# Tensor Methods for Nonlinear Matrix Completion[*]

Greg Ongie[†], Daniel Pimentel-Alarcón[‡], Laura Balzano[§], Rebecca Willett[¶], and Robert D. Nowak[‖]

**Abstract.** In the low-rank matrix completion (LRMC) problem, the low-rank assumption means that the columns (or rows) of the matrix to be completed are points on a low-dimensional linear algebraic variety. This paper extends this thinking to cases where the columns are points on a low-dimensional *nonlinear* algebraic variety, a problem we call *low algebraic dimension matrix completion* (LADMC). Matrices whose columns belong to a union of subspaces are an important special case. We propose an LADMC algorithm that leverages existing LRMC methods on a tensorized representation of the data. For example, a second-order tensorized representation is formed by taking the Kronecker product of each column with itself, and we consider higher-order tensorizations as well. This approach will succeed in many cases where traditional LRMC is guaranteed to fail because the data are low-rank in the tensorized representation but are not in the original representation. We also provide a formal mathematical justification for the success of our method. In particular, we give bounds on the rank of these data in the tensorized representation, and we prove sampling requirements to guarantee uniqueness of the solution. We also provide experimental results showing that the new approach outperforms existing state-of-the-art methods for matrix completion under a union-of-subspaces model.

**Key words.** matrix completion, tensors, algebraic varieties, unions of subspaces, subspace clustering

**AMS subject classifications.** 15A83, 14Q15, 14N20, 65F55

**DOI.** 10.1137/20M1323448

**1. Introduction.** The past decade of research on matrix completion has shown it is possible to leverage linear dependencies among columns and/or rows of a matrix to impute missing values. For instance, if each column of a matrix corresponds to a different high-dimensional data point belonging to the same low-dimensional linear subspace, then the corresponding matrix is low-rank, and missing values can be imputed using low-rank matrix completion (LRMC)

[†]Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI 53233 USA (gregory.ongie@marquette.edu).

[‡]Department of Biostatistics, University of Wisconsin–Madison, Madison, WI 53715 USA (pimentelalar@wisc.edu).

[§]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (girasole@umich.edu).

[¶]Departments of Statistics and Computer Science, University of Chicago, Chicago, IL 60637 USA (willett@g.uchicago.edu).

[‖]Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53715 USA (rdnowak@wisc.edu).

[5, 6, 40, 41, 23]. These ideas continue to impact diverse applications, such as recommender systems [26], image inpainting [20], computer vision [22], and array signal processing [43], among others.

The high-level idea of this body of work is that if the data defining the matrix have fewer degrees of freedom than the matrix dimension, that structure provides redundancy that can be leveraged to impute missing data values. However, the usual low-rank assumption is not always satisfied in practice. Extending matrix completion theory and algorithms to exploit low-dimensional nonlinear structure in data will allow missing data imputation in a far richer class of problems.

This paper describes matrix completion in the context of nonlinear algebraic varieties, a polynomial generalization of linear subspaces. More precisely, let $X \in \mathbb{R}^{d \times N}$ be a matrix whose columns lie in a low-dimensional algebraic variety $\mathcal{V} \subset \mathbb{R}^d$. Such matrices will be called *low algebraic dimension* (LAD) matrices. In the case where $\mathcal{V}$ is a low-dimensional linear variety, i.e., a linear subspace, this reduces to LRMC. We call the more general problem of completing LAD matrices *low algebraic dimension matrix completion* (LADMC).

Recently [29] proposed a new LADMC approach based on lifting the problem to a higher-dimensional representation using polynomial expansions of the columns of $X$. The algorithm in [29] can be interpreted as alternating between LRMC in the lifted representation and *unlifting* this low-rank representation back to the original representation to obtain a completion of the original matrix. This approach appears to provide good results in practice, but two problems were unresolved:

- The "unlifting" step is highly nonlinear and nonconvex, and so little can be proved about its accuracy or correctness.
- While [29] provides an intuitive explanation for the potential of the approach (based on a degrees-of-freedom argument) and why it may succeed in cases where LRMC fails, a rigorous argument is lacking.

This paper addresses both issues. To address the first point, we propose a new LADMC algorithm based on an explicit tensorization of the data. Specifically, let the *tensorized representation* of a matrix be another matrix where each column is the vectorized $p$-fold Kronecker product of the same column of the original matrix with itself.[1] If the original matrix has missing entries, then we treat an entry in its tensorized representation as missing if it cannot be formed from products of observed entries of $X$ (i.e., if one or more of the factors making up the product is missing in $X$). Our algorithm consists of three steps: (1) form the tensorized representation of the original matrix with missing entries, (2) apply standard LRMC to the tensorized representation to recover its missing entries, and (3) map each completed column of the tensorized matrix back to the original domain by applying a simple columnwise "detensorization" step based on a rank-1 (truncated) singular value decomposition (SVD). Provided the completion in step (2) is correct, step (3) is guaranteed to recover the correct original column. See Figure 1 for a schematic and Algorithm 2.1 for further details. We

---

[1]The tensorized representation could also be formally viewed as a $(p + 1)$-way tensor, for which low-rank tensor completion techniques could be brought to bear (see, e.g., [15]). However, in this work we focus on a matrix flattening of the tensorized representation, since our approach only exploits linear relationships between the columns of the tensorized representation.
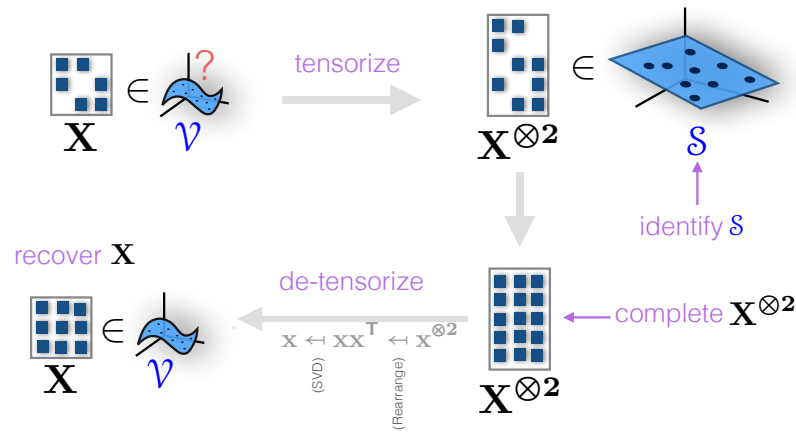
**Figure 1.** *Schematic of the proposed low-algebraic dimension matrix completion (LADMC) approach using a quadratic tensorization.* Given a matrix $\boldsymbol{X}$ with missing entries we synthesize entries of its tensorized representation $\boldsymbol{X}^{\otimes 2}$, the columnwise Kronecker product of $\boldsymbol{X}$ with itself. If the columns of $\boldsymbol{X}$ belong to an unknown variety $\mathcal{V}$ with low algebraic dimension, then $\boldsymbol{X}^{\otimes 2}$ will be low-rank, i.e., its columns belong to a low-dimensional subspace $\mathcal{S}$. Therefore, assuming the observed entries of $\boldsymbol{X}^{\otimes 2}$ are sufficient to uniquely identify $\mathcal{S}$, it is possible to complete $\boldsymbol{X}^{\otimes 2}$ using standard low-rank matrix completion (LRMC). Finally, we recover $\boldsymbol{X}$ by a simple "detensorization" procedure by applying a rank-1 truncated singular value decomposition (SVD) to each reshaped column of $\boldsymbol{X}^{\otimes 2}$.

present both iterative and noniterative versions of this algorithm and experiments showing that its performance is as good as or better than state-of-the-art methods in the popular case of the union-of-subspaces model.

To address the second point, we make several contributions towards understanding the information-theoretic requirements of LADMC. We build on recent work that gives deterministic sampling conditions for LRMC and subspace clustering with missing entries [33, 34, 35, 36, 37]. Specifically, we give conditions guaranteeing that the column space of the *tensorized representation* of the data matrix is uniquely identifiable from its *canonical projections*, i.e., projections of the subspace onto a collection of canonical basis elements. This sampling model is equivalent to an entrywise sampling model that assumes sufficiently many data columns are sampled with the same observation pattern. The main challenge in our setting is that while data may be missing uniformly at random in the original data matrix $\boldsymbol{X}$, the pattern of missing data in the tensorized domain is highly structured (not uniformly random) due to the product structure, and, in fact, only a small fraction of all patterns is possible. Hence, the canonical projections that we may observe in the tensor representation are a restrictive subset of all possible canonical projections. Nonetheless, our main results show that, under mild genericity assumptions on the underlying variety, the admissible canonical projections in the tensor space are indeed sufficient to identify the subspace in the tensorized representation.

To give a flavor of our main results, here we present a simplified version of our necessary and sufficient conditions specialized to the case of a quadratic tensorization and data belonging to a union of subspaces (a special case of an algebraic variety).

**Theorem (simplified).** *Assume that the columns of $\boldsymbol{X}$ belong to a union of $K \geq 4$ subspaces*

*drawn independently and uniformly at random from the Grassmannian of all r-dimensional subspaces in $\mathbb{R}^d$. If each column of $\boldsymbol{X}$ is observed in at least $\left\lceil \sqrt{Kr(r+1)} \right\rceil + 2$ uniformly randomly selected coordinates and if the total number of columns is sufficiently large, then unique matrix completion is possible using the proposed LADMC algorithm with a quadratic tensorization. If fewer than $\left\lceil \sqrt{Kr(r+1)} \right\rceil$ coordinates are observed in each column, then unique LRMC of the quadratic tensorized representation is impossible.*

See Corollaries 1 and 2 for a formal statement of this result and its extension to higher-order tensorizations. These formal statements are in terms of the canonical projections of a subspace, which is equivalent to an entrywise sampling model that assumes the total number of columns is sufficient for guaranteeing multiple copies of every observable sampling pattern. Also, we give more general versions of these results in Lemma 3 and Theorem 1 that hold for a broader class of algebraic varieties beyond unions of subspaces.

Roughly speaking, this result shows that a sampling rate of $m = O(\sqrt{K}r)$ entries per column is necessary and sufficient for unique identification of the column space of the tensorized representation arising from a generic union of $K$ $r$-dimensional subspaces under a quadratic tensorization. In contrast, $m = O(Kr)$ observations per column are necessary and sufficient to uniquely identify the column space of the original matrix, i.e., we see an improvement by a factor of $\sqrt{K}$ in the number of observations per column for which LADMC is information-theoretically possible relative to LRMC in the original domain.

The rest of the paper lays out our contributions as follows. Section 2 presents the algorithm in more detail, and section 2.1 describes the algebraic assumptions that are required for the tensorized representation of the data matrix to be low-rank. Then section 3 addresses sampling conditions for unique identifiability of the tensorized subspace, giving both necessary (section 3.4) and sufficient (section 3.5) conditions, including results that specialize to the union-of-subspaces model. Section 4 presents numerical results.

**1.1. Related work.** A union of subspaces (UoS) is a special type of algebraic variety [29, 47], and a number of approaches to matrix completion for a UoS model have been proposed [42, 2, 35, 3, 38, 51, 31, 49, 32, 12, 16]; see [24] for classification and comparison of these approaches for the task of subspace clustering with missing data. Most of these algorithms involve iterating between subspace clustering and completion steps, and relatively little can be guaranteed about their performance. Exceptions include [2] and [16], which give performance guarantees for algorithms based on a noniterative neighborhood search procedure. Also, recent work [9, 45] gives performance guarantees for a version of the sparse subspace clustering algorithm modified to handle missing data [51].

Our proposed LADMC approach is closely related to algebraic subspace clustering (ASC), also known as generalized principal component analysis [47, 50, 27, 44, 46]. Similar to our approach, the ASC framework models unions of subspaces as an algebraic variety and makes use of tensorizations (i.e., Veronese embeddings) of the data to identify the subspaces. However, the ASC framework has not been extended to the matrix completion setting, which is the main focus of this work.

A preliminary version of this work was published in a conference proceedings [39]. We extend the theory and algorithms in [39] to higher-order tensorizations of the data matrix;

[39] only considered quadratic tensorizations. We also correct several issues with the theory in [39]. In particular, parts (i) and (ii) of Theorem 2 of [39] are incorrect as stated; in this work we correct these results and also extend them to higher-order tensorizations (see Corollary 1). Additionally, the proof of part (iii) of Theorem 2 in [39] is incorrect; here we give a correct proof and likewise extend the result to higher-order tensorizations (see Corollary 2). Finally, we also expand the experiments section by comparing the proposed LADMC algorithm with the previously proposed variety-based matrix completion (VMC) algorithm [29] and demonstrate the proposed LADMC algorithm for matrix completion of real data.

**2. Setup and algorithm.** Suppose we observe a subset of the entries of a matrix

$$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{d \times N}$$

at locations indicated by ones in the binary matrix $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N] \in \{0, 1\}^{d \times N}$. We refer to $\boldsymbol{\Omega}$ and $\boldsymbol{\omega}_i$ as the matrix and vector *observation patterns*, respectively.

