The Veronese Hitting Subspace Method for Subspace Clustering with Missing Data

Christopher Gadzinski*

August, 2022

Abstract

The well-known problem of *low-rank matrix completion* (LRMC) asks to recover a subspace from a sample of points, with the complication that some coordinates of each point may be hidden. *Subspace clustering with missing data* (SCMD) is an extension of this problem for unions of subspaces. In a recent work [OPAB+21], Ongie, Pimentel, and collaborators proposed a way to transform SCMD into LRMC over a polynomial transformation of the dataset, which they called *low algebraic dimension matrix completion* (LADMC). Under some conditions, they showed that the two problems were formally equivalent, allowing SCMD to be solved by well-known techniques for LRMC.

We now suggest a simple improvement to LADMC called the Veronese hitting subspace (VHS) problem. We focus on the limiting case where arbitrary amounts of data are available, subject only to a bound r on the number of coordinates available per data point. Given a generic union of k subspaces of dimension d, SCMD is well-posed with $r = d + 2$. Based on computational evidence, we conjecture that the VHS problem is formally equivalent to SCMD for this value of r so long as $kd < n$. In contrast, LADMC requires $r \approx \sqrt{k}d$ coordinates per data point. In light of this improvement, we propose the VHS method as a basis for new rank minimization-based SCMD algorithms.

*Some work by this author was performed in the scope of the MobiWise project (P2020 SAICTPAC/0011/2015), co-financed by COMPETE 2020, Portugal 2020—Operational Program for Competitiveness and Internationalization (POCI), European Unions ERDF (European Regional Development Fund), and the Portuguese Foundation for Science and Technology (FCT).

1 Introduction

Suppose $X = \begin{bmatrix}\mathbf{x}_1, \ldots, \mathbf{x}_m\end{bmatrix}$ is a matrix of data points, some of whose coordinates are not observed. How can we infer relationships in this kind of "incomplete" dataset? Such questions are highly motivated by the design of *collaborative filtering* systems, where incomplete reports from many individual cases are combined "collaboratively" to learn about an underlying distribution.

Let the "incomplete observation" of a data point $\mathbf{x}_i$ be represented by a linear equation $\pi_i(\mathbf{x}_i) = \tilde{\mathbf{x}}_i$. If we believe that our data belongs to a proper subspace of $\mathbb{R}^n$, a natural way to complete X is to solve

$$\begin{aligned} \textbf{minimize} \quad & \operatorname{rank} X \\ \textbf{subject to} \quad & \forall i,\ \pi(\mathbf{x}_i) = \tilde{\mathbf{x}}_i. \end{aligned}$$

This problem, called *low-rank matrix completion* (LRMC), gives the best policy for exact completion of a rank-deficient matrix. This problem has received a large amount of research interest in the past 15 years, beginning with the discovery of a convex relaxation (nuclear norm minimization) that is formally equivalent to LRMC under practical assumptions in [CR08] and [CT09]. This observation led to efficient iterative algorithms, like [MGC11]. Methods not based on nuclear norm minimization are also available; for example, see [MW11], which proposes an algorithm to minimize Schatten p-norm of X.

Now, suppose our data is drawn from a *union of subspaces*. When we observe all coordinates of each data point, the problem of reconstructing the union of subspaces and labeling data points by subspace is called *subspace clustering* (SC). When some coordinates of each data point could be missing, as in LRMC, the reconstruction problem is called *subspace clustering with missing data* (SCMD). Both SC and SCMD have been a focus of study in the past decade. For a survey of recent SCMD techniques, see [LBY$^+$19].

One approach, developed by Ongie, Pimentel et al. in [PAOB$^+$17] and [OPAB$^+$21], focuses on the fact that unions of subspaces are algebraic varieties, and in fact are mapped into relatively low-dimensional subspaces by polynomial feature maps with degree as low as two. Their main proposal, *low algebraic dimension matrix completion*, is to "lift" an SCMD problem to polynomial feature space and solve it with LRMC.

Let us take a moment to describe the LADMC method in more detail. Define the "Veronese map"

$$\begin{aligned} &\eta \colon \mathbb{R}^n \to \mathbb{R}^{\binom{n+1}{2}} \\ &\eta(x_1, \ldots, x_n) = (x_i x_j)_{1 \le i \le j \le n}, \end{aligned}$$

which is the "feature map" for the space of quadratic forms. A subspace $V \subseteq \mathbb{R}^n$ of dimension d is mapped into a subspace of dimension $\binom{d+1}{2}$ of $\mathbb{R}^{\binom{n+1}{2}}$ by η, and so a union of subspaces is indeed sent into a relatively low-dimensional subspace under η. Furthermore, if a data point $\mathbf{x} \in \mathbb{R}^n$ is observed at some subset of k coordinates, then $\binom{k+1}{2}$ coordinates of the monomial vector $\eta(\mathbf{x})$ can be calculated. We are lead to pose the problem of choosing vectors $\mathbf{y}_i \in \mathbb{R}^{\binom{n+1}{2}}$, each agreeing with the known coordinates of $\eta(\mathbf{x}_i)$, such that the rank of $\left[\mathbf{y}_1, \ldots, \mathbf{y}_m\right]$ is minimized. If $\eta(X) = \left[\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\right]$ happens to be obtained as the solution to this problem, then the vectors $\mathbf{x}_i$ can be recovered easily by solving the equations $\eta(\mathbf{x}_i) = \mathbf{y}_i$.

However, it is not immediately clear when this rank minimization problem gives us the solution we want; it is possible for a vector $\mathbf{y}_i$ to agree with the known coordinates of $\eta(\mathbf{x}_i)$ but not be of the form $\eta(\mathbf{z})$ for *any* $\mathbf{z} \in \mathbb{R}^n$. One perspective is to ask how many coordinates need to be observed *per data point* in the original problem for $\eta(X)$ to be recovered via matrix rank minimization, assuming sufficiently many data points are observed at sufficiently diverse subsets of coordinates. In [OPAB$^+$21], the answer was shown to be around $\sqrt{k}d$ coordinates per data point. This is a promising success but reveals LADMC to be sub-optimal; it is theoretically possible to infer our union of subspaces given only $d+2$ coordinates per data point.

In this note, we propose an improved "lifting" strategy. In essence, our idea is to constrain each column $\mathbf{y}_i$ by *all* linear relations implied on $\eta(\mathbf{x}_i)$ by the observed coordinates of $\mathbf{x}_i$, some of which are not coordinate-wise constraints. For example, if the first two coordinates of $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are observed, then only three coordinates of $\eta(\mathbf{x})$ are known—namely, x_1^2, x_1x_2, and x_2^2—but we also know the ratios between x_1x_i and x_2x_i for all $i > 2$, which amounts to an additional $n-2$ linear relations. Although we do not consider the problem of implementing a numerical method here, various LRMC algorithms can be modified to use these slightly more general constraints on columns.

