{E}$
15: Return (A, E_k).

Note that E_{k+1} is uniquely determined from E_k and the process is monotone and bounded, i.e., $E_k \subseteq E_{k+1} \subseteq \mathcal{E}$. The first statement is true because the order of the iteration over missing edges in line 4 is irrelevant as we look if there is a $(r+1) \times (r+1)$

biclique in $E_k \cup (i,j)$ for each missing edge (i,j). Therefore, Algorithm 4 terminates with either $E_k = \mathcal{E}$ or $E_k \subsetneq \mathcal{E}$ and the following definition is valid.

Definition 33 *A set $E \subset \mathcal{E}$ is* minor closable in rank r *if Algorithm 4 reconstructs all the entries in positions $\mathcal{E} \setminus E$. Moreover, we say E is k-step minor closable in rank r, if Algorithm 4 terminates with k steps, i.e., $E_k = \mathcal{E}$ in line 14.*

Since each entry is uniquely determined when it is reconstructed, any minor closable set is uniquely completable.

A crucial step in Algorithm 4 is `FindAClique` in line 7. The function should return the indices of rows and columns, if an $r \times r$ biclique exists in subgraph (I, J, E'). This can be achieved in various ways. Although the worst case complexity is $O(|I|^r|J|^r)$, it can be much more efficient in practice, because many vertices can be safely pruned due to the fact that any $r \times r$ biclique may not contain vertices with degree less than r. An efficient implementation that employs a row-wise recursion of this step, proposed by Takeaki Uno, is presented in Appendix B.

We would like to note that Algorithm 3, as presented above, and all related algorithms below, need the true matrix to be generic. Probabilities for this supposition to hold can be backed out of from Remark 14.

5.3 Local completion

The circuit property can also be interpreted differently: instead of using multiple circuits to complete many different entries, one can also think of concentrating on one single entry and trying to reconstruct that as accurately as possible. Algorithm 5 describes a general strategy on how to obtain estimates of single finitely or uniquely completable entries, from noisy observations via local circuit completion.

Algorithm 5 Local completion/denoising of a single entry (k, ℓ).
Input: A set $E \subset \mathcal{E}$ of observed positions, the entry.
Output: Estimate for $A_{k\ell}$

1: Find completing (w.r.t. E) circuits $C_1, \dots, C_N$ containing (k, ℓ)
2: Compute the circuit polynomials θ_{C_i}, where the observed entries are substituted and $A_{k\ell}$ is the only unknown
3: For all i, find all solutions $a^{(i,j)}$ of θ_{C_i}.
4: Return $A_{k\ell} = f(\dots, a^{(i,j)}, \dots)$, where f is an appropriate averaging function

The idea in Algorithm 5 is to obtain many candidate solutions in step 3 and then trade them off appropriately in step 4. If all circuit polynomials θ_{C_i} have degree one, there is only one solution per polynomial, and f can be taken as the mean, or a weighted average that minimizes some loss or a variance. If there are some circuit polynomial with higher degree, then one can try to decide which solution is the right one - e.g., by clustering the $a^{(i,j)}$ and rejecting all candidate solutions except the one which contains some $a^{(i,j)}$ for the highest number of i, and then proceeding as in the degree one case. Also, one can imagine f being adaptive, e.g. including Bayesian learning methods.

For rank one, an closed explicit form is possible for the variance minimizing estimate, as it was shown in Kiraly and Theran (2013). For arbitrary rank, a first-order approximation to variance minimization can be employed to yield fast and competitive single-entry estimates; see Blythe et al. (2014) for a derivation of variance minimization in higher rank, and Blythe and Király (2015) for a practical adaptation of the algorithm to the context of athletic performance prediction.

For illustration, we give a short overview of the crucial statements in the rank one case. The proofs can be found in Kiraly and Theran (2013).

Theorem 34 *The rank one circuit graphs are exactly the simple cycles (bipartite and thus of even length). The corresponding circuit polynomials are all binomials of the form*

$$\theta_C = \prod_{\nu=1}^{L} A_{i_\nu j_\nu} - \prod_{\nu=1}^{L} A_{i_\nu j_{\nu+1}},$$

where L is an arbitrary number, $i_1, \dots, i_L$ are arbitrary disjoint numbers, and $j_1, \dots, j_L$ are arbitrary disjoint numbers, with the convention that $j_1 = j_{L+1}$. The i_ν and j_ν do not need to be disjoint from each other.

In particular, Theorem 34 implies that the circuit polynomials are all linear in every occurring variable. Moreover, the specific structure of the problem allows a further simplification:

Remark 35 *Keep the notations of Theorem 34. Write $B_{ij} := \log |A_{ij}|$. Then, the equations*

$$L_C = \sum_{\nu=1}^{L} B_{i_\nu j_\nu} - \sum_{\nu=1}^{L} B_{i_\nu j_{\nu+1}}$$

vanish on all rank one matrices.

With the elementary computation in Remark 35, matrix completion becomes estimation with linear boundary constraints. That is, the function f in step 4 of Algorithm 5 could be taken as the least squares regressor of all $B_{k\ell}$ obtained from completing circuits for (k, ℓ). The algorithm in Kiraly and Theran (2013) gives a version which takes different observation variances into account, and efficient graph theoretic observations making the computation polynomial.

We paraphrase this as Algorithm 6; more details, e.g. on how to efficiently find a basis for the set of completing circuits[3] is efficiently found, or how the kernel matrix Σ is constructed, can be found in Kiraly and Theran (2013).

5.4 Variance and error estimation

The locality of circuits also allows to obtain estimates for the reconstruction error of single missing entries obtained by the strategy in Section 5.3, independent of the method which does the actual reconstruction. The simplest estimate of this kind is obtained from a

3. This is equivalent to finding a basis for first $\mathbb{Z}$-homology of the graph G, taken as a 1-complex.

Algorithm 6 Local completion/denoising of a single entry (k, ℓ) in a rank 1 matrix.
Input: A set $E \subset \mathcal{E}$ of observed positions, observation variances σ, the position (k, ℓ).
Output: Estimate for $A_{k\ell}$

1: Find a basis $C_1, \ldots, C_N$ for the set of completing circuits (w.r.t E) for (k, ℓ)
2: Find solutions a_i for the corresponding circuit polynomials, write $b_i := \log |a_i|$
3: Compute the $(N \times N)$-path kernel matrix $\Sigma = \Sigma(E, \sigma)$ corresponding to the C_i; set $\alpha := \Sigma^{-1} \cdot \mathbf{1}$
4: Compute the weighted mean $b := \left(\sum_{i=1}^{N} \alpha_i \cdot a_i\right) / \left(\sum_{i=1}^{N} \alpha_i\right)$
5: As estimate, return $\widehat{A}_{k\ell} = \pm \exp(b)$, where the sign is determined by the sign parity of the circuits.

variational approach: say θ_C is a completing circuit (w.r.t $E \subseteq \mathcal{E}$) for the missing entry (k, ℓ). In the simplest case, where θ_C is linear in the missing entry $A_{k\ell}$, we can obtain a solving equation

$$\widehat{A}_{k\ell} = \theta_C(A_e, e \in E),$$

by solving for $A_{k\ell}$ as an unknown. A first order approximation for the standard error can be obtained by the variational approach

$$\delta \widehat{A}_{k\ell} = \sum_{e \in E} \frac{\partial \theta_C}{\partial A_e}(A_e, e \in E)\ \delta A_e.$$

The right hand side can be obtained from a suitable noise model and the observations A_e, or, if the error should be estimated independently from the A_e, from a noise model plus a sampling model for the A_e. A general strategy for entry-wise error estimation is analogous to Algorithm 5 for local completion. For rank one, it has been shown in Kiraly and Theran (2013) that the variance estimate depends only on the noise model and not on the actual observation, and takes a closed logarithmic-linear form, as it is sketched in Algorithm 7.

Algorithm 7 Error prediction for a single entry (k, ℓ), rank one.
Input: A set $E \subset \mathcal{E}$ of observed positions, observation variances σ, the position (k, ℓ).
Output: Estimate for the (log-)variance error of the estimate $\widehat{A}_{k\ell}$

1: Calculate Σ and α, as in Algorithm 6.
2: As log-variance, return $\alpha^\top \Sigma \alpha$.
3: If an estimate $\widehat{A}_{k\ell}$ is available, as standard error, return $\widehat{A}_{k\ell} \cdot \left(\exp(\alpha^\top \Sigma \alpha) - 1\right)$

Note that the log-variance error is independent of the actual estimate $\widehat{A}_{k\ell}$, therefore the variance patterns can be estimated without actually reconstructing the entries.

6. Combinatorial Completability Conditions

Through Sections 3 and 4, we have shown that for a given $E \subseteq \mathcal{E}$, both finitely completable closure (Theorem 12) and uniquely completable closure (Theorem 17) are properties of the (isomorphism type of) the associated bipartite graph $G(E)$; see also Corollary 15.

In this Section, using tools from graph and matroid theories, we relate the structural properties of the bipartite graph $G(E)$ to finite completability.

For a set of observed positions, $E \subseteq \mathcal{E}$, let $G(E) = (V, W, E)$ be a bipartite graph, where the sets of vertices V and W correspond to row and column of the observed positions; we call V and W *row vertices* and *column vertices*, respectively. We assume that $G(E)$ has no isolated vertices (those corresponding to rows or columns with no observed positions.)

As usual, we will take r, n, and m to be the rank and parameters of the ground set $\mathcal{E}$, respectively. However, since our convention for graphs is that they do not have isolated vertices, we will take care to indicate the ambient ground set.

6.1 Sparsity and independence

Suppose we want to maximize the size of the completable closure $\mathrm{cl}_r(E)$, with the number of positions to observe fixed. To do this, consider the process of constructing E one position at a time. What we need is to pick each successive entry in a way that causes $\mathrm{cl}_r(E)$ to grow. Theorem 12 implies that a position (k, l) is finitely completable from E, if and only if $\mathbf{J}_{\{(k,l)\}}$ lies in the span of $\mathbf{J}_E$. In particular, this tells us that adding such a (k, ℓ) to E will not affect the finite completability of other unobserved positions; in matroid terminology, we say (k, l) is dependent on E. We see, then, that it is wasteful to choose positions that are dependent on the already chosen positions. Therefore intuitively we need to choose the positions so that they are well spread out, which we call *rank-r sparse*; see Section 6.1.1. Rank-r sparsity implies a more classical combinatorial property, namely r-connectivity; see Section 6.1.2. Finally, in Section 6.1.3, we show by a counterexample that rank-r sparsity, though necessary, is not a sufficient condition for finite completability.