In traditional LRMC one assumes $\boldsymbol{X}$ is low-rank in order to recover the missing entries. We take a different approach in this work. Rather than completing $\boldsymbol{X}$ directly, we consider low-rank completing the *tensorized matrix*

$$\boldsymbol{X}^{\otimes p} := [\boldsymbol{x}_1^{\otimes p}, \ldots, \boldsymbol{x}_N^{\otimes p}] \in \mathbb{R}^{D \times N}.$$

Here $\boldsymbol{x}^{\otimes p}$ denotes the *p-fold tensorization* of a vector $\boldsymbol{x}$, defined as $\boldsymbol{x}^{\otimes p} := \boldsymbol{x} \otimes \cdots \otimes \boldsymbol{x}$, where $\otimes$ is the Kronecker product, and $\boldsymbol{x}$ appears $p$ times in the expression. Every tensorized vector $\boldsymbol{x}^{\otimes p}$ can be reordered into a $p$th-order symmetric tensor that is uniquely determined by $D := \binom{d+p-1}{p}$ of its entries. For example, the vector $\boldsymbol{x}^{\otimes 2}$ has the same entries as the matrix $\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} \in \mathbb{R}^{d \times d}$, which is uniquely determined by its $\binom{d+1}{2}$ upper triangular entries. Hence, with slight abuse of notation, we consider tensorized vectors $\boldsymbol{x}^{\otimes p}$ as elements of $\mathbb{R}^D$.

Additionally, given partial observations of matrix $\boldsymbol{X}$ at locations in $\boldsymbol{\Omega}$ we can synthesize observations of the tensorized matrix $\boldsymbol{X}^{\otimes p}$ at all locations indicated by $\boldsymbol{\Omega}^{\otimes p} = [\boldsymbol{\omega}_1^{\otimes p}, \ldots, \boldsymbol{\omega}_N^{\otimes p}] \in \{0, 1\}^{D \times N}$ simply by multiplying the observed entries of $\boldsymbol{X}$. In particular, if the data column $\boldsymbol{x}_i$ is observed in $m$ locations, then the tensorized data column $\boldsymbol{x}_i^{\otimes p}$ can be observed at $\binom{m+p-1}{p}$ locations indicated by ones in the binary vector $\boldsymbol{\omega}_i^{\otimes p}$. We refer to $\boldsymbol{\Omega}^{\otimes p}$ and $\boldsymbol{\omega}_i^{\otimes p}$ as the matrix and vector *tensorized observation patterns*, respectively.

Remarkably, there are situations where the original data matrix $\boldsymbol{X}$ is full-rank, but the tensorized matrix $\boldsymbol{X}^{\otimes p}$ is low-rank, owing to (nonlinear) algebraic structure of the data, described in more detail below. In these situations, $\boldsymbol{X}^{\otimes p}$ can potentially be recovered from its entries indicated by $\boldsymbol{\Omega}^{\otimes p}$ using standard LRMC algorithms.

If the LRMC step recovers $\boldsymbol{X}^{\otimes p}$ correctly, then we can uniquely recover $\boldsymbol{X}$ from $\boldsymbol{X}^{\otimes p}$. To see this, first consider the case of a quadratic tensorization ($p = 2$). Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]$ be the output from LRMC applied to the tensorized matrix. If the completion is correct, meaning we exactly recover $\boldsymbol{y}_i = \boldsymbol{x}_i^{\otimes 2}$, then we can reshape $\boldsymbol{y}_i$ into the exactly rank-1 symmetric $d \times d$ matrix $\boldsymbol{Y}_i = \boldsymbol{x}_i \boldsymbol{x}_i^{\mathsf{T}}$. Hence, we can recover $\boldsymbol{x}_i$ by computing the leading eigenvalue-eigenvector pair $(\lambda_i, \boldsymbol{u}_i)$ of $\boldsymbol{Y}_i$ and setting $\boldsymbol{x}_i = \pm\sqrt{\lambda_i}\boldsymbol{u}_i$, where we determine the sign by matching it to the observed entries of $\boldsymbol{x}_i$. If there is noise and the completion is not exact, $\boldsymbol{Y}_i$ may not be exactly rank-1, but we see empirically that the leading eigenvector provides a denoised estimate.

For higher-order tensorizations ($p \geq 3$), we can recover $\boldsymbol{x}_i$ from $\boldsymbol{y}_i$ using a similar procedure: we reshape $\boldsymbol{y}_i$ into a $d \times d^{p-1}$ and take its rank-1 truncated SVD,[2] which, similarly, will only be exact in the noiseless case.

These observations motivate our proposed algorithm, low algebraic dimension matrix completion (LADMC), summarized below in Algorithm 2.1; see Figure 1 for a schematic.

---

**Algorithm 2.1.** Low algebraic dimension matrix completion (LADMC).

**Input:** Subset of entries of data matrix $\boldsymbol{X}$.

**Tensorize:** Form new matrix $\boldsymbol{X}^{\otimes p}$ by replacing each column $\boldsymbol{x}_i$ with its $p$-fold tensor product $\boldsymbol{x}_i^{\otimes p}$ (with missing entries corresponding to any products involving missing entries in $\boldsymbol{x}_i$).

**LRMC:** Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]$ be the low-rank completion of $\boldsymbol{X}^{\otimes p}$.

**Detensorize:** Compute the best rank-1 symmetric tensor approximation $\hat{\boldsymbol{x}}_i^{\otimes p}$ of each column $\boldsymbol{y}_i$ of $\boldsymbol{Y}$ such that $\hat{\boldsymbol{x}}_i$ matches the observed entries of $\boldsymbol{x}_i$.

**Output:** Completed matrix $\hat{\boldsymbol{X}}$ whose $i$th column is $\hat{\boldsymbol{x}}_i$.

---

### 2.1. Algebraic variety models and rank of the tensorized matrix.
Here we describe in more detail the algebraic assumptions that are required for the tensorized data matrix $\boldsymbol{X}^{\otimes p}$ to be low-rank.

Suppose $\boldsymbol{X}^{\otimes p} \in \mathbb{R}^{D \times N}$ is a wide matrix, i.e., the number of data columns $N$ exceeds the tensor space dimension $D$. Then $\boldsymbol{X}^{\otimes p}$ is rank-deficient if and only if the rows of $\boldsymbol{X}^{\otimes p}$ are linearly dependent, in which case there exists a vector $\boldsymbol{v} \in \mathbb{R}^D$ such that $\boldsymbol{v}^{\mathsf{T}} \boldsymbol{x}_i^{\otimes p} = 0$ for all columns $\boldsymbol{x}_i$ of $\boldsymbol{X}$. In other words, the columns $\boldsymbol{x}_i$ belong to the zero set of the polynomial $q(\boldsymbol{x}) = \boldsymbol{v}^{\mathsf{T}} \boldsymbol{x}^{\otimes p}$. Hence, we have shown the following.

**Proposition 1.** *The tensorized matrix $\boldsymbol{X}^{\otimes p} \in \mathbb{R}^{D \times N}$ with $N \geq D$ is rank-deficient only if the columns of $\boldsymbol{X}$ belong to a* proper *algebraic variety, i.e., the common zero set of a collection of polynomials that are not all identically zero.*

In particular, we focus on the class of varieties defined by *homogeneous polynomials.*[3] A degree-$p$ homogeneous polynomial is any polynomial of the form $q(\boldsymbol{x}) = \boldsymbol{v}^{\mathsf{T}} \boldsymbol{x}^{\otimes p}$ for some vector of coefficients $\boldsymbol{v} \in \mathbb{R}^D$.

**Definition 1.** *A set $\mathcal{V} \subset \mathbb{R}^d$ is a* (real) projective variety[4] *if there exist homogeneous polynomials $q_1, \ldots, q_n$ (with possibly different degrees) such that*

$$\mathcal{V} = \{\boldsymbol{x} \in \mathbb{R}^d : q_1(\boldsymbol{x}) = \cdots = q_n(\boldsymbol{x}) = 0\}.$$

---

[2]Another approach would be to augment the tensorized representation with the original matrix prior to the LRMC step. Then we could recover $\boldsymbol{X}$ by restricting the completion to the block corresponding to the original matrix. However, empirically we found this approach to be more sensitive to noise. Also, the proposed detensorization step based on a rank-1 SVD is crucial to our iterative version of LADMC (see section 4.1).

[3]Our approach extends to varieties defined by inhomogeneous polynomials if we redefine $\boldsymbol{x}^{\otimes p}$ to be the map $\boldsymbol{x} \mapsto [\begin{smallmatrix} 1 \\ \boldsymbol{x} \end{smallmatrix}]^{\otimes p}$, i.e., augment $\boldsymbol{x}$ with a 1 before tensorization.

[4]For any homogeneous polynomial $q$, we have $q(\boldsymbol{x}) = 0$ if and only if $q(\lambda \boldsymbol{x}) = 0$ for any scalar $\lambda \neq 0$. This means the zero sets of homogeneous polynomials can be considered as subsets of *projective space*, i.e., the set of all lines through the origin in $\mathbb{R}^d$, and this is the source of the term "projective variety." For simplicity, we typically consider projective varieties as subsets of Euclidean space $\mathbb{R}^d$ unless otherwise noted.

An important fact for this work is that a union of subspaces is a projective variety, as shown in the following example.

*Example* 1 (UoS are projective varieties). Suppose $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathbb{R}^d$. Then the union of the two subspaces $\mathcal{V} := \mathcal{U} \cup \mathcal{W}$ is given by the common zero set of the collection of quadratic forms $q_{i,j}(\boldsymbol{x}) = (\boldsymbol{x}^\mathsf{T} \boldsymbol{u}_i^\perp)(\boldsymbol{x}^\mathsf{T} \boldsymbol{w}_j^\perp)$, where $\{\boldsymbol{u}_i^\perp\}$ is a basis of the orthogonal complement of $\mathcal{U}$, and $\{\boldsymbol{w}_j^\perp\}$ is a basis of the orthogonal complement of $\mathcal{W}$. Hence $\mathcal{V}$ is a projective variety determined by the common zero set of a collection of quadratic forms. More generally, a union of $K$ distinct subspaces is a projective variety defined by a collection of degree-$K$ polynomials, each of which is a product of $K$ linear factors; this fact forms the foundation of algebraic subspace clustering methods [27, 48].

Given a matrix whose columns are points belonging to a projective variety, the rank of the associated tensorized matrix is directly related to the dimension of the associated *tensorized subspace*, defined as follows.

Definition 2. *Let $\mathcal{V} \subset \mathbb{R}^d$ be a projective variety. We define the pth-order* tensorized subspace *associated with $\mathcal{V}$ by*

$$\text{(2.1)} \qquad \qquad \mathcal{S} := \text{span}\{\boldsymbol{x}^{\otimes p} : \boldsymbol{x} \in \mathcal{V}\} \subset \mathbb{R}^D,$$

*i.e., the linear span of all pth-order tensorized vectors belonging to $\mathcal{V}$.*

If the columns of a matrix $\boldsymbol{X}$ belong to a projective variety $\mathcal{V}$, then the column space of the tensorized matrix $\boldsymbol{X}^{\otimes p}$ belongs to the tensorized subspace $\mathcal{S}$, and so the dimension of the tensorized subspace is an upper bound on the rank of the tensorized matrix. In particular, if there are a total of $L$ linearly independent degree-$p$ homogeneous polynomials vanishing on $\mathcal{V}$, then $\mathcal{S}$ is a subspace of $\mathbb{R}^D$ of dimension at most $D - L$. Therefore, if there are sufficiently many such polynomials, then $\mathcal{S}$ is a low-dimensional subspace, and hence $\boldsymbol{X}^{\otimes p}$ is low-rank. One situation in which this can occur is when any homogeneous polynomial $q(\boldsymbol{x})$ of degree $p'$ smaller than the tensor order $p$ vanishes on the variety. In this case, all degree-$p$ polynomials given by the product $q(\boldsymbol{x})$ with a monomial of degree $p - p'$ will also vanish on the variety and will be linearly independent, which drives down the dimension $R$ of the tensorized subspace.

As a more concrete example, consider the special case of a union of subspaces. The linear span of all points belonging to a union of $K$ $r$-dimensional subspaces (in general position) defines a subspace of dimension $\min\{Kr, d\}$ in $\mathbb{R}^d$. However, in the tensor space these points lie in a subspace of $\mathbb{R}^D$ whose dimension relative to the tensor space dimension is potentially smaller, as shown in the following lemma.

Lemma 1. *Let $\mathcal{V} \subset \mathbb{R}^d$ be a union of $K$ $r$-dimensional subspaces, and let $\mathcal{S} \subset \mathbb{R}^D$ be its pth-order tensorized subspace. Then $\mathcal{S}$ is $R$-dimensional where*

$$\text{(2.2)} \qquad \qquad R \le \min\left\{ K\binom{r+p-1}{p}, D \right\}.$$

The proof of Lemma 1 is elementary: any $r$-dimensional subspace in $\mathbb{R}^d$ spans an $\binom{r+p-1}{p}$-dimensional subspace in the tensor space, and so points belonging a union of $K$ $r$-dimensional subspaces spans at most a $K\binom{r+p-1}{p}$-dimensional subspace in the tensor space.

