Classifying and comparing approaches to subspace clustering with missing data

Connor Lane[1], Ron Boger[1], Chong You[2], Manolis C. Tsakiris[3], Benjamin D. Haeffele[1], and René Vidal[1]

[1]Mathematical Institute for Data Science, Johns Hopkins University, USA
[2]Dept. of Electrical Engineering and Computer Sciences, University of California Berkeley, USA
[3]School of Information Science and Technology, ShanghaiTech University, China

Abstract

In recent years, many methods have been proposed for the task of subspace clustering with missing data (SCMD), and its complementary problem, high-rank matrix completion (HRMC). Given incomplete data drawn from a union of subspaces, these methods aim to simultaneously cluster each data point and recover the unobserved entries. In this work, we review the current state of this literature. We organize the existing methods into five distinct families and discuss their relative strengths and weaknesses. This classification exposes some gaps in the current literature, which we fill by introducing a few natural extensions of prior methods. Finally, we provide a thorough and unbiased evaluation of representative methods on synthetic data. Our experiments demonstrate a clear advantage for alternating between projected zero-filled sparse subspace clustering, and per-group matrix completion. Understanding why this intuitive but heuristic method performs well is an open problem for future theoretical study.

1. Introduction

Modeling high-dimensional data as a union of low-dimensional subspaces is pervasive in computer vision and data science more broadly [34, 16, 17, 31, 37]. To this end, significant progress has been made toward developing efficient subspace clustering algorithms with both strong theoretical guarantees and empirical performance, under a broad range of noisy data settings [8, 9, 22, 23, 15, 18, 40, 41, 32, 42, 38]. For much of the previous decade however, a relatively small amount of work has focused on analyzing union-of-subspaces (UoS) data with missing entries.

More recently, there has been a surge of interest in the subspace clustering with missing data (SCMD) and high-rank matrix completion (HRMC) problems. Many new methods have been proposed, based on a diverse range of principles [10, 2, 29, 39, 15, 26, 19, 7, 30, 14, 11, 12, 13]. Significant theoretical results have also emerged, establishing for example necessary and sufficient conditions for UoS identifiability [29, 27, 28], and relationships between the number of missing entries and cluster correctness [6, 33].

Because of this rapid progress however, there is now limited agreement on which of the current methods in fact perform the best, and which will be most interesting for further theoretical study. In an effort to raise consensus, we review the major existing approaches to joint SCMD+HRMC problem and evaluate them on common ground. In the process, we also observe and fill in some noticeable gaps in the set of current methods.

Paper contributions.

- We organize the set of existing methods for SCMD+HRMC into five distinct families, and discuss their relative strengths and weaknesses (Table 1).
- We introduce several natural extensions of prior methods that are currently missing from the literature.
- We perform a thorough and unbiased evaluation of representative methods on synthetic data.
- Our experiments reveal a clear advantage for an intuitive but heuristic approach: alternating between projected zero-filled sparse subspace clustering, and per-group low-rank matrix completion. We suggest that understanding this method's performance would be a valuable goal for future theoretical study.

Paper organization. In Section 2, we first introduce notation and a statement of the SCMD+HRMC problem. We then review the well-known approaches to this problem, classifying them into five distinct families (Table 1). We discuss each method (or group of closely related methods) under its own heading. Headings marked with a "(*)" contain our proposed gap-filling extensions. In Section 3, we present our evaluation of representative methods on synthetic data. We conclude in Section 4.

Family	Strengths	Weaknesses	Example methods
Alternating clustering & completion	Intuitive alternating procedure leveraging strong prior algorithms.	No associated joint optimization problem.	LRMC-SSC; (Alt) (P)ZF-EnSC+gLRMC (*) [39, 19].
Joint clustering & completion	Formulates SCMD+HRMC as a joint optimization problem.	Problems are non-convex; algorithms are more complex; relies on poorly understood self-expression based completion.	SSC-lifting [7]; S3LR [19]; SSC-SRMC [11]; SRME-MC [12].
Matrix factorization	Formulates SCMD+HRMC as a joint optimization problem; well-motivated completion.	Problems are non-convex; requires an estimate of the subspace dimension.	KSS-MD [2]; EM-MD [29]; (LR-)GSSC (*) [26].
Algebraic	Exploits the (nonlinear) low-dimensional structure in unions of subspaces; does not depend on clustering for completion.	Complex algorithms; requires a large number of samples.	VMC [25]; LADMC [30, 24]; KFMC [13].
Neighborhood based	Intuitive, non-iterative procedure.	No associated joint optimization problem; depends on correctness of nearest neighbors.	HRMC [10]; Robust HRMC [14]; (Alt) TSC+gLRMC (*) [15].

Table 1. Classification of the major approaches to SCMD+HRMC into five broad families. Methods with at least some component that is new to this work are marked with a "(*)".

2. Classifying approaches to SCMD

Notation. Matrices are denoted with upper case bold letters, $\boldsymbol{X}$; vectors with lower case bold letters, $\boldsymbol{x}$; and scalars with lower case letters, x. $\mathbb{R}^D$ denotes the set of real vectors of dimension D; $[k]$ the set of integers $1, \dots, k$; and $\boldsymbol{X} \odot \boldsymbol{Y}$ the Hadamard product. Finally, $\|\boldsymbol{X}\|_F$, $\|\boldsymbol{X}\|_*$, and $\|\boldsymbol{X}\|_1$ refer to the Frobenius norm, nuclear norm (also known as the trace norm), and entrywise ℓ_1 norm respectively.

Problem description. We assume we are given data $\boldsymbol{X} \in \mathbb{R}^{D \times N}$ with data points $\boldsymbol{x}_i \in \mathbb{R}^D$ concentrated near a union of subspaces $\bigcup_{j=1}^n \mathcal{S}_j$, each of dimension $d < D$. Moreover, we assume we have access only to a subset of $\ell_i \leq D$ entries for each $\boldsymbol{x}_i$. We let $\boldsymbol{\Omega} \in \{0,1\}^{D \times N}$ denote the indicator for the observed entries with ω_i the indicator for $\boldsymbol{x}_i$. We use $P_{\boldsymbol{\Omega}}(\cdot)$ to denote the projection onto the coordinate subspaces of the observed entries, i.e. $P_{\boldsymbol{\Omega}}(\boldsymbol{Y}) = \boldsymbol{\Omega} \odot \boldsymbol{Y}$. The zero-filled data are denoted $\bar{\boldsymbol{X}} = P_{\boldsymbol{\Omega}}(\boldsymbol{X})$.