We recall some basic terminologies from matroid theory. The rank function $\mathrm{rank}_r(E)$ of the rank r determinantal matroid is defined in Definition 13. Note that $\mathrm{rank}_r(E) \leq d_r(m, n)$, where $d_r(m, n) = r(m + n - r)$ if $m \geq r$ and $n \geq r$, $d_r(m, n) = mn$, otherwise. A set of positions $E \subseteq \mathcal{E}$ is called *independent* if $|E| = \mathrm{rank}_r(E)$. On the other hand, it is called *dependent* if $|E| > \mathrm{rank}_r(E)$. A basis B of $E \subseteq \mathcal{E}$ is a maximally independent subset of E. In addition, a basis of $\mathcal{E}$ is called a basis of the rank r determinantal matroid. A basis B of E consists of $\mathrm{rank}_r(E)$ edges. In particular, a basis B of the rank r determinantal matroid consists of $\mathrm{rank}_r(\mathcal{E}) = d_r(m, n)$ edges. A basis of E is not unique unless E is independent. A circuit $C \subseteq \mathcal{E}$ of of the rank r determinantal matroid is a minimally dependent set in the sense that for any $(i, j) \in C$, $C - \{(i, j)\}$ is an independent set; see also Definition 27.

We have the following two properties from matroid theory.

Proposition 36 *1. Let $E \subseteq \mathcal{E}$ be a set of observed positions and $B \subseteq E$ be any basis of E. Then, $\mathrm{cl}_r(B) = \mathrm{cl}_r(E)$.*

2. Let $E \subseteq \mathcal{E}$ be an independent set in the rank r determinantal matroid. Then, any $E' \subseteq E$ is independent.

In other words, (i) the finitely completable closures of E and any basis B of E are the same (ii) and an independent graph $G(E)$ cannot contain a dependent subgraph $G(E')$. Both statements arise from the fact that the rank-r determinantal matroid is a linear matroid defined by the linear independence of the rows of the Jacobian $\mathbf{J}_E$ and that the matroid closure coincides with the finitely completable closure.

6.1.1 Rank-r-sparsity

Let $G' = (V', W', E')$ be a subgraph of $G = (V, W, E)$. Since E' being independent implies a bound on the cardinality $|E'| \leq d_r(|V'|, |W'|)$, we consider the notion of *rank-r-sparsity* defined as follows.

Definition 37 *A graph $G = (V, W, E)$ is* rank-r-sparse *if, for all subgraphs $G' = (V', W', E')$ of G, it holds that $|E'| \leq d_r(|V'|, |W'|)$.*

Theorem 38 *Let $E \subseteq \mathcal{E}$ be an independent set in the rank r determinantal matroid on $[m] \times [n]$. Then $G(E)$ is rank-r-sparse.*

Proof Suppose that there is a subgraph $G' = (V', W', E')$ with $|E'| > d_r(|V'|, |W'|) \geq \mathrm{rank}_r(E')$, then this subgraph must be dependent, which contradicts Proposition 36, part 2. ■

6.1.2 Connectivity and vertex degrees

Rank r sparsity implies some other, more classical, graph theoretic properties in a straightforward way, since rank-r-sparsity is hereditary.

Corollary 39 *Let $m, n > r$, and $E \subseteq \mathcal{E}$ be the set of observed positions. If $G(E)$ contains a rank-r sparse subgraph $G(E')$ with $|E'| = d_r(m, n)$ edges, then:*

1. *$G(E)$ has minimum vertex degree at least r.*
2. *$G(E)$ is r-edge-connected.*

In particular, if E is finitely completable, it contains a basis E' (Proposition 36, part 1) with $|E'| = d_r(m, n)$ edges and $G(E')$ is rank-r sparse. Thus, E is r-edge connected.

The proof of the above corollary relies on the following lemma:

Lemma 40 *Let $E \subseteq \mathcal{E}$ be rank-r sparse with $|E| = d_r(m, n)$ edges, and $E = \cup_{i=1}^{N} E_i$ be an edge disjoint partition of E. For any set $E' \subseteq \mathcal{E}$ of edges incident to m' row and n' column vertices, we define $d_r(E') := d_r(m', n')$. Then we have*

$$d_r(m, n) \leq \sum_{i=1}^{N} d_r(E_i).$$

Proof By the assumption,

$$d_r(m, n) = |E| = \sum_{i=1}^{N} |E_i| \leq \sum_{i=1}^{N} d_r(E_i),$$

where the first equality holds because E is independent and the last inequality follows from Theorem 38. ■

Proof [Proof of Corollary 39] Since Statement 2 implies statement 1, we prove Statement 2. First, we can assume without loss of generality that E is rank-r sparse and $E' = E$ without loss of generality, because if E' is r-edge-connected, so is E.

Consider any partition $V = V_1 \cup V_2$ and $W = W_1 \cup W_2$. V_1 or W_1 can be empty (but not at the same time). This induces an edge disjoint partition $E = E_1 \cup E_2 \cup_{(i,j)\in E-E_1-E_2} \{(i,j)\}$, where E_1 and E_2 are sets of edges induced by (V_1, W_1) and (V_2, W_2), respectively. Treating each edge in $E - E_1 - E_2$ as a subgraph, we have $d_r((i,j)) = 1$. By applying Lemma 40, we have

$$|E - E_1 - E_2| \geq d_r(m,n) - d_r(E_1) - d_r(E_2). \tag{4}$$

Let $m_1 := |V_1|$, $m_2 := |V_2|$, $n_1 := |W_1|$, and $n_2 := |W_2|$. Due to symmetry, there are three situations that we need to consider. First, if $m_1, m_2, n_1, n_2 \geq r$, RHS of (4) $= r^2$. Next, if $m_1 \leq r$ and $n_2 \leq r$, RHS of (4) $= r(m+n-r) - m_1 n_1 - m_2 n_2 \geq r^2$, which is true considering maximizing the inner product between (m_1, m_2) and (n_1, n_2) subject to $m_1 + m_2 = m$ and $n_1 + n_2 = m$. Finally, if $m_1, n_1 \leq r$, RHS of (4) $= r(m_1 + n_1) - m_1 n_1 \geq r$. The minimum is obtained for $m_1 = 1$ and $n_1 = 0$, or vice versa. Therefore E is r-edge connected. ■

6.1.3 Sparsity is not sufficient

On the other hand, rank r sparsity is *not* a sufficient condition for independence in determinantal matroids. The bipartite graph defined by the following mask in rank 2 have $d_2(5,5) = 16$ edges and rank-2 sparse but not independent:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

This example amounts, graph theoretically, to gluing the graphs of two bases of the determinantal matroid together along r vertices in a way that preserves rank-r-sparsity but not independence. One can make the construction rigorous to show that, for any $r \geq 2$, there are infinitely many rank-r-sparse dependent sets in the determinantal matroid.

6.2 Circuit and stress supports

We have discussed stresses in Section 4 and circuits in Section 5. Here we show that for each circuit C, there is a corresponding stress $\mathbf{S}$ that is supported on every position of C. Here the support $S \in \mathcal{E}$ of stress $\mathbf{S}$ is defined as $S = \{(i,j) \in \mathcal{E} : S_{ij} \neq 0\}$. Moreover, using the structure of the Jacobian matrix (see Definition 4), we show that every vertex of circuit C has degree at least $r+1$. These results further imply that any finitely completable position spans vertices in the r-core (see Section 6.2.1). Furthermore, combining the above degree lower bound with the rank-r sparsity shown in the previous subsection, we show a bound on the number of circuits in the rank r determinantal matroid in Section 6.2.2. The proof of the key Theorem 41 is presented in Section 6.2.3.

Theorem 41 *For a generic $\mathbf{A} \in \mathcal{M}(m \times n, r)$, and a circuit C, the stress space $\Psi_{\mathbf{A}}(C)$ is one dimensional; thus a stress $\mathbf{S}$ of a circuit C is unique up to scalar multiplication. Moreover, the support of $\mathbf{S}$ is all of C.*

The power of Theorem 41 can be seen in the following proposition, which lower bounds the degree of a vertex in a circuit.

Proposition 42 *Let $C \subseteq E$ be a circuit in the rank r determinantal matroid. Then every vertex in the graph $G(C)$ has degree at least $r+1$ edges.*

Proof By Theorem 41, for generic $(\mathbf{U}, \mathbf{V})$, the rows of $\mathbf{J}_C$ are dependent, with the associated stress $\mathbf{S}$ supported on all the rows.

From (2), we see that any vertex j is associated with exactly r columns in $\mathbf{J}_C$. Let J be the indices of these columns. The number of non-zero rows in $\mathbf{J}_C[\cdot, J]$ is exactly the degree d of j. If we suppose $d \leq r$, the stress $\mathbf{S}$ cannot generically cancel these d columns. Therefore, it holds that $d \geq r+1$. ■

6.2.1 Where are the completable positions?

The concept of k-core is useful for narrowing down where the completable positions can be and where the circuits can lie.

We recall a concept from graph theory:

Definition 43 *Let G be a graph, and let $k \in \mathbb{N}$. The k-*core *of G, denoted $\operatorname{core}_k(G)$, is the maximal subgraph of G with minimum vertex degree k.*

In rank r, the non-trivial aspects of matrix completion occur inside the r-core.

Theorem 44 *Let $E \subseteq \mathcal{E}$,*

(i) If $(i,j) \in \mathcal{E} \setminus E$ and $(i,j) \in \operatorname{cl}_r(E)$, then the vertices i and j are in $\operatorname{core}_r(G(E))$.

(ii) Any circuit $C \subseteq E$ is contained in $\operatorname{core}_{r+1}(G(E))$.

Proof (i) We have $(i,j) \in \operatorname{cl}_r(E)$ if and only if there is a circuit $C \subseteq E \cup \{(i,j)\}$ with $(i,j) \in C$. Then (i) will follow from (ii) because for $G(C) \subseteq \operatorname{core}_{r+1}(G(E))$, we need $i \in \operatorname{core}_r(G(E))$ and $j \in \operatorname{core}_r(G(E))$.
(ii) This follows from the fact that the $(r+1)$-core is the union of all induced subgraphs with minimum degree at least $r+1$ and by Proposition 42, every C lies inside such an induced subgraph. ■

Note here that $\operatorname{ucl}_r(E) \subseteq \operatorname{cl}_r(E)$, so the same things are true for the uniquely completable closure.

