Exact Subspace Segmentation and Outlier Detection by Low-Rank Representation

Guangcan Liu
Electrical and Computer Engineering, National University of Singapore, 119077, Singapore

ELELIUG@NUS.EDU.SG

Huan Xu
Mechanical Engineering, National University of Singapore, 117575, Singapore

MPEXUH@NUS.EDU.SG

Shuicheng Yan
Electrical and Computer Engineering, National University of Singapore, 119077, Singapore

ELEYANS@NUS.EDU.SG

Editor:

Abstract

In this work, we address the following matrix recovery problem: suppose we are given a set of data points containing two parts, one part consists of samples drawn from a union of multiple subspaces and the other part consists of outliers. We do not know which data points are outliers, or how many outliers there are. The rank and number of the subspaces are unknown either. Can we detect the outliers and segment the samples into their right subspaces, efficiently and exactly? We utilize a so-called Low-Rank Representation (LRR) method to solve this problem, and prove that under mild technical conditions, any solution to LRR exactly recover the row space of the samples and detect the outliers as well. Since the subspace membership is provably determined by the row space, this further implies that LRR can perform exact subspace segmentation and outlier detection, in an efficient way.

Keywords: Low-Rank Modeling, Subspace Segmentation, Outlier Detection, Robust Estimation, Nuclear Norm Regularization

1. Introduction

This paper is about the following problem: suppose we are given a data matrix $X$, each column of which is a data point, and we know it can be decomposed as

$$X = X_0 + C_0,$$

(1)

where $X_0$ is a low-rank matrix with the column vectors drawn from a union of multiple subspaces, and $C_0$ is a column-sparse matrix that is non-zero in only a fraction of the columns. Except these mild restrictions, both components are arbitrary. In particular we do not know which columns of $C_0$ are non-zero, or how many non-zero columns there are. The rank of $X_0$ and the number of subspaces are unknown either. Can we recover the row
space of \( X_0 \), and the identities of the non-zero columns of \( C_0 \), efficiently and exactly? If so, under which conditions?

This problem is motivated from the **subspace segmentation** problem, an important problem in machine learning and computer vision that attracts tremendous amount of research effort (e.g., Costeira and Kanade, 1998; Eldar and Mishali, 2009; Elhamifar and Vidal, 2009; Fischler and Bolles, 1981; Gear, 1998; Gruber and Weiss, 2004; Liu et al., 2010c,b, 2010; Rao et al., 2010; Vidal, 2011, and many others). As often in computer vision and image processing applications, one observes data points drawn from the union of *multiple* subspaces (Ma et al., 2007, 2008). The goal of subspace segmentation is to segment the samples into their respective subspaces. Indeed, subspace segmentation can be regarded as a generalization of Principal Component Analysis (PCA) that has only one subspace. As such, similar to PCA, segmentation algorithms can be sensitive to the presence of outliers. In fact, because of the coupling between segmentation and outlier detection, robust subspace segmentation appears to be a challenging problem, and very few methods with theoretic guarantees, if any, have been proposed in literature.

Our main thrust, as we show below in Section 2.3, is the fact that the row space of the data samples \( X_0 \) determines the correct segmentation. Thus, both subspace segmentation and outlier detection can be transformed into solving Problem (1), where the column support of \( C_0 \) indicates the outliers, and the row space of \( X_0 \) gives the segmentation result of the “authentic” samples. To this end, we analyze the following convex optimization problem, termed **Low-Rank Representation (LRR)** (Liu et al., 2010b):

\[
\min_{Z,C} ||Z||_* + \lambda ||C||_{2,1}, \quad \text{s.t.} \quad X = XZ + C,
\]

where \( ||\cdot||_* \) denotes the sum of the singular values, also known as the nuclear norm (Fazel, 2002), the trace norm or the Ky Fan norm; \( ||\cdot||_{2,1} \) is called the \( \ell_{2,1} \) norm and is defined as the sum of \( \ell_2 \) norms of the columns of a matrix, and the parameter \( \lambda > 0 \) is used to balance the effects of the two parts.