Example 1 shows that a union of $K$ subspaces is always defined by a set of polynomials of degree $K$. This suggests that a $K$th-order tensorization is necessary to have low-rank in the tensorized representation and hence for successful LRMC. When $K$ is large, this may be computationally prohibitive. However, as we will see, lower tensor orders (e.g., $p = 2, 3$) still allow for completion of data belonging to UoS in scenarios where standard LRMC fails. This is because even though the UoS is always defined by degree-$K$ polynomials, several lower degree polynomials also vanish on the UoS, which is sufficient for reducing the rank in the tensorized representation enough to allow for successful LRMC in the tensor domain.

More precisely, assuming we need $O(R)$ observations per column to complete a rank-$R$ matrix, completing a matrix whose columns belong to a union of $K$ $r$-dimensional subspaces in the original space would require $O(Kr)$ observations per column, but completing the corresponding tensorized matrix would require $O(K\binom{r+p-1}{p})$ entries (products) per column, which translates to $O(K^{1/p}r)$ entries per column in the original matrix. This suggests that LADMC could succeed with far fewer observations per column than LRMC (i.e., $O(K^{1/p}r)$ versus $O(Kr)$) in the case of UoS data.

We note that Lemma 1 is a special case of a more general bound due to [8] that holds for any *equidimensional* projective variety.[5] Roughly speaking, a projective variety is *equidimensional* if at all points where the variety is smooth the local dimension of the variety (treated as a smooth manifold) is the same. The bound in [8] is posed in terms of the *degree* and *dimension* of the variety (see, e.g., [10] for definitions of these quantities). Translated into our setting, this result says that if $\mathcal{V}$ is an equidimensional projective variety of degree $K$ and dimension $r$, then its $p$th-order tensorized subspace is $R$-dimensional where $R$ obeys the same bound as in (2.2). Therefore, given a matrix whose columns belong to an equidimensional projective variety, we should expect LADMC to succeed with $O(K^{1/p}r)$ observations per column, where now $K$ is the degree of the variety and $r$ is its dimension. In other words, when the data belong to a projective variety with *high degree* and *low dimension* we expect LADMC to potentially succeed where LRMC may fail.

## 3. Theory.

### 3.1. Limitations of prior theory.
Algorithm 2.1 is primarily inspired by the ideas in [29], where an informal argument is given for the minimum number of observed entries per data column necessary for successful completion of a tensorized matrix based on the dimension of the corresponding tensorized subspace. Translated into the setting of this paper, the claim made in [29] is that in order to successfully complete a matrix $\boldsymbol{X}$ whose $p$th-order tensorized matrix $\boldsymbol{X}^{\otimes p}$ is rank $R$, we must observe at least $m_0$ entries per column of $\boldsymbol{X}$, where $m_0$ is the smallest integer such that

$$(3.1) \qquad \binom{m_0+p-1}{p} > R,$$

i.e., unique low-rank completion ought to be possible when the number of observations per column of the tensorized matrix exceeds its rank. This conjecture was based on the fact

---

[5]The result in [8] gives an upper bound on the values of the Hilbert function associated with any homogeneous unmixed radical ideal $I \subset k[x_0, \ldots, x_d]$ over a perfect field $k$. We specialize this result to the vanishing ideal of an equidimensional variety in real projective space. In particular, the dimension of the $p$th-order tensorized subspace coincides with the Hilbert function of the vanishing ideal evaluated at degree $p$.

that $R + 1$ is the minimum number of observations per column necessary to uniquely complete a matrix whose columns belong to an $R$-dimensional subspace in general position [36]. Additionally, $R + 1$ observations per column are sufficient for unique completion, assuming there are sufficiently many data columns and the observation patterns are sufficiently diverse [34]. A similar argument is given in [14], where the authors identify the minimum number of parameters needed to uniquely determine a matrix $\boldsymbol{X}$ among the set of all matrices whose tensorization has the same rank, i.e., the set $\{\boldsymbol{X} : \mathbb{R}^{d \times N} : \operatorname{rank}(\boldsymbol{X}^{\otimes p}) = R\}$. This number of parameters is conjectured to be the minimum number of observed entries necessary to recover $\boldsymbol{X}$, which matches the condition in (3.1) when considering the sampling model in this work.

However, there are two key technical issues not considered in [29] and [14] that prevent their arguments from being rigorous. One is related to the fact that the patterns of missing entries in the tensorized matrix are highly structured due to the tensor product. Consequently, the set of realizable observation patterns in the tensorized matrix is severely limited. These constraints on the observation patterns imply that existing LRMC theory (which typically requires uniform random observations) does not apply directly to tensorized representations.

The other technical issue not considered by [29] and [14] is that the tensorized subspace (i.e., the column space of the tensorized matrix) is *not* always in general position as a subspace of $\mathbb{R}^D$. For example, if an $R$-dimensional subspace is in general position, then the restriction of the subspace to any subset of $R$ canonical coordinates is $R$-dimensional (i.e., if $\boldsymbol{B} \in \mathbb{R}^{D \times R}$ is any basis matrix for the subspace, then all $R \times R$ minors of $\boldsymbol{B}$ are nonvanishing). However, generally this property does not hold for a tensorized subspace arising from a UoS, even if the subspaces in the union are in general position (see Example 3 below). General position assumptions are essential for results that describe deterministic conditions on the observation patterns allowing for LRMC [33, 35]. Hence, the direct application of these results to the LADMC setting is not possible.

For these reasons it was unclear whether unique completion via LADMC was information-theoretically possible. In fact, we prove there are cases where condition (3.1) is satisfied, but where $\boldsymbol{X}^{\otimes p}$ cannot be completed uniquely using LRMC, even with an unlimited amount of data (see Example 2 below). In the remainder of this section we derive necessary and sufficient conditions under which unique completion via LADMC is possible and compare these with condition (3.1).

**3.2. Unique identifiability of the tensorized subspace.** To simplify our results, we consider a sampling model in which we observe exactly $m$ entries per column of the original matrix. The main theoretical question we are interested in is the following.

Question 1. *What is the minimum number of observations per column, $m$, of the original matrix for which unique completion is* information-theoretically possible *with Algorithm* 2.1?

Rather than study Question 1 directly, we will study the more basic problem of the unique identifiability of the tensorized subspace (i.e., the column space of the tensorized matrix) from its projections onto subsets of canonical coordinates. This is related to Question 1 as follows: Suppose that we observe multiple columns of the original matrix $\boldsymbol{X}$ with the same observation pattern. Then we will observe the corresponding columns of the tensorized matrix $\boldsymbol{X}^{\otimes p}$ with the same tensorized observation pattern. Hence, given sufficiently many columns that are in

general position, we can compute a basis of the projection of the tensorized subspace onto coordinates specified by the tensorized observation pattern. This means that given sufficiently many data columns observed with observation patterns of our choosing, we could in principle compute any projection of the tensorized subspace onto coordinates specified by any tensorized observation pattern. Hence, we consider instead the following closely related question.

**Question 2.** *What is the minimum value of $m$ for which the tensorized subspace is* uniquely identifiable *from its projections onto all possible tensorized observation patterns arising from a sampling of $m$ entries per column in the original domain?*

To more precisely describe what we mean by *unique identifiability* of the tensorized subspace in Question 2, we introduce the following notation and definitions.

For any observation pattern $\boldsymbol{\omega} \in \{0,1\}^d$ we let $|\boldsymbol{\omega}|$ denote the total number of ones in $\boldsymbol{\omega}$. We say the tensorized observation pattern $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ is of size $m$ if $|\boldsymbol{\omega}| = m$. Note that if $\boldsymbol{v}$ is a tensorized observation pattern of size $m$, then $\boldsymbol{v}$ has $\binom{m+p-1}{p}$ ones, i.e., $|\boldsymbol{v}| = \binom{m+p-1}{p}$. For any observation pattern $\boldsymbol{v} \in \{0,1\}^D$ and any vector $\boldsymbol{y} \in \mathbb{R}^D$ let $\boldsymbol{y}_{\boldsymbol{v}} \in \mathbb{R}^{|\boldsymbol{v}|}$ denote the restriction of $\boldsymbol{y}$ to coordinates indicated by ones in $\boldsymbol{v}$. Likewise, for any subspace $\mathcal{S} \subset \mathbb{R}^D$ we let $\mathcal{S}_{\boldsymbol{v}} \subset \mathbb{R}^{|\boldsymbol{v}|}$ denote the subspace obtained by restricting all vectors in $\mathcal{S}$ to coordinates indicated by ones in $\boldsymbol{v}$ and call $\mathcal{S}_{\boldsymbol{v}}$ the *canonical projection* of $\mathcal{S}$ onto $\boldsymbol{v}$. For any subspace $\mathcal{S} \subset \mathbb{R}^D$ and any observation pattern matrix $\boldsymbol{\Upsilon} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \subset \{0,1\}^{D \times n}$ we define $\mathcal{S}(\mathcal{S}, \boldsymbol{\Upsilon})$ to be the set of all subspaces $\mathcal{S}'$ whose canonical projections onto observation patterns in $\boldsymbol{\Upsilon}$ agree with those of $\mathcal{S}$, i.e., all $\mathcal{S}'$ such that $\mathcal{S}'_{\boldsymbol{v}_i} = \mathcal{S}_{\boldsymbol{v}_i}$ for all $i = 1, \ldots, n$. We say a subspace $\mathcal{S}$ is *uniquely identifiable* from its canonical projections in $\boldsymbol{\Upsilon}$ if $\mathcal{S}(\mathcal{S}, \boldsymbol{\Upsilon}) = \{\mathcal{S}\}$.

To aid in determining whether a subspace is uniquely identifiable from a collection of canonical projections, we introduce the *constraint matrix* $\boldsymbol{A} = \boldsymbol{A}(\mathcal{S}, \boldsymbol{\Upsilon})$, defined below.

**Definition 3.** *Given a subspace $\mathcal{S} \subset \mathbb{R}^D$ and observation pattern matrix $\boldsymbol{\Upsilon} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \{0,1\}^{D \times n}$, define the* constraint matrix $\boldsymbol{A} \in \mathbb{R}^{D \times T}$ *as follows: for all $i = 1, \ldots, n$ suppose $M_i := |\boldsymbol{v}_i|$ is strictly greater than $R'_i := \dim \mathcal{S}_{\boldsymbol{v}_i}$, and let $\boldsymbol{A}_{\boldsymbol{v}_i} \in \mathbb{R}^{M_i \times (M_i - R'_i)}$ denote a basis matrix for $(\mathcal{S}_{\boldsymbol{v}_i})^\perp \subset \mathbb{R}^{M_i}$, the orthogonal complement of the canonical projection of $\mathcal{S}$ onto $\boldsymbol{v}_i$, so that $\ker \boldsymbol{A}_{\boldsymbol{v}_i}^\mathsf{T} = \mathcal{S}_{\boldsymbol{v}_i}$. Define $\boldsymbol{A}_i \in \mathbb{R}^{D \times (M_i - R'_i)}$ to be the matrix whose restriction to the rows indexed by $\boldsymbol{v}_i$ is equal to $\boldsymbol{A}_{\boldsymbol{v}_i}$ and whose restriction to rows not in $\boldsymbol{v}_i$ is all zeros. Finally, set $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n]$, which has a total of $T = \sum_{i=1}^n (M_i - R'_i)$ columns.*

The intuition here is that the orthogonal complement of each $\mathcal{S}_{\boldsymbol{v}_i}$ constrains the set of subspaces consistent with the observed projections, and $\boldsymbol{A}$ reflects the collection of these constraints across all $n$ observation patterns. The following result shows that unique identifiability of a subspace from its canonical projections is equivalent to a rank condition on the corresponding constraint matrix.

**Lemma 2.** *An $R$-dimensional subspace $\mathcal{S}$ is uniquely identifiable from canonical projections in $\boldsymbol{\Upsilon}$ if and only if $\dim \ker \boldsymbol{A}^\mathsf{T} = R$, in which case $\mathcal{S} = \ker \boldsymbol{A}^\mathsf{T}$.*

*Proof.* By construction, $\mathcal{S}' \in \mathcal{S}(\mathcal{S}, \boldsymbol{\Upsilon})$ if and only if $\mathcal{S}'_{\boldsymbol{v}_i} = \mathcal{S}_{\boldsymbol{v}_i} = \ker \boldsymbol{A}_{\boldsymbol{v}_i}^\mathsf{T}$ for all $i = 1, \ldots, n$ if and only if $\mathcal{S}' \in \ker \boldsymbol{A}^\mathsf{T}$. Hence, the set $\mathcal{S}(\mathcal{S}, \boldsymbol{\Upsilon})$ coincides with all $R$-dimensional subspaces contained in $\ker \boldsymbol{A}^\mathsf{T}$. In particular, we always have $\mathcal{S} \subset \ker \boldsymbol{A}^\mathsf{T}$ and, by linearity, $\text{span}\{\boldsymbol{x} \in \mathcal{S} : \mathcal{S} \in \mathcal{S}(\mathcal{S}, \boldsymbol{\Upsilon})\} = \ker \boldsymbol{A}^\mathsf{T}$. Hence, if $\dim \ker \boldsymbol{A}^\mathsf{T} = R$, it must be the case that $\ker \boldsymbol{A}^\mathsf{T} = \mathcal{S}$. ∎

*Remark* 1. Lemma 2 gives an empirical criterion for determining whether a subspace is uniquely identifiable: given canonical projections of a subspace $\mathcal{S}$, one can construct the constraint matrix $\boldsymbol{A}$ above and numerically check if the dimension of the null space of $\boldsymbol{A}^{\mathsf{T}}$ agrees with the subspace dimension $R$ (if this dimension is known). We will use this fact to explore the possibility of unique identifiability of tensorized subspaces arising from UoS of small dimensions (see Table 1).