6.2.2 Circuit size and counting

Combining the results in this section, we obtain bounds on the number of circuits in the rank r determinantal matroid.

Theorem 45 *Let C be a circuit in the rank r determinantal matroid with graph $G(C) = (V, W, C)$. Then $|W| \leq r(|V| - r) + 1$*

Proof Let $m' = |V|$ and $n' = |W|$. Using Proposition 42 for the lower bound and Theorem 38 for the upper bound, we have $n'(r+1) \leq |C| \leq r(m' + n' - r) + 1$. Subtracting $n'r$ from both sides, we get $|W| = n' \leq r(m' - r) + 1$. ■

Corollary 46 *The number of circuits in the rank r determinantal matroid on $[m] \times [n]$ is at most $2^{mr(m-r)+m}$.*

6.2.3 Proof of Theorem 41

First, by Definition 19, $\operatorname{rank}_r(C) = \operatorname{rank} \mathbf{J}_C = |C| - 1$. Thus the left null space of $\mathbf{J}_C$ is one dimensional.

Next, we explicitly construct a stress $\mathbf{S}$. By Theorem 29, there is a unique polynomial θ_C for each circuit C. Then taking the derivative of θ_C, we have

$$\sum_{(i,j)\in C} \left.\frac{\partial \theta_C}{\partial A_{ij}}\right|_{\Omega_C(\mathbf{A})} dA_{ij} = 0,$$

for any tangent vector $(dA_{ij})_{(i,j)\in C}$ of $\mathcal{M}(m \times n, r)$ at $\mathbf{A}$. The vector $(\partial\theta_C/\partial A_{ij})_{(i,j)\in C}|_{\Omega_C(\mathbf{A})}$ is, then, a stress for C. In addition, the coefficient of the stress is uniquely determined by the entries $\Omega_C(\mathbf{A})$. If any of the coefficients of $(\partial\theta_C/\partial A_{ij})$ were identically zero, we could remove the associated row ij of $\mathbf{J}_C$ and the left-kernel of $\mathbf{J}_{C\setminus(i,j)}$ would still be one-dimensional. Since this is a contradiction to C being a circuit, we conclude that none of the coefficients are identically zero. Since the coefficients are, in addition, rational functions in $\Omega_C(\mathbf{A})$, each of them is non-vanishing on a Zariski open subset of $\mathcal{M}(m \times n, r)$. The (finite) intersection of these sets is again open, proving that the generic support of the stress is all of C.

6.3 Completability of random masks

Up to this point we have considered the completability of a fixed mask, which we have shown to be equivalent to questions about the associated bipartite graph. We now turn to the case where the masking is sampled at random, which, by Corollary 15, implies that, generically, this is a question about *random bipartite graphs.*

6.3.1 Random graph models

A *random graph* is a graph valued random variable. We are specifically interested in two such models for bipartite random graphs:

Definition 47 *The* Erdős-Rényi *random bipartite graph $G(m, n, p)$ is a bipartite graph on m row and n column vertices vertices with each edge present with probability p, independently.*

Definition 48 *The* (d, d')-biregular *random bipartite graph* $G(m, n, d, d')$ *is the uniform distribution on graphs with* m *row vertices,* n *column ones, and each row vertex with degree* d *and each column vertex with degree* d'*.*

Clearly, we need $md = nd'$, and if $m = n$, the (d, d')-regular random bipartite graph is, in fact d-regular.

We will call a mask corresponding to a random graph a *random mask*. We now quote some standard properties of random graphs we need.

Proposition 49 (i) *Connectivity threshold. The threshold for* $G(m, n, p)$ *to become connected, w.h.p., is* $p = \Theta((m + n)^{-1} \log n)$ *(Bollobás, 2001, Theorem 7.1).*

(Minimum degree threshold) *The threshold for the minimum degree in* $G(n, n, p)$ *to reach* d *is* $p = \Theta((m + n)^{-1}(\log n + d \log \log n + \omega(1)))$*. When* $p = cn$*, w.h.p., there are isolated vertices (Bollobás, 2001, Exercise 3.2).*

(ii) *Connectivity threshold. With high probability,* $G(m, n, d, d')$ *is* d*-connected (Bollobás, 2001, Theorem 7.3.2). (Recall that we assume* $m \leq n$*) .*

(iii) *Density principle. Suppose that the expected number of edges in either of our random graph models is at most* Cn*, for constant* C*. Then for every* $\epsilon > 0$*, there is a constant* c*, depending on only* C *and* ϵ *such that, w.h.p., every subgraph of* n' *vertices spanning at least* $(1 + \epsilon)n'$ *edges has* $n' \geq cn$ *(Janson and Luczak, 2007, Lemma 5.1).*

(iv) *Emergence of the* k*-core. Define the* k*-core of a graph to be the maximal induced subgraph with minimum* k*. For each* k*, there is a constant* c_k *such that* $p = c_k/n$ *is the first-order threshold for the* k*-core to emerge. When the* k*-core emerges, it is giant and afterwards its size and number of edges spanned grows smoothly with* p *(Pittel et al., 1996).*

6.3.2 Sparser sampling and the completable closure

The lower bounds on sample size for completion of rank r *incoherent* matrices do *not* carry over verbatim to the generic setting of this paper. This is because genericity and incoherence are related, but incomparable concepts: there are generic matrices that are not incoherent (consider a very small perturbation of the identity matrix); and, importantly, the block diagonal examples showing the lower bound for incoherent completability are not generic, since many of the entries are zero.

Thus, in the generic setting, we expect sparse sampling to be more powerful. This is demonstrated experimentally in Section 7.2. In the rest of this section, we derive some heuristics for the expected generic completability behavior of sparse random masks. We are particularly interested in the question of: *when are* $\Omega(mn)$ *of the entries completable from a sparse random mask*? We call this the *completability transition*. We will conjecture that there is a sharp threshold for the completability transition, and that the threshold occurs well below the threshold for $G(n, m, p)$ to be completable.

Let c be a constant. We first consider the emergence of a circuit in $G(n, n, c/n)$. Theorem 44 implies that any circuit is a subgraph of the $(r + 1)$-core. By Theorem 12 and

Proposition 36, having a circuit is a monotone property, which occurs with probability one for graphs with more than $2rn$ edges, and thus the value

$$t_r := \sup\{t : G(n, n, t/n) \text{ is } r\text{-independent, w.h.p.}\}$$

is a constant. If we define C_r as

$$C_r := \sup\{c : \text{the } (r+1)\text{-core of } G(n, n, c/n) \text{ has average degree at most } 2r, \text{ w.h.p.}\}$$

smoothness of the growth of the $(r+1)$-core implies that we have

$$c_{r+1} \leq t_r \leq C_{r+1}$$

where we recall that c_{r+1} is the threshold degree for the $(r+1)$-core to emerge. Putting things together we get:

Proposition 50 *There is a constant t_r such that, if $c < t_r$ then w.h.p., $G(n, n, c/n)$ is r-independent, and, if $c > t_r$ then w.h.p. $G(n, n, c/n)$ contains a giant r-circuit inside the $(r+1)$-core. Moreover, t_r is at most the threshold for the $(r+1)$-core to reach average degree $2r$.*

Proposition 50 gives us some structural information about where to look for rank r circuits in $G(n, n, c/n)$: they emerge suddenly inside of the $(r+1)$-core and are all giant when they do. If rank r circuits were themselves completable, this would then yield a threshold for the completability transition. Unfortunately, the discussion in Section 6.1.3 tell us that this is not always true. Nonetheless, we conjecture:

Conjecture 51 *The constant t_r is the threshold for the completability transition in $G(n, n, c/n)$. Moreover, we conjecture that almost all of the $(r+1)$-core is completable above the threshold.*

We want to stress that the conjecture includes a conjecture about the *existence* of the threshold for the completabilty transition, which hasn't been established here, unlike the existence for the emergence of a circuit. The subtlety is that we haven't ruled out examples of r-independent graphs with no rank-r-spanning subgraph for which, nonetheless, the closure in the rank r completion matroid is giant. Conjecture 51 is explored experimentally in Sections 7.1 and 7.2. The conjectured behavior is analogous to what has been proved for distance matrices (also known as *bar-joint frameworks*) in dimension 2 in (Kasiviswanathan et al., 2011).

Our second conjecture is about $2r$-regular masks.

Conjecture 52 *With high probability $G(n, n, 2r, 2r)$ is completable. Moreover, we conjecture that it remains so, w.h.p., after removing r^2 edges uniformly at random.*

We provide evidence in Section 7.2. This behavior is strikingly different than the incoherent case, and consistent with proven results about 2-dimensional distance matrices (Jackson et al., 2007, Theorem 4.1).

6.3.3 Denser sampling and the r-closure

The conjectures above, even if true, provide only information about matrix *completability* and not matrix *completion*. In fact, the convex relaxation of Candès and Recht (2009) does not seem to do very well on $2r$-regular masks in our experiments, and the density principle for sparse random graphs implies that, w.h.p., a $2r$-regular mask has no dense enough subgraphs for our closability algorithm in Section B.1 to even get started. Thus it seems possible that these instances are quite "hard" to complete even if they are known to be completable.

If we consider denser random masks, then the closability algorithm becomes more practical. A particularly favorable case for it is when every missing entry is part of some $K^-_{r+1,r+1}$. In this case, the error propagation will be minimal and, heuristically, finding a $K^-_{r+1,r+1}$ is not too hard, even though the problem is NP-complete in general.

Define the 1-*step r-closure* of a bipartite graph G as the graph G' obtained by adding the missing edge to each $K^-_{r+1,r+1}$ in G. If the 1-step closure of G is $K_{n,n}$, we define G to be 1-*step r-closable*. We conjecture an upper bound on the threshold for 1-step r-closability.