Our analysis is simplified by the introduction of a new problem formulation, which we call the *hitting subspace* problem, leading us to call our method the *Veronese hitting subspace* (VHS) method. The conditions under which $\eta(X)$ is a solution to the VHS problem are not yet fully understood. However, computational evidence gives some reason to believe this occurs if $d+2$ coordinates per data point are given in the original dataset, assuming (as in Ongie's paper) that the dataset was sufficiently large and diverse. If true, the VHS problem provides a formally equivalent relaxation to SCMD that retains optimal economy in the coordinates per data point metric. We propose that the VHS method is worth investigating to develop new methods for SCMD.

2 The Hitting Subspace Problem

Definition 1. *Let $\mathcal{W}$ be a family of subspaces in $\mathbb{R}^n$. A* ***hitting subspace*** *for $\mathcal{W}$ is a subspace V of $\mathbb{R}^n$ so that*

$$\forall W \in \mathcal{W}, \dim V \cap W \geq 1.$$

The ***hitting subspace problem*** *is to find a hitting subspace for $\mathcal{W}$ with smallest possible dimension.*

We have named this in analogy with the *hitting set problem*, especially well-known in computer science, which asks to find a set V of minimum cardinality which intersects each member of a family $\mathcal{W}$ of finite sets.

Generically, LRMC can be interpreted as a special case of the hitting subspace problem. Indeed, the observed entries of a certain column $\mathbf{x}$, encoded as a linear constraint $\pi(\mathbf{x}) = \tilde{\mathbf{x}}$, simply force the column space V of our matrix to intersect the affine space $\pi^{-1}(\tilde{\mathbf{x}})$. Assuming that $\ker \pi$ intersects V trivially, this is equivalent to V inciding in $W = \langle \mathbf{x} \rangle + \ker \pi$. Thus, assuming that the coordinate projection associated with each column is injective on V, low-rank matrix completion is essentially equivalent to a hitting subspace problem. The converse is not true since an incident subspace W in a hitting subspace problem need not be of the form $\langle \mathbf{x} \rangle + \ker \pi$ for a coordinate projection π.

Let Gr_k be the Grassmannian manifold of k-dimensional subspaces in $\mathbb{R}^n$. This is an algebraic variety and a homogeneous space of dimension $k(n-k)$. The condition of intersecting another subspace W non-trivially gives an algebraic constraint on Gr_k.

Definition 2. *Given $W \subseteq \mathbb{R}^n$, let $\mathrm{Ind}_k(W)$ be the set*

$$\mathrm{Ind}_k(W) = \{V \in \mathrm{Gr}_k : \dim V \cap W \geq 1\}.$$

Let d be the dimension of W. $\mathrm{Ind}_k(W)$ is a proper subvariety of Gr_k exactly when $k + d \leq n$. In the algebraic geometry literature, $\mathrm{Ind}_k(W)$ is a special case of a *Schubert variety*. The following proposition characterizes the geometry of $\mathrm{Ind}_k(W)$ around a point V where $\dim V \cap W = 1$.

Proposition 1. *There is a (natural) isomorphism*

$$T_V \, \mathrm{Gr}_k \cong \mathrm{Hom}(V, \mathbb{R}^n / V).$$

Let $V \in \mathrm{Ind}_k(W)$ with $\dim V \cap W = 1$. Then $\mathrm{Ind}_k(W)$ is a regular submanifold around V, and under the isomorphism above,

$$T_V \, \mathrm{Ind}_k(W) = \{f : f(V \cap W) \subseteq W/V\}.$$

In particular,

$$\operatorname{codim} T_V \, \mathrm{Ind}_k(W) = \operatorname{codim} W/V = n - k - d + 1.$$

A proof of Proposition 1 is available in the Appendix.

For our purposes, it will be slightly more useful to think of Proposition 1 in terms of dual spaces. (Recall in that, when $V \subseteq \mathbb{R}^n$ is any subspace, $V^0 \subseteq (\mathbb{R}^n)^*$ is the *annihilator subspace* of functionals vanishing on V.)

Corollary 1. *There is a natural isomorphism*

$$T_V^* \operatorname{Gr}_k \cong V \otimes V^0.$$

Under this isomorphism, for $V \in \operatorname{Ind}_k(W)$ with $\dim V \cap W = 1$,

$$[T_V \operatorname{Ind}_k(W)]^0 \cong (V \cap W) \otimes (V^0 \cap W^0).$$

This local characterization of the sets $\operatorname{Ind}_k(W)$ lets us define the following well-posedness condition for a hitting subspace problem.

Definition 3. *Let $\mathcal{W}$ be a family of subspaces. A k-dimensional subspace V is* ***transversally determined*** *by $\mathcal{W}$ when*

$$\bigcap_{\substack{W \in \mathcal{W} \\ \dim W \cap V = 1}} T_V \operatorname{Ind}_k(W) = \{0\}.$$

Concretely, this means that

$$V \otimes V^0 = \sum_{\substack{W \in \mathcal{W} \\ \dim W \cap V = 1}} (V \cap W) \otimes (V^0 \cap W^0).$$

If V is transversally determined, then V is at least a unique *local* solution to the hitting subspace problem. Although it is possible for V to be a unique solution without being transversally determined, we will focus on transversally determined solutions here. We believe that this does not limit our analysis significantly.

Let us investigate the transversal determinacy of an LRMC problem. In this case, the incident subspaces W are of the form $\langle \mathbf{x} \rangle + K$, where $\mathbf{x}$ is an element of V and K is a kernel of some coordinate projection. Specifically, we will consider coordinate projections with a fixed rank r.

Definition 4. *Let $\mathcal{K}_r$ be the family of subspaces generated by sets of $r - n$ distinct basis vectors of $\mathbb{R}^n$.*

For V to be transversally determined by any collection of incident subspaces of the form we have described, it must of course be transversally determined by the whole family

$$\{\langle \mathbf{x} \rangle + K : \mathbf{x} \in V, K \in \mathcal{K}_r\}.$$

In this case, we will say that V is transversally determined in the *sampling limit*, given r observed coordinates per data point.

Informally, the most we can learn about V by observing its image modulo K is the space $V^0 \cap K^0$ of linear relations that hold on V and factor through the quotient map $\mathbb{R}^n \to \mathbb{R}^n/K$. If the whole annihilator V^0 can be recovered by summing together these intersections as K runs over a family $\mathcal{K}$, we will say that V is "identifiable" under $\mathcal{K}$.