**3.3. Generic unions of subspaces.** We are particularly interested in understanding Question 2 in the context of tensorized subspaces arising from a UoS, i.e., varieties of the form $\mathcal{V} = \cup_{k=1}^{K} \mathcal{U}_k$, where each $\mathcal{U}_k \subset \mathbb{R}^d$ is a linear subspace. To simplify our results, we will focus on UoS where each subspace in the union has the same dimension $r$. We will also often make the assumption that the UoS is *generic* in an algebraic-geometric sense (see, e.g., [7]). More precisely, we say a result holds for a *generic union of $K$ $r$-dimensional subspaces in $\mathbb{R}^d$* if it holds for all collections of subspaces $(\mathcal{U}_1, \ldots, \mathcal{U}_K)$ belonging to an (often unspecified) open dense subset of the product space $\mathbb{G}(r, \mathbb{R}^d) \times \cdots \times \mathbb{G}(r, \mathbb{R}^d)$, where $\mathbb{G}(r, \mathbb{R}^d)$ denotes the Grassmannian of all $r$-dimensional subspaces in $\mathbb{R}^d$. Put in probabilistic terms, this says a property holds for a generic UoS if and only if it holds with probability one for any union of random subspaces whose joint distribution is absolutely continuous with respect to the $K$-fold product of the uniform measure on $\mathbb{G}(r, \mathbb{R}^d)$. In particular, if a property holds for a generic union of $K$ $r$-dimensional subspaces, then it holds with probability one for a union of $K$ random subspaces drawn independently from the uniform measure on $\mathbb{G}(r, \mathbb{R}^d)$.

We will repeatedly make use of the following facts regarding generic UoS (see, e.g., [50]).

Proposition 2. *Let $\mathcal{S}$ be the $p$th-order tensorized subspace arising from a generic union of $K$ $r$-dimensional subspaces in $\mathbb{R}^d$. Then $R(d, K, r, p) := \dim \mathcal{S}$ is a constant that depends only on $d, K, r, p$.*

Treated as a function of $p$, the quantity $R(d, K, r, p)$ is called the *Hilbert function* of a generic UoS (also called a "generic subspace arrangement") and is studied in [50, 7, 11].

**3.4. Necessary conditions for unique identifiability of tensorized subspaces.** Here we derive necessary conditions for unique identifiability of tensorized subspaces from their canonical projections based on properties of the corresponding constraint matrix (see Definition 3). In Lemma 3 we give a general result that holds for tensorized subspaces arising from any projective variety and then specialize to the case of those arising from generic UoS in Corollary 1. In particular, we show that there are conditions under which the sampling rate (3.1) conjectured to be necessary and sufficient in [29] is indeed necessary, but we also demonstrate that it fails to be necessary and sufficient in general (see Examples 2 and 3).

First, observe that Lemma 2 implies a general necessary condition for unique identifiability of an $R$-dimensional tensorized subspace $\mathcal{S} \subset \mathbb{R}^D$: in order for $\dim \ker \boldsymbol{A}^{\mathsf{T}} = R$, the number of columns $\boldsymbol{A}$ needs to be at least $D - R$; this is achieved simply by considering matrix dimensions. This immediately gives the following result.

Lemma 3. *Let $\mathcal{V} \subset \mathbb{R}^d$ be a projective variety whose $p$th-order tensorized subspace $\mathcal{S} \subset \mathbb{R}^D$ is $R$-dimensional. Suppose we observe canonical projections of $\mathcal{S}$ onto $n$ distinct tensorized observation patterns $\boldsymbol{\Upsilon} = [\boldsymbol{v}_i, \ldots, \boldsymbol{v}_n] \subset \{0, 1\}^{D \times n}$. For all $i = 1, \ldots, n$ define $M_i := |\boldsymbol{v}_i|$ and*

$R'_i := \dim \mathcal{S}_{\boldsymbol{v}_i}$. *Then a necessary condition for $\mathcal{S}$ to be uniquely identifiable is*

$$(3.2) \qquad \sum_{i=1}^{n}(M_i - R'_i) \geq D - R.$$

Lemma 3 has several implications regarding the necessary sample complexity for tensorized subspaces arising from a union of subspaces. Consider the case where $\mathcal{S}$ is the $p$th-order tensorized subspace corresponding to a *generic* union of $K$ subspaces of dimension $r$. Suppose $\boldsymbol{\Upsilon}$ consists of all $\binom{d}{m}$ tensorized observation patterns of size $m$, i.e., each column $\boldsymbol{v}_i$ of $\boldsymbol{\Upsilon}$ has $M = \binom{m+p-1}{p}$ ones. From Lemma 2 we know that $\dim \mathcal{S}_{\boldsymbol{v}_i} = R'$ where $R' \leq R \leq K\binom{r+p-1}{p}$ and where the value of $R'$ is the same for all tensorized observation patterns $\boldsymbol{v}_i$ by genericity (for more on this point see the discussion following Theorem 1 below). This means the constraint matrix has a total of $\binom{d}{m}(M - R')$ columns, which gives the following necessary condition for unique identifiability of tensorized subspaces arising from generic UoS.

**Corollary 1.** *Let $\mathcal{V} \subset \mathbb{R}^d$ be a generic union of $K$ $r$-dimensional subspaces. Suppose its $p$th-order tensorized subspace $\mathcal{S} \subset \mathbb{R}^D$ is $R$-dimensional. Let $R' \leq R$ be the dimension of $\mathcal{S}$ projected onto any tensorized observation pattern of size $m$. Then a necessary condition for $\mathcal{S}$ to be uniquely identifiable from its canonical projections onto all possible tensorized observation patterns of size $m$ is*

$$(3.3) \qquad \binom{d}{m}(M - R') \geq D - R.$$

*where $M = \binom{m+p-1}{p}$ and $D = \binom{d+p-1}{p}$.*

Immediately from (3.3), we see that a simpler, but weaker, necessary condition for unique identifiability is $M > R'$, which is independent of the ambient dimension $d$. In fact, assuming $m > p$ and the ambient dimension $d$ is sufficiently large, then the condition in (3.3) reduces to $M > R'$. To see this, observe that $\binom{d}{m} = O(d^m)$ and $D = \binom{d+p-1}{p} = O(d^p)$, and so $(D - R)/\binom{d}{m} < 1$ for large enough $d$. Hence, in this case (3.3) reduces to $M > R'$. In the event that $R' = R$, this further reduces to the condition $m \geq m_0$ given in (3.1), the rate conjectured to be necessary and sufficient in [29].

However, the following two examples show that when some of the above assumptions are violated (e.g., when $m \leq p$ or $R' < R$) the condition given in (3.1) is neither necessary nor sufficient for unique recovery of the tensorized subspace.

*Example 2.* Suppose $\mathcal{V}$ is a generic union of two 1-D subspaces under a quadratic tensorization ($K = 2, r = 1, p = 2$). Consider all tensorized observation patterns of size $m = 2$. In this case we have $M = 3 > 2 = R' = R$, which satisfies the condition (3.1). Yet, the necessary condition (3.3) is violated in all ambient dimensions $d \geq 3$ since

$$(3.4) \qquad \binom{d}{2} \cdot \underbrace{1}_{M-R'} < \underbrace{\binom{d+1}{2}}_{D} - \underbrace{2}_{R}.$$

Hence, unique identifiability of the tensorized subspace is impossible in this case, which shows condition (3.1) is not sufficient. However, if we increase the number of observations per column

to $m = 3$, it is easy to show that the necessary condition (3.3) is always met in dimensions $d \geq 4$, and experimentally we find that the sufficient condition $\dim \ker \boldsymbol{A}^{\mathsf{T}} = R$ of Lemma 2 is also met (see Table 1).

*Example* 3. Suppose $\mathcal{V}$ is a generic union of two 2-D subspaces under a quadratic tensorization in four-dimensional ambient space ($K = 2, r = 2, p = 2, d = 4$). Suppose we consider all observation patterns of size $m = 3$. In this case $M = 6 = R$, which violates condition (3.1). However, we have $R' = 5$ since the canonical projection of the tensorized subspace onto a tensorized observation pattern of size $m = 3$ has the same dimension as a tensorized subspace arising from a generic union of two 2-D subspaces in $\mathbb{R}^3$, which has dimension 5. Hence, the necessary condition (3.3) is satisfied:

$$(3.5) \qquad 4 = \binom{4}{3} \cdot \underbrace{1}_{M-R'} = \underbrace{\binom{5}{2}}_{D} - \underbrace{6}_{R} = 4.$$

This shows that unique identification of the tensorized subspace may still be possible in this case. In the supplementary materials we prove that the sufficient condition $\dim \ker \boldsymbol{A}^{\mathsf{T}} = R$ of Lemma 2 holds in this case, which shows the tensorized subspace is uniquely identifiable. Therefore, condition (3.1) is not always necessary.

A natural question is whether the necessary condition in Corollary 1 is also sufficient, i.e., if (3.3), holds do we have unique identifiability of the tensorized subspace? Table 1 shows the results of numerical experiments that suggest this is indeed the case. In particular, we generated several generic UoS in ambient dimension $d = 12$ for a varying number of subspaces and their dimension, computed its tensorized subspace, and constructed the constraint matrix $\boldsymbol{A}$ from all possible canonical projections of the tensorized subspace onto tensorized observation patterns of size $m$. Then we searched for the minimal value of $m$ for which the necessary and sufficient condition $\dim \ker \boldsymbol{A}^{\mathsf{T}} = R$ given in Lemma 2 holds.[6] We compare this with the minimum value of $m$ for which the necessary condition (3.3) holds, and we found they agree in all cases considered.

Given the strong numerical evidence, we conjecture that the necessary condition (3.3) is also sufficient. While we do not prove this conjecture in this work, in the next section we give a sufficient condition that is only slightly stronger than the necessary condition (3.3) and orderwise optimal in terms of the number of subspaces and their dimension in many cases.

**3.5. Sufficient conditions for unique identifiability of tensorized subspaces.** In this section we identify conditions under which canonical projections onto *all possible* tensorized observation patterns arising from a sampling of $m$ coordinates in the original domain are sufficient for unique identification of the tensorized subspace. In Theorem 1 we state a general result that holds for tensorized subspaces arising from any projective variety, which we later specialize to generic UoS in Corollary 2. In particular, Corollary 2 relates the sufficient sampling rate in Theorem 1 to properties of the *Hilbert function* of generic UoS, a tool from algebraic geometry. Proposition 3 summarizes known results regarding the Hilbert function

---

[6]If the condition $\ker \boldsymbol{A}^{\mathsf{T}} = R$ holds for one random realization of a union of $K$ $r$-dimensional subspaces, then it holds generically since the condition $\ker \boldsymbol{A}^{\mathsf{T}} = R$ can be recast as a polynomial system of equations in terms of the entries of a collection of basis matrices for each subspace in the UoS.

