LOW ALGEBRAIC DIMENSION MATRIX COMPLETION

Daniel Pimentel-Alarcón

Department of Computer Science
Georgia State University
Atlanta, GA, USA

Gregory Ongie and Laura Balzano

Department of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, MI, USA

Rebecca Willett and Robert Nowak

Department of Electrical
and Computer Engineering
University of Wisconsin
Madison, WI, USA

Abstract—Low rank matrix completion (LRMC) has received tremendous attention in recent years. The low rank assumption means that the columns (or rows) of the matrix to be completed are points on a low-dimensional linear variety. This paper extends this thinking to cases where the columns are points on low-dimensional *nonlinear* algebraic varieties. While others have recently studied matrix completion in such settings, existing results focus mainly on algorithms and experiments, without supporting theory. This paper proposes a new approach to what we call *Low Algebraic-Dimension Matrix Completion* (LADMC). We propose a new LADMC algorithm that leverages existing LRMC methods on a tensorized representation of the data. We also provide a formal mathematical justification for the success of our method. In particular, the new algorithm can succeed in many cases where traditional LRMC is guaranteed to fail. We also provide experimental results showing that the new approach significantly outperforms existing state-of-the-art methods for matrix completion in many situations.

I. INTRODUCTION

Let $\mathbf{X} \in \mathbb{R}^{d \times N}$ be a matrix whose columns lie in a low-dimensional algebraic variety $\mathcal{V}$ (not necessarily linear). Such matrices will be called *low algebraic-dimension* (LAD) matrices. Matrix completion refers to the problem of recovering $\mathbf{X}$ from an observation consisting of a subset of its entries. This problem is well understood in the case of low-rank matrices, in which the columns lie in a low-dimensional subspace—this is the so-called *low rank matrix competion* (LRMC) problem [1, 2]. We call the more general problem of completing LAD matrices *low algebraic dimension matrix completion* (LADMC). *Unions of subspaces* (UoS) are a special case of algebraic varieties [3, 4], and a number of approaches to matrix completion for the UoS model have been proposed [5–11]. Most of these algorithms involve iterating between clustering and completion steps, and little can be guaranteed about their performance. However, recently [3] proposed a new 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 [3] can be interpreted as alternating between LRMC in the lifted representation and *unlifting* this low-rank representation back to the original representation (*i.e.,* solving for a *pre-image* matrix) to obtain a completion of the original matrix. This approach appears to provide good results in practice, but two problems were unresolved:

- The *unlifting* step is highly nonlinear and non-convex, and so little can be proved about its accuracy or correctness.
- While [3] provides an intuitive explanation for the potential of the approach (based on a degrees of freedom argument) and why it may succeed in cases where LRMC fails, a rigorous argument is lacking.

This paper addresses both issues. We provide a formal mathematical explanation for the success of such methods, and based on this insight we propose a new LADMC algorithm that uses a simple *unlifting* step based on the singular value decomposition (SVD). Experiments show that the new algorithm performs as well or better than state-of-the-art methods in the popular case of the UoS model.

A Mathematical Contribution to LADMC

The key to the usual LRMC problem is to learn the underlying subspace spanned by the columns of the matrix $\mathbf{X}$ from incomplete observations. Suppose that the columns of the matrix are observed in certain patterns. Let $\boldsymbol{\Omega}$ denote a binary matrix indicating all the *distinct* observed patterns in $\mathbf{X}$. The question of whether a subspace is *identifiable* from such observations is equivalent to determining whether the subspace can be reconstructed from its canonical projections onto the coordinates indicated by the columns of $\boldsymbol{\Omega}$. Essentially, the idea is that if we observe data with a certain sampling pattern multiple times, then we can use these data to reconstruct the projection of the underlying subspace onto the coordinates in the pattern. This observation has been developed and used in the analysis of LRMC and related problems [9, 12–15]

The main mathematical contribution of this paper is to generalize this to show that nonlinear varieties can be reconstructed from canonical projections. Specifically, we focus on quadratic varieties, which are linear varieties in a quadratic (tensor) representation. The tensor representation is defined as follows. Use the notation

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}^\mathsf{T}$$

to denote a column in $\mathbf{X}$. The tensor representation of this point is the vector

$$\begin{bmatrix} x_1^2 & x_1 x_2 & \cdots & x_1 x_d & x_2^2 & x_2 x_3 & \cdots & x_d^2 \end{bmatrix}^\mathsf{T},$$

i.e., the vector comprised of all unique pairwise products of the elements of $\mathbf{x}$.

The wrinkle here is that samplings in the original representation generate only a subset of all possible samplings in the tensor representation. For example, if we observe x_1 and x_2 but not x_3, then we can form the products x_1^2, $x_1 x_2$, and x_2^2, but not $x_1 x_3$, nor $x_2 x_3$, nor x_3^2, nor any other product involving x_3. This implies that the projections that we may observe in the tensor representation are restricted, and so one cannot simply apply known results [12] for the reconstruction of linear subspaces from canonical projections. Our main results show that under mild assumptions, the (possible) canonical projections in the tensor space are sufficient to identify linear subspaces (in the tensor representation), and thus (equivalently) quadratic varieties in the original domain. This shows that quadratic varieties are indeed identifiable from incomplete data. Furthermore, we derive precise information theoretic bounds on how much missingness can be tolerated in terms of the dimension of the tensor space. As we show in Theorem 1, to identify the low-dimensional subspace in the lifted space we are only required to compute the kernel of a matrix encoding the canonical projections. We show how these results apply to many interesting examples of quadratic varieties, including the popular UoS model.

Motivated by these results, we propose another new and more practical algorithm that we call LADMC (Algorithm 1) to complete data in low-dimensional quadratic varieties. Our algorithm performs LRMC on a lifted tensorized representation of the data matrix and solves the matrix pre-image problem once to return a matrix completion in the original domain. As a consequence of our theoretical results, we can guarantee that if the lifted matrix has been successfully completed with LRMC then the pre-image step is guaranteed to recover the original LAD matrix. In contrast with [3], this approach can be implemented non-iteratively (aside possibly from the subroutine used to solve the LRMC problem), although an iterative implementation appears to yield some empirical improvement.