Definition 5. *The subspace $V \subseteq U$ is* ***identifiable*** *by a family $\mathcal{K}$ when*

$$V^0 = \sum_{K \in \mathcal{K}} V^0 \cap K^0.$$

In fact, is it straightforward to see that this notion of identifiability exactly characterizes transversal determinacy of an LRMC-type problem in the sampling limit.

Proposition 2. *Let $\mathcal{K}$ be a family of subspaces with $V \cap K = 0$ for all $K \in \mathcal{K}$. Then V is transversally determined by the family*

$$\{\langle \mathbf{x} \rangle + K : \mathbf{x} \in V, K \in \mathcal{K}\}$$

iff V is identifiable by the family $\mathcal{K}$.

Proof. For each $\mathbf{x}$ and K,

$$\begin{aligned} [T_V \operatorname{Ind}_k(\langle \mathbf{x} \rangle + K)]^0 &= (V \cap (\langle \mathbf{x} \rangle + K)) \otimes (V^0 \cap (\langle \mathbf{x} \rangle + K)^0) \\ &= \langle \mathbf{x} \rangle \otimes (V + \langle \mathbf{x} \rangle + K)^0 = \langle \mathbf{x} \rangle \otimes (V^0 \cap K^0). \end{aligned}$$

Such subspaces span all of $V \otimes V^0$ iff V is identifiable under $\mathcal{K}$. □

It is easy to see that a generic subspace $V \subseteq \mathbb{R}^n$ of dimension $d < n$ is identifiable under $\mathcal{K}_r$ exactly when $d < r$, and so, in the sampling limit, LRMC is well-posed in the sense of transversal determinacy under exactly this condition. In fact, even without the tools we have introduced in this section, it is simple to argue that a generic subspace V of dimension $d < r$ will be the unique subspace of minimum dimension agreeing with its images modulo each element of $\mathcal{K}_r$. However, transversal determinacy is a convenient tool to investigate well-posedness of hitting subspace problems that are not in the LRMC form, as will be the case of the VHS problem.

3 Hitting Subspace Relaxations of SCMD

Let $\Lambda = V_1 \cup \ldots \cup V_k$ be a union of subspaces. For simplicity, we assume that every subspace has the same dimension d. Additionally, we suppose that the subspaces V_i are chosen "generically," meaning that we may exclude proper Zariski-closed sets of configurations from consideration.

Suppose points are sampled from Λ but observed modulo elements of $\mathcal{K}_r$ for some $r < n$. The problem of *subspace clustering with missing data* (SCMD) is to recover Λ from this information. In this section, we propose the new Veronese hitting subspace problem and recall its predecessor, LADMC. Both of these are attempts to encode SCMD as a hitting subspace problem over a higher-dimensional ambient space—namely, the feature space of quadratic forms on $\mathbb{R}^n$. Although higher-order feature spaces were considered in [OPAB+21], we restrict our focus to the quadratic case for simplicity.

We begin with some definitions.

Definition 6. *When V is a vector space, let SV be the symmetric tensor product $V \otimes V$, and define*

$$\begin{aligned} &\eta \colon V \to SV \\ &\eta(\mathbf{x}) = \mathbf{x} \otimes \mathbf{x}. \end{aligned}$$

When $f\colon V \to W$ is a linear map, let $Sf\colon SV \to SW$ be the unique linear map satisfying the equation

$$(Sf)(\eta(\mathbf{x})) = \eta(f(\mathbf{x}))$$

Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for V and $\{\mathbf{e}_i \otimes \mathbf{e}_j : 1 \leq i \leq j \leq n\}$ be a basis for SV. Then we can express η in coordinates as

$$\eta(x^1\mathbf{e}_1 + \ldots + x^n\mathbf{e}_n) = \sum_{1 \leq i \leq j \leq n} x^i x^j \mathbf{e}_i \otimes \mathbf{e}_j.$$

The map η is the "feature map" for quadratic forms over V, in the sense that any quadratic form $q\colon V \to \mathbb{R}$ equals a composition $\lambda \circ \eta$ for a unique linear map $\lambda\colon SV \to \mathbb{R}$. Note also that elements in the image of η—the simple tensors—generate SV. Meanwhile, Sf is characterized by the equations

$$(Sf)(\mathbf{e}_i \otimes \mathbf{e}_j) = f(\mathbf{e}_i) \otimes f(\mathbf{e}_j).$$

Note also that our definitions make S a covariant functor. Using S, we now define two natural ways to transform a subspace $V \subseteq \mathbb{R}^n$ into a subspace of $S\mathbb{R}^n$.

Definition 7. *When V is a subspace of $\mathbb{R}^n$ with inclusion $\iota_V : V \to \mathbb{R}^n$ and projection $\pi_V : \mathbb{R}^n \to \mathbb{R}^n/V$, define*

$$S(V) = \operatorname{im} S\iota_V, \quad Q(V) = \ker S\pi_V.$$

Note that the maps $S\iota_V$ and $S\pi_V$ participate in the following commutative diagram.

$$\begin{tikzcd}
V \arrow[r, "\iota_V"] \arrow[d, "\eta"'] & \mathbb{R}^n \arrow[r, "\pi_V"] \arrow[d, "\eta"'] & \mathbb{R}^n/V \arrow[d, "\eta"'] \\
SV \arrow[r, "S\iota_V"'] & S\mathbb{R}^n \arrow[r, "S\pi_V"'] & S(\mathbb{R}^n/V)
\end{tikzcd}$$

By inspection of this diagram, we find that $S(V) \subseteq Q(V)$; indeed, $S(V)$ is spanned by elements of the form $\eta(\mathbf{v})$ for $\mathbf{v} \in V$, but every such element is mapped to zero by $S\pi_V$ because $S\pi_V(\eta(\mathbf{v})) = \eta(\pi_V(\mathbf{v})) = 0$. However, this inclusion is proper when V is proper and non-zero, which we can see by examining these two subspaces in coordinates; when $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$ such that $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ is a basis for V,

$$\begin{aligned}
S(V) &= \langle \mathbf{e}_i \otimes \mathbf{e}_i : 1 \leq i, j \leq k \rangle, \text{ while} \\
Q(V) &= \langle \mathbf{e}_i \otimes \mathbf{e}_j : 1 \leq i \leq k \vee 1 \leq j \leq k \rangle.
\end{aligned}$$

Thus, when $\dim V = k$, we have

$$\dim S(V) = \binom{k+1}{2}, \ \dim Q(V) = \binom{k+1}{2} + k(n-k).$$

Viewing the dual space $(S\mathbb{R}^n)^*$ as the quadratic forms on $\mathbb{R}^n$, the annihilator $S(V)^0$ consists of the forms which vanish on V, while $Q(V)^0$ has the forms that are constant along the cosets of V. It is also helpful to note that, identifying $(S\mathbb{R}^n)^* \cong S(\mathbb{R}^n)^*$, we have $S(V^0) = Q(V)^0$.