**Table 1**

***Evidence that necessary condition*** (3.3) ***is also sufficient for unique identification of tensorized subspaces.*** *Here we identify the minimal value of m for which the tensorized subspace arising from a union of K, r-dimensional generic subspaces is uniquely identifiable from its canonical projections onto all possible tensorized observation patterns of size m. The left column gives the smallest value of m satisfying condition (3.1) that was conjectured to be necessary and sufficient in [29]. The middle column reports the smallest value of m satisfying the necessary condition (3.3). The right column reports the smallest value of m satisfying the sufficient condition ker $\boldsymbol{A}^\mathsf{T} = R$ given in Lemma 2, which is verified numerically by constructing the constraint matrix $\boldsymbol{A}$ from a randomly drawn UoS. The middle and right columns are the same and show that the necessary condition (3.3) is also sufficient in these cases. In the left column, red boxes indicate values less than the true necessary and sufficient m, and yellow indicates values more than the true necessary and sufficient m, illustrating the shortcomings of previous theory that have been addressed in this paper.*

2nd order tensorization ($p = 2$)

smallest $m$ s.t. $\binom{m+1}{2} > R$

| K\r | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 2 | 4 | 5 | 6 | 8 |
| 3 | 3 | 4 | 6 | 8 | 10 |
| 4 | 3 | 5 | 7 | 9 | 11 |
| 5 | 3 | 6 | 8 | 10 | 12 |

smallest $m$ s.t. (3.3) holds

| K\r | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 3 | 3 | 4 | 5 | 6 |
| 3 | 3 | 4 | 6 | 7 | 9 |
| 4 | 3 | 5 | 7 | 9 | 11 |
| 5 | 3 | 6 | 8 | 10 | 12 |

smallest $m$ s.t. $\dim \ker \boldsymbol{A}^\mathsf{T} = \mathrm{R}$

| K\r | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 2 | 3 | 3 | 4 | 5 | 6 |
| 3 | 3 | 4 | 6 | 7 | 9 |
| 4 | 3 | 5 | 7 | 9 | 11 |
| 5 | 3 | 6 | 8 | 10 | 12 |

3rd order tensorization ($p = 3$)

smallest $m$ s.t. $\binom{m+2}{3} > R$

| K\r | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 2 | 3 | 5 | 6 | 7 | 8 |
| 3 | 2 | 4 | 5 | 7 | 8 | 10 |
| 4 | 3 | 4 | 6 | 7 | 9 | 11 |
| 5 | 3 | 5 | 6 | 8 | 10 | 11 |
| 6 | 3 | 5 | 7 | 9 | 10 | 12 |
| 7 | 3 | 5 | 7 | 9 | 11 | 12 |

smallest $m$ s.t. (3.3) holds

| K\r | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 3 | 4 | 5 | 6 | 7 |
| 3 | 3 | 4 | 4 | 5 | 6 | 7 |
| 4 | 3 | 4 | 5 | 7 | 8 | 9 |
| 5 | 3 | 5 | 6 | 8 | 10 | 11 |
| 6 | 3 | 5 | 7 | 9 | 10 | 12 |
| 7 | 3 | 5 | 7 | 9 | 11 | 12 |

smallest $m$ s.t. $\dim \ker \boldsymbol{A}^\mathsf{T} = R$

| K\r | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 3 | 4 | 5 | 6 | 7 |
| 3 | 3 | 4 | 4 | 5 | 6 | 7 |
| 4 | 3 | 4 | 5 | 7 | 8 | 9 |
| 5 | 3 | 5 | 6 | 8 | 10 | 11 |
| 6 | 3 | 5 | 7 | 9 | 10 | 12 |
| 7 | 3 | 5 | 7 | 9 | 11 | 12 |

of generic UoS, which allows us to establish explicit sufficient sampling rates under certain restrictions on the number of subspaces, their dimension, and the tensorization order.

First, we state our main result which holds for any projective variety.

**Theorem 1.** *Let $\mathcal{V} \subset \mathbb{R}^d$ be a projective variety whose pth-order tensorized subspace $\mathcal{S}$ is R-dimensional. Suppose there exists a tensorized observation pattern $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ such that $|\boldsymbol{v}| > R$ and $\dim \mathcal{S}_{\boldsymbol{v}} = R$. Then $\mathcal{S}$ is uniquely identifiable from its canonical projections onto all possible tensorized observation patterns of size $m \geq |\boldsymbol{\omega}| + p$.*

We give the proof of Theorem 1 in Appendix A. Roughly speaking, Theorem 1 says that a sampling rate of $m \geq |\boldsymbol{\omega}| + p$ (i.e., $m$ observed entries per data column of the original matrix) is sufficient for ensuring that unique LADMC is information-theoretically *possible* (given sufficiently many columns and sufficiently diverse observation patterns). Note that Theorem 1 does not make any general position assumptions about the tensorized subspace.

By specializing to the case of tensorized subspaces generated by generic UoS, we are able to more explicitly characterize the sampling rate appearing in Theorem 1. Consider the

tensorized subspace $\mathcal{S}$ of a generic union of $K$ $r$-dimensional subspaces $\mathcal{V} \subset \mathbb{R}^d$. Recall that we define $R(d, K, r, p) = \dim \mathcal{S}$, i.e., the dimension of the tensorized subspace depends only on $d, K, r, p$ (see Proposition 2). Now, given any tensorized observation pattern $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ of size $m^*$, observe that $\mathcal{S}_{\boldsymbol{v}}$ is equal to the tensorized subspace arising from $\mathcal{V}_{\boldsymbol{\omega}} \subset \mathbb{R}^{m^*}$, the UoS restricted to the $m^*$ coordinates specified by $\boldsymbol{\omega}$. Provided $m^* > r$, $\mathcal{V}_{\boldsymbol{\omega}}$ is again a generic union of $K$ $r$-dimensional subspaces, except now it is in $m^*$-dimensional ambient space. Hence, $\mathcal{S}_{\boldsymbol{v}}$ has the same dimension as the tensorized subspace arising from generic UoS in $m^*$-dimensional ambient space, and so we have $\dim \mathcal{S}_{\boldsymbol{v}} = R(m^*, K, r, p)$ *for any* tensorized observation pattern $\boldsymbol{v}$ of size $m^* > r$. This fact, combined with Theorem 1, gives the following immediate corollary.

**Corollary 2.** *Let $\mathcal{V} \subset \mathbb{R}^d$ be a generic union of $K$ $r$-dimensional subspaces, and let $\mathcal{S} \subset \mathbb{R}^D$ be its pth-order tensorized subspace. Assume $m^*$ is such that $r < m^* \leq d$ and $R(m^*, K, r, p) = R(d, K, r, p)$. Then $\mathcal{S}$ is uniquely identifiable from its canonical projections onto all possible tensorized observation patterns of size $m \geq m^* + p$.*

The key assumption made in Corollary 2 is that $R(m^*, K, r, p) = R(d, K, r, p)$. Characterizing the set of values for which this condition holds in all generality appears to be a difficult problem (see, e.g., [7]). However, using existing results [18, 11, 7] that characterize exact values of $R(d, K, r, p)$, we can establish the following special cases.

**Proposition 3.** *Let $m_0$ be the smallest integer such that $\binom{m_0+p-1}{p} > K\binom{r+p-1}{p}$, and set $m^* = \max\{m_0, 2r\}$. Then $R(m^*, K, r, p) = K\binom{r+p-1}{p} = R(d, K, r, p)$ in the following cases:*
  (a) *$p = 2$, for any $K$, for any $r$ (i.e., any generic UoS under a quadratic tensorization).*
  (b) *$p \geq 3$, for any $K$, and $r = 1$ or $2$ (i.e., any generic union of 1-D subspaces, or any generic union of 2-D subspaces, under any higher-order tensorization).*
  (c) *$p \geq 3$, $K \leq p$, for any $r$ (i.e., any generic UoS consisting of at most $p$ $r$-dimensional subspaces under a pth-order tensorization).*

Case (a) is due to [7, Theorem 3.2], case (b) is due to [18], and case (c) is due to [11, Corollary 4.8] (see also [27, Corollary 2.16]).

The quantity $m_0$ defined in Proposition 3 is $O(K^{1/p}r)$, and hence so is $m^* = \max\{2r, m_0\}$. Therefore, Proposition 3 combined with Corollary 2 shows that a sampling rate of $m = O(K^{1/p}r + p)$ is sufficient for unique identifiability of the tensorized subspace arising from a generic union of $K$ subspaces of dimension $r$ (under one of assumptions (a)–(c) in Corollary 2). When $m^* \geq 2r$, the sampling rate identified in Corollary 2 is only slightly more than the minimal sampling rate $m_0$ given in (3.1) conjectured to be necessary and sufficient for unique identifiability in [29]. Specifically, in this case the sampling rate in Corollary 2 is $m \geq m_0 + p$. By the discussion following Corollary 1, this rate is also necessary provided $m_0 > p$ and provided the ambient dimension $d$ is sufficiently large. Hence, in these cases there is only a gap of up to $p$ observations per column between our necessary and sufficient conditions (i.e., $m \geq m_0$ versus $m \geq m_0 + p$).

In general, we conjecture that the value of $m^*$ as defined in Proposition 3 is always sufficient for ensuring $R(m^*, K, r, p) = R(d, K, r, p)$ for higher-order tensorizations $p \geq 3$. While proving this may be difficult, this condition can also be checked numerically by sampling sufficiently many points from a randomly generated UoS with the specified parameters and computing the rank of tensorized matrix. However, we reiterate that empirical evidence leads us to believe

that the necessary sampling rate for UoS identified in Corollary 1 is also sufficient, which generally is less than the rate given in Corollary 2.

### 3.6. Implications for LADMC.

**3.6.1. Sample complexity.** The above results are stated in terms of the unique identification of the tensorized subspace $S$ from canonical projections. However, unique identification of the tensorized subspace also implies unique completion of the tensorized matrix $X^{\otimes p}$ provided the observation pattern matrix $\Omega$ contains sufficiently many duplicates of each of the $\binom{d}{m}$ possible observation patterns so that all the canonical projections of the tensorized subspace can be determined.

For example, suppose each of the $N$ columns in $\Omega$ is drawn randomly from the $n = \binom{d}{m}$ possible observation patterns of size $m$ in $d$ coordinates. Then, by a variant of the coupon collector's problem, with high probability $\Omega$ will contain $R$ copies of each observation pattern provided the number of columns $N = n \log n + (R-1)n \log \log n + O(n)$, which reduces to $N = O(Rd^m \log d)$. If every subset of $R$ columns of $X$ is in general position, it will be possible to determine the canonical projections of $S$ from the columns of $X^{\otimes p}$. Once $S$ is recovered, then it is possible to uniquely complete the matrix by projecting each incomplete column of $X^{\otimes p}$ onto $S$ and performing the "detensorization" step of LADMC (step 4 of Algorithm 2.1).

While the above argument establishes a sufficient number of columns to uniquely complete an LAD matrix with high probability, we believe this is a vast overestimate of how many columns are truly necessary and sufficient for successful completion with LADMC. For example, a naive extension of the results in [34] would lead one to believe that $N \geq (R+1)(D-R)$ columns are necessary and sufficient for unique recovery of $S$, which is far less than the estimate given in the previous paragraph. However, the tensorization process violates many of the genericity assumptions in [34], which prevents the direct extension of these results to the present setting. Nevertheless, empirically we observe that LADMC often successfully recovers synthetic variety data with the necessary minimal number observations per column (selected uniformly at random) provided there are $N = O(R(D-R))$ columns, and we conjecture this is the true necessary and sufficient orderwise number of columns needed for recovery with LADMC (see Figure 3 and section 4 for more discussion on this point).

**3.6.2. Tightness of bounds.** In the special case of a union of $K$ subspaces of dimension $r$, Corollary 2 shows that $m = O(K^{1/p}r + p)$ observations per data column are information-theoretically sufficient for unique identification of the tensorized subspace and hence unique completion of the original matrix (given sufficiently many data columns and under some restrictions on $p, K$, and $r$). In contrast, the information-theoretic requirement for the success of subspace clustering with missing data (SCMD), which is mathematically equivalent to matrix completion under a UoS model, is $m = r + 1$ observations per data column [35]. If $p = O(1)$, i.e., the tensor order is fixed and not allowed to scale with the number of subspaces, this shows that the information-theoretic necessary sample complexity of LADMC is orderwise suboptimal by a factor of $K^{1/p}$. However, if the tensor order $p$ scales with the number of subspaces $K$ as $p = O(\log K)$, then we have $m = O(r + \log K)$, which is nearly orderwise optimal. Nonetheless, even with fixed and low tensor orders (e.g., $p = 2, 3$), empirically we find that LADMC performs equally as well as or better than most state-of-the-art SCMD methods on UoS data (see Figure 2).

**Table 2**

*Sample, memory, and computational complexity of algorithms for completing a $d \times N$ matrix whose columns belong to a union of $K$ $r$-dimensional subspaces. We compare rates for low-rank matrix completion (LRMC) via the iterative singular value hard thresholding (ISVHT) algorithm [19], low algebraic dimension matrix completion (LADMC) via Algorithm 2.1, and the iterative variant (iLADMC) using a pth-order tensorization and the kernel-based variety matrix completion (VMC) algorithm of [29].*

| | LRMC | LADMC | iLADMC | VMC |
|---|---|---|---|---|
| Minimum necessary sampling rate (observations/column) | $O(Kr)$ | $O(K^{1/p}r)$ | $\leq O(K^{1/p}r)$ | $\leq O(K^{1/p}r)$ |
| Memory requirements (bytes) | $O(dN)$ | $O(d^pN)$ | $O(d^pN)$ | $O(N^2)$ |
| Computational complexity (flops/iteration) | $O(dNKr)$ | $O(d^pNKr^p)$ | $O(d^pNKr^p)$ | $O(N^2Kr^p)$ |

**4. Experiments.** In the following experiments we demonstrate the performance of the proposed LADMC algorithm (Algorithm 2.1) on real and synthetic data having low algebraic dimension, and we empirically verify the information-theoretic sampling requirements for LADMC for data belonging to a UoS model given in section 3.

**4.1. Implementation details: LADMC, iLADMC, and VMC.** In our implementation of LADMC (Algorithm 2.1) we use the iterative singular value hard thresholding (ISVHT) algorithm [19] to perform LRMC in the tensorized domain. The "detensorization" step of Algorithm 2.1 is performed by a rank-1 truncated SVD of each reshaped tensorized column, where the sign ambiguity is resolved by matching with the sign of the observed entry having maximum absolute value in each column.