Additionally, we demonstrate the proposed LADMC algorithm on UoS data, which shows competitive or improved matrix completion performance as compared to other state-of-the-art matrix completion methods, including approaches that are based on explicitly estimating UoS models. In contrast with these approaches, our LADMC algorithm does not require any prior knowledge of the number of subspaces or their dimension.

II. Setup and Algorithm

Let $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ be a matrix whose columns lie in a low-dimensional variety $\mathcal{V}$ (not necessarily linear). Let $D := \binom{d+1}{2}$. For any column $\mathbf{x} \in \mathbb{R}^d$, define its *tensorized column* $\mathbf{x}^{\otimes 2} \in \mathbb{R}^D$ as

$$\mathbf{x}^{\otimes 2} := (\mathbf{x} \otimes \mathbf{x})_D,$$

where $\otimes$ denotes Kronecker product and $(\cdot)_D$ denotes restriction of the Kronecker product to its D unique entries, *i.e.,* we delete $\binom{d}{2}$ duplicate entries that arise from taking the Kronecker product of a vector with itself. We call

$$\mathbf{X}^{\otimes 2} = [\mathbf{x}_1^{\otimes 2} \cdots \mathbf{x}_N^{\otimes 2}]$$

the *tensorized matrix*, and call the subset of $\mathbb{R}^D$ to which these matrices belong the *tensor space*.

Definition 1. *Let* $\mathbf{V}$ *be a collection of vectors in* $\mathbb{R}^D$. *Given* $\mathbf{V}$*, we define the* quadratic variety $\mathcal{V} \subset \mathbb{R}^d$ *as*

$$\mathcal{V} := \left\{ \mathbf{x} \in \mathbb{R}^d : \mathbf{v}^\mathsf{T} \mathbf{x}^{\otimes 2} = 0 \text{ for every } \mathbf{v} \in \mathbf{V} \right\}.$$

We call this variety quadratic *because every element* $\mathbf{x} \in \mathcal{V}$ *satisfies quadratic polynomial constraints of the form*

$$v_1 x_1^2 + v_2 x_1 x_2 + v_3 x_1 x_3 + \cdots + v_D x_d^2 = 0.$$

In other words, $\mathbf{V}$ *contains the coefficients of the polynomial constraints that define* $\mathcal{V}$.

Notice that by definition, if $\mathbf{x} \in \mathcal{V}$, then $\mathbf{x}^{\otimes 2} \in \ker \mathbf{V}^\mathsf{T} =: \mathcal{S}$, which is a linear subspace of $\mathbb{R}^D$. In words, low-dimensional *quadratic* varieties are low-dimensional *linear* varieties in the tensor space.

This is the main idea motivating this work. More precisely, if the columns of $\mathbf{X}$ are points on a quadratic variety $\mathcal{V}$ whose corresponding tensorized subspace $\mathcal{S}$ is low-dimensional, then the tensorized matrix $\mathbf{X}^{\otimes 2}$ is low-rank and may be recovered from a subset of its entries using standard LRMC techniques.

Furthermore, if LRMC indeed recovers $\mathbf{X}^{\otimes 2}$ from a subset of its entries, then we can always recover $\mathbf{X}$ from $\mathbf{X}^{\otimes 2}$. To see this, let $\mathbf{x}^{\otimes 2} \in \mathbb{R}^D$ be a column of $\mathbf{X}^{\otimes 2}$. Since $\mathbf{x}^{\otimes 2} = (\mathbf{x} \otimes \mathbf{x})_D$, we can populate the entries of the symmetric matrix $\mathbf{x} \mathbf{x}^\mathsf{T} \in \mathbb{R}^{d \times d}$ with the entries of $\mathbf{x}^{\otimes 2}$. This way $\mathbf{x} \mathbf{x}^\mathsf{T}$ is rank one, and its leading singular vector is precisely $\mathbf{x}$ up to a choice of sign, which can be determined from any non-zero entry of $\mathbf{x}$. These observations motivate our proposed algorithm, summarized in Algorithm 1.

Algorithm 1: Low-Algebraic-Dimension Matrix Completion (LADMC).

---

1 **Input:** Subset of entries of data matrix $\mathbf{X}$.
2 **Tensorize:** Form new matrix $\mathbf{X}^{\otimes 2}$ by replacing each Column $\mathbf{x}_i$ with tensor product $\mathbf{x}_i \otimes \mathbf{x}_i$ (with missing entries corresponding to missing data in $\mathbf{x}_i$).
3 **LRMC:** Let $\hat{\mathbf{X}}^{\otimes 2}$ be the low-rank completion (obtained from the observed entries in $\mathbf{X}^{\otimes 2}$).
4 **SVD:** Let $\hat{\mathbf{x}}_i$ be principle singular vector of square symmetric matrix formed from i^{th} column of $\hat{\mathbf{X}}^{\otimes 2}$.
5 **Output:** completed matrix $\hat{\mathbf{X}}$ whose i^{th} column is $\hat{\mathbf{x}}_i$.

Remarkably, there are situations where the original data matrix $\mathbf{X}$ is full rank, but the tensorized matrix $\mathbf{X}^{\otimes 2}$ is low-rank, owing to the (nonlinear) algebraic structure of the data. In such cases, the LADMC matrix can succeed, while LRMC will fail.

For example, a union of K r-dimensional subspaces is a special case of a nonlinear variety. Specifically, note that a union of K rank-r subspaces (in general position) lies in a subspace of $\mathbb{R}^d$ of dimension $\min(Kr, d)$ in $\mathbb{R}^d$. However, in the tensor space, these points lie in a subspace of $\mathbb{R}^D$ of dimension $R = \min(K\binom{r+1}{2}, D)$ in $\mathbb{R}^D$; see the appendix for a proof. The basic intuition is this: if we need $O(R)$ observations (non-missing entries) per column to complete a rank-R matrix, then completing a union of K r-dimensional subspaces in the original space would require $O(Kr)$ observations per column, but completing the corresponding tensorized matrix would require $O(K\binom{r+1}{2})$ observations per column, $O(\sqrt{K}r)$ observations per column in the original space. In cases like this, LADMC matrix can succeed with far fewer observations per column than LRMC.