Conjecture 53 *There is a constant $C > 0$ such that, if $p = Cn^{-2/(r+2)} \log n$ then, w.h.p., $G(n, n, p)$ is 1-step r-closable.*

7. Experiments

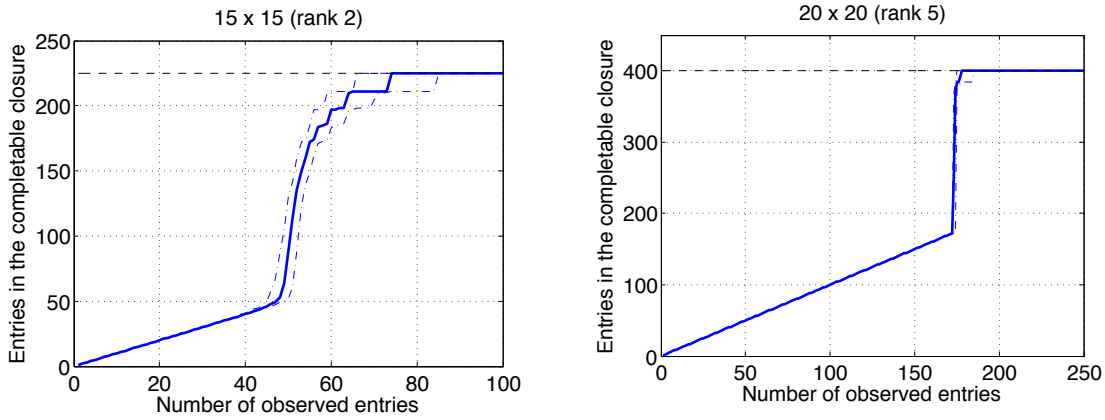
In this section we will investigate the set of entries that are finitely completable from a set of given entries. In Section 3 we have seen that the finitely completable closure $\mathrm{cl}_r(E)$ does not depend on the values of the observed entries but only on their positions E. First, we check the set of completable entries for synthetic random positions and empirically investigate the completability phase transitions in terms of the number of known entries, as described in Section 6.3. We also check the number of completable entries for MovieLens data set in terms of the putative rank. Then, we present experiments on actual reconstruction and algorithm-independent error estimation in the case of rank one matrices.

7.1 Randomized algorithms for completability

For a quantitative analysis, we perform experiments to investigate how the expected number of completable entries is influenced by the number of known entries. In particular, Section 6.3 suggests that a phase transition between the state where only very few additional entries can be completed and the state where a large set of entries can be completed should take place at some point. Figure 2 shows that this is indeed the case when slowly increasing the number of known entries: first, the set of completable entries is roughly equal to the set of known entries, but then, a sudden phase transition occurs and the set of completable entries quickly reaches the set of all entries.

7.2 Phase transitions

Figure 3 shows phase transition curves of various conditions for 100×100 matrices at rank 3. We consider uniform sampling model here. More specifically, we generated random 100×100 masks with various number of edges by first randomly sampling the order of



(a) Results for $m = 15, n = 15, r = 2$

(b) Results for $m = 20, n = 20, r = 5$

Figure 2: Expected number of completable entries (in rank r) versus the number of known entries where the positions of the known entries are uniformly randomly sampled in an $(m \times n)$-matrix. The expected number of completable entries was estimated for each data points from repeated calculations of the completable closure (200 for $r = 2$, and 20 for $r = 5$). The blue solid line is the median, the blue dotted lines are the 1st and 3rd quartiles. The black dotted line is the total number of entries, $m \cdot n$.

edges (using MATLAB `randperm` function) and adding 100 entries at a time from 100 to 6000 sequentially. In this way, we made sure to preserve the monotonicity of the properties considered here. This experiment was repeated 100 times and averaged to obtain estimates of success probabilities. The conditions plotted are (a) minimum degree at least r, (b) r-connected, (c) completable at rank r, (d) minor closable in rank r (e) nuclear norm successful, and (f) one-step minor closable. For nuclear norm minimization (e), we used the implementation of the algorithm in (Tomioka et al., 2010) which solves the minimization problem

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{subject to} \quad X_{ij} = A_{ij} \quad \forall (i,j) \in E,$$

where $\|\mathbf{X}\|_* = \sum_{j=1}^{r} \sigma_j(\mathbf{X})$ is the nuclear norm of X. The success of nuclear norm minimization is defined as the relative error $\|\hat{\mathbf{X}} - \mathbf{A}\|_F / \|\mathbf{A}\|_F$ less than 0.01.

The success probabilities of the (a) minimum degree, (b) r-connected, and (c) completable are almost on top of each other, and exceeds chance (probability 0.5) around $|E| \simeq 1,000$. The success probability of the (d) minor closable curve passes through 0.5 around $|E| \simeq 1,300$. Therefore the r-closure method is nearly optimal. On the other hand, the nuclear norm minimization required about $2,200$ entries to succeed with probability larger than 0.5.

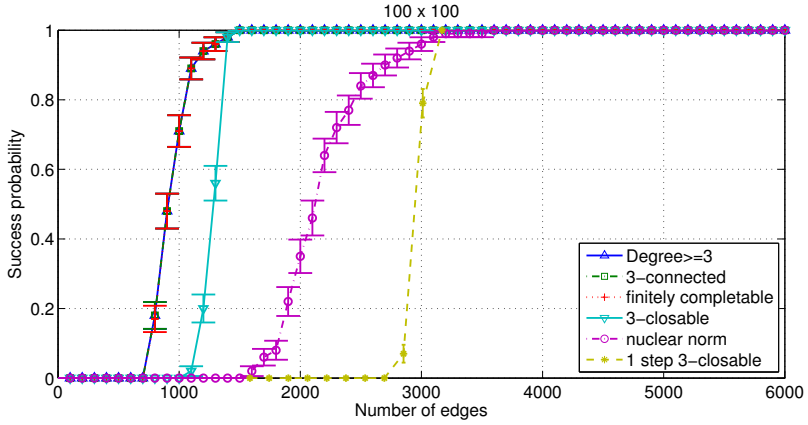


Figure 3: Phase transition curves of various conditions for 100×100 matrices at rank 3.

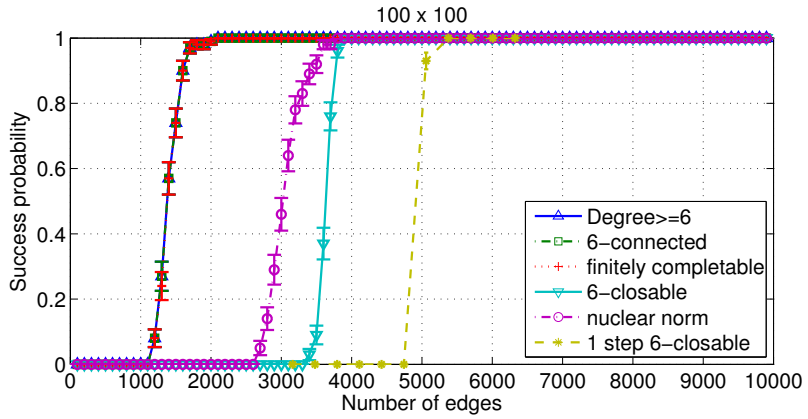


Figure 4: Phase transition curves of various conditions for 100×100 matrices at rank 6.

Figure 4 shows the same plot as above for 100×100 matrices at rank 6. The success probabilities of the (a) minimum degree, (b) r-connected, (c) completable are again almost the same, and exceeds chance probability 0.5 around $|E| \simeq 1,400$. On the other hand, the number of entries required for minor closability is at least $3,700$. This is because the masks that we need to handle around the optimal sampling density is so large and sparse that we cannot hope to find a 6×6 biclique required by the minor clusre algorithm to even get started. The nuclear norm minimization required about $3,100$ samples.

Figure 5 shows the phase transition from a non-completable mask to a completable mask for almost $2r$-regular random masks. Here we first randomly sampled $2r$-regular $(n \times n)$- masks using Steger & Wormald algorithm (Steger and Wormald, 1999). Next we randomly permuted the edges included in the mask and the edges not included in the mask independently and concatenated them into a single list of edges. In this way, we obtained a length mn ordered list of edges that become $2r$-regular exactly at the $2rn$th edge. For

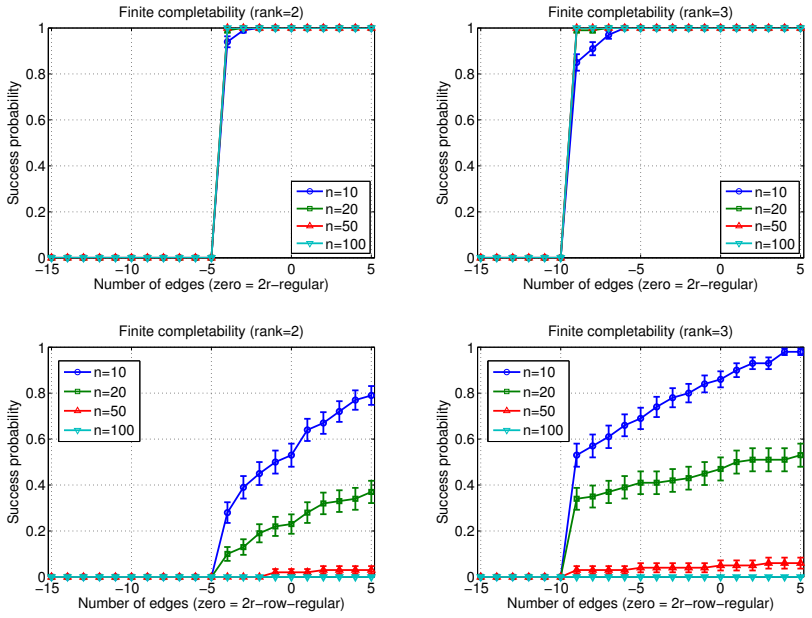


Figure 5: Phase transition in an almost regular mask.

each ordered list sampled this way, we took the first $2rn + i$ edges and checked whether the mask corresponding to these edges was completable for $i = -15, -14, \ldots, 5$. This procedure was repeated 100 times and averaged to obtain a probability estimate. In order to make sure that the phase transition is indeed caused by the regularity of the mask, we conducted the same experiment with row-wise $2r$-regular masks, i.e., each row of the mask contained exactly $2r$ entries while the number of non-zero entries varied from a column to another.

