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 a 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 not in the original representation. We also provide a formal mathematical justification for the success of our method. In particular, we give bounds of 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

1. Introduction.

The past decade of research on matrix completion has shown it is possible to leverage linear relationships among columns (or rows) of a matrix to impute missing values. If each column of a matrix corresponds to a different high-dimensional data point belonging to a low-dimensional linear subspace, then the corresponding matrix is low-rank and missing values can be imputed using low-rank matrix completion [4, 5, 35, 36, 19]. These ideas continue to impact diverse applications such as recommender systems [22], image inpainting [17], computer vision [18], and array signal processing [38], among others.

The high-level idea of this body of work is that if the data defining the matrix belongs to a structure having fewer degrees of freedom than the entire dataset, that structure provides redundancy that can be leveraged to complete the matrix. However, the typical linear subspace 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

*Submitted to the editors May 26, 2020. This work is an extension of [34]

Funding: R.W., R.N. and G.O. were supported in part by AFOSR FA9550-18-1-0166, DOE DE-AC02-06CH11357, NSF OAC-1934637 and NSF DMS-1930049. L.B. was supported in part by ARO W911NF1910027, NSF CCF-1845076 and IIS-1838179, and the IAS Charles Simonyi Endowment.

†Department of Statistics, University of Chicago, Chicago, IL (gongie@uchicago.edu)

‡Department of Biostatistics, University of Wisconsin, Madison, WI (pimentelalar@wisc.edu)

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

¶Departments of Statistics and Computer Science, University of Chicago, Chicago, IL (willett@uchicago.edu)

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

polynomial generalization of classical linear subspaces. In this setting, we develop and analyze novel algorithms for imputing missing values under an algebraic variety model and derive new bounds on the amount of missing data that can be accurately imputed.

More precisely, let $\mathbf{X} \in \mathbb{R}^{\mathrm{d}\times\mathrm{N}}$ be a matrix whose columns lie in a low-dimensional algebraic variety $\mathcal{V} \subset \mathbb{R}^{\mathrm{d}}$. 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 subspace, this reduces to low-rank matrix completion (LRMC). We call the more general problem of completing LAD matrices *low algebraic dimension matrix completion* (LADMC).

Recently [25] proposed a new LADMC approach based on lifting the problem to a higher-dimensional representation (*e.g.,* tensor or polynomial expansions of the columns of $\mathbf{X}$). The algorithm in [25] 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:

- While [25] 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.
- The *unlifting* step is highly nonlinear and non-convex, and so little can be proved about its accuracy or correctness.

This paper addresses both issues. We provide sampling conditions in the original representation that guarantee uniqueness of the low-rank solution in the lifted (tensorized) representation. We also propose a new LADMC algorithm that uses a simple unlifting step based on the singular value decomposition (SVD), which is guaranteed to recover the original LAD matrix if the LRMC step succeeds. In contrast with [25], the LADMC algorithm proposed in this work can be implemented non-iteratively (besides the subroutine used to solve the LRMC problem). Experiments show that the new algorithm performs as well or better than state-of-the-art methods in the popular case of the union of subspaces model, and outperforms the algorithm proposed in [25] for the same task. We also propose an iterative version of the algorithm that alternates between solving LRMC in the tensorized representation and unlifting steps, which appears to yield additional empirical improvement.

1.1. Mathematical Contribution to LADMC. The main mathematical contribution of this paper is to generalize the deterministic sampling conditions for low-rank matrix completion and subspace clustering with missing entries [28, 29, 30, 31, 32] to the LADMC setting. In line with [28], we give conditions guaranteeing the column space of the tensorized data matrix is uniquely identifiable from its *canonical projections*, *i.e.,* projections of the subspace onto a collection of canonical basis elements. In particular, assuming a model where we observe exactly m entries per column of a data matrix whose columns belong to an algebraic variety, our results identify necessary and sufficient values of m for which unique identification of the column space of the tensorized matrix is information-theoretically *possible* (*i.e.,* provided there are sufficiently many data columns and the observation patterns are sufficiently diverse).

To achieve this result, one cannot simply apply known results for the reconstruction of linear subspaces from canonical projections such as [28]. The challenge here is that the observation patterns (*i.e.,* locations of the observed entries) in the original representation, when tensorized, generate only a small subset of all possible observation patterns in the tensor representation. Hence, the canonical projections that we may observe in the tensor representation are a restrictive subset of all possible canonical projections. Our main results show that under mild genericity assumptions on the underlying variety, the admissible canonical projections in the tensor space are sufficient to identify the subspace in the tensorized representation. Furthermore, we derive precise information theoretic bounds on how many missing entries can be tolerated in terms of the dimension of the subspace in the tensorized representation.

1.2. Related Work.

Unions of subspaces (UoS) are a special case of algebraic varieties [25, 42], and a number of approaches to matrix completion for a UoS model have been proposed [37, 1, 30, 2, 33, 46, 26, 44, 27, 11, 13]; see [20] for classification and comparison of these approaches for the task of subspace clustering with missing data. Most these algorithms involve iterating between subspace clustering and completion steps, and relatively little can be guaranteed about their performance. Exceptions include [1] and [13], which give performance guarantees for algorithms based on a non-iterative neighborhood search procedure. Also, recent work [8, 40] gives performance guarantees for a version of the sparse subspace clustering algorithm modified to handle missing data [46].

Our proposed LADMC approach is closely related to algebraic subspace clustering (ASC), also known as generalized principal component analysis [42, 45, 23, 39, 41]. 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 [34]. We extend the theory and algorithms in [34] to higher order tensorizations of the data matrix; [34] only considered quadratic tensorizations. We also correct several issues with the theory in [34]. In particular, parts (i) and (ii) of Theorem 2 of [34] are incorrect as stated; in this work we correct this result and also extend it to higher order tensorizations (see Corollary 1). Additionally, the proof of part (iii) of Theorem 2 in [34] 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 VMC algorithm [25], 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

$$\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_\mathrm{N}] \in \mathbb{R}^{\mathrm{d}\times\mathrm{N}}$$

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

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

$$\mathbf{X}^{\otimes\mathrm{p}} := [\mathbf{x}_1^{\otimes\mathrm{p}}, ..., \mathbf{x}_\mathrm{N}^{\otimes\mathrm{p}}] \in \mathbb{R}^{\mathrm{D}\times\mathrm{N}}.$$

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

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

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

If the LRMC step recovers $\mathbf{X}^{\otimes \mathrm{p}}$ correctly, then we can uniquely recover $\mathbf{X}$ from $\mathbf{X}^{\otimes \mathrm{p}}$. To see this, first consider the case of a quadratic tensorization ($\mathrm{p} = 2$). Let $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_{\mathrm{N}}]$ be the output from LRMC applied to the tensorized matrix. If the completion is correct, then $\mathbf{y}_i = \mathbf{x}_i^{\otimes 2}$, and we can reshape $\mathbf{y}_i$ into the rank-1 symmetric $\mathrm{d} \times \mathrm{d}$ matrix $\mathbf{Y}_i = \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}$. Hence, we can recover $\mathbf{x}_i$ by the computing the leading eigenvalue-eigenvector pair $(\lambda_i, \mathbf{u}_i)$ of $\mathbf{Y}_i$ and setting $\mathbf{x}_i = \pm\sqrt{\lambda_i}\mathbf{u}_i$, where we determine the sign by matching it to the observed entries of $\mathbf{x}_i$. For higher-order tensorizations ($\mathrm{p} \geq 3$), we can recover $\mathbf{x}_i$ from $\mathbf{y}_i$ using a similar procedure: we reshape $\mathbf{y}_i$ into a $\mathrm{d} \times \mathrm{d}^{\mathrm{p}-1}$ and take its rank-one truncated SVD.