We also test an iterative version of LADMC (iLADMC), where we perform a small number of ISVHT iterations in the tensorized domain, map back to the original domain by the columnwise rank-1 SVD detensorization step, fill in the observed entries of the matrix, and repeat these steps until convergence. In the experiments below we ran 30 iterations of ISVHT for iLADMC. Periodically performing the detensorization step amounts to a projection onto the space of matrices in the tensorized space with the necessary tensorized structure, i.e., each column is a vector of the form $\boldsymbol{x}^{\otimes p}$. While we have no theory to show that an iterative approach should outperform LADMC, empirically we find that iLADMC converges much faster than LADMC (in terms of the number of ISVHT steps, which is the main computational bottleneck) and succeeds in completing matrices at lower sampling rates than plain LADMC.

In an earlier work [29] we introduced an algorithm called variety-based matrix completion (VMC) designed to achieve the same goal as LADMC and iLAMDC. In particular, VMC attempts to minimize the nonconvex Schatten-$q$ quasi-norm ($0 < q < 1$) of the tensorized matrix $\boldsymbol{X}^{\otimes p}$ using an iterative reweighted least squares approach [28]. The VMC algorithm is most similar to iLADMC, since it also enforces the tensorized structure at each iteration.

Each of these approaches has different trade-offs with respect to sample complexity, memory complexity, and computational complexity. In Table 2 we compare these complexities for completing data belonging to a union of subspaces. Note that LADMC, iLADMC, and VMC have similar sample complexities, while LADMC/iLADMC have lower memory and computational complexity when the dimension of the data points raised to the tensor order $d^p$ is less than the number of data columns $N$.

**4.2. Sample complexity of union-of-subspaces data.** Figure 2 shows the performance of the LADMC and iLADMC algorithms against competing methods for the recovery of

synthetic union-of-subspaces data with missing entries. We generated $d \times N$ data matrices whose columns belong to a union of $K$ subspaces, each of dimension $r$, and sampled $m$ entries in each column, selected uniformly at random. We used the settings $d = 15$, $N = 50K$, $r = 2$ for varying measurements $m$ and number of subspaces $K$ and measured the fraction of successful completions over 25 random trials for each pair $(m, K)$. We judged the matrix to be successfully completed if the normalized root mean square error $\|\mathbf{X} - \mathbf{X_0}\|_F / \|\mathbf{X_0}\|_F$ was less than $10^{-4}$, where $\mathbf{X}$ is the recovered matrix and $\mathbf{X_0}$ is the ground truth matrix and $\| \cdot \|_F$ denotes the Frobenius norm. Here we compared against low-rank matrix completion (LRMC) via iterative singular value hard thresholding (ISVHT) [19] in the original matrix domain, and three methods based on subspace clustering: sparse subspace clustering with entrywise zero fill [51] followed by LRMC on each identified cluster (SSC+EWZF), the expectation-maximization (EM) algorithm proposed in [38], and the group-sparse subspace clustering (GSSC) algorithm [31] followed by LRMC on each cluster. For the subspace clustering algorithms that required an estimate of the number of subspaces and their dimension we used the exact values. The EM and GSSC algorithms were initialized with the subspace clustering obtained by SSC-EWZF. Any remaining free parameters in these algorithms were set via cross-validation. For LADMC and iLADMC we used a quadratic tensorization ($p = 2$), and LRMC steps for these algorithms were performed via ISVHT with the rank threshold parameter set to the exact rank of the tensorized matrix.
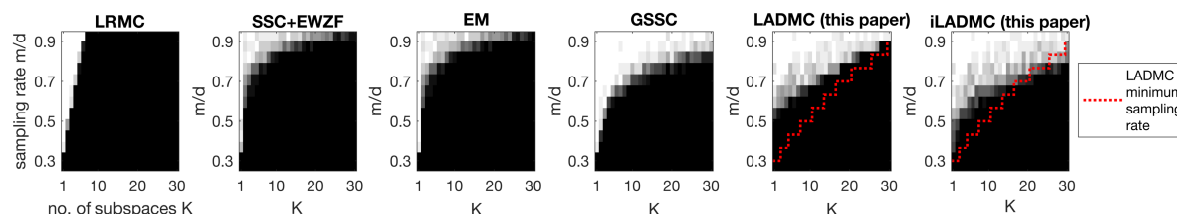


**Figure 2.** *Phase transitions for matrix completion of synthetic UoS data. We simulated data belonging to $K$ subspaces and sampled each column of the data matrix at a rate of $m/d$ and performed matrix completion using LRMC, state-of-the-art subspace clustering based algorithms (SSC+EWZF, GSSC, EM), and the proposed LADMC and iLADMC algorithms with quadratic tensorizations. Grayscale values indicate the fraction of random trials where the matrix was successfully recovered; white is $100\%$ success and black is $100\%$ failure. The red dashed line indicates the minimal information-theoretic sampling rate $m/d = O(\sqrt{K})$ needed for LRMC to succeed in the tensorized domain as specified by Corollary 1.*

We find that LADMC is able to successfully complete the data when the number of measurements per column in the tensorized domain exceeds the information-theoretic bounds established in Corollary 1, as indicated by the red dashed line in Figure 2. This is a substantial extension over standard LRMC: for these settings, LADMC is able to complete data matrices drawn from up to $K = 30$ subspaces, whereas LRMC is limited to data drawn from fewer than $K = 7$ subspaces. However, for LADMC there is a small gap between the information-theoretic bound and the true phase transition, which is most apparent where the number of subspaces and the sampling rate are low (lower-left of the plot), but this gap closes as the number of subspaces and sampling rate increase (upper-right of the plot). We hypothesize that this is due to insufficiently many data columns (see Figure 3 and the discussion below). This gap is less pronounced for iLADMC, and, in fact, in the upper-right of the plot iLADMC
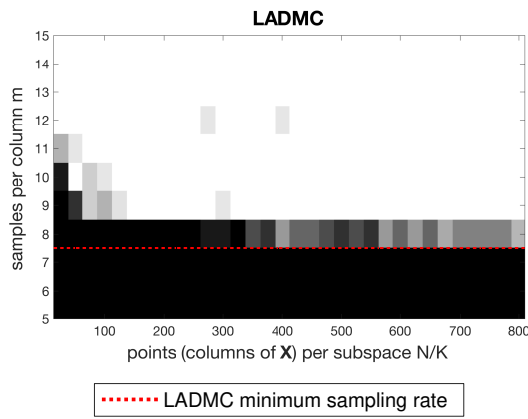
**Figure 3.** *Effect of number of data columns per subspace $N/K$ on the overall LADMC recovery of synthetic UoS data ($K = 10$ subspaces of dimension $r = 5$ in a $d = 15$ dimensional space). As the number of columns per subspace increases, the probability of exact recovery approaches 1 for $m \geq 8$, the necessary minimum number of samples per column identified by Corollary 1.*
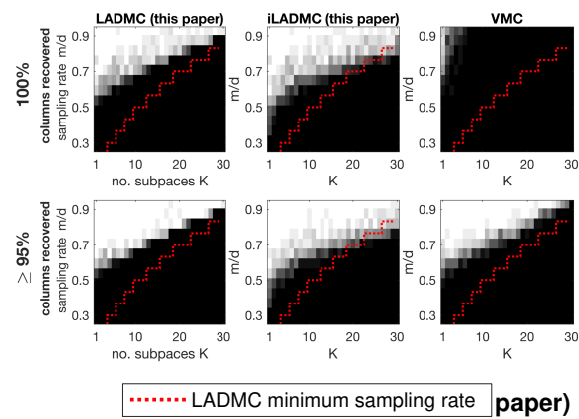


**Figure 4.** *Relative performances of LADMC and iLADMC compared with VMC on synthetic UoS data. LADMC and iLADMC recover with high probability at $0.7$ sampling and VMC often fails (top row). They perform similarly when comparing recovery of at least 95% of the columns (bottom row).*

shows recovery below the LADMC information-theoretic bound. We conjecture that this is because iLADMC is enforcing extra structural constraints that are not accounted for in the theory, which may reduce the sampling rate necessary to monitor. iLADMC reduces both of the necessary number of measurements and the number of iterations. We also observe that the performance of iLADMC and LADMC are comparable with a performing subspace clustering-based algorithm, which in this case is GSSC.

In Figure 3 we investigate the effect of the number of data columns per subspace on the overall recovery performance of LADMC for synthetic UoS data. Here we use the same settings as in the previous experiment but fix the number of subspaces to be $K = 10$ and vary the number of columns per subspace, $N/K$, and the number of random measurements per column, $m$. In this case, the tensorized subspace has rank $R = 30$, and the necessary minimum number of observations per column according to Corollary 1 is $m = 8$. Observe that as the number of columns per subspace increases, the probability of exact recovery is approaching one for $m \geq 8$, as predicted by Corollary 1. We conjecture that the minimum number of columns per subspace needed for exact recovery is $N/K = O(R(D - R)/K)$ (see section 3.6.1). Assuming the constant in the orderwise expression is one, we have $N/K \approx 270$. Note that we do see exact recovery at $m = 9$ samples per column when $N/K = 270$ and partial success at $m = 8$ with double or triple the columns, as predicted.

**4.3. Comparison with VMC.** In Figure 4 we compare the relative performances of LADMC and iLADMC with VMC for the same synthetic UoS data as in Figure 2. One drawback of VMC observed in [29] is that it often failed to complete a small proportion of the data columns correctly, even at high sampling rates on synthetic data. Consistent with the results in [29], we find that VMC and LADMC/iLADMC perform similarly when comparing probability of

recovering at least 95% columns. However, both LADMC and iLADMC recover 100% of the data columns correctly above the minimum sampling rate, whereas VMC mostly fails under this stricter recovery criterion. This shows that LADMC/iLADMC could have some empirical benefits over VMC if high accuracy solutions are desired.

**4.4. Higher-order tensorizations.** In Figure 5 we experimentally verify the predicted minimal sampling rate for UoS data with higher-order tensorizations specified in Corollary 1. In this work we do not pursue higher-order $p \geq 3$ LADMC with Algorithm 2.1, due to poor scalability with respect to the ambient dimension $d$ and a lack of an efficient implementation of the detensorization step, which prohibited us from investigating the phase transition behavior of LADMC over a reasonable range of the number of subspaces $K$. Instead, we verify our predictions using the VMC algorithm [29], for which the sufficient conditions of Corollary 2 also hold (although the necessary conditions of Corollary 1 may not hold). We find that the phase transition recovery follows the dependence $m = O(K^{1/p})$ for tensor orders $p = 2, 3$ as predicted by Corollaries 1 and 2.
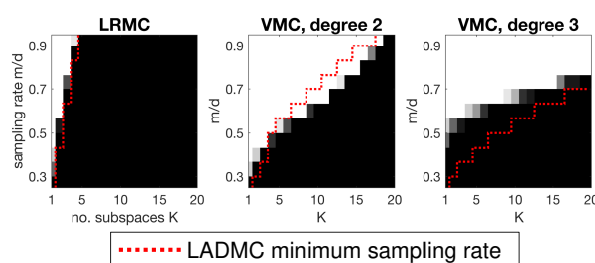


**Figure 5.** *Phase transitions for matrix completion for union of subspaces using no tensorization (LRMC), 2nd-order tensorization (VMC, degree 2), and 3rd-order tensorization (VMC, degree 3). The phase transition follows closely the LADMC minimum sampling rate established in Corollary 1, which is $m = O(K^{1/p})$ where $K$ is the number of subspaces and $p$ is the tensor order. Here the ambient dimension is $d = 15$ and the subspace dimension is $r = 3$. (Figure adapted from [29]).*

**4.5. Experiments on real data.** In Table 3 we illustrate the performances of LADMC and iLADMC on the following three real-world datasets:[7] the Oil Flow dataset introduced in [4], the Jester-1 recommender systems dataset [17], and the MNIST digit recognition dataset introduced in [25]. We chose these datasets to demonstrate the feasibility of LADMC on a variety of data sources and because they had sufficiently small row dimension to make LADMC to be computationally practical. For the Oil Flow and MNIST datasets we simulated missing data by randomly subsampling each data column uniformly at a rate of 15%-25% training-validation-test split of the data. For the Jester dataset we used ten randomly selected ratings of each user for training, nine random ratings for validation, and the remainder for testing. To establish baselines we compared with filling the missing entries with the mean of the observed entries within each column (Mean-fill), and with LRMC via nuclear

---

[7]The following datasets are available online: Oil Flow (http://inverseprobability.com/3PhaseData.html), Jester-1 (http://goldberg.berkeley.edu/jester-data/), MNIST (http://yann.lecun.com/exdb/mnist/). For computational reasons, we reduced the size of the MNIST dataset by selecting a random subset of 20,000 images and downsampling each image by a factor of two in both dimensions.