Using the nuclear-norm based approach to tackle the subspace segmentation problem is not a completely new idea. In Liu et al. (2010b), the authors showed that if there is no outlier, then the formulation

\[
\min_Z ||Z||_*, \quad \text{s.t.} \quad X = XZ,
\]

exactly solves the subspace segmentation problem. They further conjectured that in the presence of corruptions, the formulation (2) may be helpful. However, no theoretic analysis was offered. In contrast, we show that under mild conditions, both the row space of \( X_0 \) and the column support of \( C_0 \) can be recovered by solving Problem (2). Thus, one can simultaneously perform subspace segmentation and outlier detection in an efficient way. While our analysis shares similar features as previous work in Robust Principal Component Analysis (RPCA) including Candès et al. (2009); Xu et al. (2010), it is complicated by the fact that the variable \( Z \) is left-multiplied by a dictionary matrix \( X \), and (perhaps more significantly) by the fact that the dictionary itself is contaminated by outliers. Also, it is worth noting that the problem of recovering row space with column-wise corruptions essentially cannot be addressed by existing RPCA methods (Torre and Black, 2001; Xu et al., 2010), which are designed for recovering the column space with column-wise corruptions. In this regard,
LRR also has a unique role in solving the RPCA problem under the context of corrupted features (i.e., row-wise corruptions); that is, one can recover the column space with row-wise corruptions by solving the following transposed version of (2):

$$\min_{Z,C} ||Z||_* + \lambda ||C||_{2,1}, \text{ s.t. } X^T = X^T Z + C.$$ 

As discussed above, existing RPCA methods (e.g., Xu et al., 2010) that focus on recovering the column space with column-wise corruption are fundamentally unable to address this problem.

The remainder of this paper is organized as follows. Section 2 introduces some preliminaries for reading this paper. The main results of this paper are presented and proven in Section 3 and Section 4, respectively. Section 5 presents the experimental results and Section 6 concludes this paper.

2. Preliminaries

For ease of reading, we introduce in this section some preliminaries, including the usage of mathematical notations, the concept of independent subspaces, the role of row space in subspace segmentation, and some previous results about recovering row space by LRR.

2.1 Summary of Notations

Capital letters such as $M$ are used to represent matrices, and accordingly, $[M]_i$ denotes the $i$-th column vector of $M$. Letters $U$, $V$, $I$ and their variants (complements, subscripts, etc.) are reserved for column space, row space and column support, respectively. There are four associated projection operators we use throughout. The projection onto the column space, $U$, is denoted by $P_U$ and given by $P_U(M) = UU^T M$, and similarly for the row space $P_V(M) = MVV^T$. Sometimes, we need to apply $P_V$ on the left side of a matrix. This special operator is denoted by $P_V^L$ and given by $P_V^L(\cdot) = VV^T(\cdot)$. The matrix $P_I(M)$ is obtained from $M$ by setting column $[M]_i$ to zero for all $i \notin I$. Finally, $P_T$ is the projection to the space spanned by $U$ and $V$, and given by $P_T(\cdot) = P_U(\cdot) + P_V(\cdot) - P_U P_V(\cdot)$. Note that $P_T$ depends on both $U$ and $V$, and we suppress this notation wherever it is clear which $U$ and $V$ we are using. The complementary operators, $P_U^\perp$, $P_V^\perp$, $P_T^\perp$, $P_V^L$ and $P_T^c$ are defined as usual (e.g., Xu et al., 2010). The same notation is also used to represent a subspace of matrices: e.g., we write $M \in P_U$ for any matrix $M$ that satisfies $P_U(M) = M$.

Five matrix norms are used: $||M||_*$ is the nuclear norm, $||M||_{2,1}$ is the sum of the $\ell_2$ norms of the columns $[M]_i$, $||M||_{2,\infty}$ is the largest $\ell_2$ norm of the columns, and $||M||_F$ is the Frobenius norm. The largest singular value of a matrix (i.e., the spectral norm) is $||M||$, and the smallest positive singular value is denoted by $\sigma_{\text{min}}(M)$. The only vector norm used is $||\cdot||_2$, the $\ell_2$ norm. Depending on the context, $I$ is either the identity matrix or the identity operator, and $e_i$ is the $i$-th standard basis vector.