These observations motivate our proposed algorithm, Low Algebraic Dimension Matrix Completion (LADMC), summarized below in Algorithm 2.1.

Algorithm 2.1 Low Algebraic Dimension Matrix Completion (LADMC).

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

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

LRMC: Let $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_{\mathrm{N}}]$ be the low-rank completion of $\mathbf{X}^{\otimes \mathrm{p}}$.

De-tensorize: Compute the best rank-one symmetric tensor approximation $\hat{\mathbf{x}}_i^{\otimes \mathrm{p}}$ of each column $\mathbf{y}_i$ of $\mathbf{Y}$ such that $\hat{\mathbf{x}}_i$ matches the observed entries of $\mathbf{x}_i$.

Output: Completed matrix $\hat{\mathbf{X}}$ whose i^{th} column is $\hat{\mathbf{x}}_{\mathrm{i}}$.

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 $\mathbf{X}^{\otimes \mathrm{p}}$ to be low-rank.

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

$q(\mathbf{x}) = \mathbf{v}^\mathsf{T}\mathbf{x}^{\otimes \mathrm{p}}$. Hence, we have shown the following:

Proposition 1. *The tensorized matrix* $\mathbf{X}^{\otimes \mathrm{p}} \in \mathbb{R}^{\mathrm{D}\times\mathrm{N}}$ *with* $\mathrm{N} \geq \mathrm{D}$ *is rank deficient if and only if the columns of* $\mathbf{X}$ *belong to an* algebraic variety, i.e., *the common zero set of a collection of polynomials.*

In particular, we focus on the class of varieties defined by *homogeneous polynomials*[1]. A degree-p homogeneous polynomial is any polynomial of the form $q(\mathbf{x}) = \mathbf{v}^\mathsf{T}\mathbf{x}^{\otimes \mathrm{p}}$, for some vector of coefficients $\mathbf{v} \in \mathbb{R}^{\mathrm{D}}$.

Definition 1. *A set* $\mathcal{V} \subset \mathbb{R}^{\mathrm{d}}$ *is a* (real) projective variety[2] *if there exists homogeneous polynomials* $q_1, ..., q_n$ *(with possibly different degrees) such that*

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

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

Example 1 (Unions of subspaces are projective varieties). *Suppose* $\mathcal{U}$ *and* $\mathcal{W}$ *are subspaces of* $\mathbb{R}^{\mathrm{d}}$. *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* $q_{i,j}(\mathbf{x}) = (\mathbf{x}^\mathsf{T}\mathbf{u}_i^\perp)(\mathbf{x}^\mathsf{T}\mathbf{w}_j^\perp)$, *where* $\{\mathbf{u}_i^\perp\}$ *is a basis of the orthogonal complement of* $\mathcal{U}$ *and* $\{\mathbf{w}_j^\perp\}$ *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 [23, 43].*

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* p^{th}*-order* tensorized subspace *associated with* $\mathcal{V}$ *by*

$$\mathcal{S} := \mathrm{span}\{\mathbf{x}^{\otimes \mathrm{p}} : \mathbf{x} \in \mathcal{V}\} \subset \mathbb{R}^{\mathrm{D}} \tag{2.1}$$

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

If the columns of a matrix $\mathbf{X}$ belong to a projective variety $\mathcal{V}$, then the column space of the tensorized matrix $\mathbf{X}^{\otimes \mathrm{p}}$ 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 $\mathbb{R}^{\mathrm{D}}$ of dimension at most $\mathrm{D} - \mathrm{L}$. Therefore, if there are sufficiently many such polynomials then $\mathcal{S}$ is a low-dimensional subspace, and hence $\mathbf{X}^{\otimes \mathrm{p}}$ is low-rank.

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

[2] For any homogeneous polynomial q we have $q(\mathbf{x}) = 0$ if and only if $q(\lambda\mathbf{x}) = 0$ 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 $\mathbb{R}^{\mathrm{d}}$, and this is the source of the term "projective variety". For simplicity, we typically consider projective varieties as subsets of Euclidean space $\mathbb{R}^{\mathrm{d}}$, unless otherwise noted.

For 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 $\min\{\mathrm{Kr}, \mathrm{d}\}$ in $\mathbb{R}^{\mathrm{d}}$. However, in the tensor space these points lie in a subspace of $\mathbb{R}^{\mathrm{D}}$ whose dimension relative to the tensor space dimension is potentially smaller, as shown in the following lemma.

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

$$\mathrm{R} \leq \min\left\{\mathrm{K}\binom{\mathrm{r}+\mathrm{p}-1}{\mathrm{p}}, \mathrm{D}\right\}. \tag{2.2}$$

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

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

In fact, Lemma 1 is a special case of a more general bound due to [7] that holds for any *equidimensional* projective variety[3]. Roughly speaking, a projective variety is *equidimensional* if the local dimension of the variety (treated as a smooth manifold) is everywhere the same. The bound in [7] is posed in terms of the *degree* and *dimension* of the variety (see, *e.g.,* [9] for definitions of these quantities). Translated to our setting, this result says if $\mathcal{V}$ is a 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 that LADMC will succeed with $O(\mathrm{K}^{1/\mathrm{p}}\mathrm{r})$ observations per column, where now K is the degree of the variety and r is its dimension. In other words, we expect LADMC will succeed in the case of data belonging to a projective variety with *high degree* and *low dimension*.

3. Theory.

3.1. Limitations of prior theory.

Algorithm 2.1 is primarily inspired by the ideas in [25]. In [25], 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 to the setting of this paper, the claim made in [25] is that in order to successfully complete a matrix $\mathbf{X}$ whose p^{th}-order tensorized matrix $\mathbf{X}^{\otimes \mathrm{p}}$ is rank R, we must observe at least m_0 entries per column of $\mathbf{X}$, where m_0 is the

[3] The result in [7] gives an upper bound on the values of the Hilbert function associated with any homogeneous unmixed radical ideal $I \subset k[x_0, ..., x_d]$ 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.

smallest integer such that

$$\binom{m_0+p-1}{p} > R, \tag{3.1}$$

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 that $R+1$ is the necessary minimum number of observations per column to uniquely complete a matrix whose columns belong to a R-dimensional subspace in general position [31]. Additionally, $R+1$ observations per column is sufficient for unique completion assuming there are sufficiently many data columns and the observation patterns are sufficiently diverse [29].