In Figure 5, the phase transition curves for different n at rank 2 and 3 are shown. The two plots in the top part show the results for the $2r$-regular masks, and the two plots in the bottom show the same results for the $2r$-row-wise regular masks. For the $2r$-regular masks, the success probability of completability sharply rises when the number of edges exceeds $2rn - r^2$ ($i = -4$ for $r = 2$ and $i = -9$ for $r = 3$); the phase transition is already rather sharp for $n = 10$ and for $n \geq 20$ it becomes almost zero or one. On the other hand, the success probabilities for the $2r$-row-wise regular masks grow rather slowly and approach zero for large n. This is natural, since it is likely for large n that there is some column with non-zero entries less than r, which violates the necessary conditions in Corollary 39.

7.3 Completability of the MovieLens data set

This section is devoted to studying a well-known data set - the MovieLens data published by GroupLens - with the methods developed in this paper. We demonstrated how the algorithms given above can be used to make statements about the sets of entries which are (a) completable, (b) uniquely completable, and (c) not completable with any algorithm.

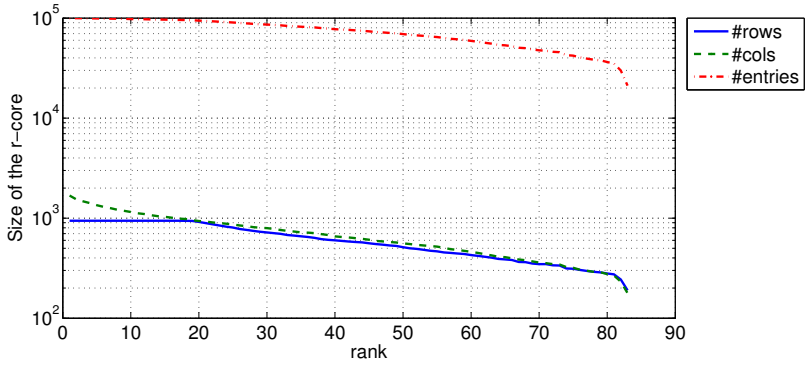


Figure 6: Size of the r-core of the MovieLens 100k data set for varying r. For each rank r, the figure shows the number of rows (solid blue), the number of columns (dashed green), and the number of entries (dash-dotted red) in the r-core of the mask corresponding to the observed entries of the MovieLens 100k data set. The biggest rank with non-empty r-core is $r = 83$.

The underlying data set for the following analyses is the MovieLens 100k data set. By convention, columns will correspond to the 1682 movies, while the rows will correspond to the 943 users in the data set.

For growing rank r, the r-core of the MovieLens data set was computed by the algorithm which is standard in graph theory - by Theorem 44 only the missing entries in the r-core can be completed, and any entry not contained in the r-core is not completable by any algorithm. Figure 6 shows the size (columns, rows, entries) of the r-core of the MovieLens data for growing r.

Under rank 18, the vast majority of the entries are in the r-core, and so is the majority of the rows, while some columns with very few entries are removed with increasing r. At rank $r = 18$, the number of columns in the r-core attains the number of rows in the r-core; above rank 18, the number of rows and columns in the r-core diminish exponentially with the same speed. Above rank 79, the r-core rapidly starts to shrink, with $r = 83$ being the biggest rank with non-empty r-core.

For growing rank r, the finitely completable closure $\mathrm{cl}_r(E)$ in the MovieLens data set was identified in the following way: First, it was checked with Algorithm 1 whether the 83-core was r-completable. If not, the completable entries in the 83-core were computed by an implementation of Algorithm 1. Then, the minor closure of the completed 83-cores was computed by Algorithm 4; by Theorem 44, it was sufficient to check for completable entries in the r-core. Note that the positions of the completable entries were also computed in the process.

Figure 7 shows the number of completable entries in the MovieLens data set for growing r determined in this way.

An interesting thing to note is the inflection point at rank $r = 18$. It corresponds to the phase transition in Figure 6 where the r-core starts to shrink exponentially and

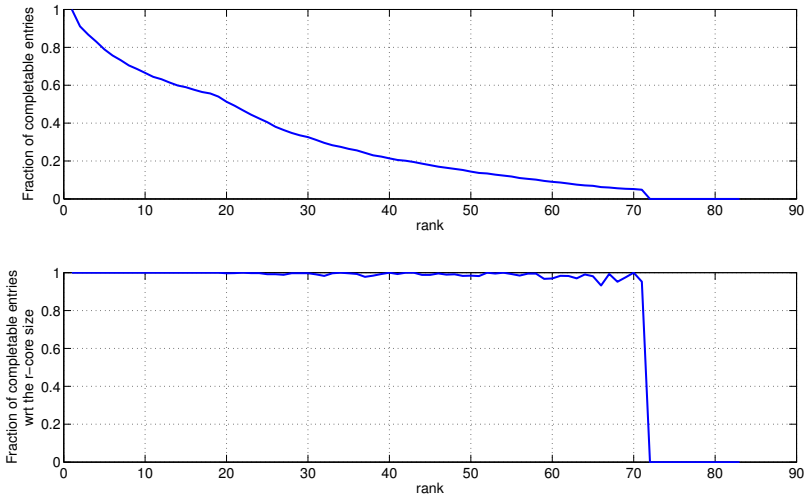


Figure 7: Number of completable entries in the MovieLens 100k data set for varying r; observed entries are not counted as completable, only completable entries which are not observed. For each rank r, the upper figure shows the number of completable entries, as a fraction of all missing entries. The lower figure shows the number of completable entries, as a fraction of the missing entries in the r-core. For $r \geq 84$, the r-core is empty, thus no missing entries can be completed, see Figure 6.

simultaneously in rows and columns. At rank $r = 72$ and above, no missing entry in the 83-core can be completed.

7.4 Entry-wise completion and error prediction

In the rest of the experiments, we recapitulate some results from Kiraly and Theran (2013) on entry-wise reconstruction and error prediction for rank one matrices.

To test reconstruction, we generated 10 random masks of size 50×50 with 200 entries sampled uniformly and a random (50×50) matrix of rank one. The multiplicative noise was chosen entry-wise independent, with variance $\sigma_i = (i-1)/10$ for each entry. Figure 9(a) compares the Mean Squared Error (MSE) for three algorithms: Nuclear Norm (using the implementation Tomioka et al. (2010)), OptSpace (Keshavan et al., 2010), and Algorithm 6. It can be seen that on these masks, Algorithm 6 is competitive with the other methods and even outperforms them for low noise.

Figure 9(b) compares the error of each of the methods with the variance predicted by Algorithm 7 each time the noise level changed. The figure shows that for any of the algorithms, the mean of the actual error increases with the predicted error, showing that the error estimate is useful for a-priori prediction of the actual error - independently of the particular algorithm. Note that by construction of the data this statement holds in particular for entry-wise predictions. Furthermore, in quantitative comparison Algorithm 7 also outperforms the other two in each of the bins. The qualitative reversal between the

algorithms in Figures 9(b) (a) and (b) comes from the different error measure and the conditioning on the bins.

7.5 Universal error estimates

For three different masks, we calculated the predicted minimum variance for each entry of the mask. The mask sizes are all 140×140. The noise was assumed to be i.i.d. Gaussian multiplicative with $\sigma_e = 1$ for each entry. Figure 8 shows the predicted a-priori minimum variances for each of the masks. The structure of the mask affects the expected error. Known entries generally have least variance, and it is less than the initial variance of 1, which implies that the (independent) estimates coming from other paths can be used to successfully denoise observed data. For unknown entries, the structure of the mask is mirrored in the pattern of the predicted errors; a diffuse mask gives a similar error on each missing entry, while the more structured masks have structured error which is determined by combinatorial properties of the completion graph.

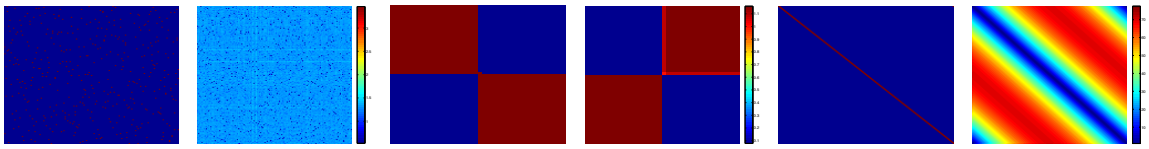


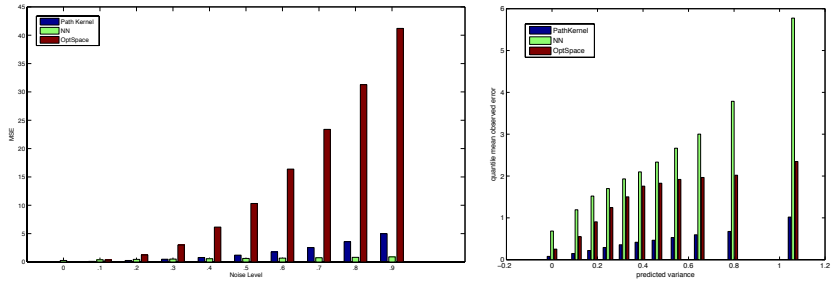
Figure 8: The figure shows three pairs of masks and predicted variances. A pair consists of two adjacent squares. The left half is the mask which is depicted by red/blue heatmap with red entries known and blue unknown. The right half is a multicolor heatmap with color scale, showing the predicted variance of the completion. Variances were calculated by our implementation of Algorithm 7.

8. Discussion and Outlook

In this paper we have demonstrated the usefulness and practicability of the algebraic combinatorial approach for matrix completion, by deriving reconstructability statements, and actual reconstruction algorithms for single missing entries. Our theory allows to treat the positions of the observations separately from the entries themselves. As a prominent model feature, we are able to separate the sampling scheme from algebraic and combinatorial conditions for reconstruction and explain existing reconstruction bounds by the combinatorial phase transition for the uniform random sampling scheme.