The task of subspace clustering with missing data (SCMD) is then to cluster the $\bar{\boldsymbol{x}}_i$ according to subspace membership. The task of high-rank matrix completion (HRMC), assuming a union of subspaces model, is to recover the unobserved entries of $\boldsymbol{X}$. In this work, we refer to the joint clustering and completion task as SCMD+HRMC.

2.1. Alternating subspace clustering and per-group completion

The majority of existing methods for subspace clustering with complete data follow a self-expressive approach, in which one searches for a matrix $\boldsymbol{C} \in \mathbb{R}^{N \times N}$ satisfying $\boldsymbol{X} \approx \boldsymbol{X}\boldsymbol{C}$. By choosing an appropriate regularization, one can promote $\boldsymbol{C}$ to be *subspace-preserving*. That is, $c_{ij} = 0$ if $\boldsymbol{x}_i$ and and $\boldsymbol{x}_j$ are drawn from different subspaces. A segmentation of the data can then be obtained by applying spectral clustering to an affinity $\boldsymbol{W} = |\boldsymbol{C}| + |\boldsymbol{C}^\top|$. This problem can be formulated as

$$\min_{\boldsymbol{C}} \frac{\lambda}{2}\|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \theta(\boldsymbol{C}) \text{ s.t. } \operatorname{diag}(\boldsymbol{C}) = \boldsymbol{0}, \quad (1)$$

where $\lambda > 0$ is a penalty parameter and $\theta(\cdot)$ is a regularizer. For example, $\theta(\boldsymbol{C}) = \|\boldsymbol{C}\|_1$, $\|\boldsymbol{C}\|_F^2$, $\|\boldsymbol{C}\|_*$, or $\gamma\|\boldsymbol{C}\|_1 + (1-\gamma)\|\boldsymbol{C}\|_F^2$ for sparse subspace clustering (SSC) [9], least-squares regression (LSR) [23], low-rank subspace clustering [35] and low-rank representation (LRR) [22], and elastic-net subspace clustering (EnSC) [40] respectively.

(P)ZF-SC+gLRMC. The self-expressive methods can be immediately extended to the missing data case by working with the zero-filled $\bar{\boldsymbol{X}}$. This approach combined with sparse regularization is referred to as zero-filled SSC (ZF-SSC), and was first studied experimentally in [39]. More recently, theoretical conditions on the maximum tolerable missing entry rates were established in [6, 33].

A seemingly more attractive alternative to zero-filling is to first apply low-rank matrix completion (LRMC) to $\bar{\boldsymbol{X}}$ by solving the following convex problem [3, 20]

$$\min_{\boldsymbol{Y}} \|\boldsymbol{Y}\|_* \text{ s.t. } P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \boldsymbol{0}. \quad (2)$$

One can then substitute the solution $\boldsymbol{Y}_{\text{MC}}$ for $\boldsymbol{X}$ in (1). In the intended regime where $\boldsymbol{X}$ is full-rank, however, we should not expect this strategy to add much benefit. We refer to this method as LRMC-SSC, where the prefix indicates initialization by LRMC.

These naive approaches can be improved by also projecting the self-expressive differences onto the pattern of observed entries, yielding for example the SSC with entrywise zero-filling (SSC-EWZF) method proposed in [39]:

$$\min_{\boldsymbol{C}} \frac{\lambda}{2}\|P_{\boldsymbol{\Omega}}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{X}}\boldsymbol{C})\|_F^2 + \|\boldsymbol{C}\|_1 \text{ s.t. } \operatorname{diag}(\boldsymbol{C}) = \boldsymbol{0}. \quad (3)$$

We refer to this method as projected zero-filled SSC (PZF-

Algorithm 1 Alt SC+gLRMC algorithm framework

Input: Observed data $\bar{\boldsymbol{X}}$; indicator for observed entries $\boldsymbol{\Omega}$; $\texttt{maxit} \geq 0$.
Initialize completion $\boldsymbol{Y}_0 \leftarrow \bar{\boldsymbol{X}}$ or $\boldsymbol{Y}_0 \leftarrow \boldsymbol{Y}_{\text{MC}}$.
Compute affinity $\boldsymbol{W}_0$ given $\boldsymbol{Y}_0$, e.g. by PZF-SSC.
Spectral clustering on $\boldsymbol{W}_0$ to get $\boldsymbol{Q}_0$.
for $k = 1, \ldots, \texttt{maxit}$ **do**
    Update each group $\boldsymbol{Y}_{k-1} \operatorname{diag}((\boldsymbol{Q}_k)_i)$ by gLRMC.
    Repeat steps 3-4
    if $\boldsymbol{Q}_k$ unchanged (up to label permutation): **break**
Return: $\boldsymbol{Q}_k$, $\boldsymbol{Y}_k$, $\boldsymbol{W}_k$

SSC), following [33]. The purpose of the projection operator is merely to discount the meaningless self-expressive errors over the zero-filled unobserved entries.

To extend these methods to the joint SCMD+HRMC problem, one can follow subspace clustering with *per-group* LRMC (gLRMC). Given a segmentation $\boldsymbol{Q} \in \{0,1\}^{N \times n}$ with $\boldsymbol{Q}\mathbf{1} = \mathbf{1}$ from spectral clustering, one solves

$$\min_{\boldsymbol{Y}} \sum_{i=1}^{n} \|\boldsymbol{Y} \operatorname{diag}(\boldsymbol{q}_i)\|_* \text{ s.t. } P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \mathbf{0}. \quad (4)$$

We indicate methods using this approach with a "+gLRMC" suffix, e.g. PZF-SSC+gLRMC.

Alt (P)ZF-SC+gLRMC (*). A natural further generalization is to repeatedly alternate between subspace clustering and completion. In [19], Li and co-authors consider one variant of this approach. After initializing the completion $\boldsymbol{Y}$ by LRMC, they alternate between standard SSC and gLRMC (Alt LRMC-SSC+gLRMC). In this work, we also consider a variant that initializes by zero-filling and alternates between PZF-EnSC and gLRMC (Alt PZF-EnSC+gLRMC). Both of these alternating methods as well as many others are instances of the general Algorithm 1. In particular, each of the previously discussed methods can be represented as special cases of 1 with $\texttt{maxit} \in \{0, 1\}$.