Finally, we describe the relevance of the quadratic feature space to the study of unions of subspaces. When $\mathcal{V} \subseteq S\mathbb{R}^n$ is a subspace, $\eta^{-1}(\mathcal{V})$ can be understood as the loci of a family of quadratic polynomials on $\mathbb{R}^n$. Let us call this a **quadratic set**. As it turns out, unions of subspaces in $\mathbb{R}^n$ are frequently quadratic sets. For example, we have the following sufficient condition.

Proposition 3. *A union $\Lambda = V_1 \cup \ldots \cup V_k$ of d-dimensional subspaces in general position is a quadratic set if $n \geq kd$.*

Proof of Proposition 3. Any hyperplane or union of two hyperplanes is the zero set of some quadratic form. From this, we deduce easily that any subspace or union of two subspaces is a quadratic set.

Suppose now that $k > 2$. For each $i = 1, \ldots, k$, define

$$\Lambda_i = V_i \cup \sum_{j \neq i} V_j.$$

We claim that $\Lambda = \bigcap_i \Lambda_i$. An intersection of quadratic sets is a quadratic set, so this will prove the theorem.

It is clear that $\Lambda \subseteq \bigcap_i \Lambda_i$. On the other hand, if $\mathbf{x} \in \bigcap_i \Lambda_i$ but $\mathbf{x} \notin \Lambda$, then

$$\mathbf{x} \in \bigcap_i \sum_{j \neq i} V_j.$$

In particular, $\mathbf{x} \in \sum_i V_i$. By hypothesis that none of the subspaces V_i intersect (generically true) and that $n \geq kd$, we can find a basis for $\sum_i V_i$ that partitions into bases for each subspace V_i. If $\mathbf{x} \in \sum_{j \neq i} V_j$ for a given i, then the coefficients of x in the basis vectors for V_i are zero. Since this is true for each i, we conclude that $x = 0$, which is a contradiction. This proves that $\bigcap_i \Lambda_i \subseteq \Lambda$, so indeed $\Lambda = \bigcap_i \Lambda_i$. □

Now, we return to the problem of SCMD. When the union Λ is a quadratic set, as in the previous proposition, it is fully encoded by the subspace $\mathcal{V} = \langle \eta(\Lambda) \rangle$. It is also obvious that, when enough vectors $\{\mathbf{x}_i\}$ are sampled in general position from Λ, $\mathcal{V}$ will be the linear span of $\{\eta(\mathbf{x}_i)\}$. How can we recover $\mathcal{V}$ when the points $\mathbf{x}_i$ are only observed at some coordinates, encoded by the equations $\pi_i(\mathbf{x}_i) = \tilde{\mathbf{x}}_i$?

The idea of recovering a union of subspaces by fitting a subspace to a polynomial transformation of a dataset was introduced in [VMS03] and is known as generalized principal component analysis. However, GPCA cannot be applied easily in the case of partially observed data. We are led to pose the harder problem of determining a minimal-dimensional subspace $\mathcal{V}$ that has non-trivial intersection with each image $\eta(\mathbf{x}_i + \ker \pi_i)$. This was done in [OWNB17], and an algorithm to solve such problems was proposed. However, there may be a computational benefit in relaxing our optimization problem to one of an LRMC or hitting subspace type, for which many effective algorithms are available.

This idea, in the form of LADMC, was defined first in [PAOB+17] and later clarified in [OPAB+21]. In our language, the LADMC problem can be stated in the following way.

Definition 8. ***Low algebraic dimension matrix completion (LADMC)*** *is the low-rank matrix completion problem for the matrix* $[\mathbf{y}_1, \ldots, \mathbf{y}_m]$*, with columns* $\mathbf{y}_i \in S\mathbb{R}^n$*, subject to the equations*

$$S\pi_i(\mathbf{y}_i) = \eta(\tilde{\mathbf{x}}_i),\ i = 1, \ldots, m.$$

If we assume that the projections $S\pi_i$ are injective on $\mathcal{V}$, then LADMC can also be understood as the hitting subspace problem for the subspaces

$$\langle \eta(\mathbf{x}_i) \rangle + \ker S\pi_i = \langle \eta(\mathbf{x}_i) \rangle + Q(\ker \pi_i).$$

In fact, we will see in Proposition 6 that LADMC cannot succeed in a regime where the projections $S\pi_i$ are not injective on $\mathcal{V}$, so this conversion to a hitting subspace problem is justified. On the other hand, from the point of view of hitting subspace problems, the following problem—our original suggestion—arises as the tightest possible hitting subspace-type relaxation of the constraints that $\mathcal{V}$ intersect the sets $\eta(\langle \mathbf{x}_i \rangle + \ker \pi_i)$ non-trivially.

Definition 9. *The* ***Veronese hitting subspace (VHS) problem*** *is the hitting subspace problem for the family*

$$\{S(\langle \mathbf{x}_i \rangle + \ker \pi_i) : i = 1, \ldots, m\}.$$

For any $K = \ker \pi$, note that

$$S(\langle \mathbf{x} \rangle + K) \subseteq \langle \eta(\mathbf{x}) \rangle + Q(K),$$

since the right-hand side are the solutions $\mathbf{y}$ to the equation

$$S\pi(\mathbf{y}) = S\pi(\eta(\mathbf{x})),$$

while the left-hand side is generated by elements of the form $\eta(\mathbf{z})$ with $\mathbf{z}$ satisfying $\pi(\mathbf{z}) = \pi(\mathbf{x})$. Thus, LADMC is a relaxation of VHS. Indeed, let $K \in \mathcal{K}_r$ be $(n-r)$-dimensional. Then the incident subspace $S(\langle \mathbf{x} \rangle + K)$ used by the VHS problem will have codimension

$$\dim S(\langle \mathbf{x} \rangle + K)^0 = \dim Q((\langle \mathbf{x} \rangle + K)^0) = \binom{r}{2} + (r-1)(n-r+1).$$

in $S\mathbb{R}^n$. By comparison, the incident subspace $\langle \eta(\mathbf{x}) \rangle + Q(K)$ used by LADMC has only codimension

$$\dim Q(K)^0 - 1 = \dim S(K^0) - 1 = \binom{r+1}{2} - 1 = \binom{r}{2} + r - 1.$$

From our discussion in Section 2—specifically, Proposition 1—we know that an incident subspace of codimension c will constrain $\mathcal{V}$ by

$$\max\{c + 1 - \dim \mathcal{V}, 0\}$$

degrees of freedom. Thus, the extra $(r-1)(n-r)$ codimensions possessed by the subspaces $S(\langle \mathbf{x} \rangle + K)$ may give the VHS method a significant advantage over LRMC.