The discussed framework provides the foundation for a number of **novel matrix completion strategies for the practitioner**:

- The presented algorithms allow for **entry-wise error estimates** which are independent of the method. More precisely, as it has been studied by Kiraly and Theran (2013) for rank 1, the algorithm of actual reconstruction can be separated from the question whether the entry is reconstructible, and with which error, allowing the com-



(a) mean squared errors (b) error vs. predicted variance

Figure 9: For 10 randomly chosen masks and 50×50 true matrix, matrix completions were performed with Nuclear Norm (green), OptSpace (red), and Algorithm 6 (blue) under multiplicative noise with variance increasing in increments of 0.1. For each completed entry, minimum variances were predicted by Algorithm 7. 9(a) shows the mean squared error of the three algorithms for each noise level, coded by the algorithms' respective colors. 9(b) shows a bin-plot of errors (y-axis) versus predicted variances (x-axis) for each of the three algorithms: for each completed entry, a pair (predicted error, true error) was calculated, predicted error being the predicted variance, and the actual prediction error being the squared logarithmic error (i.e., $(\log|a_{true}| - \log|a_{predicted}|)^2$ for an entry a). Then, the points were binned into 11 bins with equal numbers of points. The figure shows the mean of the errors (second coordinate) of the value pairs with predicted variance (first coordinate) in each of the bins, the color corresponds to the particular algorithm; each group of bars is centered on the minimum value of the associated bin.

bination of any reconstruction algorithm with reconstruction bounds obtained from our framework.

- The presented ideas allow **completion/denoising of single entries** in the practically relevant case where only one entry or a subset of all entries should be reconstructed or denoised. A rank one method has been presented by Kiraly and Theran (2013), the case of rank 2 and higher is studied by Blythe et al. (2014).
- The use of circuits for reconstruction pave the way for **local completion/denoising**, that is, a good reconstruction can be obtained from a small combinatorial neighborhood of entries which can be determined from the theory (and which is not necessarily a submatrix), allowing to avoid processing of the whole matrix - which is especially desirable if the matrix is huge.

In our new setting, we are also left with a number of major **open questions**:

- Characterize all **circuits and circuit polynomials** in rank 2 or higher.
- Give a sufficient and necessary combinatorial **criterion for unique completability**.

- Give an **efficient**[4] **algorithm certifying for unique completability** when given the positions of the observed entries (or, more generally, one which computes the number of solutions).
- Prove the **phase transition bound for the completable core** (the phase transition bound for completability has been shown in Király and Theran, 2013).
- Explain the existing **guarantees for whole matrix reconstruction MSE in terms of single entry expected error**, for the various sampling models in literature (an explanation for rank one can be inferred from Kiraly and Theran, 2013).

Finally, our presented results suggest a number of **future directions**:

- Problems such as **matrix completion under further constraints** such as for **symmetric matrices**, **distance matrices** or **kernel matrices**, are closely related to the ones we consider here, and can be treated by similar techniques. Under a phase transition aspect, these models were studied by Király and Theran (2013); for general matroids, the theory in (Király et al., 2013) yields a starting point.
- **Completion of tensors** is a natural generalization of matrix completion and accessible to the techniques presented here or in (Király and Theran, 2013; Király et al., 2013).
- We have essentially shown matrix completion to be an **algebraic manifold learning problem**. This makes it accessible to the kernel/ideal learning techniques presented in (Király et al., 2014).
- The algebraic theory used to infer genericity and identifiability is largely independent of the matrix completion setting and can be applied in a very general context of **compressed sensing, identifiability and inverse problems that are algebraic**. For a more detailed discussion and some related problems, see Section 3.4.

Summarizing, we argue that recognizing and exploiting algebra and combinatorics in machine learning problems is beneficial from the practical and theoretical perspectives. When it is present, methods using underlying algebraic and combinatorial structures yield sounder statements and more practical algorithms than can be obtained when ignoring it, conversely algebra and combinatorics can profit from the various interesting structure surfacing in machine learning problems. Therefore all involved fields can only profit from a more widespread interdisciplinary collaboration with and between each other.

Acknowledgments

We thank Andriy Bondarenko, Winfried Bruns, Eyke Hüllermeyer, Mihyun Kang, Yusuke Kobayashi, Martin Kreuzer, Cris Moore, Klaus-Robert Müller, Kazuo Murota, Kiyohito Nagano, Zvi Rosen, Raman Sanyal, Bernd Sturmfels, Sumio Watanabe, Volkmar Welker, and Günter Ziegler for valuable discussions. RT is partially supported by MEXT KAKENHI

4. say polynomial time, success with high probability

22700138, the Global COE program "The Research and Training Center for New Development in Mathematics", FK by Mathematisches Forschungsinstitut Oberwolfach (MFO), and LT by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 247029-SDModels. This research was partially carried out at MFO, supported by FK's Oberwolfach Leibniz Fellowship.

Appendix A. Algebraic Geometry Fundamentals

This section collects some algebraic geometric tools used in the main corpus.

A.1 Algebraic Genericity

We will briefly review the concept of genericity for our purposes. Intuitively, algebraic genericity describes that some statements holds for almost all objects, with the exceptions having an algebraic structure. The following results will be stated for algebraic varieties over the real or complex numbers, that is, over the field $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

Definition 54 *Let $\mathcal{Y} \subseteq \mathbb{K}^n$ be an algebraic variety. Let P be some property of points $y \in \mathcal{Y}$. Write $P(\mathcal{Y}) = \{y \in \mathcal{Y} \,:\, y \text{ has property } P\}$, and $\neg P(\mathcal{Y}) = \mathcal{Y} \setminus P(\mathcal{Y})$.*

(i) *We call P an* open condition *if $P(\mathcal{Y})$ is a Zariski open subset of $\mathcal{Y}$.*

(ii) *We call P a* Zariski-generic condition *if there is an open dense subset $U \subseteq \mathcal{Y}$ such that $U \subseteq P(\mathcal{Y})$.*

(iii) *We call P a* Hausdorff-generic condition *if $\neg P(\mathcal{Y})$ is a $\mathcal{Y}$-Hausdorff zero set.*

The different types of conditions above can be put in relation to each other:

Proposition 55 *Keep the notation of Definition 54.*

(i) *If P is a Zariski-generic condition, then P is a Hausdorff-generic condition as well.*

(ii) *Assume $\mathcal{Y}$ is irreducible, and $P(\mathcal{Y})$ is non-empty. If P is an open condition, then it is a Zariski-generic condition.*

(iii) *Assume $\mathcal{Y}$ is irreducible, and $P(\mathcal{Y})$ is constructible in the Zariski topology, i.e., can be written as finite union and intersection of open and closed sets. If P is a Hausdorff-generic condition, then it is a Zariski-generic condition as well.*

Proof (i) follows from the fact that Zariski closed sets of smaller Krull dimension are Hausdorff zero sets.
(ii) follows from the fact that non-empty Zariski open sets are dense in an irreducible algebraic set.
(iii) as $P(\mathcal{Y})$ is Zariski-constructible, it will have positive Hausdorff measure if and only if it contains a non-empty (relatively Zariski) open set. The statement then follows from (ii). ■

Furthermore, Hausdorff-genericity is essentially states that the condition holds, universally with probability one:

Proposition 56 *Keep the notation of Definition 54. The following are equivalent:*

(i) *P is a Hausdorff-generic condition.*

(ii) *For all Hausdorff-continuous random variables X taking values in $\mathcal{Y}$, the statement $P(X)$ holds with probability one.*

Proof (i)$\Leftrightarrow$ (ii) follows from taking Radon-Nikodym derivatives. ■

All relevant properties and conditions which are referenced from the main corpus describe (a) irreducible varieties - in this case, the determinantal variety, and (b) are Zariski-constructible. Therefore, by Proposition 55, all three definitions agree for the purpose of this paper. The terminology used in the paper can be given as follows in the above definitions:

Definition 57 *Let $\mathcal{Y} \subseteq \mathbb{K}^n$ be an algebraic variety. Let P be some property of points $y \in \mathcal{Y}$. We say "a generic $y \in \mathcal{Y}$ has property P" if P is a Hausdorff-generic condition for points in $\mathcal{Y}$.*

A.2 Open Conditions and Generic Properties of Morphisms

In this section, we will summarize some algebraic geometry results used in the main corpus. The following results will always be stated for algebraic varieties over $\mathbb{C}$.

Proposition 58 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a morphism of algebraic varieties (over any field). Then, if $\mathcal{X}$ is irreducible, so is $f(\mathcal{X})$. In particular, if f is surjective, and $\mathcal{X}$ is irreducible, then $\mathcal{Y}$ also is.*

Proof This is classical and follows directly from the fact that morphisms of algebraic varieties are continuous in the Zariski topology. ■

Theorem 59 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a morphism of algebraic varieties. The function*

$$\mathcal{Y} \to \mathbb{N}, \quad y \mapsto \dim f^{-1}(y)$$

is upper semicontinuous in the Zariski topology.

Proof This follows from (Grothendieck and Dieudonné, 1966, Théorème 13.1.3). ■

Proposition 60 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a morphism of algebraic varieties, with $\mathcal{Y}$ be irreducible. Then, there is an open dense subset $V \subseteq \mathcal{Y}$ such that $f : U \to V$, where $U = f^{-1}(V)$, is a flat morphism.*

Proof This follows from (Grothendieck and Dieudonné, 1965, Théorème 6.9.1). ■

Theorem 61 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a morphism of algebraic varieties. Let $d, \nu \in \mathbb{N}$. Then, the following are open conditions for $y \in \mathcal{Y}$:*

(i) $\dim f^{-1}(y) \leq d$.