The motivation behind these methods is that by performing several iterations, the algorithms will be able to refine the self-expression $\boldsymbol{C}$ based on the progressively more accurate completion. We further predict that retaining the projection beyond the first application of gLRMC may help prevent incorrect completions from derailing this progress.

The strength of this family of methods is its intuitive basic algorithm, founded on strong principles from self-expressive subspace clustering and low-rank matrix completion. A limitation, however, is that it is not associated with any formal optimization problem. As a result, theoretical analysis will be more difficult.

2.2. Joint self-expressive subspace clustering and completion

In contrast to the intuitive but heuristic alternating methods, several algorithms have been proposed that integrate self-expressive subspace clustering and matrix completion into a unified optimization problem. These problems obey the following general form

$$\begin{aligned} \min_{\boldsymbol{Y},\boldsymbol{C}} \lambda\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{C}\| + \theta(\boldsymbol{Y}, \boldsymbol{C}) \\ \text{s.t. } P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \mathbf{0},\ \operatorname{diag}(\boldsymbol{C}) = \mathbf{0} \end{aligned} \quad (5)$$

where $\|\cdot\|$ denotes a general "norm" in an abuse of notation, e.g. $\|\cdot\| = \frac{1}{2}\|\cdot\|_F^2$, and $\theta(\cdot,\cdot)$ is a general regularizer acting jointly on $\boldsymbol{Y}$ and $\boldsymbol{C}$.

Having access to an explicit optimization problem enables analysis of the corresponding algorithms. The specific problem (5) has at least two weaknesses, however. First, the problem is non-convex due to the product $\boldsymbol{Y}\boldsymbol{C}$ in the self-expressive term. Second, the joint optimization introduces a new dependence between the completion $\boldsymbol{Y}$ and the self-expression term. Intuitively, if $\boldsymbol{C}$ correctly captures the linear relationships among the data points, and there are not too many missing entries, then perhaps this *self-expressive based completion* will be sufficient to recover the missing entries. Compared to LRMC however, the performance of this completion approach is poorly understood.

S3LR. Among this family, the S3LR method of [19] is the most closely related to the alternating algorithms from the previous section. The main motivation is to combine the three components of Algorithm 1: (1) self-expression, (2) spectral clustering, and (3) gLRMC, into a single optimization problem. Specifically, they propose to optimize

$$\begin{aligned} \min_{\boldsymbol{C},\boldsymbol{Y},\boldsymbol{Q}} \lambda\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{C}\|_1 + \sum_{j=1}^{n} \|\boldsymbol{Y} \operatorname{diag}(\boldsymbol{q}_j)\|_* \\ + \gamma(\alpha\|\Theta(\boldsymbol{Q}) \odot \boldsymbol{C}\|_1 + \|\boldsymbol{C}\|_1) \\ \text{s.t. } \operatorname{diag}(\boldsymbol{C}) = \mathbf{0},\ P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \mathbf{0},\ \boldsymbol{Q} \in \mathcal{Q} \end{aligned} \quad (6)$$

where $\mathcal{Q} \triangleq \{\boldsymbol{Q} \in \{0,1\}^{N \times n} \mid \boldsymbol{Q}\mathbf{1} = \mathbf{1},\ \operatorname{rank}(\boldsymbol{Q}) = n\}$. $\Theta(\boldsymbol{Q}) \in \mathbb{R}^{N \times N}$ is defined by $(\Theta(\boldsymbol{Q}))_{ij} = 1/2\|\boldsymbol{Q}_{i,\cdot} - \boldsymbol{Q}_{j,\cdot}\|_2^2$. Thus, the term involving $\Theta(\boldsymbol{Q})$ in fact represents the spectral clustering objective. Similarly, the first and fourth terms correspond to SSC, while the second term is precisely the gLRMC objective.

Importantly, these similarities do not imply that the same Algorithm 1 can be used to optimize (6). Rather, there are extra dependencies between the variables, e.g. $\boldsymbol{Y}$ on $\boldsymbol{C}$, $\boldsymbol{Q}$ on $\boldsymbol{Y}$, which necessitate a more complex algorithm. Given a candidate segmentation $\boldsymbol{Q}$, the authors optimize (6) with respect to $\boldsymbol{C}$ and $\boldsymbol{Y}$ using linearized ADMM [21]. They then update the segmentation $\boldsymbol{Q}$ by spectral clustering on $\boldsymbol{W} = |\boldsymbol{C}| + |\boldsymbol{C}^\top|$ (ignoring $\boldsymbol{Q}$'s dependence on $\boldsymbol{Y}$).

SSC-lifting. In [7], Elhamifar considers a more direct ex-

tension of SSC to the joint SCMD+HRMC problem. His method, SSC-lifting, optimizes the following problem

$$\min_{\boldsymbol{Y},\boldsymbol{C}} \|\boldsymbol{C}\|_0 \text{ s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C},\ P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \boldsymbol{0},\ \operatorname{diag}(\boldsymbol{C}) = \boldsymbol{0}. \tag{7}$$

To optimize this complex non-convex problem, Elhamifar proposes a convex relaxation over a set of $N\sum_{i=1}^{N}(D-\ell_i)$ lifted variables. Although the global optimum of the convex relaxation can technically be found in polynomial time, it is nonetheless very expensive to solve. In particular, merely evaluating the objective requires calculating the singular values of a $(D-\ell_i)\times N$ matrix, for every $i = 1,\dots,N$.

SC-SEMC & SRME-MC. In [11], Fan and Chow consider a natural generalization of SSC-lifting, where the exact self-expressive constraint is relaxed and the ℓ_0 norm on $\boldsymbol{C}$ is replaced by one of three popular self-expressive penalties: $\|\cdot\|_1$, $\|\cdot\|_F^2$, or $\|\cdot\|_*$. Specifically, the authors optimize

$$\min_{\boldsymbol{Y},\boldsymbol{C}} \frac{\lambda}{2}\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{C}\|_F^2 + \|\boldsymbol{C}\|_q \quad \text{s.t. } P_{\boldsymbol{\Omega}}(\boldsymbol{Y} - \bar{\boldsymbol{X}}) = \boldsymbol{0},\ \operatorname{diag}(\boldsymbol{C}) = \boldsymbol{0}. \tag{8}$$

where the "norm" q is allowed to be one of the previous three choices[1]. Rather than consider a convex relaxation, the authors attempt to minimize the non-convex objective directly using linearized ADMM [21]. This results in a more efficient algorithm compared to SSC-lifting, yet one whose convergence properties are less well understood. We refer to these methods as self-expressive based matrix completion (SEMC) methods.