III. PREAMBLE AND ASSUMPTIONS

Algorithm 1 is primarily inspired by the ideas in [3]. However, the key technical issue that was unresolved in [3] is related to the fact that the pattern of missing entries in the tensorized matrix is highly structured (due to the tensor product). Consequently, the missingness patterns in the tensorized matrix are not generated uniformly at random (even if the original missingness patterns are).

Moreover, the set of possible sampling patterns per column is highly constrained. For instance, consider a data matrix where each column is observed in, say, m locations. Then each column of the tensorized matrix is observed in $M = \binom{m+1}{2}$ locations (*i.e.,* we can compute only M of the $D = \binom{d+1}{2}$ possible products in each column). If we were sampling directly in the tensor space, there would be $\binom{D}{M}$ possible sampling patterns for each column. However, because we are sampling in the original space, there are in fact only $\binom{d}{m}$ possible sampling patterns per column. These constraints on the sampling patterns imply that existing LRMC theory (which typically requires uniform sampling) does not apply directly to the tensorized representations.

In addition, typical LRMC methods also rely on incoherence, which is a parameter indicating how aligned a subspace is to the canonical axes. Incoherence guarantees that the information is well-spread over all coordinates. However, it is unclear (and generally unverifiable) whether the subspace containing $\mathbf{X}^{\otimes 2}$ will be incoherent.

For these reasons it was not clear (until now) whether LRMC algorithms would succeed when applied to the tensorized data. In fact, as shown by the following example, there are cases where $\mathbf{X}^{\otimes 2}$ cannot be low-rank completed, even with an unlimited amount of data, sampled randomly and ideal incoherence.

Example 1. *Suppose $\mathbf{X}$ lies in the union of $K = 2$ subspaces in general position, each of dimension $r = 1$, in ambient dimension $d = 3$. Then $\mathbf{X}^{\otimes 2}$ lies in a single subspace of dimension $R = K\binom{r+1}{2} = 2$, in ambient dimension $D = \binom{d+1}{2} = 6$.*

Information-theoretically, one must observe at least $r + 1$ entries per column to complete $\mathbf{X}$ [9]. So suppose $\mathbf{X}$ is observed on exactly $r + 1 = 2$ entries per column. It follows that there are only 3 possible samplings for any column of $\mathbf{X}$, and consequently there are only 3 possible samplings for any column of $\mathbf{X}^{\otimes 2}$. The respective possible samplings are indicated by the columns in the following matrices:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

This way, each column of $\mathbf{X}^{\otimes 2}$ will be observed on exactly 3 entries, thus satisfying the minimum $R + 1 = 3$ observations per column required for low-rank completion [13]. If the number of columns N is sufficiently large, at some point there will be enough observations to account for the number of degrees of freedom in a $D \times N$ rank-R matrix.

However, having enough observations is necessary for completion, but in general insufficient. The observations need to be in the right *locations [13]. In fact, one may use the results in [12] to verify that even if $\mathbf{X}^{\otimes 2}$ is maximally incoherent (ideal scenario), and even with an unlimited number of columns, $\mathbf{X}^{\otimes 2}$ cannot be recovered.*

We point out that given enough observations, it is information-theoretically possible to recover $\mathbf{X}$ (for example using subspace clustering with missing data [9]). However, it is impossible to recover $\mathbf{X}$ by low-rank completing $\mathbf{X}^{\otimes 2}$.

Hence it was unclear (until now) whether a low-rank matrix in the tensorized domain is even *identifiable* from these constrained sampling patterns. In this paper we prove that low-rank matrices that arise from tensorized LAD structures are indeed identifiable from tensorized sampling patterns.

Of course, without further assumptions on $\mathbf{X}$ and the observed entries, completion may be impossible. To see this consider the simple example where the columns of $\mathbf{X}$ lie in a variety $\mathcal{V}$ that is only supported on the jth coordinate (perfectly aligned with the jth canonical axis). Then it would be impossible to recover $\mathbf{X}$ unless all the columns were observed on the jth row. In most completion applications this would be unlikely. For example, in a movie recommender system like Netflix, this would require that *all* the users with similar preferences watched (and rated) the same movie. To rule out scenarios like these, typical completion results require incoherence and uniform sampling. However, as discussed above, these assumptions do not hold in our

setting, so it was unclear whether $\mathbf{X}^{\otimes 2}$ could be low-rank completed.

To avoid these issues, and show that $\mathbf{X}^{\otimes 2}$ can indeed be low-rank completed, our main theoretical results will assume that $\mathcal{V}$ is *generic*. More precisely:

(A1) Recall from Definition 1 that $\mathbf{V}$ contains the coefficients of the polynomials that determine $\mathcal{V}$. Assume that the entries in $\mathbf{V}$ are drawn independently with respect to an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$.

Essentially, **A1** requires that the coefficients $\mathbf{V}$ of the polynomials that determine $\mathcal{V}$ are generic. This guarantees that $\mathcal{V}$ is in general position. Varieties of this sort do not have a particular orientation with respect to the canonical axes. Notice that by definition, $\mathcal{V}$ is the set of points $\mathbf{x} \in \mathbb{R}^d$ such that $\mathbf{x}^{\otimes 2} \in \ker \mathbf{V}^{\mathsf{T}} =: \mathcal{S}$, which is a linear subspace of $\mathbb{R}^D$. This way, $\mathcal{V}$ is also determined by $\mathcal{S}$. Equivalently, **A1** requires that $\mathcal{S}$ is in general position. Similar sorts of genericity assumptions have been used to study standard LRMC and related problems [9, 12–18].

As mentioned in Section II, unions of subspaces (UoS) are a special type of variety (see [3] for more details). UoS are very relevant in modern applications and have received considerable attention [5–11]. However, they do not follow **A1**. Luckily, as we show in Section V, all our results also apply to UoS that satisfy the following general position assumption:

(A2) $\mathcal{V}$ is a variety corresponding to the union of K linear subspaces drawn independently at random with respect to uniform measure over the Grassmann manifold of r-dimensional subspaces in $\mathbb{R}^d$.