We reserve letters $X$, $Z$, $C$ and their variants (complements, subscripts, etc.) for the data matrix (also the dictionary), coefficient matrix (in LRR) and outlier matrix, respectively. The SVD of $X_0$ and $X$ are $U_0 \Sigma_0 V_0^T$ and $U_X \Sigma_X V_X^T$, respectively. We use $\mathcal{I}$ to denote the column support of $C_0$, $d$ the ambient data dimension, $n$ the total number of data points in $X$, $\gamma \triangleq |Z_0|/n$ the fraction of outliers, and $r_0$ the rank of $X_0$. For a convex function
Liu, Xu and Yan

If \( f: \mathbb{R}^{m \times m'} \to \mathbb{R} \), we say that \( Y \) is a subgradient of \( f \) at \( M \), denoted as \( Y \in \partial f(M) \), if and only if \( f(M') \geq f(M) + \langle M' - M, Y \rangle, \forall M' \). We also adopt the conventions of using \( \text{span}(M) \) to denote the linear space spanned by the columns of a matrix \( M \), using \( y \in \text{span}(M) \) to denote that a vector \( y \) belongs to the space \( \text{span}(M) \), and using \( Y \in \text{span}(M) \) to denote that all column vectors of \( Y \) belong to \( \text{span}(M) \). A list of notations can be found in Appendix B for convenience of readers.

2.2 Independent Subspaces

The concept of independence will be used in our analysis. Its definition is as follows:

**Definition 1** A collection of \( k \) (\( k \geq 2 \)) subspaces \( \{S_1, S_2, \cdots, S_k\} \) are independent if \( S_i \cap \sum_{j \neq i} S_j = \{0\} \) for \( i = 1, \cdots, k \).

A closely related concept is pairwise disjointness, which means there is no intersection between any two subspaces, i.e., \( S_i \cap S_j = \{0\}, \forall i \neq j \). It is easy to see that when there are only two subspaces (i.e., \( k = 2 \)), independence is equivalent to pairwise disjointness. On the other hand, when \( k > 2 \), independence is a sufficient condition for pairwise disjointness, but not necessary.

2.3 Relation Between Row Space and Segmentation

The subspace memberships of the authentic samples are determined by the row space \( V_0 \). Indeed, as shown in Costeira and Kanade (1998); Gear (1998), when subspaces are independent, \( V_0 V_0^T \) forms a block-diagonal matrix: the \((i, j)\)-th entry of \( V_0 V_0^T \) can be non-zero only if the \( i \)-th and \( j \)-th samples are from the same subspace. Hence, this matrix, termed as **Shape Iteration Matrix (SIM)** (Gear, 1998), has been widely used for subspace segmentation (Costeira and Kanade, 1998; Gear, 1998; Vidal, 2011). Previous approaches simply compute the SVD of the data matrix \( X = U \Sigma X V_X^T \) and then use \( |V_X V_X^T| \) for subspace segmentation. However, in the presence of outliers, \( V_X \) can be far away from \( V_0 \) and thus the segmentation using such approaches may be inaccurate. In contrast, we show that LRR can recover \( V_0 V_0^T \) even when the data matrix \( X \) is corrupted by outliers.

In practice, the subspaces may not be independent. As one would expect, in this case \( V_0 V_0^T \) is not necessarily block-diagonal, since when the subspaces have nontrivial intersections, some samples may belong to multiple subspaces simultaneously. Nevertheless, recovering \( V_0 V_0^T \) is still of interest to subspace segmentation. Indeed, numerical experiments have shown that, as long as the subspaces are pairwise disjoint (but not independent), \( V_0 V_0^T \) is close to be block-diagonal (Liu et al., 2010a), as exemplified in Figure 1. Note that the analysis in this paper focuses on when \( V_0 V_0^T \) can be recovered, and hence does not rely on whether or not the subspaces are independent.