However, we have not yet been able to clarify the VHS problem's advantage rigorously. In the next section, we present all that is currently known.

4 Questions and Partial Results

As in our simple analysis of LRMC in Section 2, we consider the problem of success in the "sampling limit." That is, we ask how many coordinates must be observed per data point for either LADMC or the VHS problem to be well-posed (in the sense of transversal determinacy) if arbitrarily many data points are observed with each possible subset of coordinates.

Definition 10. *For an arbitrary set $\Lambda \subseteq \mathbb{R}^n$, and let $\mathcal{V} = \langle \eta(\Lambda) \rangle$.*

1. $\mathrm{VHS}(\Lambda)$ *is the smallest $r \in \{1, \ldots, n\}$ for which $\mathcal{V}$ is transversally determined by*

$$\{S(\langle \mathbf{x} \rangle + K) : K \in \mathcal{K}_r, \mathbf{x} \in \Lambda\}.$$

2. $\mathrm{LADMC}(\Lambda)$ *is the smallest $r \in \{1, \ldots, n\}$ for which $\mathcal{V}$ is transversally determined by*

$$\{\langle \eta(\mathbf{x}) \rangle + Q(K) : K \in \mathcal{K}_r, \mathbf{x} \in \Lambda\}.$$

In the trivial case where $r = n$, both the VHS problem and LADMC obviously transversally determine $\mathcal{V}$. Furthermore, if either problem is determined with K running over $\mathcal{K}_r$, it is also determined with K running over $\mathcal{K}_{r'}$ for any $r' \geq r$. Thus, the functions VHS and LADMC are well-defined and fully characterize the behavior of our problems in the sampling limit. However, while unknown coordinates of data points can be completed so long as $\mathcal{V}$ is transversally determined—since by assumption of transversal determinacy the incident subspace of $S\mathbb{R}^n$ corresponding to each data point has one-dimensional intersection with $\mathcal{V}$—the union Λ itself cannot necessarily be inferred from $\mathcal{V}$ without additional assumptions. One sufficient assumption is that Λ is a quadratic set, as defined above. To understand the theoretical applicability of our subspace hitting-based methods, we propose the following questions.

Question 1. *When is the union of subspaces Λ a quadratic set?*

Question 2. *For Λ a quadratic set, what are* $\mathrm{LADMC}(\Lambda)$ *and* $\mathrm{VHS}(\Lambda)$*?*

In answer to Question 1, we have only our sufficient condition of $kd \leq n$ in Proposition 3. Question 2 is the topic of the rest of this section.

Let us begin with an easier question, the answer to which we understand well: for what value of r is recovery of Λ from a partially observed dataset possible in the sampling limit by *any* method? In other words, when is Λ determined by its images under rank-r coordinate projections? Every such image is trivial when $r \leq d$, so inference is impossible unless $r > d$. On the other hand, the following proposition shows that $r = d + 2$ is already enough.

Proposition 4. *Let $\Lambda = V_1 \cup \ldots \cup V_k$ be a generic union of d-dimensional subspaces. There is a fixed function that, given the images of Λ under every subspace in $\mathcal{K}_{d+2}$, returns Λ.*

For simplicity we assume in the following proof that this function depends on k and d. Obviously, this restriction can be lifted.

Proof. For each coordinate subspace K, let π_K be the projection modulo K. For each $K \in \mathcal{K}_{d+1}$, let $H(K) = \pi_K^{-1}(\pi_K(\Lambda))$, which is a union of k hyperplanes in $\mathbb{R}^n$. Each hyperplane of $H(K)$ contains a unique subspace V_i. Given two coordinate subspaces $K, L \in \mathcal{K}_{d+1}$, construct a bijection $\rho_K^L \colon H(K) \to H(L)$ by putting hyperplanes that contain the same subspaces in correspondence.

Together, the sets $H(K)$ for $K \in \mathcal{K}_{d+1}$ and the bijections ρ_K^L for $K, L \in \mathcal{K}_{d+1}$ determine Λ, since

$$\Lambda = \bigcup_{W \in H(K)} \bigcap_{L \in \mathcal{K}_{d+1}} \rho_K^L(W).$$

It is obvious that the images of Λ modulo elements of $\mathcal{K}_{d+1}$ are determined by its images modulo elements of $\mathcal{K}_{d+2}$. To conclude the proof, it remains to argue that the assignments ρ_K^L are also determined by this information.

Suppose K and L differ in only one basis vector so that $K \cap L \in \mathcal{K}_{d+2}$. Then ρ_K^L can be determined from the image $\pi_{K\cap L}(\Lambda)$ by corresponding hyperplanes that are images of the same subspace of $\pi_{K\cap L}(\Lambda)$ under the natural projections

$$\begin{tikzcd} \mathbb{R}^n/(K \cap L) \arrow[d] \arrow[dr] & \\ \mathbb{R}^n/K & \mathbb{R}^n/L\,. \end{tikzcd}$$

Since the whole set $\mathcal{K}_{d+1}$ is spanned by paths whose edges connect pairs of coordinate subspaces differing in one coordinate, we are done. $\square$

When $d = 1$ and $k \geq 3$, it can be argued that no fewer than $d + 2 = 3$ coordinates per data point let us reconstruct the union. However, due to computational experiments described below, we believe the statement of the previous proposition remains true with $\mathcal{K}_{d+1}$ instead of $\mathcal{K}_{d+2}$ when $d \geq 2$.

We now return to Question 2. How do LADMC(Λ) and VHS(Λ) compare to the essentially ideal value of $d + 2$? Our first result is that LADMC(Λ) $\approx \min\{\sqrt{k}d, n\}$ for any generic union of subspaces. This was already shown in [OPAB+21]. Furthermore, since the hitting subspaces used in the VHS problem are strictly smaller than those used in LADMC, it is clear that VHS(Λ) $\leq$ LADMC(Λ). Unfortunately, we were not able to prove a better bound on VHS(Λ). In particular, the kind of argument used for LADMC does not seem to generalize to the VHS problem. However, based on computer verification of transversal determinacy in simple cases, we conjecture VHS(Λ) $\leq d + 2$ so long as $kd \leq n$.

We begin with LADMC. As a problem of LRMC type, it is amenable to the kind of analysis we performed in Section 2.

Proposition 5. LADMC(Λ) $\leq r$ *if and only if* $\mathcal{V}$ *is identifiable under*

$$Q(\mathcal{K}_r) = \{Q(K) : K \in \mathcal{K}_r\}$$

and $S\pi_K$ *is injective on* $\mathcal{V}$ *for each* $K \in \mathcal{K}_r$.