With these assumptions we are ready to show that low-rank matrices that arise from tensorized LAD structures are indeed identifiable from tensorized sampling patterns.

IV. MAIN RESULTS

First observe that the problem of completing the rank-R matrix $\mathbf{X}^{\otimes 2}$ is tantamount to identifying the R-dimensional subspace $\mathcal{S}$ spanned by its columns [13], and this is how we will approach it.

Let $\boldsymbol{\Omega} = [\boldsymbol{\omega}_1 \cdots \boldsymbol{\omega}_n] \in \{0,1\}^{d \times n}$ be the matrix containing all the *distinct* sampling patterns of the columns in $\mathbf{X}$. Define $\boldsymbol{v}_i := (\boldsymbol{\omega}_i \otimes \boldsymbol{\omega}_i)_D$, and let $\boldsymbol{\Upsilon} := [\boldsymbol{v}_1 \cdots \boldsymbol{v}_n] \in \{0,1\}^{D \times n}$ indicate all the distinct sampling patterns of $\mathbf{X}^{\otimes 2}$.

Next notice that $L \geq R$ columns in $\mathbf{X}$ observed on rows $\boldsymbol{\omega}_i$ will generate $L \geq R$ columns in $\mathbf{X}^{\otimes 2}$ observed on rows $\boldsymbol{v}_i$. If $\mathcal{V}$ is generic (*i.e.*, satisfies **A1** or **A2**) and the columns in $\mathbf{X}$ are drawn generically from $\mathcal{V}$ (*i.e.*, according

to an absolutely continuous distribution with respect to the Lebesgue measure on $\mathcal{V}$), then the tensorized columns will span $\mathcal{S}_{\boldsymbol{v}_i}$, the projection of $\mathcal{S}$ onto the coordinates indicated in $\boldsymbol{v}_i$. In other words, $L \geq R$ columns of $\mathbf{X}$ observed on $\boldsymbol{\omega}_i$ contain the same information about $\mathcal{S}$ as the canonical projection $\mathcal{S}_{\boldsymbol{v}_i}$.

With this in mind, we will derive necessary and sufficient conditions on the sampling pattern $\boldsymbol{\Omega}$ to guarantee that $\mathcal{S}$ can be uniquely recovered from the canonical projections indicated in $\boldsymbol{\Upsilon}$, *i.e.*, $\{\mathcal{S}_{\boldsymbol{v}_i}\}_{i=1}^n$. See Figure 1 to build some intuition.

The key insight is that observing a projection onto more than R coordinates places constraints on what $\mathcal{S}$ may be. For example, if we observe a projection onto $R + 1$ coordinates, then not all R-dimensional subspaces will be consistent with this projection. If we observe more projections, then even fewer subspaces will be consistent with them. In effect, each coordinate, in addition to the first R, places one constraint that an R-dimensional subspace must satisfy in order to be consistent with the projections. The projections onto different coordinates may or may not produce redundant constraints. The main result in [12] is a simple condition on the set of constraints (resulting from all the projections) that is necessary and sufficient to guarantee that only one subspace satisfies all the constraints. This in turn provides a simple condition to identify the subspace.

To state these results, we introduce the matrix $\check{\boldsymbol{\Upsilon}}$ that encodes the set of all constraints in a way that allows us to easily express the necessary and sufficient condition.

Let $j_1^{(i)}, \ldots, j_{m_i}^{(i)}$ denote the indices of the m_i nonzero entries in the ith column of $\boldsymbol{\Upsilon}$. Define $\check{\boldsymbol{\Upsilon}}_i$ as the $D \times (m_i - R)$ matrix, whose κ^{th} column has the value 1 in rows $j_1^{(i)}, \ldots, j_R^{(i)}, j_{R+\kappa}^{(i)}$, and zeros elsewhere. For example,
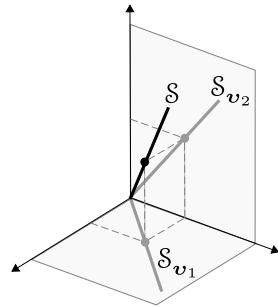

Fig. 1. When can $\mathcal{S}$ be identified from its canonical projections $\{\mathcal{S}_{\boldsymbol{v}_i}\}_{i=1}^n$?

if $j_1^{(i)} = 1, j_2^{(i)} = 2, \ldots, j_{m_i}^{(i)} = m_i$, then

$$\check{\Upsilon}_i = \begin{bmatrix} \mathbf{1} \\ \hline \mathbf{I} \\ \hline \mathbf{0} \end{bmatrix} \begin{array}{l} \} R \\ \} m_i - R \\ \} D - m_i, \end{array}$$
$$\underbrace{}_{m_i - R}$$

where $\mathbf{1}$ denotes a block of all 1's and $\mathbf{I}$ the identity matrix. Finally, define $\check{\Upsilon} := [\check{\Upsilon}_1 \cdots \check{\Upsilon}_n]$.

This way, each column of $\check{\Upsilon}$ indicates the coordinates involved in one constraint placed by the projections. But not only that; $\check{\Upsilon}$ also allows to write all constraints as a simple, linear matrix equation. To see this, let $\check{v}_j$ denote the jth column in $\check{\Upsilon}$. Lemma 1 below shows that $\mathcal{S}_{\check{v}_j}$ is a hyperplane, *i.e.*, an R-dimensional subspace in $\mathbb{R}^{R+1}$. As such, it is characterized by its orthogonal direction, which we will call $\mathbf{a}_{\check{v}_j}$. More precisely, let $\mathbf{a}_{\check{v}_j} \in \mathbb{R}^{R+1}$ be a nonzero vector in $\ker \mathcal{S}_{\check{v}_j}$, and let $\mathbf{a}_j$ be the vector in $\mathbb{R}^D$ with the entries of $\mathbf{a}_{\check{v}_j}$ in the locations of $\check{v}_j$, and zeros elsewhere. This way, any subspace that agrees with the projection $\mathcal{S}_{\check{v}_j}$ must be orthogonal to $\mathbf{a}_{\check{v}_j}$. Let $\mathbf{A}$ be the matrix formed with $\{\mathbf{a}_i\}$ as columns. Then $\ker \mathbf{A}^\mathsf{T}$ contains all the subspaces consistent with the projections. If $\dim \ker \mathbf{A}^\mathsf{T} = R$, then there is only one subspace that agrees with the projections, namely $\mathcal{S} = \ker \mathbf{A}^\mathsf{T}$ (which by definition is also equal to $\ker \mathbf{V}^\mathsf{T}$; it follows that $\mathbf{A}$ actually spans the same subspace as $\mathbf{V}$). This way $\mathbf{A}$ encodes the information of the projections $\{\mathcal{S}_{v_i}\}_{i=1}^n$.