2.4 Relation Between Row Space and LRR

To better illustrate our intuition, we begin with the “ideal” case where there is no outlier in the data: i.e., \( X = X_0 \) and \( C_0 = 0 \). Thus, the LRR problem reduces to \( \min_{Z} \|Z\|_*, \text{ s.t. } X_0 = X_0 Z \). As shown in Liu et al. (2010a), this problem has a unique solution \( Z^* = V_0 V_0^T \), i.e., the solution of LRR identifies the row space of \( X_0 \) in this special case. Thus, when the data
are contaminated by outliers, it is natural to consider Problem (2). The following lemma, implied by Theorem 4.3 of Liu et al. (2010a), sheds insight on when LRR recovers the row space.

**Lemma 1** For any optimal solution \((Z^*, C^*)\) to the LRR problem (2), we have that

\[ Z^* \in \mathcal{P}_{V_X}^L, \]

i.e., \(Z^* \in \text{span} \left(X^T\right)\), where \(V_X\) is the row space of \(X\).

The above lemma states that the optimal solution (with respect to the variable \(Z\)) to LRR always locates within the row space of \(X\). This provides us an important clue on the conditions for recovering \(V_0V_0^T\) by \(Z^*\).

### 3. Settings and Results

In this section we present our main result: under mild assumptions detailed below, LRR can *exactly* recover both the row space of \(X_0\) (i.e., the true SIM that encodes the subspace memberships of the samples) and the columns support of \(C_0\) (i.e., the identities of the outliers) from \(X\).

While several articles, e.g., Candès and Recht (2009); Candès et al. (2009); Xu et al. (2010), have shown that the nuclear norm regularized optimization problems are powerful in dealing with corruptions including missed observations and outliers, it is considerably more challenging to establish the success conditions of LRR. This is partly due to the
bilinear interaction between the corrupted matrix $X = X_0 + C_0$ and the unknown $Z$ in the equation $X_0 + C_0 = (X_0 + C_0)Z + C$, which is essentially a matrix recovery task under a noisy dictionary, a topic not studied in literature to the best of our knowledge. Moreover, our goal is to recover row space from column-wise corruptions. This is a new task not addressed by previous RPCA and matrix recovery methods that mainly focus on recovering column space (Candès et al., 2009; Candès and Plan, 2010; Candès and Recht, 2009; Devlin et al., 1981; Torre and Black, 2001; Wright et al., 2009; Xu et al., 2010), and hence calls for new analysis tools.

3.1 Problem Settings

We discuss in this subsection three conditions sufficient for LRR to succeed. Note that these conditions also reveal how the outliers and samples are defined in LRR.

3.1.1 A Necessary Condition for Exact Recovery

Suppose $(Z^*, C^*)$ is an optimal solution to (2), then Lemma 1 concludes that the column space of $Z^*$ is a subspace of $V_X$. Hence, for $Z^*$ (or a part of $Z^*$) to exactly recover $V_0$, $V_0$ must be a subspace of $V_X$, i.e., the following is a necessary condition:

$$V_0 \in \mathcal{P}_V^L.$$ (3)

To show how the above assumption can hold, we establish the following lemma which show that (3) can be satisfied when the outliers are independent to the samples (the proof is presented in Appendix A.1).

**Lemma 2** If span $(C_0)$ and span $(X_0)$ are independent to each other, i.e., span $(C_0) \cap$ span $(X_0) = \{0\}$, then (3) holds.

3.1.2 Relatively Well-Definedness

As we discussed earlier, one technical challenge to the analysis of LRR comes from the bilinear interaction between the corrupted matrix $X = X_0 + C_0$ and the unknown $Z$ in the equation $X = XZ + C$. In fact, because the (outlier corrupted) data matrix $X$ is used as the dictionary, certain conditions to ensure that the dictionary is “well-behaved” appear to be necessary. In particular, we need the following relatively well-defined (RWD) condition.