**Table 3**

*Matrix completion results on real data.*

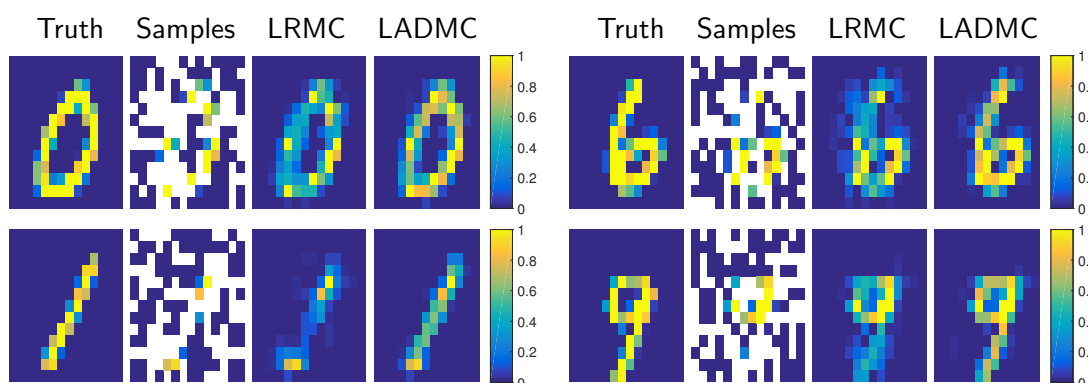| Dataset | Size | Samples | Completion RMSE on test set | | | |
|---------|------|---------|-----------|------|-------|--------|
|         |      |         | Mean-fill | LRMC | LADMC | iLADMC |
| Oil flow | 12×1000 | 50% | 0.237 | 0.164 | 0.155 | **0.127** |
| Jester-1 | 100×24983 | 18% | 4.722 | **4.381** | 4.420 | 4.394 |
| MNIST | 196×20000 | 50% | 0.309 | 0.210 | **0.187** | **0.187** |



**Figure 6.** *Representative examples of matrix completion on MNIST dataset. Here we randomly remove 50% of the pixels in each MNIST image and attempt to jointly recover the missing pixels of all images by LRMC and LADMC using a quadratic tensorization ($p = 2$).*

norm minimization [41], which outperformed LRMC via singular value iterative hard thresholding [19] on these datasets. For the LRMC routine within LADMC we set the rank cutoff $R$ to the value that gave the smallest completion error on the validation set, and we use the same rank cutoff $R$ for iLADMC. For all methods, we report the root mean square error (RMSE) of the completion on the test set. We find that LADMC/iLADMC give significantly lower RMSE on the Oil Flow and MNIST datasets relative to the baselines; iLADMC gives lower RMSE than LADMC on the Oil Flow dataset, but performs similarly to LADMC on the others. Figure 6 illustrates the improvement of LADMC over LRMC on a selection of examples from the MNIST dataset. We see fewer differences between LRMC and LADMC/iLADMC on the Jester-1 dataset, where LADMC/iLADMC give nearly the same RMSE as LRMC. Because of the lower sampling rate for the Jester-1 dataset, the rank cutoff $R$ in LADMC was kept small to avoid overfitting, and we suspect in this case that LADMC is fitting a linear subspace to the data, which would explain the similar performance to LRMC.

**5. Conclusion.** The theory and algorithms presented in this paper give new insight into conducting matrix completion when the matrix columns correspond to points on a nonlinear algebraic variety, including union-of-subspaces as a special case. Unlike most matrix completion methods assuming a union-of-subspace model, the proposed approach does not necessitate an intermediate subspace clustering step that can be fragile in the presence of missing data.

The theoretical guarantees in this work focus on unique identifiability of the tensorized subspace from canonical projections, i.e., we assume we observe multiple columns with each possible observation pattern. This assumption is not always met in practice, yet the proposed

LADMC algorithm nevertheless performs well empirically. An important avenue for future study is to establish conditions for unique completion of partially sampled data matrices.

In the experimental portion of this work we primarily focused on LADMC with a quadratic tensorization. Yet, we also showed that our approach and results generalize to LADMC with higher-order tensorizations. In principle, this extension would facilitate the completion of data belonging to a richer class of varieties and with more missing data. However, the computational complexity of LADMC scales as $O(d^p)$, where $d$ is the ambient (data) dimension and $p$ is the tensor order, making our approach computationally challenging for even modest data dimensions $d$.

One potential solution is to use a kernelized algorithm, such as VMC [29], that avoids explicitly constructing a large-scale tensorized matrix. Unfortunately, kernelized approaches have complexity and storage requirements that scale quadratically with the number of data columns, making such an approach computationally challenging for large datasets.

We are actively investigating memory and computationally efficient algorithms that allow more efficient extensions of the LADMC approach for higher-order tensorizations. Along these lines, recent work investigates efficient online algorithms for a class of nonlinear matrix completion problems that includes the LADMC model [13]. An alternative approach would be to incorporate matrix sketching techniques to allow for more efficient computation of the SVD in the tensorized domain. For example, recent work that focuses on sketches for matrices with Kronecker product structure, such as TensorSketch [30] and its modifications [1] or the Kronecker fast Johnson–Lindenstrauss transform [21], could potentially be adapted to our setting.

**Appendix A. Proof of Theorem 1.** We prove Theorem 1 by showing we can construct an observation pattern matrix $\boldsymbol{\Upsilon}^* \in \{0,1\}^{D \times (D-R)}$ such that the resulting *constraint matrix* $\boldsymbol{A} = \boldsymbol{A}(\mathcal{S}, \boldsymbol{\Upsilon}^*) \in \mathbb{R}^{D \times (D-R)}$ satisfies Lemma 2, i.e., $\dim \ker \boldsymbol{A}^\mathsf{T} = R$.

Note that in the tensor domain we observe projections of the tensorized subspace onto subsets of $M = \binom{m+p-1}{p}$ coordinates where $M$ may be larger than $R + 1$. However, from any canonical projection of the tensorized subspace onto $M > R + 1$ coordinates we can also recover its canonical projections onto any subset of $R + 1$ coordinates of the $M$ coordinates. That is, if we observe one canonical projection $\mathcal{S}_{\boldsymbol{v}}$ with $|\boldsymbol{v}| = M$, then we also have access to all canonical projections $\mathcal{S}_{\boldsymbol{v}'}$ where $\boldsymbol{v}'$ is any observation pattern with $\operatorname{supp}(\boldsymbol{v}') \subset \operatorname{supp}(\boldsymbol{v})$ and $|\boldsymbol{v}'| = R + 1$.

To express this fact more succinctly, we introduce some additional notation. For any observation pattern matrix $\boldsymbol{\Upsilon} \in \{0,1\}^{D \times n}$ whose columns all have greater than $R$ nonzeros, let $\widehat{\boldsymbol{\Upsilon}}$ denote the matrix of observation patterns having exactly $R + 1$ nonzeros that can be generated from the observation patterns in $\boldsymbol{\Upsilon}$. For example, if $D = 4$, $R = 1$, then from the two 3-dimensional projections indicated in $\boldsymbol{\Upsilon}$ below we obtain the five 2-dimensional projections indicated in $\widehat{\boldsymbol{\Upsilon}}$ below:

$$\boldsymbol{\Upsilon} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \mapsto \widehat{\boldsymbol{\Upsilon}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Recall that every coordinate in tensor space is associated with an ordered tuple $(k_1, \ldots, k_p)$ satisfying $1 \leq k_1 \leq \cdots \leq k_p \leq d$, where each $k_i$ indicates one of the $d$ coordinates in the original space. We assume that coordinates in tensor space are ordered such that for all $m = 1, \ldots, d$, the first $M = \binom{m+p-1}{p}$ coordinates correspond to all tuples $1 \leq k_1 \leq \cdots \leq k_d \leq m$. We call this the *standard ordering.*

We now show that if $\mathbf{\Upsilon}$ consists of all tensorized observation patterns of a certain size, then the expanded observation pattern matrix $\widehat{\mathbf{\Upsilon}}$ contains several submatrices having a useful canonical form.

**Lemma 4.** *Fix a tensor order $p \geq 2$. Suppose the columns of $\mathbf{\Upsilon}$ are given by all $\binom{d}{m}$ tensorized observation patterns of size $m \geq m^* + p$ where $m^*$ is the smallest integer such that $M^* := \binom{m^*+p-1}{p} > R$, and let $\widehat{\mathbf{\Upsilon}}$ be its expanded observation pattern matrix having exactly $R+1$ ones per column. Then any permutation of the first $M^*$ rows of $\widehat{\mathbf{\Upsilon}}$ has a submatrix of the form*

$$
(A.1) \qquad \mathbf{\Upsilon}^\star = \begin{bmatrix} \mathbf{1} \\ \hline \mathbf{I} \end{bmatrix} \begin{array}{l} \left.\vphantom{\mathbf{1}}\right\} R \\ \left.\vphantom{\mathbf{I}}\right\} D - R, \end{array}
$$

*where $\mathbf{1}$ is the $R \times (D-R)$ matrix of all ones, and $\mathbf{I}$ is the $(D-R) \times (D-R)$ identity.*

*Proof.* Let $\boldsymbol{v}_j^*$ denote the $j$th column of $\mathbf{\Upsilon}^\star$ in (A.1), whose first $R$ entries are ones and $(R+j)$th entry is one while the rest of its entries are zero. We show that $\boldsymbol{v}_j^*$ is a column $\widehat{\boldsymbol{v}}$ in the expanded matrix $\widehat{\mathbf{\Upsilon}}$. Let $(k_1, \ldots, k_p)$ be the ordered tuple corresponding to the $(R+j)$th coordinate in the tensor space. Note that a column $\boldsymbol{v}$ in $\mathbf{\Upsilon}$ has nonzero $(R+j)$th entry if and only if the corresponding observation pattern $\boldsymbol{\omega}$ is nonzero in entries $k_1, k_2, \ldots, k_p$. Let $\boldsymbol{\omega}$ be any column of $\mathbf{\Omega}$ such that all entries indexed by $\{1, \ldots, m^*, k_1, \ldots, k_p\}$ are equal to one. By construction, $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ has ones in its first $M^*$ entries and must also have a one at the $(R+j)$th entry. This shows that $\boldsymbol{v}_j^*$ is a column of the expanded matrix $\widehat{\mathbf{\Upsilon}}$ and thus of any permutation of the first $M^*$ rows of $\widehat{\mathbf{\Upsilon}}$. Since this is true for every $j = 1, \ldots, D-R$, we know that $\mathbf{\Upsilon}$ will produce a matrix $\widehat{\mathbf{\Upsilon}}$ containing $\mathbf{\Upsilon}^\star$ as in (A.1) (and likewise for any permutation of the first $M^*$ rows of $\widehat{\mathbf{\Upsilon}}$). ∎

Now we are ready to give the proof of Theorem 1.

*Proof of Theorem 1.* First, we will permute the tensorized coordinate system into a convenient arrangement. Assume there exists a tensorized observation pattern $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ such that $\dim \mathcal{S}_{\boldsymbol{v}} = R$. Define $m^* = |\boldsymbol{\omega}|$. Without loss of generality, we may permute coordinates in the original domain such that the first $m^*$ entries of $\boldsymbol{\omega}$ are ones. Under the standard ordering of tensor coordinates, this means the first $M^* = \binom{m^*+p-1}{p}$ entries of $\boldsymbol{v} = \boldsymbol{\omega}^{\otimes p}$ are ones. Since $\dim \mathcal{S}_{\boldsymbol{v}} = R$, there exists an observation pattern $\boldsymbol{v}'$ with $\operatorname{supp}(\boldsymbol{v}') \subset \operatorname{supp}(\boldsymbol{v})$ having exactly $R$ ones such that $\dim \mathcal{S}_{\boldsymbol{v}'} = R$. We may permute the first $M^*$ coordinates in tensor space so that $\boldsymbol{v}'$ has all its ones in the first $R$ coordinates. Thus, the restriction of $\mathcal{S}$ to the first $R$ coordinates is $R$-dimensional (i.e., $\dim \mathcal{S}_{\boldsymbol{v}'} = R$).

Now suppose we observe canonical projections of $\mathcal{S}$ onto all tensorized observation patterns

of size $m \geq m^* + p$, which we collect into a matrix $\boldsymbol{\Upsilon} \in \{0,1\}^{D \times \binom{d}{m}}$. Then by Lemma 4 there exists a submatrix $\boldsymbol{\Upsilon}^* \in \{0,1\}^{D \times (D-R)}$ of the expanded matrix $\widehat{\boldsymbol{\Upsilon}}$ having the form (A.1). Hence, from canonical projections of $\mathcal{S}$ onto observation patterns in $\boldsymbol{\Upsilon}$ we can derive all canonical projections of $\mathcal{S}$ onto observation patterns in $\boldsymbol{\Upsilon}^*$.