Our main result, summarized in the next theorem, gives the necessary and sufficient conditions on the given projections to guarantee that $\mathcal{S}$ can be identified. This in turn provides an identifiability condition for LADMC. The proof is given in Section V.

Theorem 1. *Assume $\mathcal{V}$ satisfies **A1** or **A2**, with $\mathcal{S}$ being an R-dimensional subspace of $\mathbb{R}^D$. Then with probability 1, $\mathcal{S}$ can be uniquely recovered from the canonical projections indicated in Υ if and only if there is a matrix $\Upsilon^\star$ formed with $D - R$ columns of $\check{\Upsilon}$, such that every matrix formed with a subset of η columns of $\Upsilon^\star$ has at least $\eta + R$ nonzero rows. Furthermore, this condition holds if and only if $\dim \ker \mathbf{A}^\mathsf{T} = R$, whence $\mathcal{S} = \ker \mathbf{A}^\mathsf{T}$.*

Theorem 1 gives exact conditions on the original samplings Ω to identify the tensor subspace $\mathcal{S}$. We show in Section V that there are plenty of sampling patterns satisfying these conditions.

Theorem 1 also gives provable algorithm for identifying $\mathcal{S}$ from canonical projections: if the data sampling satisfies

Theorem 2, then from the canonical projections on those coordinates we can identify $\mathcal{S}$ uniquely in closed form as $\mathcal{S} = \ker \mathbf{A}^\mathsf{T}$. In this case, any algorithm for finding the null space of $\mathbf{A}^\mathsf{T}$ is a provably correct algorithm for identifying the subspace $\mathcal{S}$.

As consequence of Theorem 1 we also obtain exact information-theoretic sampling rates required for LADMC, summarized in the next theorem. To present these, first recall that $\mathcal{S}$ is R-dimensional, so it can only be recovered if Υ has more than R observations per column [12]. It is easy to see that a column in Ω with m observations will produce a column in Υ with $\binom{m+1}{2}$ observations. It follows that columns in Ω need to be observed on *at least* $m = \ell$ entries, where the critical point (lower bound) ℓ is the smallest integer such that $\binom{\ell+1}{2} \geq R$. Surprisingly, the next theorem shows that $m = \ell$, and even $m = \ell + 1$, are in general insufficient (even when observing *all* possible columns with $\ell + 1$ observations!) and hence $m = \ell + 2$ is in general necessary. Also surprising, the theorem shows that $m = \ell + 2$ is sufficient.

Theorem 2. *Assume $\mathcal{V}$ satisfies **A1** or **A2**, with $\mathcal{S}$ being an R-dimensional subspace of $\mathbb{R}^D$. Suppose Ω contains all possible $\binom{d}{m}$ subsets of m of the d coordinates in the original data space. Let ℓ be the smallest integer such that $\binom{\ell+1}{2} \geq R$.*

(i) *If $m < \ell$, then $\mathcal{S}$ cannot be uniquely determined.*
(ii) *There exist cases with $m = \ell$ and $m = \ell + 1$ where $\mathcal{S}$ cannot be uniquely determined.*
(iii) *If $m \geq \ell + 2$, then $\mathcal{S}$ can be uniquely determined.*

The proof of (i) follows by the simple argument above. The proof of (ii) follows by counterexample (Example 1 above). The proof of (iii) is given in Section V. It follows by showing that the conditions of Theorem 1 are met if Ω contains all the sampling patterns with $m = \ell + 2$ observations per column.

V. Proof

In order to prove Theorem 1, we will use the results in [12]. These results show that a subspace *in general position* can be recovered from a subset of its canonical projections (see Figure 1 to build some intuition). Furthermore, they determine exactly which projections are needed.

More precisely, [12] requires that the subspace is drawn according to the uniform measure over the Grassmannian. This is to guarantee that all its projections have full-dimension, or equivalently, that all the minors of its bases are full-rank. Intuitively, this guarantees that all its projections provide enough information about the subspace, so that these projections can be *glued* together to recover the whole subspace.

Recall that $\mathcal{S}$ is the R-dimensional subspace of $\mathbb{R}^D$ containing all the columns in $\mathbf{X}^{\otimes 2}$. Observe that $\mathcal{S}$ is determined by the tensorized products of $\mathcal{V}$, and hence some of its coordinates have some structured relationships. In other words, it is not drawn uniformly over the Grassmannian. However, the next lemma shows that in fact, all its projections have full-dimension, and hence all the results in [12] are applicable.

Lemma 1. *Assume $\mathcal{V}$ satisfies* **A1** *or* **A2**. *Let υ be a subset of $\{1, \ldots, D\}$ with exactly R elements. Let $\mathcal{S}_\upsilon$ denote the restriction of $\mathcal{S}$ to the coordinates in υ. Then $\dim \mathcal{S}_\upsilon = R$.*

Proof. Let $\mathbf{U}$ denote a basis for the R-dimensional subspace $\mathcal{S}$. It suffices to show that every $R \times R$ submatrix of $\mathbf{U}$ is full-rank with probability 1.

First consider the case when $\mathcal{V}$ satisfies **A1**. Recall that the span of $\mathbf{V}$ the orthogonal complement of $\mathcal{S}$. Since the elements of $\mathbf{V}$ are continuously distributed and i.i.d., the subspace spanned by $\mathbf{V}$ is continuously distributed with respect to uniform measure on the Grassmannian. It follows that the complementary subspace spanned by $\mathbf{U}$ is as well, and therefore every $R \times R$ submatrix of $\mathbf{U}$ is full-rank with probability 1.