**Definition 2** The dictionary $X$ generated by $X = X_0 + C_0$, with SVD $X = U_X \Sigma_X V_X^T$ and $X_0 = U_0 \Sigma_0 V_0^T$, is said to be RWD (with regard to $X_0$) with parameter $\beta$ if

$$\|\Sigma_X^{-1}V_X^TV_0\| \leq \frac{1}{\beta\|X\|}.$$ (4)

For LRR to succeed, the RWD parameter $\beta$ can not be too small. Notice that $\beta$ can be loosely bounded by

$$\beta \geq \frac{1}{\text{cond}(X)},$$

where $\text{cond}(X)$ is the condition number of $X$. This condition ensures that the dictionary is not too “narrow” and allows for a reasonable recovery of the row space of $X_0$.
Subspace Segmentation & Outlier Detection by Low-Rank Representation

Figure 2: Plotting the RWD parameter $\beta = 1/(\|X\|/\sum_{X}^{-1}V^{T}V_{0})$ as a function of the relative magnitude $\|C_{0}\|/\|X_{0}\|$. These results are averaged from 100 random trials. In those experiments, the outlier fraction is fixed to be $\gamma = 0.5$, and the outlier magnitude is varied for investigation. The matrices $X_{0}$ and $C_{0}$ are generated in a similar way as in Section 5.

where $\text{cond}(X) = \|X\|/\sigma_{\min}(X)$ is the condition number of $X$. This implies that $\beta = 1$ when $X$ is “perfectly well-defined” (e.g., $r_{0} = 1$ and $C_{0} = 0$). However, when $X$ is severely singular (e.g., due to the presence of outliers), this bound is too loose to guarantee RWD holds. In this case, we can apply the following bound, which essentially states that the RWD parameter $\beta$ is reasonably large when the outliers are not too large. See Appendix A.2 for the proof.

Lemma 3 If span $(C_{0})$ and span $(X_{0})$ are independent to each other, then

$$\beta \geq \frac{\sin(\theta)}{\text{cond}(X_{0})(1 + \|C_{0}\|/\|X_{0}\|)},$$

where $\text{cond}(X_{0}) = \|X_{0}\|/\sigma_{\min}(X_{0})$ is the condition number of $X_{0}$, and $\theta > 0$ is the smallest principal angle between span $(C_{0})$ and span $(X_{0})$.

Remark 1 To ensure that $\beta$ is reasonably large, the above lemma states that the outlier magnitude (comparing to the sample magnitude) should not be too large. This is verified by our numerical experiments, as shown in Fig.2.

Remark 2 To ensure that $\beta$ is reasonably large, the above lemma also states that the principal angle $\theta$ should be notably large; that is, the outliers in LRR are restricted to the data points which are notably far way from the underlying subspaces. This conclusion is consistent with the experimental observations reported in Liu et al. (2010a), which shows that LRR can distinguish between the outliers (corresponding to large $\theta$) and the corrupted samples (corresponding to small $\theta$), where a corrupted sample is sampled from the subspaces, but does not exactly lie on the underlying subspaces due to the corruptions.
3.1.3 Incoherence

Finally, as now standard (Candès and Recht, 2009; Candès et al., 2009; Xu et al., 2010), we require the incoherence condition to hold, to avoid the issue of un-identifiability. As an extreme example, consider the case where the data matrix $X_0$ is non-zero in only one column. Such a matrix is both low-rank and column-sparse, thus the problem is unidentifiable. To make the problem meaningful, the low-rank matrix $X_0$ cannot itself be column-sparse. This is ensured via the following incoherence condition.

**Definition 3** The matrix $X_0 \in \mathbb{R}^{d \times n}$ with SVD $X_0 = U_0 \Sigma_0 V_0^T$, rank $(X_0) = r_0$ and $(1-\gamma)n$ of whose columns are non-zero, is said to be column-incoherent with parameter $\mu$ if

$$\max_i \|V_0^T e_i\|_2 \leq \frac{\mu r_0}{(1-\gamma)n},$$

where $\{e_i\}$ are the standard basis vectors.

Thus if $V_0$ has a column aligned with a coordinate axis, then $\mu = (1-\gamma)n/r_0$. Similarly, if $V_0$ is perfectly incoherent (e.g., if $r_0 = 1$ and every non-zero entry of $V_0$ has magnitude $1/\sqrt{(1-\gamma)n}$), then $\mu = 1$.