In [12], Fan and Chow propose a variant of SSC-SEMC called "Sparse Representation with Missing Entries and Matrix Completion" (SRME-MC), which includes additional nuclear norm regularization term $\alpha\|\boldsymbol{Y}\|_*$ in (8). When the data are low rank, this added regularization should improve completion performance. In the intended full-rank regime, however, it is unclear what if any benefit should be expected.

In summary, the joint self-expressive clustering and completion methods each benefit from a unified optimization problem. The cost in return is more complex, non-convex optimization. The self-expressive based completion also requires further understanding.

2.3. Structured matrix factorization methods

The family of structured matrix factorization methods represents an alternative approach to formulating SCMD+HRMC as a unified optimization problem. The motivation is analogous to that for classic sparse and low-rank recovery. By seeking a compact representation of the data, one can expect to solve the seemingly under-determined inverse problem and recover the missing entries. In sparse recovery, the representation is compact with respect to a fixed dictionary; in low-rank recovery, it is compact with respect to the set of rank-1 matrices. The challenge for methods here is to identify the corresponding set of appropriate atomic factors for the union-of-subspaces setting.

KSS-MD & EM. In [2], Balzano and co-authors adapt the well-known k-subspaces method to the missing data setting (KSS-MD). The optimization problem can be written as

$$\min_{\boldsymbol{Q},\{(\boldsymbol{U}_j,\boldsymbol{V}_j)\}_{j=1}^n} \left\| P_{\boldsymbol{\Omega}}\Big(\bar{\boldsymbol{X}} - \sum_{j=1}^{n} \boldsymbol{U}_j\boldsymbol{V}_j \operatorname{diag}(\boldsymbol{q}_j)\Big)\right\|_F^2 \quad \text{s.t. } \boldsymbol{U}_j^\top\boldsymbol{U}_j = \boldsymbol{I} \text{ for } j = 1,\dots,n,\ \boldsymbol{Q} \in \{0,1\}^{N\times n},\ \boldsymbol{Q}\boldsymbol{1} = \boldsymbol{1}, \tag{9}$$

where the $\boldsymbol{U}_j \in \mathbb{R}^{D\times d}$ are orthogonal bases, $\boldsymbol{V}_j \in \mathbb{R}^{d\times N}$ contain coefficients representing each data point according to each subspace, and the segmentation $\boldsymbol{Q} \in \mathbb{R}^{N\times n}$ assigns data points to unique subspaces. To optimize (9), the authors implement an efficient (linear time) online algorithm based on the Grassmannian rank one update subspace estimation (GROUSE) algorithm for tracking individual incomplete subspaces [1]. A related approach is proposed in [29], where Pimentel-Alarcon and co-authors introduce an EM algorithm for Gaussian mixtures, adapted to handle missing data, which can be viewed as an extension of KSS-MD to more general covariances beyond scaled identity.

In both works, the data are represented explicitly as a union of low-dimensional subspaces using a structured factorization. E.g. $\boldsymbol{X} \approx \sum_j \boldsymbol{U}_j\boldsymbol{V}_j \operatorname{diag}(\boldsymbol{q}_j)$ for KSS-MD. Since the factorization exactly matches the underlying structure of the data, one expects strong recovery performance. However, as with k-means and traditional k-subspaces, the optimization of (9) as well as EM are susceptible to poor local minima. Furthermore, they both require an estimate of the subspace dimension, $\hat{d}$.

(LR-)GSSC (*). In [26], Pimentel-Alarcon and co-authors introduce another factorization based method which they call group-sparse subspace clustering (GSSC). This method optimizes the following problem (using slightly different notation than [26])

$$\min_{\{(\boldsymbol{U}_j,\boldsymbol{V}_j)\}} \frac{\lambda}{2}\left\| P_{\boldsymbol{\Omega}}\Big(\bar{\boldsymbol{X}} - \sum_{j=1}^{n}\boldsymbol{U}_j\boldsymbol{V}_j\Big)\right\|_F^2 + \sum_{j=1}^{n}\|\boldsymbol{V}_j\|_{2,1} \quad \text{s.t. } \sum_{j=1}^{n}\|\boldsymbol{U}_j\|_F^2 \le 1, \tag{10}$$

where $\boldsymbol{V}_j \in \mathbb{R}^{d\times N}$ and $\|\boldsymbol{V}_j\|_{2,1} = \sum_i \|(\boldsymbol{V}_j)_i\|_2$. The regularization on the $\boldsymbol{V}_j$ can be understood as promoting group sparsity. Ideally for every i, only one of the $(\boldsymbol{V}_j)_i$ will be non-zero across $j = 1,\dots,n$. This way, the assignment of points to subspaces is represented implicitly in the non-zero column supports of the $\boldsymbol{V}_j$.

One limitation of GSSC is that it also requires an esti-

[1] The diagonal constraint is omitted when $\|\cdot\|_q = \|\cdot\|_*$.

mate of the subspace dimension, $\hat{d}$. Here we extend GSSC to include low-rank regularization as a way to partially address this issue. Our LR-GSSC variant optimizes

$$\min_{\{(\boldsymbol{U}_j,\boldsymbol{V}_j)\}} \frac{\lambda}{2}\Big\|P_{\boldsymbol{\Omega}}\Big(\bar{\boldsymbol{X}} - \sum_{j=1}^{n}\boldsymbol{U}_j\boldsymbol{V}_j\Big)\Big\|_F^2 + \sum_{j=1}^{n}\Big(\gamma\|\boldsymbol{U}_j\|_{2,1} + \|\boldsymbol{V}_j\|_{2,1}\Big). \tag{11}$$

Using the rotation invariance of the products $\boldsymbol{U}_j\boldsymbol{V}_j$, one can show that $\ell_{2,1}$ regularization on the $\boldsymbol{U}_j$ is in fact equivalent to nuclear norm regularization.

Both problems are optimized locally by exact alternating minimization—first minimizing with respect to the $\boldsymbol{V}_j$ while fixing the $\boldsymbol{U}_j$, and then similarly for updating the $\boldsymbol{U}_j$.

2.4. Algebraic methods

The structured factorization methods search for a compact representation of the data in the original ambient space. By contrast, the algebraic methods seek a low-rank representation in an embedded space of higher dimension.