(ii) *f is unramified over y.*

(iii) *f is unramified over y, and the number of irreducible components of $f^{-1}(y)$ equals ν.*

In particular, if f is surjective, then the following is an open property as well:

(iv) *f is unramified over y, and $\left|f^{-1}(y)\right| = \nu$, for some $\nu \in \mathbb{N}$.*

Proof (i) follows from (Grothendieck and Dieudonné, 1965, Corollaire 6.1.2).
(ii) follows from (Grothendieck and Dieudonné, 1966, Théorème 12.2.4(v)).
(iii) follows from (Grothendieck and Dieudonné, 1966, Théorème 12.2.4(vi)).
(iv) follows from (i), applied in the case $\dim f^{-1}(y) \leq 0$ which is equivalent to $\dim f^{-1}(y) = 0$ due to surjectivity of f, and (iii). ■

Corollary 62 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a generically unramified and surjective morphism of algebraic varieties, with $\mathcal{Y}$ be irreducible. Then, there are unique $d, \nu \in \mathbb{N}$ such that the following sets are Zariski closed, proper subsets of $\mathcal{Y}$ (and therefore Hausdorff zero sets):*

(i) $\{y \ : \ \dim f^{-1}(y) \neq d\}$

(ii) $\{y \ : \ f \text{ is ramified at } y\}$

(iii) $\{y \ : \ f \text{ is ramified at } y\} \cup \{y \ : \ \left|f^{-1}(y)\right| \neq \nu\}$

Proof This is implied by Theorem 61 (i), (ii) and (iii), using that a non-zero open subset of the irreducible variety $\mathcal{Y}$ must be open dense, therefore its complement in $\mathcal{Y}$ a closed and a proper subset of $\mathcal{Y}$. ■

Proposition 63 *Let $f : \mathcal{X} \to \mathcal{Y}$ be a morphism of algebraic varieties, with $\mathcal{Y}$ irreducible. Then, the following are equivalent:*

(i) *f is unramified over y and $\left|f^{-1}(y)\right| = \nu$.*

(ii) *There is a Borel open neighborhood $U \subseteq \mathcal{Y}$ of $y \in U$, such that f is unramified over U and $\left|f^{-1}(z)\right| = \nu$ for all $z \in U$.*

(iii) *There is a Zariski open neighborhood $U \subseteq \mathcal{Y}$ of $y \in U$, dense in $\mathcal{Y}$, such that f is unramified over U and $\left|f^{-1}(z)\right| = \nu$ for all $z \in U$.*

Proof The equivalence is implied by Corollary 62 and the fact that $\mathcal{Y}$ is irreducible. Note that either condition implies that f is generically unramified due to Theorem 61 (ii) and irreducibility of $\mathcal{Y}$. ■

A.3 Real versus Complex Genericity

We derive some elementary results how generic properties over the complex and real numbers relate. While some could be taken for known results, they appear not to be folklore - except maybe Lemma 65. In any case, they seem not to be written up properly in literature known to the authors. A first version of the statements below has also appeared as part of Király and Ehler (2014).

Definition 64 *Let $\mathcal{X} \subseteq \mathbb{C}^n$ be a variety. We define the* real part *of $\mathcal{X}$ to be $\mathcal{X}_\mathbb{R} := \mathcal{X} \cap \mathbb{R}^n$.*

Lemma 65 *Let $\mathcal{X} \subseteq \mathbb{C}^n$ be a variety. Then, $\dim \mathcal{X}_\mathbb{R} \leq \dim \mathcal{X}$, where $\dim \mathcal{X}_\mathbb{R}$ denotes the Krull dimension of $\mathcal{X}_\mathbb{R}$, regarded as a (real) subvariety of $\mathbb{R}^n$, and $\dim \mathcal{X}$ the Krull dimension of $\mathcal{X}$, regarded as subvariety of $\mathbb{C}^n$.*

Proof Let $k = n - \dim \mathcal{X}$. By (Mumford, 1999, Section 1.1), $\mathcal{X}$ is contained in some complete intersection variety $\mathcal{X}' = \mathrm{V}(f_1, \ldots, f_k)$. That is $(f_1, \ldots, f_k)$ is a complete intersection, with $f_i \in \mathbb{C}[X_1, \ldots, X_n]$ and $\dim \mathcal{X}' = \dim \mathcal{X}$, such that f_i is a non-zero divisor modulo $f_1, \ldots, f_{i-1}$. Define $g_i := f_i \cdot f_i^*$, one checks that $g_i \in \mathbb{R}[X_1, \ldots, X_n]$, and define $\mathcal{Y} := \mathrm{V}(g_1, \ldots, g_k)$ and $\mathcal{Y}_\mathbb{R} := \mathcal{Y} \cap \mathbb{R}^n$. The fact that f_i is a non-zero divisor modulo $f_1, \ldots, f_{i-1}$ implies that g_i is a non-zero divisor modulo $g_1, \ldots, g_{i-1}$; since $g_i \cdot h \cong 0$ modulo $g_1, \ldots, g_{i-1}$ implies $f_i \cdot (h \cdot f_i^*) \cong 0$ modulo $f_1, \ldots, f_{i-1}$. Therefore, $\dim \mathcal{Y}_\mathbb{R} \leq \dim \mathcal{X}$; by construction, $\mathcal{X}' \subseteq \mathcal{Y}$, and $\mathcal{X} \subseteq \mathcal{X}'$, therefore $\mathcal{X}_\mathbb{R} \subseteq \mathcal{Y}_\mathbb{R}$, and thus $\dim \mathcal{X}_\mathbb{R} \leq \dim \mathcal{Y}_\mathbb{R}$. Combining it with the above inequality yields the claim. ■

Definition 66 *Let $\mathcal{X} \subseteq \mathbb{C}^n$ be a variety. If $\dim \mathcal{X} = \dim \mathcal{X}_\mathbb{R}$, we call $\mathcal{X}$* observable over the reals*. If $\mathcal{X}$ equals the (complex) Zariski-closure of $\mathcal{X}_\mathbb{R}$, we call $\mathcal{X}$* defined over the reals*.*

Proposition 67 *Let $\mathcal{X} \subseteq \mathbb{C}^n$ be a variety.*

(i) *If $\mathcal{X}$ is defined over the reals, then $\mathcal{X}$ is also observable over the reals.*

(ii) *The converse of (i) is false.*

(iii) *If $\mathcal{X}$ irreducible and observable over the reals, then $\mathcal{X}$ is defined over the reals.*

Proof (i) Let $k = n - \dim \mathcal{X}_\mathbb{R}$. By (Mumford, 1999, Section 1.1), $\mathcal{X}_\mathbb{R}$ is contained in some complete intersection variety $\mathcal{X}' = \mathrm{V}(f_1, \ldots, f_k)$, with $f_i \in \mathbb{R}[X_1, \ldots, X_n]$ a complete intersection. By an argument, analogous to the proof of Lemma 65, one sees that the f_i are a complete intersection in $\mathbb{C}[X_1, \ldots, X_n]$ as well. Since the Zariski-closure of $\mathcal{X}_\mathbb{R}$ and $\mathcal{X}$ are equal, it holds that $f_i \in \mathrm{I}(\mathcal{X})$. Therefore, $\mathcal{X} \subseteq \mathrm{V}(f_1, \ldots, f_k)$, which imples $\dim \mathcal{X} \leq n - k$, and by definition of k, as well $\dim \mathcal{X} \leq \dim \mathcal{X}_\mathbb{R}$. With Lemma 65, we obtain $\dim \mathcal{X}_\mathbb{R} = \dim \mathcal{X}$, which was the statement to prove.
(ii) It suffices to give a counterexample: $\mathcal{X} = \{1, i\} \subseteq \mathbb{C}$. Alternatively (in a context where $\varnothing$ is not a variety) $\mathcal{X} = \{(1, x) \,:\, x \in \mathbb{C}\} \cup \{(i, x) \,:\, x \in \mathbb{C}\} \subseteq \mathbb{C}^2$.
(iii) By definition of dimension, Zariski-closure preserves dimension. Therefore, the closure $\overline{\mathcal{X}_\mathbb{R}}$ is a sub-variety of $\mathcal{X}$, with $\dim \overline{\mathcal{X}_\mathbb{R}} = \dim \mathcal{X}$. Since $\mathcal{X}$ is irreducible, equality $\overline{\mathcal{X}_\mathbb{R}} = \mathcal{X}$ must hold. ■

Theorem 68 *Let $\mathcal{X} \subseteq \mathbb{C}^n$ be an irreducible variety which is observable over the reals, let $\mathcal{X}_\mathbb{R}$ be its real part. Let P be an algebraic property. Assume that a generic $x \in \mathcal{X}$ is P. Then, a generic $x \in \mathcal{X}_\mathbb{R}$ has property P as well.*

Proof Since P is an algebraic property, the P points of $\mathcal{X}$ are contained in a proper sub-variety $\mathcal{Z} \subseteq \mathcal{X}$, with $\dim \mathcal{Z} \lneq \dim \mathcal{X}$. Since $\mathcal{X}$ is observable over the reals, it holds $\dim \mathcal{X} = \dim \mathcal{X}_\mathbb{R}$. By Lemma 65, $\dim \mathcal{Z}_\mathbb{R} \leq \dim \mathcal{Z}$. Putting all (in-)equalities together, one obtains $\dim \mathcal{Z}_\mathbb{R} \lneq \dim \mathcal{X}_\mathbb{R}$. Therefore, the $\mathcal{Z}_\mathbb{R}$ is a proper sub-variety of $\mathcal{X}_\mathbb{R}$; and the P points of $\mathcal{X}_\mathbb{R}$ are contained in it - this proves the statement. ■

A.4 Algebraic Properties of the Masking

We conclude with checking the conditions previously discussed in the specific case of the masking:

Proposition 69 *For $E \subseteq \mathcal{E}$, consider the determinantal variety $\mathcal{M}(m \times n, r)$ (over $\mathbb{C}$), and the masking*

$$\Omega : \mathcal{M}(m \times n, r) \to \mathbb{C}^{|E|}, \quad \mathbf{A} \mapsto \{A_e, e \in E\}.$$

(i) *The determinantal variety $\mathcal{M}(m \times n, r)$ is irreducible.*

(ii) *The determinantal variety $\mathcal{M}(m \times n, r)$ is observable over the reals.*

(iii) *The determinantal variety $\mathcal{M}(m \times n, r)$ is defined over the reals.*

(iv) *The variety $\Omega(\mathcal{M}(m \times n, r))$ is irreducible.*

(v) *The map Ω is generically unramified.*

Proof (i) follows from Proposition 58, applied to the surjective map

$$\Upsilon : \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times r} \to \mathcal{M}(m \times n, r),\ (U, V) \mapsto UV^\top,$$

and irreducibility of affine space $\mathbb{C}^{m \times r} \times \mathbb{C}^{n \times r}$.
(ii) follows from considering the map Υ over the reals, observing that the rank its Jacobian is not affected by this.
(iii) follows from (i), (ii) and Proposition 67 (iii).
(iv) follows from (i) and Proposition 58, applied to Ω.
(v) follows from the fact that Ω is a coordinate projection, therefore linear. ■

Appendix B. Advanced Algorithm for Minor Closure

Takeaki Uno
uno@nii.jp
National Institute of Informatics
Tokyo 101-8430, Japan

B.1 Closability and the r-Closure

A crucial step in minor closure algorithm 4 is to find a $r \times r$ biclique in a (sub)graph(V, W, E).

Algorithm 8 can find an $d_1 \times d_2$ biclique efficiently based on two ideas: (i) iterate over row vertices and (ii) work only on the (d_2, d_1)-core; here (d_2, d_1)-core is the maximal subgraph of (V, W, E) that the degrees of the row and column vertices are at least d_2 and d_1, respectively.