3.2 The Main Result

In the following theorem, we present our main result: under mild technical conditions, any solution $(Z^*, C^*)$ to (2) exactly recovers the row space of $X_0$ and the column support of $C_0$ simultaneously.

**Theorem 1** Suppose a given data matrix $X$ is generated by $X = X_0 + C_0$, where $X_0$ is of rank $r_0$, $X$ has RWD parameter $\beta$ and $X_0$ has incoherence parameter $\mu$. Suppose $C_0$ is supported on $\gamma n$ columns. Let $\gamma^*$ be such that

$$\frac{\gamma^*}{1-\gamma^*} = \frac{324\beta^2}{49(11+4\beta)^2\mu r_0},$$

then LRR with parameter $\lambda = \frac{3}{\|X\| \sqrt{\gamma^* n}}$ strictly succeeds, as long as $\gamma \leq \gamma^*$ and (3) holds. Here, LRR “strictly succeeds” means that any optimal solution $(Z^*, C^*)$ to (2) satisfies

$$U^*(U^*)^T = V_0 V_0^T \quad \text{and} \quad I^* = I_0,$$

where $U^*$ is the column space of $Z^*$, and $I^*$ is the column support of $C^*$.

Theorem 1 indeed states that the fraction of outliers that LRR can successfully handle, namely $\gamma^*$, depends on the rank $r_0$ (the lower the better), the RWD parameter $\beta$ (the larger the better), and the incoherence parameter $\mu$ (the smaller the better).

Recall that as discussed in the introduction, LRR can be used to solve PCA tasks with feature-wise corruption by solving a transposed version of Problem (2). Hence, Theorem 1 also provides a theoretical guarantee in this setup.

4. Proof of Theorem 1

In this section, we present the detailed proofs of our main result, Theorem 1.
4.1 Roadmap of the Proof

In this subsection we provide an outline for the proof of Theorem 1. The proof follows three main steps.

1. **Equivalent Conditions:** Identify the necessary and sufficient conditions (called equivalent conditions), for any pair \((Z', C')\) to produce the exact results \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\). For any feasible pair \((Z', C')\) that satisfies \(X = XZ' + C'\), let the SVD of \(Z'\) as \(U'\Sigma'V'^T\) and the column support of \(C'\) as \(I'\). In order to produce the exact results \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\), as this is nothing but \(U'\) is a subspace of \(V_0\) and \(Z'\) is a subset of \(I_0\). On the other hand, it can be proven that \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\) are sufficient to ensure \(U'U'^T = V_0 V'_0\) and \(I' = I_0\). So, the exactness described in \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\) can be equally transformed into two constraints: \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\), which we will use to construct an oracle problem to facilitate the proof.

2. **Dual Conditions:** For a candidate pair \((Z', C')\) that respectively has the desired row space and column support, identify the sufficient conditions for \((Z', C')\) to be an optimal solution to the LRR problem \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\). These conditions are call dual conditions. For the pair \((Z', C')\) that satisfies \(X = XZ' + C'\), \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\), let the SVD of \(Z'\) as \(U'\Sigma'V'^T\) and the column-normalized version of \(C'\) as \(H'\). That is, column \([H']_i = \frac{[C']_i}{\|C'\|_2}\) for all \(i \in I_0\), and \([H']_i = 0\) for all \(i \notin I_0\) (note that the column support of \(C'\) is \(I_0\)). Furthermore, define \(\mathcal{P}_{T'}(\cdot) = \mathcal{P}_{U'}(\cdot) + \mathcal{P}_{V'}(\cdot) - \mathcal{P}_{U'}\mathcal{P}_{V'}(\cdot)\). With these notations, it can be proven that \((Z', C')\) is an optimal solution to LRR if there exists a matrix \(Q\) that satisfies

\[
\begin{align*}
\mathcal{P}_{T'}(X^T Q) &= U'V'^T, \\
\mathcal{P}_{I_0}(Q) &= \lambda H', \\
\|X^T Q - \mathcal{P}_{T'}(X^T Q)\| < 1, \\
\|Q - \mathcal{P}_{I_0}(Q)\|_{2, \infty} < \lambda.
\end{align*}
\]

Although the LRR problem \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\) may have multiple solutions, it can be further proven that any solution has the desired row space and column support, provided the above conditions have been satisfied. So, the left job is to prove the above dual conditions, i.e., construct the dual certificates.