However, there are two key technical issues not considered in [25] that prevent the argument above 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 [25] 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 $\mathbf{B} \in \mathbb{R}^{D\times R}$ is any basis matrix for the subspace, then all $R \times R$ minors of $\mathbf{B}$ are non-vanishing). However, generally this property does not hold for tensorized subspaces arising from union of subspaces, even if the subspaces in the union are in general position (see Example 3 below). General position assumptions are essential to results that describe deterministic conditions on the observation patterns allowing for LRMC [28, 30]. 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 $\mathbf{X}^{\otimes p}$ 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 address in this work 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 $\mathbf{X}$ with the same observation pattern. Then we will observe the corresponding columns of the tensorized matrix $\mathbf{X}^{\otimes p}$ 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 $\boldsymbol{\omega} \in \{0,1\}^{\mathrm{d}}$ we let $|\boldsymbol{\omega}|$ denote the total number of ones in $\boldsymbol{\omega}$. We say the tensorized observation pattern $\boldsymbol{\upsilon} = \boldsymbol{\omega}^{\otimes \mathrm{p}}$ is of size m if $|\boldsymbol{\omega}| = \mathrm{m}$. Note that if $\boldsymbol{\upsilon}$ is a tensorized observation pattern of size m, then $\boldsymbol{\upsilon}$ has $\binom{\mathrm{m+p-1}}{\mathrm{p}}$ ones, *i.e.,* $|\boldsymbol{\upsilon}| = \binom{\mathrm{m+p-1}}{\mathrm{p}}$. For any observation pattern $\boldsymbol{\upsilon} \in \{0,1\}^{\mathrm{D}}$ and any vector $\mathbf{y} \in \mathbb{R}^{\mathrm{D}}$ let $\mathbf{y}_{\boldsymbol{\upsilon}} \in \mathbb{R}^{|\boldsymbol{\upsilon}|}$ denote the restriction of $\mathbf{y}$ to coordinates indicated by ones in $\boldsymbol{\upsilon}$. Likewise, for any subspace $\mathcal{S} \subset \mathbb{R}^{\mathrm{D}}$ we let $\mathcal{S}_{\boldsymbol{\upsilon}} \subset \mathbb{R}^{|\boldsymbol{\upsilon}|}$ denote the subspace obtained by restricting all vectors in $\mathcal{S}$ to coordinates indicated by ones in $\boldsymbol{\upsilon}$, and call $\mathcal{S}_{\boldsymbol{\upsilon}}$ the *canonical projection* of $\mathcal{S}$ onto $\boldsymbol{\upsilon}$. For any subspace $\mathcal{S} \subset \mathbb{R}^{\mathrm{D}}$ and any observation pattern matrix $\boldsymbol{\Upsilon} = [\boldsymbol{\upsilon}_1 \ \dots \ \boldsymbol{\upsilon}_{\mathrm{n}}] \subset \{0,1\}^{\mathrm{D}\times\mathrm{n}}$ we define $\mathscr{S}(\mathcal{S}, \boldsymbol{\Upsilon})$ 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 $\mathcal{S}'_{\boldsymbol{\upsilon}_{\mathrm{i}}} = \mathcal{S}_{\boldsymbol{\upsilon}_{\mathrm{i}}}$ for all $\mathrm{i} = 1, \dots, \mathrm{n}$. We say a subspace $\mathcal{S}$ is *uniquely identifiable* from its canonical projections in $\boldsymbol{\Upsilon}$ if $\mathscr{S}(\mathcal{S}, \boldsymbol{\Upsilon}) = \{\mathcal{S}\}$.

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

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

The intuition here is that the orthogonal complement of each $\mathcal{S}_{\boldsymbol{\upsilon}_i}$ constrains the set of subspaces consistent with the observed projections, and $\mathbf{A}$ 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* $\dim\ker \mathbf{A}^{\mathsf{T}} = \mathrm{R}$*, in which case* $\mathcal{S} = \ker \mathbf{A}^{\mathsf{T}}$*.*

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

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* $\mathbf{A}$ *above and numerically check if the dimension of the null space of* $\mathbf{A}^\mathsf{T}$ *agrees with the subspace dimension* R*, if it is known. We will use this fact to explore the possibility of unique identifiability of tensorized subspaces arising from unions of subspaces 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 *union of subspaces* (UoS), *i.e.,* varieties of the form $\mathcal{V} = \cup_{k=1}^{\mathrm{K}} \mathcal{U}_k$, where each $\mathcal{U}_k \subset \mathbb{R}^{\mathrm{d}}$ 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*. More precisely, we say $\mathcal{V}$ is a *generic union of* K r*-dimensional subspaces in* $\mathbb{R}^{\mathrm{d}}$ if the collection of subspaces $(\mathcal{U}_1, ..., \mathcal{U}_{\mathrm{K}})$ making up the union belong to an (often unspecified) open dense subset of the product space $\mathbb{G}(\mathrm{r}, \mathbb{R}^{\mathrm{d}}) \times \cdots \times \mathbb{G}(\mathrm{r}, \mathbb{R}^{\mathrm{d}})$, where $\mathbb{G}(\mathrm{r}, \mathbb{R}^{\mathrm{d}})$ denotes the Grassmannian of all r-dimensional subspaces in $\mathbb{R}^{\mathrm{d}}$. In particular, if a result holds for a generic union of K r-dimensional subspaces in $\mathbb{R}^{\mathrm{d}}$, then it holds with probability 1 for a union of K random subspaces drawn independently from any absolutely continuous probability distribution on the Grassmannian $\mathbb{G}(\mathrm{r}, \mathbb{R}^{\mathrm{d}})$.

We will repeatedly make use of the following facts regarding generic UoS (see, *e.g.,* [45]):

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

Treated as a function of p, the quantity $\mathrm{R}(\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p})$ is called the *Hilbert function* of a generic UoS (also called a "generic subspace arrangement"), and is studied in [45, 6, 10].

3.4. Necessary conditions for unique identifiability of tensorized subspaces.

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

Lemma 3. *Let* $\mathcal{V} \subset \mathbb{R}^{\mathrm{d}}$ *be a projective variety whose* p^{th}*-order tensorized subspace* $\mathcal{S} \subset \mathbb{R}^{\mathrm{D}}$ *is* R*-dimensional. Suppose we observe canonical projections of* $\mathcal{S}$ *onto* n *unique tensorized observation patterns* $\boldsymbol{\Upsilon} = [\boldsymbol{v}_i, ..., \boldsymbol{v}_n] \subset \{0,1\}^{\mathrm{D}\times n}$*. For all* $i = 1, ..., n$ *define* $\mathrm{M}_i := |\boldsymbol{v}_i|$ *and* $\mathrm{R}_i' := \dim \mathcal{S}_{\boldsymbol{v}_i}$*. Then a necessary condition for* $\mathcal{S}$ *to be uniquely identifiable is*

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

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 $\binom{\mathrm{d}}{\mathrm{m}}$ tensorized observation patterns of size m, *i.e.,* each column $\boldsymbol{v}_i$ of $\boldsymbol{\Upsilon}$ has $\mathrm{M} = \binom{\mathrm{m}+\mathrm{p}-1}{\mathrm{p}}$ ones. From Lemma 2 we know that $\dim \mathcal{S}_{\boldsymbol{v}_i} = \mathrm{R}'$ where $\mathrm{R}' \leq \mathrm{R} \leq \mathrm{K}\binom{\mathrm{r}+\mathrm{p}-1}{\mathrm{p}}$ and where the value of R' is the same for all tensorized observation patterns $\boldsymbol{v}_i$ by genericity. This means the constraint matrix has a total of $\binom{\mathrm{d}}{\mathrm{m}}(\mathrm{M} - \mathrm{R}')$ columns, which gives the following necessary condition for unique identifiability of tensorized subspaces arising from generic UoS:

Corollary 1. *Let $\mathcal{V} \subset \mathbb{R}^{\mathrm{d}}$ be a generic union of* K r*-dimensional subspaces. Suppose its* p^{th}*-order tensorized subspace* $\mathcal{S} \subset \mathbb{R}^{\mathrm{D}}$ *is* R*-dimensional. Let* $\mathrm{R}' \leq \mathrm{R}$ *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*

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

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

Immediately from (3.3), we see that a simpler, but weaker, necessary condition for unique identifiability is $\mathrm{M} > \mathrm{R}'$, which is independent of the ambient dimension d. In fact, assuming $\mathrm{m} > \mathrm{p}$ and the ambient dimension d is sufficiently large, then the condition in (3.3) reduces to $\mathrm{M} > \mathrm{R}'$. To see this, observe that $\binom{\mathrm{d}}{\mathrm{m}} = O(\mathrm{d}^{\mathrm{m}})$ and $\mathrm{D} = \binom{\mathrm{d}+\mathrm{p}-1}{\mathrm{p}} = O(\mathrm{d}^{\mathrm{p}})$ and so $(\mathrm{D}-\mathrm{R})/\binom{\mathrm{d}}{\mathrm{m}} < 1$ for large enough d. Hence, in this case (3.3) reduces to $\mathrm{M} > \mathrm{R}'$. In the event that $\mathrm{R}' = \mathrm{R}$, this further reduces to the condition $\mathrm{m} \geq \mathrm{m}_0$ given in (3.1), the rate conjectured to be necessary and sufficient in [25].

However, the following two examples show that when some of the above assumptions are violated (*e.g.,* when $\mathrm{m} \leq \mathrm{p}$ or $\mathrm{R}' < \mathrm{R}$) 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 (*$\mathrm{K} = 2, \mathrm{r} = 1, \mathrm{p} = 2$*). Suppose we consider all tensorized observation patterns of size* $\mathrm{m} = 2$*. In this case we have* $\mathrm{M} = 3 > 2 = \mathrm{R}' = \mathrm{R}$*, which satisfies the condition (3.1). Yet, the necessary condition (3.3) is violated in all ambient dimensions* $\mathrm{d} \geq 3$ *since*

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

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 $\mathrm{m} = 3$*, it is easy to show the necessary condition (3.3) is always met in dimensions* $\mathrm{d} \geq 4$*, and experimentally we find that the sufficient condition* $\dim\ker \mathbf{A}^{\mathsf{T}} = \mathrm{R}$ *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 4-dimensional ambient space (*$\mathrm{K} = 2, \mathrm{r} = 2, \mathrm{p} = 2, \mathrm{d} = 4$*). Suppose we consider all observation patterns of size* $\mathrm{m} = 3$*. In this case* $\mathrm{M} = 6 = \mathrm{R}$*, which violates condition (3.1). However, we have* $\mathrm{R}' = 5$ *since the canonical projection of the tensorized subspace onto a tensorized observation pattern of size* $\mathrm{m} = 3$ *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:*

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

This shows that unique identificaiton of the tensorized subspace may still be possible in this case. In the supplementary materials we prove that the sufficient condition $\dim\ker \mathbf{A}^{\mathsf{T}} = \mathrm{R}$

of Lemma 2 holds in this case, which shows the tensorized subspace is uniquely identifiable. Therefore, condition (3.1) is not always necessary.

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.** $\dim\ker \mathbf{A}^\top = R$

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.** $\dim\ker \mathbf{A}^\top = R$

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

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 observations patterns of size m. The left-most table gives the smallest value of m satisfying condition (3.1) that was conjectured to be necessary and sufficient in [25]. The middle table reports the smallest value of m satisfying the necessary condition (3.3). The right-most table reports the smallest value of m satisfying the sufficient condition $\ker \mathbf{A}^\top = R$ given in Lemma 2, which is verified numerically by constructing the constraint matrix $\mathbf{A}$ from a randomly drawn UoS. The middle and right-most tables are the same, showing the necessary condition (3.3) is also sufficient in these cases. In the left-most tables, 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.

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 a generic UoS in ambient dimension $d = 12$ for a varying number of subspaces and their dimension, computed their tensorized subspace, and constructed the constraint matrix $\mathbf{A}$ 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 $\dim\ker \mathbf{A}^\top = R$ given in Lemma 2 holds[4]. We compare this with the

[4]If the condition $\ker \mathbf{A}^\top = R$ holds for one random realization of a union of K r-dimensional subspaces, then it holds generically since the condition $\ker \mathbf{A}^\top = R$ can be recast as a polynomial system of equations in

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 the number of subspaces and their dimension in many cases.

3.5. Sufficient conditions for unique identifiability of tensorized subspaces. This section presents a sufficient condition for unique identifiability of the tensorized subspace. First, we state a result that holds for general projective varieties and then specialize to the case of UoS.

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

We give the proof of Theorem 1 in Appendix A. Roughly speaking, Theorem 1 says that a sampling rate of $\mathrm{m} \geq |\boldsymbol{\omega}| + \mathrm{p}$ (*i.e.,* m observed entries per data column of the original matrix) is sufficient to ensure 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 $\mathcal{V} \subset \mathbb{R}^{\mathrm{d}}$. Recall that we define $\mathrm{R}(\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p}) = \dim \mathcal{S}$, *i.e.,* the dimension of the tensorized subspace depends only on $\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p}$ (see Proposition 2). Now, given any tensorized observation pattern $\boldsymbol{\upsilon} = \boldsymbol{\omega}^{\otimes \mathrm{p}}$ of size m^*, observe that $\mathcal{S}_{\boldsymbol{\upsilon}}$ is equal to the tensorized subspace arising from $\mathcal{V}_{\boldsymbol{\omega}} \subset \mathbb{R}^{\mathrm{m}^*}$, the UoS restricted to the m^* coordinates specified by $\boldsymbol{\omega}$. Provided $\mathrm{m}^* > \mathrm{r}$, $\mathcal{V}_{\boldsymbol{\omega}}$ is again a generic union of K r-dimensional subspaces except now in m^*-dimensional ambient space. Hence, $\mathcal{S}_{\boldsymbol{\upsilon}}$ has the same dimension as the tensorized subspace arising from a generic UoS in m^*-dimensional ambient space, and so we have $\dim \mathcal{S}_{\boldsymbol{\upsilon}} = \mathrm{R}(\mathrm{m}^*, \mathrm{K}, \mathrm{r}, \mathrm{p})$ *for any* tensorized observation pattern $\boldsymbol{\upsilon}$ of size $\mathrm{m}^* > \mathrm{r}$. This fact combined with Theorem 1 gives the following immediate corollary.

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

The key assumption made in Corollary 2 is that $\mathrm{R}(\mathrm{m}^*, \mathrm{K}, \mathrm{r}, \mathrm{p}) = \mathrm{R}(\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p})$. Characterizing the set of values for which this condition holds in all generality appears to be a difficult problem (see, e.g., [6]). However, using existing results [15, 10, 6] that characterize exact values of $\mathrm{R}(\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p})$ we can establish the following special cases:

Proposition 3. *Let m_0 be the smallest integer such that $\binom{\mathrm{m}_0+\mathrm{p}-1}{\mathrm{p}} > \mathrm{K}\binom{\mathrm{r}+\mathrm{p}-1}{\mathrm{p}}$ and set $\mathrm{m}^* = \max\{\mathrm{m}_0, 2\mathrm{r}\}$. Then $\mathrm{R}(\mathrm{m}^*, \mathrm{K}, \mathrm{r}, \mathrm{p}) = \mathrm{K}\binom{\mathrm{r}+\mathrm{p}-1}{\mathrm{p}} = \mathrm{R}(\mathrm{d}, \mathrm{K}, \mathrm{r}, \mathrm{p})$ in the following cases:*

*(a) $\mathrm{p} = 2$, for any K, for any r (*i.e.*, any generic UoS under a quadratic tensorization)*

---

terms of the entries of a collection of basis matrices for each subspace in the UoS.