Next consider the case when $\mathcal{V}$ satisifies **A2**. This case is a bit more complicated, since the variety corresponding to a union of subspaces is not fully generic. Let $\mathbf{u}_1, \ldots, \mathbf{u}_{Kr}$ be a basis for the original data matrix (*i.e.*, $\mathbf{u}_1, \ldots, \mathbf{u}_r$ is a basis for the first subspace, $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_{2r}$ is a basis for the second, and so on). Then tensor products of the form $\mathbf{u}_i \otimes \mathbf{u}_j$, where i and j are indices belonging to one of the K subspaces, form a basis for $\mathcal{S}$. Let $\mathbf{U}$ denote a matrix with R columns equal to these tensorized basis elements.

Note that there exists a rank deficient $R \times R$ submatrix of $\mathbf{U}$ if and only if there exists a a vector $\boldsymbol{\alpha}$ with R nonzero elements and satisfies $\mathbf{U}^\mathsf{T} \boldsymbol{\alpha} = \mathbf{0}$. We will show that this is impossible.

Let $\{u_{ij}\}_{i=1,j=1}^{d,Kr}$ denote all the elements $\mathbf{u}_1, \ldots, \mathbf{u}_{Kr}$. Observe that the tensorized basis has elements of the form $\{u_{i,j}u_{\ell,m}\}$. Form the vectors

$$\mathbf{v}_{ij} = \begin{bmatrix} 1 \\ u_{ij} \\ u_{ij}^2 \end{bmatrix},$$

and define $\mathbf{v}_j = \otimes_i \mathbf{v}_{ij}$. Note that $\mathbf{v}_j$ contains all the products in the jth column of $\mathbf{U}$ (and more). In other words, $\mathbf{U}$ is a subset of the rows in $\mathbf{V} := [\mathbf{v}_1 \cdots \mathbf{v}_{D-R}]$. Now consider the expectation

$$\mathbb{E}\mathbf{V}\mathbf{V}^\mathsf{T} = \sum_j \mathbb{E}\mathbf{v}_j\mathbf{v}_j^\mathsf{T} = \sum_j \bigotimes_i \mathbb{E}\mathbf{v}_{ij}\mathbf{v}_{ij}^\mathsf{T},$$

where we bring the expectation inside the product since the u_{ij} are assumed to be independent. Since the u_{ij} are absolutely continuous random variables, each $\mathbb{E}\mathbf{v}_{ij}\mathbf{v}_{ij}$ is full-rank (see Theorem 1 in [19]). It follows by standard Kronecker product properties that $\mathbb{E}\mathbf{V}\mathbf{V}^\mathsf{T}$ is full-rank, and thus $\mathbb{E}\mathbf{U}\mathbf{U}^\mathsf{T}$ is also full-rank.

Now suppose there exists a vector $\boldsymbol{\alpha}$ with R nonzero elements and satisfies $\mathbf{U}^\mathsf{T} \boldsymbol{\alpha} = \mathbf{0}$. This is equivalent to the condition $\boldsymbol{\alpha}^\mathsf{T} \mathbb{E}\mathbf{U}\mathbf{U}^\mathsf{T} \boldsymbol{\alpha} = 0$ (*i.e.*, $\mathbb{E}Y^2 = 0$ if and only if $Y = 0$ with probability 1), which implies that $\mathbb{E}\mathbf{U}\mathbf{U}^\mathsf{T}$ is rank deficient, contradicting the conclusion above.

$\square$

Equipped with Lemma 1, the proof of Theorem 1 follows unchanged from the proofs of Theorems 1 and 3 in [12]. We now give the proof of Theorem 2.

Proof. (Theorem 2) The proof of (i) follows by the simple argument above Theorem 2. Example 1 shows (ii). We now give the proof of (iii).

There exist plenty of samplings $\check{\boldsymbol{\Upsilon}}$ satisfying the conditions of Theorem 2. For instance, it is easy to see (and one can verify using Algorithm 1) that the following sampling, where $\mathbf{1}$ denotes a block of all 1's and $\mathbf{I}$ denotes the identity matrix, satisfies these conditions.

$$\boldsymbol{\Upsilon}^\star = \left[\begin{array}{c} \mathbf{1} \\ \hline \mathbf{I} \end{array} \right] \begin{array}{l} \left.\right\} R \\ \left.\right\} D - R. \end{array} \qquad (1)$$

To show (iii), it suffices to show that the sampling $\boldsymbol{\Omega}$ (containing all possible $\binom{d}{m}$ subsets of m of the d coordinates in the original data space) will generate a sampling $\check{\boldsymbol{\Upsilon}}$ containing a sampling $\boldsymbol{\Upsilon}^\star$ satisfying the conditions of Theorem 1. In particular, we will show that $\check{\boldsymbol{\Upsilon}}$ will contain the sampling $\boldsymbol{\Upsilon}^\star$ in (1).

To this end, recall that each coordinate in the tensor space $\mathbb{R}^D$ corresponds to two coordinates in the original space $\mathbb{R}^d$. Let $i = 1, \ldots, D - R$ index the columns of $\boldsymbol{\Upsilon}^\star$ in (1), and let $j, k \in \{1, \ldots, d\}$ be the coordinates of $\mathbb{R}^d$ corresponding to the $(R + i)^{\text{th}}$ coordinate in $\mathbb{R}^D$.

It is easy to see that $\boldsymbol{\upsilon}_i \in \boldsymbol{\Upsilon}$ is nonzero in row $(R + i)$ if and only if $\boldsymbol{\omega}_i$ is nonzero in both rows j and k. Since $\boldsymbol{\Omega}$ contains all possible columns with $m = \ell + 2$ samples, it contains a column $\boldsymbol{\omega}_i$ whose first ℓ samples are in the first ℓ rows, and whose last 2 samples are in rows j and k.