Proof. LADMC$(\Lambda) \le r$ when $\mathcal{V}$ is transversally determined under

$$\mathcal{F}_r = \{\langle \eta(\mathbf{x})\rangle + Q(K) : K \in \mathcal{K}_r, \mathbf{x} \in \Lambda\},$$

meaning that

$$\sum_{\substack{W \in \mathcal{F}_r \\ \dim W \cap \mathcal{V} = 1}} (\mathcal{V} \cap W) \otimes (\mathcal{V}^0 \cap W^0) = \mathcal{V} \otimes \mathcal{V}^0.$$

Let $W = \langle \eta(\mathbf{x})\rangle + Q(K)$. Since $\eta(\mathbf{x}) \in \mathcal{V}$,

$$W \cap \mathcal{V} = (\langle \eta(\mathbf{x})\rangle + Q(K)) \cap \mathcal{V} = \langle \eta(\mathbf{x})\rangle$$

exactly when $Q(K)$ has zero-dimensional intersection with $\mathcal{V}$, meaning $S\pi_K$ is injective on $\mathcal{V}$. Furthermore, by genericity, this holds for one subspace $K \in \mathcal{K}_r$ exactly when it holds for all of them. In this case, we have (as in the proof of Proposition 2) that

$$\mathcal{V} \cap W \otimes (\mathcal{V}^0 \cap W^0) = \langle \eta(\mathbf{x})\rangle \otimes Q(K)^0.$$

These subspaces span $\mathcal{V} \otimes \mathcal{V}^0$ exactly when $\mathcal{V}$ is $Q(\mathcal{K}_r)$-identifiable. □

The injectivity of $S\pi_K$ on $\mathcal{V}$ is necessary for transversal determinacy. Assuming this holds, we have shown that Λ can be recovered from its rank-r projections by LADMC exactly when the space $\mathcal{V}^0$ of quadratic relations holding on Λ is generated by polynomials involving r coordinates.

When does this happen? Since

$$\mathcal{V} = \langle \eta(\Lambda)\rangle = S(V_1) + \ldots + S(V_k),$$

we know that $\dim \mathcal{V} \le k\binom{d+1}{2}$. Thus, the image of $\mathcal{V}$ under $S\pi$ is proper, for a projection π onto r coordinates, so long as $k\binom{d+1}{2} < \binom{r+1}{2}$. In this case, at least one polynomial involving r coordinates exists in $\mathcal{V}^0$. Actually, an argument first introduced in [PAOB+17] shows that $\mathcal{V}$ is $Q(\mathcal{K}_r)$-identifiable in essentially this situation. We reproduce it here, with slightly different language. A proof can be found in the Appendix.

Proposition 6. *Suppose $2d \le r \le n$. If $\mathcal{V}$ is $Q(\mathcal{K}_r)$-identifiable, then*

$$k\binom{d+1}{2} < \binom{r+1}{2}. \tag{1}$$

Conversely, if (1) holds as a non-strict inequality, then $\mathcal{V}$ is $Q(\mathcal{K}_{r'})$-identifiable for $r' = \min\{r+2, n\}$.

From this it follows that LADMC$(\Lambda) \approx \min\{\sqrt{k}d, n\}$ for large k and d. Also note that, while the definition of LADMC(Λ) involves the dimension n of the ambient space of Λ, LADMC(Λ) essentially does not depend on n when $k\binom{d+1}{2} < \binom{n+1}{2}$, and in particular when $kd \le n$.

Next we turn to the case of VHS, which is our original contribution. Using our criteria for transversal determinacy, we wrote a small program in the Julia language to experimentally compute VHS(Λ) for small randomly generated unions Λ.[1] We found, for example, the following values

[1]Source is available here: `https://github.com/cgadski/vhs_number`.

of VHS(Λ) when Λ is a union of k subspaces of dimension d in $\mathbb{R}^{kd}$. (Our program is not yet well optimized and was unable to compute some positions of the following table, which we have left blank.)

k	**d**					
	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	2	3	4	5	6	7
3	3	3	4	5	6	
4	3	3	4	5		
5	3	3	4			
6	3	3				

Compare this with the *lower bounds* for LADMC(Λ) that follow from Proposition 6—namely, the smallest values r for which $\binom{r+1}{2} > k\binom{d+1}{2}$, and for which $2d \leq r$.

k	**d**					
	1	**2**	**3**	**4**	**5**	**6**
1	2					
2	2					
3	3	4	6	8	10	
4	3	5	7	9		
5	3	6	8			
6	4	6				

Apparently, VHS(Λ) does not grow with k like LADMC(Λ) does. Our conjecture is that the behavior we have observed experimentally generalizes for all k, d, and n.

Conjecture 1. *When $kd \leq n$,* VHS$(\Lambda) \leq d + 2$.

One possible approach to characterizing VHS(Λ) is to relate it to identifiability of $\mathcal{V}$ under some family of subspaces, as we did above for LADMC(Λ). Unfortunately, unlike LADMC, the VHS problem is not of LRMC type. However, note that

$$S(\langle \mathbf{x} \rangle + K) \supseteq \langle \eta(\mathbf{x}) \rangle + S(K),$$

so VHS$(\Lambda) \leq r$ implies $\mathcal{V}$ being identifiable under the family

$$S(\mathcal{K}_r) = \{S(K) : K \in \mathcal{K}_r\}.$$

Proposition 7. *If* VHS$(\Lambda) \leq r$*, then* Λ *is* $S(\mathcal{K}_r)$*-identifiable.*

As the following proposition shows, $S(\mathcal{K}_r)$-identifiability can be characterized in a similar way to $O(\mathcal{K}_r)$-identifiability. (A proof is deferred to the Appendix.) Unfortunately, $S(\mathcal{K}_r)$-identifiability turns out to be vacuous for large n, so nothing like the converse to Proposition 7 can hold in general. The best we can say is that Proposition 7 does not rule out our conjecture, since $\mathcal{V}$ will be $S(\mathcal{K}_{d+2})$-identifiable for $d + 2 \leq n$ when $n \geq (k + 1)d/2$, which is true whenever $n \geq kd$.