(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* p^{th}*-order tensorization)*

Case (a) is due to [6, Theorem 3.2], case (b) is due to [15], and case (c) is due to [10, Corollary 4.8] (see also [23, Corollary 2.16]).

The quantity m_0 defined in Proposition 3 is $O(K^{1/p}r)$, 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 the 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 [25]. 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 the value of m^* as defined in Proposition 3 is always sufficient to ensure $R(m^*, K, r, p) = R(d, K, r, p)$ 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 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 $\mathcal{S}$ from canonical projections. However, unique identification of the tensorized subspace also implies unique completion of the tensorized matrix $\mathbf{X}^{\otimes p}$ provided the observation pattern matrix $\mathbf{\Omega}$ 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 $\mathbf{\Omega}$ 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 $\mathbf{\Omega}$ 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 $N = O(R d^m \log d)$. If every subset of R columns of $\mathbf{X}$ is in general position, it will be possible to determine the canonical projections of $\mathcal{S}$ from the columns of $\mathbf{X}^{\otimes p}$. Once $\mathcal{S}$ is recovered, then it is possible to uniquely complete the matrix by projecting each incomplete column of $\mathbf{X}^{\otimes p}$ onto $\mathcal{S}$, and performing the "de-tensorization" step of LADMC (Step 4 of Algorithm 1).

While the above argument establishes a sufficient number of columns to uniquely complete a 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 [29] would lead one to believe that $\mathrm{N} \geq (\mathrm{R}+1)(\mathrm{D}-\mathrm{R})$ columns are necessary and sufficient for unique recovery of $\mathcal{S}$, which is far less than the estimate given in the previous paragraph. However, the tensorization process violates many of the genericity assumptions in [29], 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 $\mathrm{N} = O(\mathrm{R}(\mathrm{D}-\mathrm{R}))$ columns, and we conjecture this is the true necessary and sufficient orderwise number of columns needed for recovery with LADMC (see Figure 2 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 $\mathrm{m} = O(\mathrm{K}^{1/\mathrm{p}}\mathrm{r} + \mathrm{p})$ observations per data column are sufficient for LADMC to succeed given sufficiently many data columns (under some restrictions on p, K and r). In contrast, the information-theoretic requirements for subspace clustering with missing data (SCMD), which is mathematically equivalent to matrix completion under a union of subspaces (UoS) model, to succeed is $\mathrm{m} = \mathrm{r}+1$ observations per data column [30]. If $\mathrm{p} = O(1)$, *i.e.,* the tensor order is fixed and not allowed to scale with the number of subspaces, this shows that the necessary sample complexity of LADMC is order-wise suboptimal by a factor of $\mathrm{K}^{1/\mathrm{p}}$. However, if the tensor order p scales with the number of subspaces K as $\mathrm{p} = O(\log \mathrm{K})$ then we have $\mathrm{m} = O(\mathrm{r} + \log \mathrm{K})$, which is nearly orderwise optimal. Nonetheless, even with fixed and low tensor orders (e.g., $\mathrm{p} = 2, 3$), empirically we find that LADMC performs equally well or better than most state-of-the-art SCMD methods on UoS data (see Figure 1).

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 empirically verify the information-theoretic sampling requirements for LADMC for unions of subspaces data given in Section 3.

4.1. Implementation details: LADMC, iLADMC, and VMC.

In our implementation of LADMC (Algorithm 2.1) we use iterative singular value hard thresholding (ISVHT) algorithm [16] to perform LRMC in the tensorized domain. The “de-tensorization” 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 ISVHT in the tensorized domain for a small number of iterations, map back to the original domain by the rank-1 SVD de-tensorization step, fill in the observed entries of the matrix, and repeat until convergence. In the experiments below we ran 30 inner iterations ISVHT for iLADMC. Periodically performing the de-tensorization 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 $\mathbf{x}^{\otimes \mathrm{p}}$. While we have no theory to show 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 [25] 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 minimizes the non-convex Schatten-q quasi-norm ($0 < q < 1$) of the tensorized matrix $\mathbf{X}^{\otimes p}$ using an iterative reweighted least squares approach [24]. The VMC algorithm is most similar to iLADMC, since it also enforces the tensorized structure at each iteration.

4.2. Sample complexity of union of subspaces data. Figure 1 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) [16] in the original matrix domain, and three methods based on subspace clustering: sparse subspace clustering (SSC) with entry-wise zero fill [46] followed by LRMC on each identified cluster (SSC+EWZF), the expectation-maximization (EM) algorithm proposed in [33], and the group-sparse subspace clustering algorithm [26] followed by LRMC on each cluster (GSSC). The subspace clustering algorithms were passed the exact rank and number of subspaces. 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.

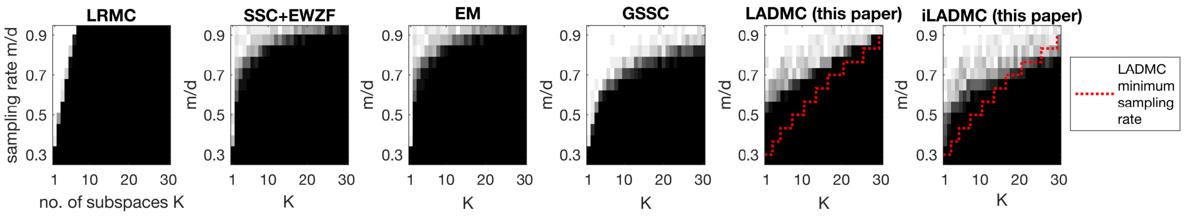

Figure 1: Phase transitions for matrix completion of synthetic union of subspaces data. We simulated data belonging to K subspaces and sampled each column of the data matrix at a rate m/d, and perform matrix completion using LRMC, state-of-the-art subspace clustering based algorithms (SSC+EWZF, GSSC, EM), and the proposed LADMC and iLADMC algorithms with quadratic tensorizations. Grayscale values indicate the fraction of random trials where the matrix were 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

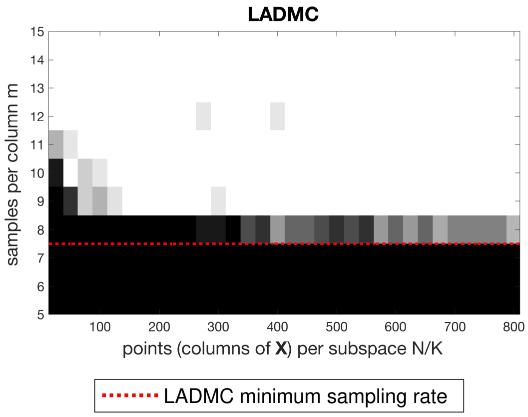

Figure 2: Effect of number of data columns per subspace N/K on probability of exact LADMC recovery of synthetic UoS data (K = 10 subspaces of dimension r = 2 in d = 15 dimensional space). As the number of columns per subspace increases the probability of exact recovery is approaching 1 for m ≥ 8, the necessary minimum number of samples per column identified by Corollary 1.