Recall that a column in $\boldsymbol{\Omega}$ with ℓ observations will produce a column in $\boldsymbol{\Upsilon}$ with $\binom{\ell+1}{2} \geq R$ observations. It follows that $\boldsymbol{\omega}_i$ will produce a column $\boldsymbol{\upsilon}_i$ with samples in the first R rows and in row $(R + i)$. In other words, the column $\boldsymbol{\omega}_i$ will produce the ith column of $\boldsymbol{\Upsilon}^\star$ in (1).

Since is true for every $i = 1, \ldots, D - R$, we know that $\boldsymbol{\Omega}$ will produce a matrix $\check{\boldsymbol{\Upsilon}}$ containing $\boldsymbol{\Upsilon}^\star$ as in (1). Since $\boldsymbol{\Upsilon}^\star$ satisfies the conditions of Theorem 1, it follows that $\mathcal{S}$ can be uniquely recovered, as claimed.

$\square$

VI. EXPERIMENTS

The following experiments demonstrate the performance of the proposed LADMC algorithm (Alg 1) on synthetic data having low algebraic dimension. For the LRMC step within
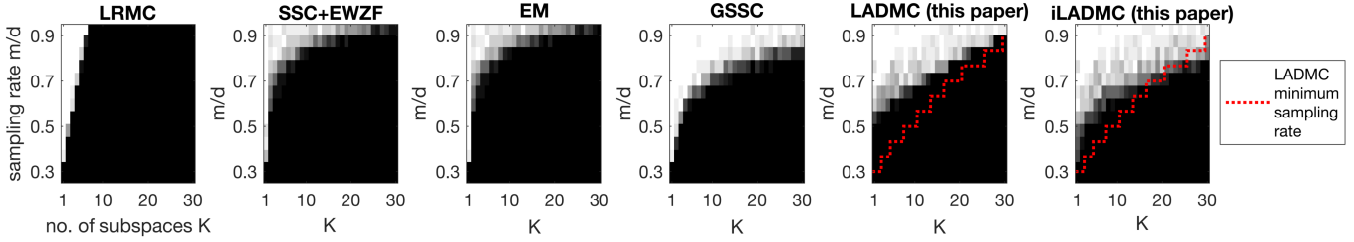
Fig. 2. Phase transitions for matrix completion of synthetic union of subspaces data. We simulated data belonging to K subspaces and sampled each column of the data matrix at a rate m/d, and perform matrix completion using LRMC, state-of-the-art subspace clustering based algorithms (SSC+EWZF, GSSC, EM), and the proposed LADMC and iLADMC algorithms. 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 needed for LRMC to succeed in the tensorized domain.

LADMC we use the singular value iterative hard thresholding algorithm, also known as singular value projection [20]. In addition to LADMC we also test an iterative version of the approach (iLADMC) where we run the iterative singular value hard thresholding step of LADMC in the tensorized domain for T iterations, map back to the original domain by the rank-1 SVD pre-image step, fill in the known entries of the matrix, and repeat until convergence. In the experiments below we ran T = 30 *inner* iterations of singular value hard thresholding, and as many *outer* iterations of LADMC until convergence. 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 iterative hard thresholding steps, which is the main computational bottleneck) and succeeds in completing matrices at lower sampling rates than plain LADMC.

Figure 2 shows the performance of LADMC and iLADMC against competing approaches 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. We sampled m entries in each column, selected uniformly at random, and attempted to recover the missing entries using our proposed LADMC algorithm. 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. Here we compared against LRMC via iterative hard thresholding minimization in the original matrix domain, and three methods based on subspace clustering: sparse subspace clustering (SSC) with entry-wise zero fill [8] followed by LRMCon each identified cluster (SSC+EWZF), the expectation-maximization (EM) algorithm proposed in [7], and the group-sparse subspace clustering algorithm [10] 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.

We find that LADMC is able to successfully complete the data with high probability when the number of measurements per column in the tensorized domain exceeds the information-theoretic bounds, as indicated by the red dashed line in Figure 2. This is a substantial extension over standard LRMC: for these settings, LADMC is able to complete data matrices drawn from up to K = 30 subspaces, whereas LRMC is limited to data drawn from less than K = 7 subspaces. However, for LADMC there 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). This gap is less pronounced for iLADMC, and in fact iLADMC shows recovery below the LADMC information-theoretic bound. 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.
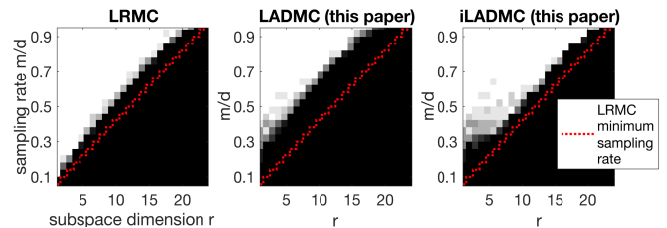

Fig. 3. Phase transitions for matrix completion of matrices whose columns are drawn from a single subspace of dimension r.

In Figure 3 we compare the performance of LADMC versus LRMC when the data is drawn from a *single* subspace of dimension $r \le d$ (using the settings d = 25, r = 1, ..., 25, N = 500). Here we find that the performance of LADMC is only slightly worse than LRMC. In fact, iLADMC shows a similar phase transition as LRMC except at very low

sampling rates where the performance degrades. This shows that LADMC and iLADMC can be expected to perform similarly well as LRMC in the case where the data is low-rank.

VII. 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 non-linear algebraic variety, a special case of which is a union of subspaces. Unlike most completion methods for unions of subspaces, the proposed approach does not necessitate an intermediate subspace clustering step that can be fragile in the presence of missing data.

The theoretical guarantees currently focus on canonical projections – *i.e.,* the assumption that we observe multiple columns with each possible sampling 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 completion of partially sampled data matrices.

In addition, while this paper focused on quadratic tensorized representations of the data, the ideas easily generalize to higher order tensorizations. This extension would facilitate the completion of higher order varieties in the presence of more missing data. Algorithmically, the LADMC algorithm would only need to have the final step (SVD) changed to finding a rank-one approximation of a multi-dimensional tensor; this topic has received concerted attention in recent years. However, the computational complexity scales as $O(d^q)$, where q is the tensor order; we are actively investigating memory- and computationally efficient algorithms.