Proposition 8. *If $\langle \eta(\Lambda) \rangle$ is $S(\mathcal{K}_r)$-identifiable, then*

$$k\binom{d+1}{2} < \binom{r+1}{2} + r(n-r). \tag{2}$$

Conversely, if (2) *holds as a non-strict inequality, then V is $S(\mathcal{K}_{r'})$-identifiable with $r' = \min\{r+1, n\}$. In particular, V is $S(\mathcal{K}_{d+2})$-identifiable as long as $n \geq (k+1)d/2$ and $d+2 \leq n$.*

5 Conclusions and Future Work

In this paper, we introduced the VHS method as a potential tool for new SCMD algorithms. Our suggestion is based on a problem formulation (the hitting subspace problem) that can be viewed as a generalization of LRMC to the case of arbitrary linear constraints per data point. Based on a well-posedness criterion for the hitting subspace problem, we considered the question of how many coordinates per data point the VHS method needs to succeed at SCMD with enough samples. Based on a computational experiment, we have conjectured that this number may be essentially the smallest possible, which would improve significantly on the results previously obtained for LADMC.

We propose two topics for further inquiry based on our work. First, our conjecture on the function $\mathrm{VHS}(\Lambda)$ could be explored, providing more concrete guarantees on the success of the VHS method. Second, our method could be paired with software to solve the underlying hitting subspace problem, and the success of our method on SCMD problems in practice could be explored. Computational experiments were already performed on LADMC in [OPAB$^+$21] and had favorable results, and a VHS method-based algorithm should not do worse. A key question on this front is whether our ideal limit of $d+2$ coordinates per data point is sufficient for VHS-based inference in problems of moderate size; transversal determinacy might not be enough for practical rank minimization algorithms to succeed.

Appendix: Proofs

Proof of Proposition 1. Recall that $W \subseteq \mathbb{R}^n$ is a subspace of dimension d and $V \subseteq \mathbb{R}^n$ is a subspace of dimension k.

Consider the action of $\mathrm{GL}(n)$ on Gr_k by direct images, and let

$$\varphi_V \colon \mathrm{GL}(n) \to \mathrm{Gr}_k$$

be the map $g \mapsto gV$. At the identity, the differential $d\varphi_V$ is a surjective map from $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ onto $T_V \mathrm{Gr}_k$. It is also easy to see that its kernel coincides with the maps f that stabilize V in the sense that $f(V) \subseteq V$. Where ι_V and π_V are the inclusion and projection maps for V, $\ker d\varphi_V$ is also the kernel of the map

$$\begin{aligned}&\eta_V \colon \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^n) \to \mathrm{Hom}(V, \mathbb{R}^n/V)\\ &\eta_V(f) = \pi_V \circ f \circ \iota_V.\end{aligned}$$

This establishes an isomorphism $T_V \operatorname{Gr}_k \cong \operatorname{Hom}(V, \mathbb{R}^n/V)$ as the unique linear map completing the diagram

$$\begin{tikzcd}
\operatorname{Hom}(\mathbb{R}^n, \mathbb{R}^n) \arrow[r, "\eta_V"] \arrow[dr, "d\varphi_V"'] & \operatorname{Hom}(V, \mathbb{R}^n/V) \\
& T_V \operatorname{Gr}_k . \arrow[u, dashed]
\end{tikzcd}$$

Now, let $\ell = n - k - d + 1$, and let

$$\{v_1, \ldots, v_k, w_2, \ldots, w_d, x_1, \ldots, x_\ell\}$$

be a basis for $\mathbb{R}^n$ such that $\{v_1, \ldots, v_k\}$ is a basis for V and $\{v_1, w_2, \ldots, w_d\}$ is a basis for W. We will interpret the subspaces V and W as multivectors, putting $V = v_1 \wedge \ldots \wedge v_k$ and $W = v_1 \wedge w_2 \wedge \ldots \wedge w_d$, and define

$$\begin{aligned}
&F\colon \operatorname{GL}(n) \to \Lambda^{k+d}(\mathbb{R}^n) \\
&F(g) = gV \wedge W.
\end{aligned}$$

For $g \in \operatorname{GL}(n)$, we have $\varphi_V(g) \in \operatorname{Ind}_k(W)$ exactly when $F(g) = 0$. Actually, in a neighborhood of the identity $e \in \operatorname{GL}(n)$, it is enough for the coefficients of the multivectors $b_i = x_i \wedge V \wedge W$ to vanish from $F(g)$, considering the basis for $\Lambda^{k+d}(\mathbb{R}^n)$ associated with our basis for $\mathbb{R}^n$. Indeed, the projection of $F(g)$ onto b_i is non-zero exactly when the projection of $gV + W$ onto the coordinates

$$\{x_i, v_1, \ldots, v_k, w_2, \ldots, w_d\}$$

is full-dimensional, and $gV + W$ has full-dimensional projection onto the coordinates $\{v_1, \ldots, v_k, w_2, \ldots, w_d\}$ for g in a neighborhood of e.

We define $\pi\colon \Lambda^{k+d}(\mathbb{R}^n) \to \mathbb{R}^\ell$ to project onto the coordinates b_i, and consider the map $\pi \circ F$. We claim that $d(\pi \circ F)_e$ has full rank, so that the equation $(\pi \circ F)(g) = 0$ defines a submanifold in a neighborhood of e. We conclude that $\operatorname{Ind}_k(W)$ is a submanifold in a neighborhood of V, and an element $f \in T_V \operatorname{Gr}_k$ belongs to $T_V \operatorname{Ind}_k(W)$ exactly when there is a map $A \in \operatorname{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ with $d(\pi \circ F)(A) = 0$ and $\eta_V(A) = f$.

We compute that

$$\begin{aligned}
dF_e(A) &= \frac{d}{dt}_{t=0} e^{tA} V \wedge W = \frac{d}{dt} \left(\bigwedge_{i=1}^{k} e^{tA} v_i \right) \wedge W \\
&= \sum_{i=1}^{k} (-1)^{i+1} A v_i \wedge \bigwedge_{\substack{j=1 \\ j \neq i}}^{k} v_j \wedge W = A v_1 \wedge v_2 \wedge \ldots \wedge v_k \wedge W,
\end{aligned}$$

where the last equality follows because $v_1 \wedge W = 0$. Furthermore, π is clearly injective on the image of dF_e, and $d(\pi \circ F)_e$ has maximal rank ℓ. We also conclude that $dF(A) = 0$ exactly when $S(V \cap W) \subseteq V + W$. In terms of a tangent vector $\eta_V(A) \in \operatorname{Hom}(V, \mathbb{R}^n/V)$ of $\operatorname{Ind}_k(W)$, this means that

$$\eta_V(A)(V \cap W) \subseteq W/V.$$

This concludes the proof. $\square$

The following proofs of Propositions 6 and Proposition 8 depend on the following lemma derived from Theorem 3.2 in [CCG12].