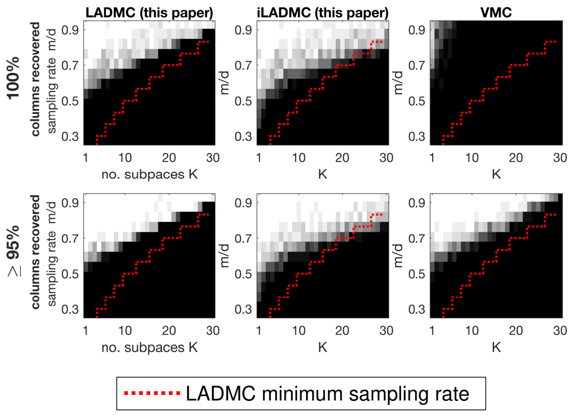

Figure 3: Relative performance of LADMC and iLADMC compared with VMC [25] for recovery of synthetic UoS data. LADMC and iLADMC succeed with high probability at recovering all the columns where VMC often fails (top row). The algorithms perform similarly when comparing the probability of recovering at least 95% columns (bottom row).

established in Corollary 1, as indicated by the red dashed line in Figure 1. 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 less 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 sampling rate is low (lower-left of the plot), but this gap closes as the number of subspaces and sampling rate increases (upper-right of the plot). We hypothesize this is due to insufficiently many data columns (see Figure 2 and the discussion below). This gap is less pronounced for iLADMC, and in fact, in the upper-right of the plot iLADMC shows recovery below the LADMC information-theoretic bound. We conjecture this is because iLADMC is enforcing extra nonlinear constraints that are not accounted for in our theory, which may reduce the sample complexity relative to non-iterative LADMC, both in terms of necessary number of data columns and the number of samples per column. We also observe that the performance of LADMC and iLADMC is competitive with the best performing subspace clustering-based algorithm, which in this case is GSSC.

In Figure 2 we investigate the effect of the number of data columns per subspace in 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. The minimum number columns per subspace needed for exact recovery we conjecture to be $N/K = O(R(D-R)/K)$ (see Section 3.6.1). Assuming the constant in the order-wise expression to be 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 two- or three-fold more columns, as predicted.

4.3. Comparison with VMC. In Figure 3 we compare the relative performance of LADMC and iLADMC with VMC for the same synthetic unions of subspaces data as in Figure 1. One drawback of VMC observed in [25] 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 [25], we find that VMC and LADMC/iLADMC perform similarly when comparing probability of recovering at least 95% columns. However, LADMC and iLADMC both recover 100% of the data columns correctly above the minimum sampling rate, whereas VMC mostly fails under this more strict 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 4 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 de-tensorization step, which prohibited us from investigating the phase transition behavior of LADMC over a reasonable range of the number of subspace K. Instead, we verify our predictions using VMC algorithm [25], 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.

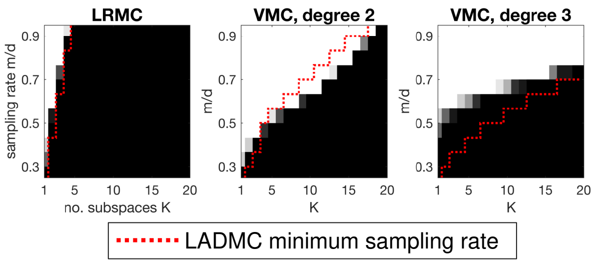

Figure 4: Phase transitions for matrix completion for unions of subspaces using no tensorization (LRMC), 2nd order tensorization (VMC, degree 2), and 3rd order tensorizaiton (VMC, degree 3). The phase transition follows closely the LADMC minimum sampling rate established in Corollary 1, which is $m = O(K^{1/p})$ 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 [25]).

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**

Table 2: Matrix completion results on real data

4.5. Experiments on real data. Here we illustrate the performance of LADMC and iLADMC on three real world datasets[5]: the Oil Flow dataset introduced in [3], the Jester-1 recommender systems dataset [14], and the MNIST digit recognition dataset introduced in [21]. We chose these datasets to demonstrate the feasibility of LADMC on a variety of data sources, and because they had sufficiently small row dimension for LADMC/iLADMC to be computationally practical. For the Oil Flow and MNIST datasets we simulate missing data by randomly subsampling each data column uniformly at random, using a 50%-25%-25% training-validation-test split of the data. For the Jester-1 dataset we used 18 randomly selected ratings of each user for training, 9 randomly selected ratings for validation and the remainder for testing. As baselines we compare with filling the missing entries with the mean of the observed entries within each column (Mean-fill), and with LRMC via nuclear norm minimization [36], which outperformed LRMC via singular value iterative hard thresholding [16] 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 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 gives 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 5 illustrates the improvement of LADMC over LRMC on a selection of examples from the MNIST dataset. We see less differences between LRMC and LADMC/iLADMC on the Jester-1 dataset, where LADMC/iLADMC give nearly the same RMSE as LRMC. Because of 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 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

[5]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.

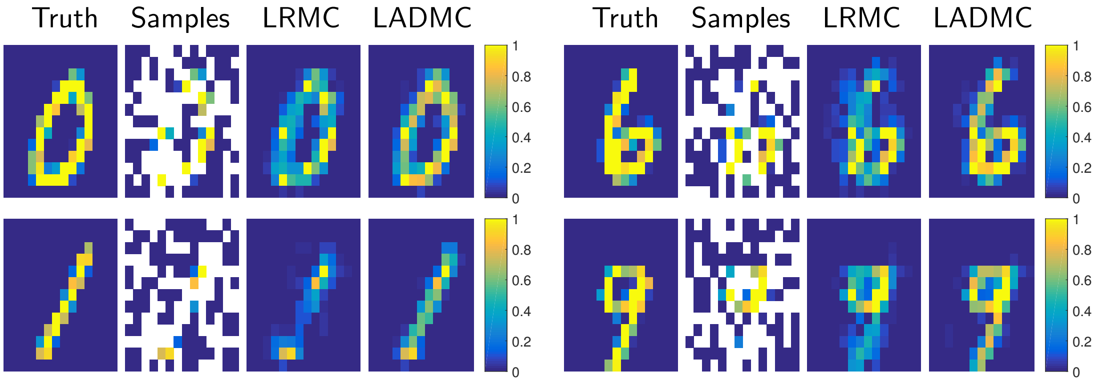

Figure 5: Representative examples of matrix completion on MNIST dataset. Here we randomly remove 50% of the pixels in each MNIST image and attempt to jointly recover the missing pixels of all images by low-rank matrix completion (LRMC) and low algebraic dimension matrix completion (LADMC) using a quadratic tensorization ($p = 2$).

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 are 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 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 computational challenging for even modest data dimensions d.

One potential solution is to use a kernelized algorithm like in [25] that avoids the construction of the 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 big datasets with many datapoints. 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 [12].

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* $\mathbf{A} = \mathbf{A}(\mathcal{S}, \boldsymbol{\Upsilon}^*) \in \mathbb{R}^{D\times(D-R)}$ satisfies Lemma 2, *i.e.,* $\dim\ker \mathbf{A}^\mathsf{T} = R$.