VMC & LADMC. In [25, 30], Ongie, Pimentel-Alarcon, and co-authors follow the algebraic subspace approach originally proposed in [36], which exploits the fact that unions of low-dimensional subspaces are algebraic varieties often admitting vanishing polynomials of low enough degree. Thus, the image of the data under the Veronese map of small degree p (sending $\boldsymbol{x}_i$ to the vector of all unique monomials of degree p) is likely to be low-rank in this larger dimension embedded space. When sufficiently many data points are present, this embedded subspace is identifiable, enabling the recovery of the missing entries.

The two methods based on this principle are variety-based matrix completion (VMC) and low algebraic dimension matrix completion (LADMC). In the former, the authors use the kernel trick to easily extend to higher order embeddings, whereas the latter represents the embedded data explicitly. This effectively limits the method to degree $p = 2$, but in exchange the authors obtain a much simpler algorithm with only linear complexity in the number of data points N (compared to quadratic for VMC). Moreover, for a generic union of n d-dimensional subspaces in $\mathbb{R}^D$, the degree $p = 2$ embedded data will have rank at most $n\binom{d+1}{2}$ [25, 5]. This is often much less than the embedded ambient dimension $\binom{D+1}{2}$, justifying the use LADMC with $p = 2$.

Crucially, after embedding into higher dimension, the number of data points required for correct recovery becomes $O(d^pD^p)$ [25]. This combined with the methods' computational cost makes them difficult to apply in high dimensional settings.

Method	Parameter
LRMC-SSC	$\lambda_0 \in \{5, 10, 20, \ldots, 320\}$
(Alt) PZF-EnSC+gLRMC	$\lambda_0 \in \{5, 10, 20, \ldots, 320\}$, $\gamma \in \{0.5, 0.6, 0.7, 0.8, 0.9, 0.99\}$
S3LR	$\lambda \in \{0.01, 0.1, \ldots, 100\}$, $\gamma \in \{10^{-5}, 10^{-4}, \ldots, 0.1\}$, $\alpha = 1$
SRME-MC	$\lambda \in \{5, 10, 20, \ldots, 160\}$, $\alpha \in \{0.01, 0.1, \ldots, 100\}$
(LR-)GSSC	$\frac{\hat{d}}{D} \in \{0.02, 0.04, 0.1, 0.2, \ldots, 0.6\}$, $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 0.1\}$, $\gamma \in \{10^{-5}, 0.01, \ldots, 100\}$
LADMC-SSC	$\lambda_0 \in \{5, 10, 20, \ldots, 320\}$
(Alt) TSC+gLRMC	$\frac{qn}{N} \in \{0.05, 0.1, 0.15, \ldots, 0.3\}$

Table 2. Parameter choices for evaluated SCMD+HRMC methods. See problem formulations for definitions in Section 2. For SSC and EnSC based methods, λ_0 denotes a scaled λ such that $\lambda_0 > 1$ ensures all columns of the optimal $\boldsymbol{C}$ are non-zero [40, 39].

2.5. Neighborhood based methods

We conclude with the neighborhood based methods, which represent perhaps the first group of methods considered for the SCMD+HRMC problem. They all rely on the principle that even in the presence of missing entries, nearest neighbors can still be identified.

HRMC & Robust HRMC. In the early work [10], Balzano, Eriksson and co-authors propose an intuitive, non-iterative neighborhood based method that they call High-rank Matrix Completion (HRMC). Their method executes the following four steps: (1) choose a random subset of seed data points and identify the nearest neighbors for each seed, (2) fit local subspaces to each neighborhood using LRMC, (3) prune all but n local subspaces, discarding those that lie in the span of two or more other subspaces, (4) assign each data point to its nearest subspace and complete by orthogonal projection. The authors prove that their procedure can recover the unobserved entries with high probability provided $N \geq O(D^{\log(D)})$. This was one of the first results establishing conditions for correct recovery in the HRMC setting. However, tighter necessary and sufficient conditions for union of subspace identifiability ($N \geq O(dn)$) have since been established [28]. More recently in [14], Gao and co-authors also improve upon HRMC by introducing more robust sub-routines for each of the four above steps.

(Alt) TSC+gLRMC (*). Another example of a neighborhood based method is threshold subspace clustering (TSC) [15]. This efficient and robust method applies spectral clustering to a weighted q-nearest neighbor graph. Neighborhoods are defined according to cosine-angles: $\theta_{ij} = |\langle \bar{\boldsymbol{x}}_i, \bar{\boldsymbol{x}}_j \rangle| / \|\boldsymbol{\omega}_j \odot \bar{\boldsymbol{x}}_i\|_2 \|\boldsymbol{\omega}_i \odot \boldsymbol{x}_j\|_2$. The affinity edge weights for neighbors i and j are then given by $w_{ij} = \exp(-2\arccos(\theta_{ij}))$. Finally, the affinity is symmetrized.

As with the methods in Section 2.1, TSC can be extended to the joint SCMD+HRMC problem by combining with gLRMC in an alternating fashion. This minor but natural extension has not been previously considered.

3. Synthetic experiments

Experiment set-up. We generated synthetic data lying near a union of subspaces in the following manner. First, we sampled n d-dimensional subspaces in $\mathbb{R}^D$ uniformly at random by drawing $\boldsymbol{U}_j \in \mathbb{R}^{D\times d}$ with standard Gaussian entries and orthogonalizing. We then generated data for each subspace, $\boldsymbol{X}_j \in \mathbb{R}^{D\times N_j}$, as

$$\begin{aligned} \boldsymbol{X}_j = \boldsymbol{U}_j\boldsymbol{V}_j + \boldsymbol{E}_j,\ (\boldsymbol{V}_j)_i \sim \mathcal{N}(\boldsymbol{0}_d, d^{-1}\boldsymbol{I}_d), \\ (\boldsymbol{E}_j)_i \sim \mathcal{N}(\boldsymbol{0}_D, \sigma^2 D^{-1}\boldsymbol{I}_D), \end{aligned} \tag{12}$$

where $\boldsymbol{V}_j \in \mathbb{R}^{d\times N_j}$, and $\boldsymbol{E}_j \in \mathbb{R}^{D\times N_j}$. The $\boldsymbol{X}_j$ were then concatenated to form the complete data matrix $\mathbb{R}^{D\times N} \ni \boldsymbol{X} = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_j]$. We fix $N_1 = \cdots = N_j$, and use $N_j \triangleq N/n$ in an abuse of notation. Finally, for each data point $\boldsymbol{x}_i$, we sample exactly $\ell > 0$ observed entries uniformly at random, following e.g. [26, 33].