For $j = 1, \ldots, D - R$, let $\boldsymbol{v}_j^*$ be the $j$th column of $\boldsymbol{\Upsilon}^*$. Since $\boldsymbol{v}_j^*$ has exactly $R+1$ ones, and the restricted subspace $\mathcal{S}_{\boldsymbol{v}_j^*}$ is at most $R$-dimensional, the orthogonal complement of the restricted subspace $\mathcal{S}_{\boldsymbol{v}_j^*}$ is positive dimensional, and so there exists at least one nonzero constraint vector $\boldsymbol{a}_j \in (\mathcal{S}_{\boldsymbol{v}_j^*})^\perp$. Let $\boldsymbol{a}_j^* \in \mathbb{R}^d$ be the vector whose restriction to $\boldsymbol{v}_j$ is equal to $\boldsymbol{a}_j$ and zero in its other entries. Then consider the constraint matrix $\boldsymbol{A}^* = [\boldsymbol{a}_1^*, \ldots, \boldsymbol{a}_{D-R}^*]$, which has the same dimensions as $\boldsymbol{\Upsilon}^*$ and is such that an entry of $\boldsymbol{A}^*$ is nonzero only if the corresponding entry of $\boldsymbol{\Upsilon}^*$ is nonzero. In particular, this means that

$$(A.2) \qquad \boldsymbol{A}^* = \begin{bmatrix} \boldsymbol{A}_0^* \\ \boldsymbol{A}_1^* \end{bmatrix} \in \mathbb{R}^{D \times (D-R)},$$

where $\boldsymbol{A}_1^* \in \mathbb{R}^{(D-R) \times (D-R)}$ is a diagonal matrix. To finish the proof it suffices to show that the diagonal entries of $\boldsymbol{A}_1^*$ are all nonzero, since this would imply $\operatorname{rank}(\boldsymbol{A}^*) = D - R$, and hence $\dim \ker[(\boldsymbol{A}^*)^\mathsf{T}] = R$, which by Lemma 2 implies that the subspace $\mathcal{S}$ is uniquely identifiable.

Showing that the diagonal entries of $\boldsymbol{A}_1^*$ are all nonzero is equivalent to showing the constraint vector $\boldsymbol{a}_j^*$ is nonzero at entry $(R+j)$ for all $j = 1, \ldots, D-R$. Suppose, by way of contradiction, that $\boldsymbol{a}_j^*$ were zero at entry $(R+j)$. This means that the nonzero support of $\boldsymbol{a}_j^*$ is contained in the first $R$ coordinates. Let $\boldsymbol{B} \in \mathbb{R}^{D \times R}$ be a basis matrix for the tensorized subspace $\mathcal{S}$, $\boldsymbol{v}'$ be the $D \times 1$ vector with first $R$ rows equal to 1 and the remainder equal to 0, and $\boldsymbol{B}_{\boldsymbol{v}'} \in \mathbb{R}^{R \times R}$ be the matrix composed of the first $R$ rows of $\boldsymbol{B}$. By definition, $\boldsymbol{a}_j^* \in \ker \boldsymbol{B}^\mathsf{T}$, and so $\boldsymbol{B}^\mathsf{T} \boldsymbol{a}_j^* = (\boldsymbol{B}_{\boldsymbol{v}'})^\mathsf{T} (\boldsymbol{a}_j^*)_{\boldsymbol{v}'} = \boldsymbol{0}$. Since $\boldsymbol{a}_j^* \neq 0$ by definition and because the non-zero support of $\boldsymbol{a}_j^*$ is the same as the nonzero support of $(\boldsymbol{a}_j^*)_{\boldsymbol{v}'}$ by hypothesis, we have $(\boldsymbol{a}_j^*)_{\boldsymbol{v}'} \neq \boldsymbol{0}$. This implies $(\boldsymbol{B}_{\boldsymbol{v}'})^\mathsf{T} \in \mathbb{R}^{R \times R}$ is rank-deficient, and hence so is $\boldsymbol{B}_{\boldsymbol{v}'}$ or, equivalently, $\dim \mathcal{S}_{\boldsymbol{v}'} < R$, which is a contradiction. Hence $\boldsymbol{a}_j^*$ is nonzero at entry $(R+j)$ for all $j = 1, \ldots, D-R$, and so $\boldsymbol{A}_1^*$ is nonzero at every entry along its diagonal, which completes the proof. ∎

## REFERENCES

[1] T. D. Ahle, M. Kapralov, J. B. T. Knudsen, R. Pagh, A. Velingker, D. P. Woodruff, and A. Zandieh, *Oblivious sketching of high-degree polynomial kernels*, in Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2020, pp. 141–160, https://doi.org/10.1137/1.9781611975994.9.

[2] B. Eriksson, L. Balzano, and R. Nowak, *High-rank matrix completion*, in Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, PMLR 22, 2012, pp. 373–381, http://proceedings.mlr.press/v22/eriksson12.html.

[3] L. Balzano, A. Szlam, B. Recht, and R. Nowak, *K-subspaces with missing data*, in Proceedings of the 2012 IEEE Statistical Signal Processing Workshop, Ann Arbor, MI, 2012, pp. 612–615.

[4] C. M. BISHOP AND G. D. JAMES, *Analysis of multiphase flows using dual-energy gamma densitometry and neural networks*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 327 (1993), pp. 580–593.

[5] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), 717.

[6] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.

[7] E. CARLINI, M. V. CATALISANO, AND A. V. GERAMITA, *Subspace arrangements, configurations of linear spaces and the quadrics containing them*, J. Algebra, 362 (2012), pp. 70–83.

[8] M. CHARDIN, *Une majoration de la fonction de Hilbert et ses conséquences pour l'interpolation algébrique*, Bull. Soc. Math. France, 117 (1989), pp. 305–318.

[9] Z. CHARLES, A. JALALI, AND R. WILLETT, *Subspace clustering with missing and corrupted data*, in Proceedings of the 2018 IEEE Data Science Workshop, Lausanne, 2018, pp. 180–184.

[10] D. COX, J. LITTLE, AND D. O'SHEA, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd ed., Undergrad. Texts Math., Springer, 2007.

[11] H. DERKSEN, *Hilbert series of subspace arrangements*, J. Pure Appl. Algebra, 209 (2007), pp. 91–98.

[12] E. ELHAMIFAR, *High-rank matrix completion and clustering under self-expressive models*, in Advances in Neural Information Processing Systems (NIPS), Vol. 29, Curran Associates, Inc., 2016, pp. 73–81.

[13] J. FAN AND M. UDELL, *Online high rank matrix completion*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, 2019, pp. 8690–8698.

[14] J. FAN, Y. ZHANG, AND M. UDELL, *Polynomial matrix completion for missing data imputation and transductive learning*, in Proceedings of the AAAI Conference on Artificial Intelligence, 34 (04), 2020, pp. 3842–3849, https://ojs.aaai.org/index.php/AAAI/article/view/5796.

[15] S. GANDY, B. RECHT, AND I. YAMADA, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, 27 (2011), 025010.

[16] P. GAO, M. WANG, J. H. CHOW, M. BERGER, AND L. M. SEVERSKY, *Missing data recovery for high-dimensional signals with nonlinear low-dimensional structures*, IEEE Trans. Signal Process., 65 (2017), pp. 5421–5436.

[17] K. GOLDBERG, T. ROEDER, D. GUPTA, AND C. PERKINS, *Eigentaste: A constant time collaborative filtering algorithm*, Information Retrieval, 4 (2001), pp. 133–151.

[18] R. HARTSHORNE AND A. HIRSCHOWITZ, *Droites en position générale dans l'espace projectif*, in Algebraic Geometry, Springer, 1982, pp. 169–188.

[19] P. JAIN, R. MEKA, AND I. S. DHILLON, *Guaranteed rank minimization via singular value projection*, in Advances in Neural Information Processing Systems (NIPS), Vol. 23, Curran Associates., Inc., 2010, pp. 937–945.

[20] K. H. JIN AND J. C. YE, *Annihilating filter-based low-rank Hankel matrix approach for image inpainting*, IEEE Trans. Image Process., 24 (2015), pp. 3498–3511.

[21] R. JIN, T. G. KOLDA, AND R. WARD, *Faster Johnson-Lindenstrauss Transforms via Kronecker Products*, preprint, https://arxiv.org/abs/1909.04801, 2019.

[22] R. KENNEDY, L. BALZANO, S. J. WRIGHT, AND C. J. TAYLOR, *Online algorithms for factorization-based structure from motion*, Comput. Vision Image Understanding, 150 (2016), pp. 139–152.

[23] R. H. KESHAVAN, A. MONTANARI, AND S. OH, *Matrix completion from a few entries*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2980–2998.

[24] C. LANE, R. BOGER, C. YOU, M. TSAKIRIS, B. HAEFFELE, AND R. VIDAL, *Classifying and comparing approaches to subspace clustering with missing data*, in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop, Seoul, South Korea, 2019, pp. 669–677.

[25] Y. LECUN, L. BOTTOU, Y. BENGIO, AND P. HAFFNER, *Gradient-based learning applied to document recognition*, Proc. IEEE, 86 (1998), pp. 2278–2324.

[26] J. LEE, S. KIM, G. LEBANON, AND Y. SINGER, *Local low-rank matrix approximation*, in Proceedings of the 30th International Conference on Machine Learning, PMLR, 2013, pp. 82–90.

[27] Y. MA, A. Y. YANG, H. DERKSEN, AND R. FOSSUM, *Estimation of subspace arrangements with applications in modeling and segmenting mixed data*, SIAM Rev., 50 (2008), pp. 413–458, https://doi.org/10.1137/060655523.

[28] K. Mohan and M. Fazel, *Iterative reweighted algorithms for matrix rank minimization*, J. Mach. Learn. Res., 13 (2012), pp. 3441–3473.

[29] G. Ongie, R. Willett, R. D. Nowak, and L. Balzano, *Algebraic variety models for high-rank matrix completion*, in Proceedings of the 34th International Conference on Machine Learning, PMLR, 2017, pp. 2691–2700.

[30] N. Pham and R. Pagh, *Fast and scalable polynomial kernels via explicit feature maps*, in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, 2013, pp. 239–247.

[31] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. D. Nowak, and R. Willett, *Group-sparse subspace clustering with missing data*, in Proceedings of the 2016 IEEE Statistical Signal Processing Workshop, Palma de Mallorca, 2016, pp. 1–5.

[32] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. D. Nowak, and R. Willett, *Mixture regression as subspace clustering*, in Proceedings of the 2017 International Conference on Sampling Theory and Applications, Tallinn, 2017, pp. 456–459.

[33] D. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, in Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, 2015, pp. 2191–2195.

[34] D. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *A characterization of deterministic sampling patterns for low-rank matrix completion*, IEEE J. Selected Topics Signal Process., 10 (2016), pp. 623–636.

[35] D. Pimentel-Alarcón and R. Nowak, *The information-theoretic requirements of subspace clustering with missing data*, in Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, JMLR, 2016, pp. 802–810.

[36] D. Pimentel-Alarcón and R. D. Nowak, *A converse to low-rank matrix completion*, in Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, 2016, pp. 96–100.

[37] D. Pimentel-Alarcón and R. D. Nowak, *Random consensus robust PCA*, in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR 54, 2017, pp. 344–352, http://proceedings.mlr.press/v54/pimentel-alarcon17a.html.

[38] D. Pimentel-Alarcón, R. D. Nowak, and L. Balzano, *On the sample complexity of subspace clustering with missing data*, in Proceedings of the IEEE Workshop on Statistical Signal Processing, Gold Coast, VIC, 2014, pp. 280–283.

[39] D. Pimentel-Alarcón, G. Ongie, L. Balzano, R. Willett, and R. D. Nowak, *Low algebraic dimension matrix completion*, in Proceedings of the 55th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, 2017, pp. 790–797.

[40] B. Recht, *A simpler approach to matrix completion*, J. Mach. Learn. Res., 12 (2011), pp. 3413–3430.

[41] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501, https://doi.org/10.1137/070697835.

[42] M. Soltanolkotabi, *Algorithms and Theory for Clustering and Nonconvex Quadratic Programming*, Ph.D. thesis, Stanford University, Stanford, CA, 2014.

[43] S. Sun, W. U. Bajwa, and A. P. Petropulu, *MIMO-MC radar: A MIMO radar approach based on matrix completion*, IEEE Trans. Aerospace Electron. Syst., 51 (2015), pp. 1839–1852.

[44] M. Tsakiris and R. Vidal, *Algebraic clustering of affine subspaces*, IEEE Trans. Pattern Anal. Mach. Intell., 40 (2018), pp. 482–489.

[45] M. Tsakiris and R. Vidal, *Theoretical analysis of sparse subspace clustering with missing entries*, in Proceedings of the 35th International Conference on Machine Learning (IMCL), Stockholm, JMLR, 2018, pp. 4975–4984.

[46] M. C. Tsakiris and R. Vidal, *Filtrated algebraic subspace clustering*, SIAM J. Imaging Sci., 10 (2017), pp. 372–415, https://doi.org/10.1137/16M1083451.

[47] R. Vidal, Y. Ma, and S. Sastry, *Generalized principal component analysis (GPCA)*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 1945–1959.

[48] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, Springer, 2016.

[49] T. Wu and W. U. Bajwa, *Learning the nonlinear geometry of high-dimensional data: Models and algorithms*, IEEE Trans. Signal Process., 63 (2015), pp. 6229–6244.

[50] A. Y. Yang, S. Rao, A. Wagner, Y. Ma, and R. M. Fossum, *Hilbert functions and applications to the estimation of subspace arrangements*, in Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05), Vol. 1, Beijing, 2005, pp. 158–165.

[51] C. Yang, D. Robinson, and R. Vidal, *Sparse subspace clustering with missing entries*, in Proceedings of the 32nd International Conference on Machine Learning (ICML), JMLR, 2015, pp. 2463–2472.