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 $\mathcal{S}_{\boldsymbol{\upsilon}}$ with $|\boldsymbol{\upsilon}| = \mathrm{M}$ then we also have access to all canonical projections $\mathcal{S}_{\boldsymbol{\upsilon}'}$ where $\boldsymbol{\upsilon}'$ is any observation pattern with $\mathrm{supp}(\boldsymbol{\upsilon}') \subset \mathrm{supp}(\boldsymbol{\upsilon})$ and $|\boldsymbol{\upsilon}'| = \mathrm{R}+1$.

To express this fact more succinctly, we introduce some additional notation. For any observation pattern matrix $\boldsymbol{\Upsilon} \in \{0,1\}^{\mathrm{D}\times \mathrm{n}}$ whose columns all have greater than R nonzeros, let $\widehat{\boldsymbol{\Upsilon}}$ denote the matrix of observation patterns having exactly $\mathrm{R}+1$ non-zeros that can be generated from the observation patterns in $\boldsymbol{\Upsilon}$. For example, if $\mathrm{D}=4$, $\mathrm{R}=1$, 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 $(\mathrm{k}_1, ..., \mathrm{k}_\mathrm{p})$ satisfying $1 \le \mathrm{k}_1 \le \cdots \le \mathrm{k}_\mathrm{p} \le \mathrm{d}$, 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 $\mathrm{m} = 1, ..., \mathrm{d}$, the first $\mathrm{M} = \binom{\mathrm{m}+\mathrm{p}-1}{\mathrm{p}}$ coordinates correspond to all tuples $1 \le \mathrm{k}_1 \le ... \le \mathrm{k}_d \le m$. We call this the *standard ordering.*

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

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

$$\boldsymbol{\Upsilon}^\star = \begin{bmatrix} \mathbf{1} \\ \hline \mathbf{I} \end{bmatrix} \begin{matrix} \}\,\mathrm{R} \\ \Big\}\,\mathrm{D}-\mathrm{R}. \end{matrix} \tag{A.1}$$

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

Proof. Let $\boldsymbol{\upsilon}^*_\mathrm{j}$ denote the jth column of $\boldsymbol{\Upsilon}^\star$ in (A.1), whose first R entries are ones and $(\mathrm{R}+\mathrm{j})^{\mathrm{th}}$ entry is one while the rest of its entries are zero. We show that $\boldsymbol{\upsilon}^*_\mathrm{j}$ is a column $\widehat{\boldsymbol{\upsilon}}$ in the expanded matrix $\widehat{\boldsymbol{\Upsilon}}$. Let $(\mathrm{k}_1, ..., \mathrm{k}_\mathrm{p})$ be the ordered tuple corresponding to the $(\mathrm{R}+\mathrm{j})^{\mathrm{th}}$ coordinate in the tensor space. Note that a column $\boldsymbol{\upsilon}$ in $\boldsymbol{\Upsilon}$ has nonzero $(\mathrm{R}+\mathrm{j})^{\mathrm{th}}$ entry if and only if the corresponding observation pattern $\boldsymbol{\omega}$ is nonzero in entries $\mathrm{k}_1, \mathrm{k}_2, ..., \mathrm{k}_\mathrm{p}$. Let $\boldsymbol{\omega}$ be any column of $\boldsymbol{\Omega}$ such that all entries indexed by $\{1, ..., \mathrm{m}^*, \mathrm{k}_1, ..., \mathrm{k}_\mathrm{p}\}$ are equal to one. By construction $\boldsymbol{\upsilon} = \boldsymbol{\omega}^{\otimes \mathrm{p}}$ has ones in its first M^* entries and must also have a one at the $(\mathrm{R}+\mathrm{j})$th entry. This shows that $\boldsymbol{\upsilon}^*_\mathrm{j}$ is a column of the expanded matrix $\widehat{\boldsymbol{\Upsilon}}$, and thus of any permutation of the first M^* rows of $\widehat{\boldsymbol{\Upsilon}}$. Since this is true for every $\mathrm{j} = 1, \ldots, \mathrm{D}-\mathrm{R}$, we know that $\boldsymbol{\Upsilon}$ will produce a matrix $\widehat{\boldsymbol{\Upsilon}}$ containing $\boldsymbol{\Upsilon}^\star$ as in (A.1) (and likewise for any permutation of the first M^* rows of $\widehat{\boldsymbol{\Upsilon}}$). ■

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 an tensorized observation pattern $\boldsymbol{\upsilon} = \boldsymbol{\omega}^{\otimes \mathrm{p}}$ such that $\dim \mathcal{S}_{\boldsymbol{\upsilon}} = \mathrm{R}$. Define $\mathrm{m}^* = |\boldsymbol{\omega}|$. 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 $\mathrm{M}^* = \binom{\mathrm{m}^*+\mathrm{p}-1}{\mathrm{p}}$ entries of $\boldsymbol{\upsilon} = \boldsymbol{\omega}^{\otimes \mathrm{p}}$ are ones. Since $\dim \mathcal{S}_{\boldsymbol{\upsilon}} = \mathrm{R}$, there exists an observation pattern $\boldsymbol{\upsilon}'$ with $\mathrm{supp}(\boldsymbol{\upsilon}') \subset \mathrm{supp}(\boldsymbol{\upsilon})$ having exactly R ones such that $\dim \mathcal{S}_{\boldsymbol{\upsilon}'} = \mathrm{R}$. We may permute the first M^* coordinates in tensor space so that $\boldsymbol{\upsilon}'$ 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.,* $\dim \mathcal{S}_{\boldsymbol{\upsilon}'} = \mathrm{R}$).

Now suppose we observe canonical projections of $\mathcal{S}$ onto all tensorized observation patterns of size $\mathrm{m} \geq \mathrm{m}^* + \mathrm{p}$, which we collect into a matrix $\boldsymbol{\Upsilon} \in \{0,1\}^{\mathrm{D}\times\binom{\mathrm{d}}{\mathrm{m}}}$. Then by Lemma 4 there exists a submatrix $\boldsymbol{\Upsilon}^* \in \{0,1\}^{\mathrm{D}\times(\mathrm{D}-\mathrm{R})}$ 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 $j = 1, \ldots, \mathrm{D}-\mathrm{R}$, let $\boldsymbol{\upsilon}_j^*$ be the jth column of $\boldsymbol{\Upsilon}^*$. Since $\boldsymbol{\upsilon}_j^*$ has exactly $\mathrm{R}+1$ ones, and the restricted subspace $\mathcal{S}_{\boldsymbol{\upsilon}_j^*}$ is at most R dimensional, the orthogonal complement of the restricted subspace $\mathcal{S}_{\boldsymbol{\upsilon}_j^*}$ is positive dimensional, and so there exists at least one non-zero constraint vector $\mathbf{a}_j \in (\mathcal{S}_{\boldsymbol{\upsilon}_j^*})^\perp$. Let $\mathbf{a}_j^* \in \mathbb{R}^d$ be the vector whose restriction to $\boldsymbol{\upsilon}_j$ is equal to $\mathbf{a}_j$ and zero in its other entries. Then consider the constraint matrix $\mathbf{A}^* = [\mathbf{a}_1^*, ..., \mathbf{a}_{\mathrm{D}-\mathrm{R}}^*]$, which has the same dimensions as $\boldsymbol{\Upsilon}^*$ and is such that an entry of $\mathbf{A}^*$ is nonzero only if the corresponding entry of $\boldsymbol{\Upsilon}^*$ is nonzero. In particular, this means that

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_0^* \\ \mathbf{A}_1^* \end{bmatrix} \in \mathbb{R}^{\mathrm{D}\times(\mathrm{D}-\mathrm{R})} \tag{A.2}$$

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

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

REFERENCES

[1] L. Balzano, B. Eriksson, and R. Nowak, *High rank matrix completion and subspace clustering with missing data*, in Artificial Intelligence and Statistics, 2012.