A naive approach for finding an $r \times r$ biclique might be to iterate over edges in E and for each edge $(v, w) \in E$ whose nodes have at least $r - 1$ neighbors, check whether the subgraph induced by $(N(w), N(v))$ contains an $r-1 \times r-1$ biclique. Instead our algorithm iterates over nodes in V and for each node $v \in V$, check whether the subgraph induced by $(V', N(v))$ contains an $r - 1 \times r$ biclique, where V' is defined by removing all previously attempted nodes and the current v from V. Iterating over row vertices results in smaller number of iterations because $|V| \leq |E|$, and allows us to avoid double checking, because previously attempted nodes can be removed from V'. Concentrating on the (d_2, d_1)-core is natural, because no $d_1 \times d_2$ biclique contains row or column vertex with degree less than d_2 or d_1, respectively. We present the pruning step for finding the (d_2, d_1)-core in Algorithm 9.

Algorithm 8 `FindAClique`$((V, W, E), d_1, d_2)$

Inputs: bipartite graph (V, W, E), size of the bipartite clique to be found $d_1 \times d_2$.
Output: vertex sets of a clique (I, J).
$(V, W, E) \leftarrow$ `FindCore`$((V, W, E), d_2, d_1)$.
if $|V| < d_1$ or $|W| < d_2$ **then**
    Return $(\emptyset, \emptyset)$.
end if
$V' \leftarrow V$.
for each $v \in V$ **do**
    if $d_1 = 1$ and $|N(v)| \geq d_2$ **then**
        Return $(\{v\}, N(v))$.
    end if
    $V' \leftarrow V' \backslash \{v\}$, $W' \leftarrow N(v)$, $E' \leftarrow (V' \times W') \cap E$.
    $(I', J') \leftarrow$ `FindAClique`$((V', W', E'), d_1 - 1, d_2)$.
    if $|I'| > 0$ and $|J'| > 0$ **then**
        Return $(I' \cup \{v\}, J')$.
    end if
end for
Return $(\emptyset, \emptyset)$.

References

Evrim Acar, Daniel M. Dunlavy, and Tamara G. Kolda. Link prediction on evolving data using matrix and tensor factorizations. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*, pages 262–269. IEEE, 2009.

Elizabeth S. Allman, Catherine Matias, and John A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics*, 37(6A):3099–3132, 2009.

Algorithm 9 `FindCore`$((V, W, E), d_1, d_2)$

Inputs: bipartite graph (V, W, E), minimum degree d_1 for the row vertices and d_2 for the column vertices.
Output: pruned bipartite graph (V, W, E).
while true **do**
  $V' \leftarrow \{v \in V : |N(v)| < d_1\}$.
  $W' \leftarrow \{w \in W : |N(w)| < d_2\}$.
  if $V' = \emptyset$ and $W' = \emptyset$ **then**
    Return $(V, W, (V \times W) \cap E)$.
  end if
  $V \leftarrow V \backslash V'$.
  $W \leftarrow W \backslash W'$.
end while

Andreas Argyriou, Craig A. Micchelli, Massimiliano Pontil, and Yi Ying. A spectral regularization framework for multi-task structure learning. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in NIPS 20*, pages 25–32. MIT Press, Cambridge, MA, 2008.

Donald Bamber. How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, 29(4):443 – 473, 1985.

Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*. MIT Press, 2014.

Duncan Blythe and Franz J. Király. Prediction and quantification of individual athletic performance. *arXiv e-prints*, 2015. arXiv 1505.01147.

Duncan Blythe, Franz J. Király, and Louis Theran. Algebraic combinatorial methods for low-rank matrix completion with application to athletic performance prediction. *arXiv e-prints*, 2014. arXiv 1406.2864.

Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

Winfried Bruns and Udo Vetter. *Determinantal Rings*. Springer-Verlag New York, Inc., 1988.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics. The Journal of the Society for the Foundations of Computational Mathematics*, 9(6):717–772, 2009. ISSN 1615-3375.

Emmanuel J. Candès and Terence Tao. The power of convex relaxation: near-optimal matrix completion. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 56(5):2053–2080, 2010.

Robert Connelly. Generic global rigidity. *Discrete Comput. Geom.*, 33(4):549–563, 2005. ISSN 0179-5376. doi: 10.1007/s00454-004-1124-4. URL `http://dx.doi.org/10.1007/s00454-004-1124-4`.

Andreas Dress and László Lovász. On some combinatorial properties of algebraic matroids. *Combinatorica*, 7(1):39–48, 1987.

Rina Foygel and Nathan Srebro. Concentration-based guarantees for low-rank matrix reconstruction. *arXiv e-prints*, 2011. arXiv:1102.3923.

Andrew Goldberg, Xiaojin Zhu, Benjamin Recht, Jun-Ming Xu, and Robert Nowak. Transduction with matrix completion: Three birds with one stone. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems (NIPS) 23*, pages 757–765. 2010.

Alexander Grothendieck and Jean Dieudonné. Éléments de géométrie algébrique IV, deuxième partie. *Publications mathématiques de l'IHÉS*, 24, 1965.

Alexander Grothendieck and Jean Dieudonné. Éléments de géométrie algébrique IV, troisième partie. *Publications mathématiques de l'IHÉS*, 28, 1966.

GroupLens. Movielens 100k data set. Available online at `http://grouplens.org/datasets/movielens/`; as downloaded on November 27th 2012.

Daniel Hsu, Sham Kakade, and Percy Liang. Identifiability and unmixing of latent parse trees. In P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems (NIPS) 25*, pages 1520–1528. 2012.

Bill Jackson, Brigitte Servatius, and Herman Servatius. The 2-dimensional rigidity of certain families of graphs. *Journal of Graph Theory*, 54(2):154–166, 2007. ISSN 0364-9024.

Svante Janson and Malwina J. Luczak. A simple solution to the k-core problem. *Random Structures Algorithms*, 30(1-2):50–62, 2007.

Shiva Prasad Kasiviswanathan, Cristopher Moore, and Louis Theran. The rigidity transition in random graphs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1237–1252, Philadelphia, PA, 2011. SIAM.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 56 (6):2980–2998, 2010.

Franz J. Király and Martin Ehler. The algebraic approach to phase retrieval and explicit inversion at the identifiability threshold. *arXiv e-prints*, 2014. arXiv:1402.4053.

Franz J. Király and Louis Theran. Coherence and sufficient sampling densities for reconstruction in compressed sensing. *arXiv e-prints*, 2013. arXiv 1302.2767.

Franz J. Kiraly and Louis Theran. Error-minimizing estimates and universal entry-wise error bounds for low-rank matrix completion. *Advances in Neural Information Processing Systems (NIPS) 26*, pages 2364–2372, 2013.

Franz J. Király, Louis Theran, Ryota Tomioka, and Takeaki UNo. The algebraic combinatorial approach for low-rank matrix completion. *arXiv e-prints*, 2013. arXiv 1211.4116v3.

Franz J. Király, Martin Kreuzer, and Louis Theran. Learning with cross-kernel matrices and ideal PCA. *arXiv e-prints*, 2014. arXiv:1406.2646.

Franz J. Király, Zvi Rosen, and Louis Theran. Algebraic matroids with graph symmetry. *arXiv e-prints*, 2013. arXiv:1312.3777.

Krysztof Kurdyka, Patrice Orro, and S. Simon. Semialgebraic Sard theorem for generalized critical values. *Journal of Differential Geometry*, 56(1):67–92, 2000.

Troy Lee and Adi Shraibman. Matrix completion from any given set of observations. In *Advances in Neural Information Processing Systems (NIPS) 26*, pages 1781–1787, 2013.

Adam Mahdi, Nicolette Meshkat, and Seth Sullivant. Structural identifiability of viscoelastic mechanical systems. *PLOS ONE*, 9(2):e86411, 2014.

Raghu Meka, Prateek Jain, and Inderjit S. Dhillon. Guaranteed rank minimization via singular value projection. *arXiv e-prints*, 2009. arXiv:0909.5457.

Aditya K. Menon and Charles Elkan. Link prediction via matrix factorization. *Machine Learning and Knowledge Discovery in Databases*, pages 437–452, 2011.

Nicolette Meshkat, Seth Sullivant, and Marisa Eisenberg. Identifiability results for several classes of linear compartment models. *arXiv e-prints*, 2014. arXiv:1410.8587.

John Milnor. *Singular points of complex hypersurfaces*. Annals of Mathematics Studies, No. 61. Princeton University Press, Princeton, N.J., 1968.

David Mumford. *The Red Book of Varieties and Schemes*. Lecture Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 1999.

Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics*, 39(2), 2011.

Sahand Negahban and Martin J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13:1665–1697, 2012.

James Oxley. *Matroid theory*, volume 21 of *Oxford Graduate Texts in Mathematics*. Oxford University Press, Oxford, second edition, 2011.

Boris Pittel, Joel Spencer, and Nicholas Wormald. Sudden emergence of a giant k-core in a random graph. *Journal of Combinatorial Theory. Series B*, 67(1):111–151, 1996.

Ruslan Salakhutdinov and Nathan Srebro. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems (NIPS) 23*, pages 2056–2064. 2010.

Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the Association for Computing Machinery*, 27(4):701–717, October 1980.

Amit Singer and Mihai Cucuringu. Uniqueness of low-rank matrix completion by rigidity theory. *SIAM Journal on Matrix Analysis and Applications*, 31(4):1621–1641, 2010.

Nathan Srebro and Adi Shraibman. Rank, trace-norm and max-norm. In *Proceedings of the 18th Annual Conference on Learning Theory (COLT)*, pages 545–560. Springer, 2005.

Nathan Srebro, Jason D. M. Rennie, and Tommi S. Jaakkola. Maximum-margin matrix factorization. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems (NIPS) 17*, pages 1329–1336. MIT Press, Cambridge, MA, 2005.

Angelika Steger and Nicholas Wormald. Generating random regular graphs quickly. *Combinatorics Probability and Computing*, 8(4):377–396, 1999.

Ryota Tomioka, Kohei Hayashi, and Hisashi Kashima. On the extension of trace norm to tensors. In *NIPS Workshop on Tensors, Kernels, and Machine Learning*, 2010.