VIII. ACKNOWLEDGEMENTS

REFERENCES

[1] E. Candès and T. Tao, *The power of convex relaxation: near-optimal matrix completion*, IEEE Transactions on Information Theory, 2010.
[2] B. Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research, 2011.
[3] G. Ongie, R. Willett, R. Nowak and L. Balzano, *Algebraic Variety Models for High-Rank Matrix Completion*, International Conference on Machine Learning, 2017.
[4] R. Vidal, Y. Ma, and S. Sastry, *Generalized principal component analysis (GPCA)*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005.
[5] B. Eriksson, L. Balzano and R. Nowak, *High-rank matrix completion and subspace clustering with missing data*, Artificial Intelligence and Statistics, 2012.
[6] L. Balzano, R. Nowak, A. Szlam and B. Recht, *k-Subspaces with missing data*, IEEE Statistical Signal Processing, 2012.
[7] D. Pimentel-Alarcón, L. Balzano and R. Nowak, *On the sample complexity of subspace clustering with missing data*, IEEE Statistical Signal Processing, 2014.
[8] C. Yang, D. Robinson and R. Vidal, *Sparse subspace clustering with missing entries*, International Conference on Machine Learning, 2015.
[9] D. Pimentel-Alarcón and R. Nowak, *The information-theoretic requirements of subspace clustering with missing data*, International Conference on Machine Learning, 2016.
[10] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. Nowak, R. Willett, *Group-sparse subspace clustering with missing data*, IEEE Statistical Signal Processing, 2016.
[11] D. Pimentel-Alarcón, L. Balzano, R. Marcia, R. Nowak, R. Willett, *Mixture regression as subspace clustering*, International Conference on Sampling Theory and Applications, 2017.
[12] D. Pimentel-Alarcón, N. Boston and R. Nowak, *Deterministic conditions for subspace identifiability from incomplete sampling*, IEEE International Symposium on Information Theory, 2015.
[13] D. Pimentel-Alarcón, N. Boston and R. Nowak, *A characterization of deterministic sampling patterns for low-rank matrix completion*, IEEE Journal of Selected Topics in Signal Processing, 2016.
[14] D. Pimentel-Alarcón and R. Nowak, *A converse to low-rank matrix completion*, IEEE International Symposium on Information Theory, 2016.
[15] D. Pimentel-Alarcón and R. Nowak, *Random consensus robust PCA*, Artificial Intelligence and Statistics, 2017.
[16] F. Király, L. Theran and R. Tomioka, *The algebraic combinatorial approach for low-rank matrix completion*, Journal of Machine Learning Research, 2015.
[17] M. Ashraphijuo, X. Wang and V. Aggarwal, *A characterization of sampling patterns for low-rank multiview data completion problem*, IEEE International Symposium on Information Theory, 2017.
[18] M. Ashraphijuo, V. Aggarwal and X. Wang, *A characterization of sampling patterns for low-tucker-rank tensor completion problem*, IEEE International Symposium on Information Theory, 2017.
[19] R. Nowak and B. Van Veen, *Random and pseudorandom inputs for volterra filter identification*, IEEE Transactions on Signal Processing, 1994.
[20] P. Jain, R. Meka, and I. Dhillon, *Guaranteed rank minimization via singular value projection*, Neural Information Processing Systems Conference, 2010.

APPENDIX

Nonlinear Algebraic Representations of Unions of Subspaces.
For simplicity of presentation, we focus on the setting where $K\binom{r+1}{2} \le D$. Consider an r-dimensional subspace of $\mathbb{R}^d$. Points in this space can be represented by $\mathbf{x} = \mathbf{U}\boldsymbol{\theta}$, where columns of $\mathbf{U} \in \mathbb{R}^{d \times r}$ are a basis for the subspace. For a matrix $\mathbf{U}$, let

$$\mathbf{U}^{\otimes 2} := (\mathbf{U} \otimes \mathbf{U})_D.$$

Note that $\mathbf{x}^{\otimes 2} = \mathbf{U}^{\otimes 2}\boldsymbol{\theta}^{\otimes 2}$. The dimensions of $\boldsymbol{\theta}^{\otimes 2}$ and $\mathbf{x}^{\otimes 2}$ are $\binom{r+1}{2}$ and $\binom{d+1}{2}$, respectively (after eliminating repeats, as described in the introduction).

Now consider the set of vectors $\{\mathbf{U}^{\otimes 2}\boldsymbol{\theta}^{\otimes 2} : \boldsymbol{\theta} \in \mathbb{R}^r\}$. This set spans a $\binom{r+1}{2}$ subspace of the (ambient) $\binom{d+1}{2}$ space. This follows simply by noting that $\mathbf{U}^{\otimes 2}$ is full rank (since $\mathbf{U}$ is full rank) and the set $\{\boldsymbol{\theta}^{\otimes 2} : \boldsymbol{\theta} \in \mathbb{R}^r\}$ spans $\binom{r+1}{2}$ dimensions, since the monomials are linearly independent.

Now consider two r-dimensional subspaces in $\mathbb{R}^d$, spanned by $\mathbf{U}$ and $\mathbf{V}$, respectively, and assume the subspaces intersect transversally; *i.e.,* the columns of $\mathbf{U}$ and $\mathbf{V}$ are linearly independent. It follows that the columns of $\mathbf{U}^{\otimes 2}$ and $\mathbf{V}^{\otimes 2}$ are linearly independent, as well. Thus, the set

$$\{\mathbf{U}^{\otimes 2}\boldsymbol{\theta}^{\otimes 2} : \boldsymbol{\theta} \in \mathbb{R}^r\} \cup \{\mathbf{V}^{\otimes 2}\boldsymbol{\theta}^{\otimes 2} : \boldsymbol{\theta} \in \mathbb{R}^r\}$$

spans a $2\binom{r+1}{2}$ dimensional subspace of the ambient $\binom{d+1}{2}$ dimensional space. More generally, if $\mathbf{x}$ belongs to a union of K subspaces, then the span of $\mathbf{x}^{\otimes 2}$ is a $K\binom{r+1}{2}$ dimensional subspace of the ambient $\binom{d+1}{2}$ dimensional space.