[2] L. Balzano, A. Szlam, B. Recht, and R. Nowak, *K-subspaces with missing data*, in IEEE Statistical Signal Processing Workshop, 2012, pp. 612–615.

[3] 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.

[4] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), p. 717.

[5] E. J. Candès and T. Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transactions on Information Theory, 56 (2010), pp. 2053–2080.

[6] E. Carlini, M. V. Catalisano, and A. V. Geramita, *Subspace arrangements, configurations of linear spaces and the quadrics containing them*, Journal of Algebra, 362 (2012), pp. 70–83.

[7] M. Chardin, *Une majoration de la fonction de Hilbert et ses conséquences pour l'interpolation algébrique*, Bulletin de la Société Mathématique de France, 117 (1989), pp. 305–318.

[8] Z. Charles, A. Jalali, and R. Willett, *Subspace clustering with missing and corrupted data*, in 2018 IEEE Data Science Workshop, 2017. arXiv preprint arXiv:1707.02461.

[9] D. Cox, J. Little, and D. O'shea, *Ideals, varieties, and algorithms*, vol. 3, Springer, 2007.

[10] H. Derksen, *Hilbert series of subspace arrangements*, Journal of Pure and Applied Algebra, 209 (2007), pp. 91–98.

[11] E. Elhamifar, *High-rank matrix completion and clustering under self-expressive models*, in Neural Information Processing Systems (NeurIPS), 2016, pp. 73–81.

[12] J. Fan and M. Udell, *Online high rank matrix completion*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8690–8698.

[13] 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 Transactions on Signal Processing, 65 (2017), pp. 5421–5436.

[14] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, *Eigentaste: A constant time collaborative filtering algorithm*, Information Retrieval, 4 (2001), pp. 133–151.

[15] R. Hartshorne and A. Hirschowitz, *Droites en position générale dans l'espace projectif*, in Algebraic geometry, Springer, 1982, pp. 169–188.

[16] P. Jain, R. Meka, and I. S. Dhillon, *Guaranteed rank minimization via singular value projection*, in Neural Information Processing Systems (NeurIPS), 2010, pp. 937–945.

[17] K. H. Jin and J. C. Ye, *Annihilating filter-based low-rank hankel matrix approach for image inpainting*, IEEE Transactions on Image Processing, 24 (2015), pp. 3498–3511.

[18] R. Kennedy, L. Balzano, S. J. Wright, and C. J. Taylor, *Online algorithms for factorization-based structure from motion*, Computer Vision and Image Understanding, 150 (2016), pp. 139–152.

[19] R. H. Keshavan, A. Montanari, and S. Oh, *Matrix completion from a few entries*, IEEE Transactions on Information Theory, 56 (2010), pp. 2980–2998.

[20] 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 International Conference on Computer Vision Workshops, 2019.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, 86 (1998), pp. 2278–2324.

[22] J. Lee, S. Kim, G. Lebanon, and Y. Singer, *Local low-rank matrix approximation*, in International Conference on Machine Learning, 2013, pp. 82–90.

[23] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum, *Estimation of subspace arrangements with applications in modeling and segmenting mixed data*, SIAM review, 50 (2008), pp. 413–458.

[24] K. Mohan and M. Fazel, *Iterative reweighted algorithms for matrix rank minimization*, The Journal of Machine Learning Research, 13 (2012), pp. 3441–3473.

[25] G. Ongie, R. Willett, R. D. Nowak, and L. Balzano, *Algebraic variety models for high-rank matrix completion*, in International Conference on Machine Learning (ICML), 2017, pp. 2691–2700.

[26] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. D. Nowak, and R. Willett, *Group-sparse subspace clustering with missing data*, in IEEE Statistical Signal Processing Workshop, 2016, pp. 1–5.
[27] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. D. Nowak, and R. Willett, *Mixture regression as subspace clustering*, in Sampling Theory and Applications, 2017, pp. 456–459.
[28] D. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, in IEEE International Symposium on Information Theory (ISIT), 2015, pp. 2191–2195.
[29] D. Pimentel-Alarcón, N. Boston, and R. D. Nowak, *A characterization of deterministic sampling patterns for low-rank matrix completion*, IEEE Journal of Selected Topics in Signal Processing, 10 (2016), pp. 623–636.
[30] D. Pimentel-Alarcón and R. Nowak, *The information-theoretic requirements of subspace clustering with missing data*, in International Conference on Machine Learning (ICML), 2016, pp. 802–810.
[31] D. Pimentel-Alarcón and R. D. Nowak, *A converse to low-rank matrix completion*, in IEEE International Symposium on Information Theory (ISIT), 2016, pp. 96–100.
[32] D. Pimentel-Alarcón and R. D. Nowak, *Random consensus robust PCA*, in Artificial Intelligence and Statistics, 2017, pp. 344–352.
[33] D. Pimentel-Alarcón, R. D. Nowak, and L. Balzano, *On the sample complexity of subspace clustering with missing data*, in IEEE Statistical Signal Processing Workshop, 2014, pp. 280–283.
[34] D. Pimentel-Alarcón, G. Ongie, R. Willett, R. D. Nowak, and L. Balzano, *Low algebraic dimension matrix completion*, in Allerton Conference on Communication, Control, and Computing, 2017.
[35] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 12 (2011), pp. 3413–3430.
[36] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM review, 52 (2010), pp. 471–501.
[37] M. Soltanolkotabi, *Algorithms and theory for clustering and nonconvex quadratic programming*, PhD thesis, Stanford University, Stanford, CA, USA., 2014.
[38] S. Sun, W. U. Bajwa, and A. P. Petropulu, *MIMO-MC radar: A MIMO radar approach based on matrix completion*, IEEE Transactions on Aerospace and Electronic Systems, 51 (2015), pp. 1839–1852.
[39] M. Tsakiris and R. Vidal, *Algebraic clustering of affine subspaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2017).
[40] M. Tsakiris and R. Vidal, *Theoretical analysis of sparse subspace clustering with missing entries*, in International Conference on Machine Learning (IMCL), 2018.
[41] M. C. Tsakiris and R. Vidal, *Filtrated algebraic subspace clustering*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 372–415.
[42] R. Vidal, Y. Ma, and S. Sastry, *Generalized principal component analysis (GPCA)*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (2005), pp. 1945–1959.
[43] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, Springer, 2016.
[44] T. Wu and W. U. Bajwa, *Learning the nonlinear geometry of high-dimensional data: Models and algorithms*, IEEE Transactions on Signal Processing, 63 (2015), pp. 6229–6244.
[45] A. Y. Yang, S. Rao, A. Wagner, Y. Ma, and R. M. Fossum, *Hilbert functions and applications to the estimation of subspace arrangements*, in IEEE International Conference on Computer Vision (ICCV), vol. 1, 2005, pp. 158–165.
[46] C. Yang, D. Robinson, and R. Vidal, *Sparse subspace clustering with missing entries*, in International Conference on Machine Learning (ICML), 2015, pp. 2463–2472.