We considered four synthetic data settings: (1) small $n = 5$, small $d = 5$, small $D = 25$; (2) large $n = 20$, small $d = 5$, small $D = 25$; (3) small $n = 5$, large $d = 25$, large $D = 100$; (4) large $n = 25$, small $d = 5$, large $D = 100$. All datasets were full rank, and in all but the first setting the subspaces were guaranteed to be non-independent. In addition, for each setting we varied the number of points per group relative to the subspace dimension: $N_j/d \in \{2, 4, 6, 8, 10\}$. We included a small amount of noise, $\sigma = 0.001$. Finally, we repeated each setting for 20 random trials.

In problems with missing data, it is common to observe rapid phase transitions in performance as the number of observed entries increases. It is therefore crucial to evaluate over a closely spaced range of ℓ values. This can be computationally expensive, however, especially when D is large. To overcome this, we used a binary search strategy to identify the narrow range of ℓ values containing the phase shift for every algorithm and setting. More specifically, we used binary search to find the minimum threshold $\hat{\ell}_{0.05}$ such that for all $\ell > \hat{\ell}_{0.05}$, the achieved clustering error was no greater than 5%. We then evaluated 11 choices for ℓ within an interval centered on $\hat{\ell}_{0.05}$ of radius no greater than $d/5$. In total, at least 13 ℓ values were tested for each method and setting: 11 central values and the two extremes $\ell = d, D$.

Comparison methods. Using this set-up, we compared the following methods across all five families.

- **Alternating subspace clustering and completion**: LRMC-SSC, PZF-EnSC+gLRMC, Alt PZF-EnSC+gLRMC (*).
- **Joint self-expressive clustering and completion**: S3LR, SRME-MC.
- **Matrix factorization**: GSSC, LR-GSSC (*).
- **Algebraic**: LADMC-SSC.
- **Neighborhood**: TSC+gLRMC, Alt TSC+gLRMC (*).

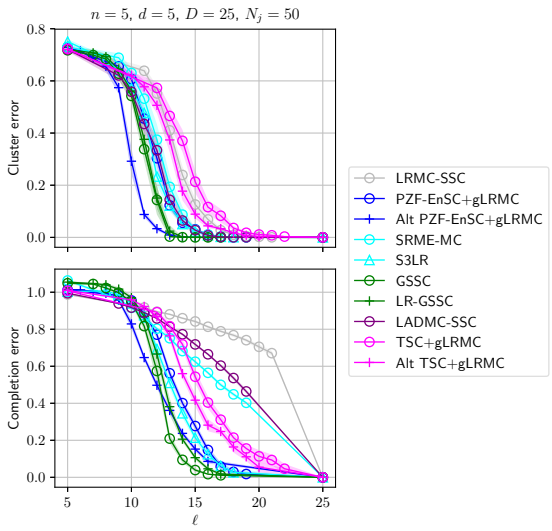

Figure 1. Mean clustering and completion error across methods for a single setting ($n = 5$, $d = 5$, $D = 25$) and choice of $N_j = 30 = 6d$. Methods belonging to the same family share the same color. The shaded error regions represent 95% confidence intervals over the 20 random trials.

Besides the baseline method LRMC-SSC, the selected methods are among the strongest performing representatives from each class. In addition, we included three extensions proposed in the current paper: Alt PZF-EnSC+gLRMC, Alt TSC+gLRMC, and LR-GSSC.

We compared all methods in terms of both clustering error and completion error. Clustering error is defined to be the fraction of misclassified points, up to a permutation of the labels. Completion error is defined to be the relative Frobenius distance between true and recovered unobserved entries: $\|P_{\boldsymbol{\Omega}^c}(\boldsymbol{Y} - \boldsymbol{X})\|_F / \|P_{\boldsymbol{\Omega}^c}(\boldsymbol{X})\|_F$. To present our results more compactly, we also report aggregate clustering and completion errors across ℓ. Our aggregate metrics are:

(1) The 5% error threshold $\hat{\ell}_{0.05}$ defined above.

(2) The (weighted) average completion error, defined as follows. Let L denote the total number of tested ℓ values and ξ_k the completion error achieved for ℓ_k. The average completion error is then

$$\frac{1}{(D-d)} \sum_{k=1}^{L-1} \xi_k(\ell_{k+1} - \ell_k). \tag{13}$$

All methods were provided the true number of subspaces n, while none were given the true dimension d. We tuned the performance of each method over a fixed set 10 pa-

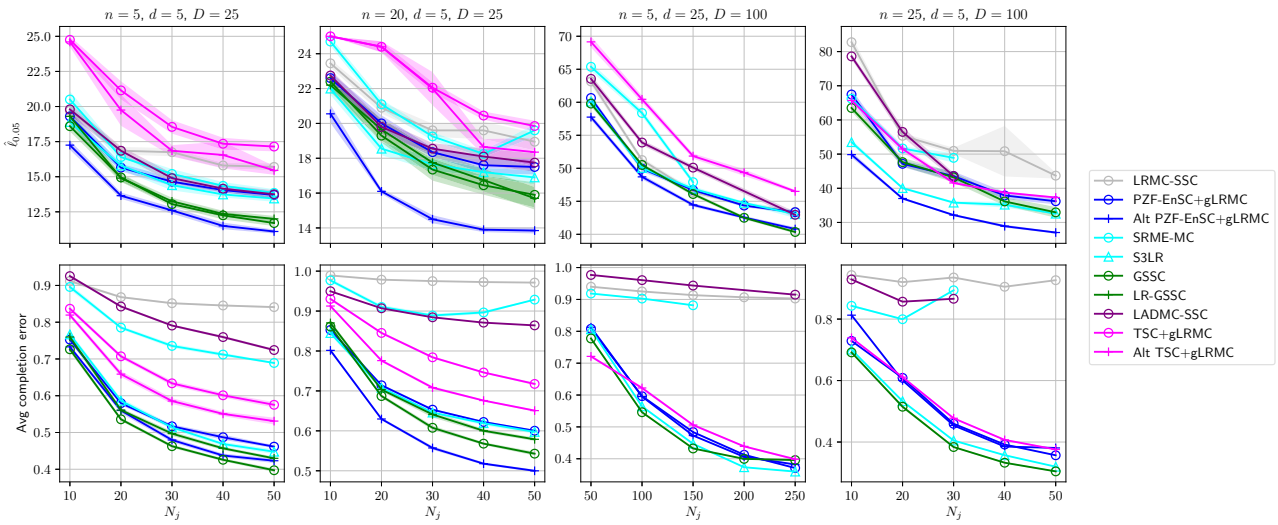

Figure 2. Comparison of aggregate clustering and completion errors between methods for each setting. Each column of dots corresponds to one experiment as shown in Figure 1. The shaded error regions represent 95% confidence intervals over the 20 random trials.