Lemma 1. *Let $\Lambda = V_1 \cup \ldots \cup V_k \subseteq \mathbb{R}^n$ be a generic union of k subspaces with dimensions $d_1 \geq d_2 \geq \ldots \geq d_k$, with $d_1 + d_2 \leq n$. Then*

$$\dim\langle\eta(\Lambda)\rangle = \max\left\{\sum_i \binom{d_i+1}{2}, \binom{n+1}{2}\right\}.$$

In the following, Λ is a union of subspaces and $\mathcal{V} = \langle\eta(\Lambda)\rangle$. Also let $\{\mathbf{e}^1, \ldots, \mathbf{e}^n\}$ be a basis for $(\mathbb{R}^n)^*$, so that $\{\mathbf{e}^j \otimes \mathbf{e}^i : 1 \leq i \leq j \leq n\}$ is a basis for $(S\mathbb{R}^n)^*$. We make use of the following simple fact.

Lemma 2. *Let $V, W \subseteq \mathbb{R}^n$ be subspaces with $V \cap W = 0$, and let $\{b^1, \ldots, b^k\}$ be covectors spanning a complementary subspace to W^0. Then V^0 is generated by elements of the sets $W^0 + \langle b^i\rangle$.*

Proof of Proposition 6. Let $K \in \mathcal{K}_r$ be a coordinate subspace, which we assume w.l.o.g. is the kernel of the projection map π_K onto the first r coordinates, so that $K = \langle \mathbf{e}_{r+1}, \ldots, \mathbf{e}_n\rangle$ and $K^0 = \langle \mathbf{e}^1, \ldots, \mathbf{e}^r\rangle$. The subspace

$$Q(K)^0 = \langle \mathbf{e}^i \otimes \mathbf{e}^j : i, j \in \{1, \ldots, r\}\rangle$$

corresponds to the quadratic forms which can be written in terms of the first r coordinates, and the intersection $\mathcal{V}^0 \cap O(K)^0$ can be viewed as the subset of these that vanish on the projection $\pi_K(\Lambda)$.

Since $2d \leq r$, we can apply Lemma 1 to the image $\pi_K(\Lambda)$, a union of subspaces in $\mathbb{R}^r$. We find that no quadratic forms vanish on $\pi_K(\Lambda)$ if (1) does not hold. In this case, $\mathcal{V}$ is clearly not $O(\mathcal{K}_r)$-identifiable. Conversely, suppose (1) holds non-strictly. Then, Lemma 1 tells us that $\mathcal{V}$ and $\langle\eta(\pi_K(\Lambda))\rangle$ have the same dimension. Since $\langle\eta(\pi_K(\Lambda))\rangle = (S\pi_K)(V)$, we conclude that $Q(K)$, which by definition is the kernel of $S\pi_K$, intersects $\mathcal{V}$ trivially.

Applying Lemma 2, $\mathcal{V}$ is cut out by covectors belonging to

$$Q(K)^0 + \langle \mathbf{e}^i \otimes \mathbf{e}^j\rangle$$

as $\mathbf{e}^i \otimes \mathbf{e}^j$ ranges over the basis of $(S\mathbb{R}^n)^*$. Any such extension of $Q(K)^0$ is a subspace of $Q(K')^0$, where K' is derived from K by the eventual omission of the basis vectors $\mathbf{e}_i$ and $\mathbf{e}_j$. A viable subspace K' can always be found in the family $\mathcal{K}_{r'}$ with $r' = \min\{r+2, n\}$, so $\mathcal{V}$ is $Q(\mathcal{K}_{r'})$-identifiable. $\square$

Proof of Proposition 8. Let $K \in \mathcal{K}_r$ be as in the proof above. The annihilator $S(K)^0$ corresponds to the quadratic forms that vanish on K, and so $\mathcal{V}^0 \cap S(K)^0$ are the forms vanishing on $\Lambda \cup K$. Applying Lemma 1 to this union, we find that $V^0 \cap S(K)^0$ is empty unless

$$k\binom{d+1}{2} + \binom{n-r+1}{2} < \binom{n+1}{2}.$$

In fact, this is equivalent to (2) because

$$\binom{n+1}{2} - \binom{n-r+1}{2} = \binom{r+1}{2} + r(n-r).$$

Conversely, if (2) holds non-strictly, then Lemma 1 guarantees that $S(K)$ must intersect $\mathcal{V}$ trivially. Applying Lemma 2, we find that V^0 is generated by covectors from the spaces

$$S(K)^0 + \langle \mathbf{e}^i \otimes \mathbf{e}^j \rangle.$$

Every such extension of $S(K)^0$ is a subspace of $S(K')^0$, where K' is derived from K by the eventual omission of either $\mathbf{e}_i$ or $\mathbf{e}_j$. A viable subspace K' can always be found in the family $S(\mathcal{K}_{r'})$ with $r' = \min\{r+1, n\}$, and so $\mathcal{V}$ is $S(\mathcal{K}_{r'})$-identifiable. $\square$

References

[CCG12] Enrico Carlini, Maria Virginia Catalisano, and Anthony V. Geramita. Subspace arrangements, configurations of linear spaces and the quadrics containing them. *Journal of Algebra*, 362:70–83, July 2012.

[CR08] Emmanuel J. Candes and Benjamin Recht. Exact low-rank matrix completion via convex optimization. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 806–812, September 2008.

[CT09] Emmanuel J. Candes and Terence Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion. Technical Report arXiv:0903.1476, arXiv, March 2009. arXiv:0903.1476 [cs, math] type: article.

[LBY+19] Connor Lane, Ron Boger, Chong You, Manolis Tsakiris, Benjamin Haeffele, and Rene Vidal. Classifying and Comparing Approaches to Subspace Clustering with Missing Data. pages 0–0, 2019.

[MGC11] Shiqian Ma, Donald Goldfarb, and Lifeng Chen. Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1):321–353, June 2011.

[MW11] A. Majumdar and R. K. Ward. Some empirical advances in matrix completion. *Signal Processing*, 91(5):1334–1338, May 2011.

[OPAB+21] Greg Ongie, Daniel Pimentel-Alarcn, Laura Balzano, Rebecca Willett, and Robert D. Nowak. Tensor Methods for Nonlinear Matrix Completion. *SIAM Journal on Mathematics of Data Science*, 3(1):253–279, January 2021.

[OWNB17] Greg Ongie, Rebecca Willett, Robert D. Nowak, and Laura Balzano. Algebraic variety models for high-rank matrix completion. In *International Conference on Machine Learning*, pages 2691–2700. PMLR, 2017.

[PAOB+17] Daniel Pimentel-Alarcn, Gregory Ongie, Laura Balzano, Rebecca Willett, and Robert Nowak. Low algebraic dimension matrix completion. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 790–797, October 2017.

[VMS03] R. Vidal, Yi Ma, and S. Sastry. Generalized principal component analysis (GPCA). In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I, June 2003. ISSN: 1063-6919.