rameter configurations. We generated the configurations by sampling uniformly from a grid of pre-defined values. The range for each parameter was based on manual experimentation and recommendations from the original papers (Table 2). Random sampling was used rather than exhaustive grid search to balance tuning effort across methods with different numbers of parameters. In addition, random parameter sampling is more effective when some parameters are more important than others. In both $D = 25$ settings, we replicated our results on a second batch of 10 random configurations, suggesting that this amount of tuning is sufficient to obtain stable performance.

The best parameter configurations were selected based on average completion error. Although tuning based on cluster error is more common in the literature, this approach translates more easily into practice. With real datasets one can always hold out some observed entries as a validation set, while in general no true cluster labels will be available.

Clustering and completion error phase transitions. In Figure 1, we report mean clustering and completion error for each method as a function of the number of observed entries. For simplicity, we restrict to the small n, small d synthetic data setting ($n = 5$, $d = 5$, $D = 25$), and $N_j = 50 = 10d$. For each curve, we also represent the 95% confidence interval around the mean as a shaded region.

Looking first at clustering error, we observe that all methods undergo a phase transition between $\ell = 10 = 2d$ and $\ell = 15 = 3d$. Within this region, the ten methods can be divided into six groups between which there appear to be reliable differences in performance. Ordered from worst to best, they are: (TSC+gLRMC), (LRMC-SSC), (Alt TSC+gLRMC), (SRME-MC), (SRME-MC/LADMC-SSC/PZF-EnSC/S3LR), (LR-GSSC/GSSC), (Alt PZF EnSC+gLRMC).

We observe a largely consistent pattern in completion error, with the following exceptions. First, unlike with clustering, Alt PZF-EnSC+gLRMC does not achieve the best completion for every ℓ. Instead, GSSC surpasses it as soon as their clustering errors become comparable $\ell = 13$. Second, the completion performances for LRMC-SSC, SRME-MC, and LADMC-SSC are each significantly worse relative to the other methods, compared to their clustering. The poor completion performance for LRMC-SSC is not surprising since the data are full-rank. Similarly, although the embedded data are low-rank for LADMC-SSC, the value for N in this experiment is far fewer than the claimed $O((Dd)^2)$ sample complexity [25]. Moreover, we observe experimentally that the tensorized data have significantly larger μ-coherence (3.07 ± 0.04 CI for the complete tensorized data, compared to 1.46 ± 0.03 CI for Gaussian data of the same size and rank) [4]. Nonetheless, LADMC achieves slightly better completion than LRMC, showing that it is benefiting somewhat from exploiting the algebraic structure. Finally, the poor performance of SRME-MC raises further issues for the idea of self-expressive based completion.

Somewhat surprisingly, these results suggest a significant benefit to performing multiple iterations of alternating subspace clustering and per-group completion. Two of the largest overall differences between methods are observed for TSC+gLRMC $\rightarrow$ Alt TSC+gLRMC and PZF-EnSC+gLRMC $\rightarrow$ Alt PZF-EnSC+gLRMC. The results also give strong support for GSSC, although we observe no difference in clustering between the original method and the proposed low-rank regularized variant. Importantly, neither

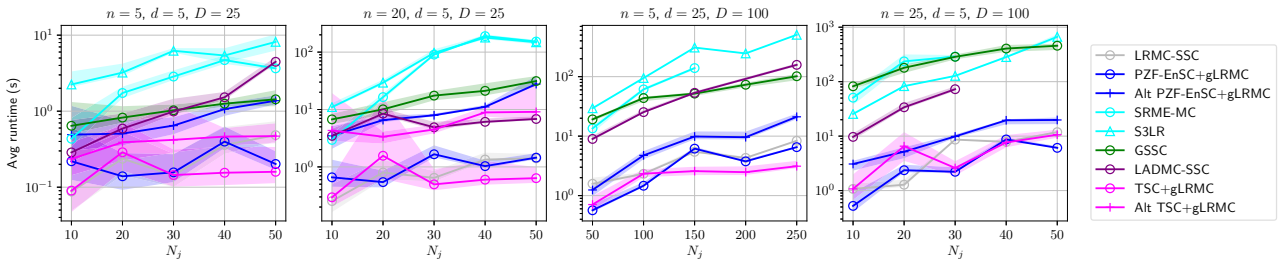

Figure 3. Comparison of mean runtime in seconds between methods for each setting. The edges of the shaded error regions correspond to the minimum and maximum observed runtimes across the 20 random trials. Note the log scaling.

method is provided the true d, suggesting that in either form GSSC is robust to small errors in estimated dimension.

Average clustering and completion error for varied N_j. In Figure 2, we report the aggregate metrics $\hat{\ell}_{0.05}$ and average completion error for every setting. Each column of dots in Figure 2 corresponds to a single experiment as shown in Figure 1. In the two high-dimension settings ($D = 100$), some methods are excluded for some larger N_j due to excessive runtime. Overall, the results are largely consistent with what was previously observed. Again, Alt PZF-EnSC+gLRMC displays a significant clustering advantage, particularly when n is large. GSSC is often the second-best in clustering, and slightly better than Alt PZF-EnSC+gLRMC in completion.

Runtime analysis. Finally, we compare the runtimes of each algorithm in Figure 3. First, we observe that the two joint self-expressive methods, S3LR and SRME-MC are consistently among the slowest methods. This reflects the additional complexity of the algorithms arising from the unified objective. LADMC-SSC and GSSC are also expensive, particularly for large D. The cost for LADMC-SSC arises from the need to compute and factorize an $O(D^2 \times N)$ matrix of embedded data points. The poor runtime for GSSC is likely due to the choice of algorithm (exact alternating minimization). Importantly, the best performing method in terms of clustering, Alt PZF-EnSC+gLRMC, also has manageable runtime.

4. Conclusions

We reviewed the state of the art for joint subspace clustering with missing data and high-rank matrix completion. We categorized the existing methods into five families, and in the process proposed several natural but previously unexamined extensions of prior algorithms. In our evaluation on synthetic data, we demonstrated superior clustering performance Alt PZF-EnSC+gLRMC. Explaining why this intuitive but heuristic method performs so well is an open challenge for future work.

Acknowledgements. This work was supported by NSF grants 1618485, 1618637, and 1704458.

References

[1] Laura Balzano, Robert Nowak, and Benjamin Recht. Online identification and tracking of subspaces from highly incomplete information. In *2010 48th Annual allerton conference on communication, control, and computing (Allerton)*, pages 704–711. IEEE, 2010.

[2] Laura Balzano, Arthur Szlam, Benjamin Recht, and Robert Nowak. K-subspaces with missing data. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pages 612–615. IEEE, 2012.

[3] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[4] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

[5] 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, 2012.

[6] Zachary Charles, Amin Jalali, and Rebecca Willett. Sparse subspace clustering with missing and corrupted data. In *2018 IEEE Data Science Workshop (DSW)*, pages 180–184. IEEE, 2018.

[7] Ehsan Elhamifar. High-rank matrix completion and clustering under self-expressive models. In *Advances in Neural Information Processing Systems*, pages 73–81, 2016.

[8] Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797, 2009.

[9] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

[10] Brian Eriksson, Laura Balzano, and Robert Nowak. High-rank matrix completion and subspace clustering with missing data. *arXiv preprint arXiv:1112.5629*, 2011.

[11] Jicong Fan and Tommy WS Chow. Matrix completion by least-square, low-rank, and sparse self-representations. *Pattern Recognition*, 71:290–305, 2017.

[12] Jicong Fan and Tommy WS Chow. Sparse subspace clustering for data with missing entries and high-rank matrix completion. *Neural Networks*, 93:36–44, 2017.

[13] Jicong Fan and Madeleine Udell. Online high rank matrix completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8690–8698, 2019.

[14] Pengzhi Gao, Meng Wang, Joe H Chow, Matthew Berger, and Lee M Seversky. Missing data recovery for high-dimensional signals with nonlinear low-dimensional structures. *IEEE Transactions on Signal Processing*, 65(20):5421–5436, 2017.

[15] Reinhard Heckel and Helmut Bölcskei. Robust subspace clustering via thresholding. *IEEE Transactions on Information Theory*, 61(11):6320–6342, 2015.

[16] Jeffrey Ho, Ming-Hsuan Yang, Jongwoo Lim, Kuang-Chih Lee, and David Kriegman. Clustering appearances of objects under varying illumination conditions. In *CVPR (1)*, pages 11–18, 2003.

[17] Wei Hong, John Wright, Kun Huang, and Yi Ma. Multiscale hybrid linear models for lossy image representation. *IEEE Transactions on Image Processing*, 15(12):3655–3671, 2006.

[18] Chun-Guang Li and Rene Vidal. Structured sparse subspace clustering: A unified optimization framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 277–286, 2015.

[19] Chun-Guang Li and René Vidal. A structured sparse plus structured low-rank framework for subspace clustering and completion. *IEEE Transactions on Signal Processing*, 64(24):6557–6570, 2016.

[20] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[21] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems*, pages 612–620, 2011.

[22] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):171–184, 2012.

[23] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan. Robust and efficient subspace segmentation via least squares regression. In *European conference on computer vision*, pages 347–360. Springer, 2012.

[24] Greg Ongie, Laura Balzano, Daniel Pimentel-Alarcón, Rebecca Willett, and Robert D Nowak. Tensor methods for nonlinear matrix completion. *arXiv preprint arXiv:1804.10266*, 2018.

[25] Greg Ongie, Rebecca Willett, Robert D Nowak, and Laura Balzano. Algebraic variety models for high-rank matrix completion. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2691–2700. JMLR. org, 2017.

[26] Daniel Pimentel-Alarcón, Laura Balzano, Roummel Marcia, R Nowak, and Rebecca Willett. Group-sparse subspace clustering with missing data. In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, pages 1–5. IEEE, 2016.

[27] Daniel Pimentel-Alarcón, Nigel Boston, and Robert D Nowak. Deterministic conditions for subspace identifiability from incomplete sampling. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2191–2195. IEEE, 2015.

[28] Daniel Pimentel-Alarcon and Robert Nowak. The information-theoretic requirements of subspace clustering with missing data. In *International Conference on Machine Learning*, pages 802–810, 2016.

[29] Daniel Pimentel-Alarcón, R Nowak, and Laura Balzano. On the sample complexity of subspace clustering with missing data. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 280–283. IEEE, 2014.

[30] Daniel Pimentel-Alarcón, 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. IEEE, 2017.

[31] Shankar Rao, Roberto Tron, Rene Vidal, and Yi Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1832–1845, 2009.

[32] Mahdi Soltanolkotabi, Emmanuel J Candes, et al. A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4):2195–2238, 2012.

[33] Manolis C Tsakiris and Rene Vidal. Theoretical analysis of sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pages 4982–4991, 2018.

[34] René Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011.

[35] René Vidal and Paolo Favaro. Low rank subspace clustering (LRSC). *Pattern Recognition Letters*, 43:47–61, 2014.

[36] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1945–1959, 2005.

[37] René Vidal, Roberto Tron, and Richard Hartley. Multiframe motion segmentation with missing data using powerfactorization and gpca. *International Journal of Computer Vision*, 79(1):85–105, 2008.

[38] Yu-Xiang Wang and Huan Xu. Noisy sparse subspace clustering. *The Journal of Machine Learning Research*, 17(1):320–360, 2016.

[39] Congyuan Yang, Daniel Robinson, and René Vidal. Sparse subspace clustering with missing entries. In *International Conference on Machine Learning*, pages 2463–2472, 2015.

[40] Chong You, Chun-Guang Li, Daniel P. Robinson, and René Vidal. Oracle based active set algorithm for scalable elastic net subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3928–3937, 2016.

[41] Chong You, Daniel P. Robinson, and René Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3918–3927, 2016.

[42] Chong You and René Vidal. Geometric conditions for subspace-sparse recovery. In *International Conference on Machine learning*, pages 1585–1593, 2015.