3. **Dual Certificates:** Show that the dual conditions can be satisfied, i.e., construct the dual certificates.

The construction of dual certificates mainly concerns a matrix \(Q\) that satisfies the dual conditions. However, since the dual conditions also depend on the pair \((Z', C')\), we actually need to obtain three matrices, \(Z', C'\) and \(Q\). This is done by considering an alternate optimization problem, often called the “oracle problem”. The oracle problem arises by imposing the success conditions as additional constraints in \(\mathcal{P}_{V_0}^{L}(Z') = Z' \text{ and } \mathcal{P}_{I_0}(C') = C'\):

**Oracle Problem:** \(\min_{Z,C} \|Z\|_* + \lambda \|C\|_{2,1}\)

\(X = XZ + C, \mathcal{P}_{V_0}^{L}(Z) = Z, \mathcal{P}_{I_0}(C) = C.\)

While it is not practical to solve the oracle problem since \(V_0\) and \(I_0\) are both unknown, it significantly facilitate our proof. Note that the above problem is always feasible, as
(V_0V_0^T, C_0) is feasible. Thus, an optimal solution, denoted as (\hat{Z}, \hat{C}), exists. Observe that because of the two additional constraints, (\hat{Z}, \hat{C}) satisfies (7). Therefore, to show Theorem 4 holds, it suffices to show that (\hat{Z}, \hat{C}) is the optimal solution to LRR. With this perspective, we would like to use (\hat{Z}, \hat{C}) to construct the dual certificates. Let the SVD of \hat{Z} be \hat{U}\hat{\Sigma}\hat{V}^T, and the column-normalized version of \hat{C} be \hat{H}. It is easy to see that there exists an orthonormal matrix \hat{V} such that \hat{U}\hat{V}^T = V_0\hat{V}^T, where V_0 is the row space of X_0. Moreover, it is easy to show that P_{\hat{O}}(\cdot) = P_{V_0}^L(\cdot), P_{\hat{V}}(\cdot) = P_{\hat{V}}(\cdot), and hence the operator P_{\hat{T}} defined by \hat{U} and \hat{V}, obeys P_{\hat{T}}(\cdot) = P_{V_0}^L(\cdot) + P_{\hat{V}}(\cdot) - P_{V_0}^L P_{\hat{V}}(\cdot).

Finally, the dual certificates are finished by constructing Q as follows:

\[ Q_1 \triangleq \lambda P_{V_0}^L (X^T \hat{H}), \]
\[ Q_2 \triangleq \lambda P_{V_0}^L P_{\hat{V}}^T P_{\hat{V}} (I + \sum_{i=1}^{\infty} (P_{\hat{V}} P_{\hat{V}}^T P_{\hat{V}})^i) P_{\hat{V}} (X^T \hat{H}), \]
\[ Q \triangleq U_X \Sigma X^{-1} V_X^T (V_0 \hat{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2), \]

where \( U_X \Sigma X V_X^T \) is the SVD of the data matrix X.

4.2 Equivalent Conditions

Before starting the main proofs, we introduce the following lemmas, which are well-known and will be used multiple times in the proof.

**Lemma 4** For any column space U, row space V and column support I, the following holds.

1. Let the SVD of a matrix M be U\Sigma V^T, then \( \partial \| M \|_* = \{ UV^T + W | P_T(W) = 0, \| W \| \leq 1 \} \).

2. Let the column support of a matrix M be I, then \( \partial \| M \|_{2,1} = \{ H + L | P_I(H) = H, [H]_i = |M|_i / \| [M]_i \|_2, \forall i \in I ; P_I(L) = 0, \| L \|_{2,\infty} \leq 1 \} \).

3. For any matrices M and N of consistent sizes, we have \( P_I(MN) = MP_I(N) \).

4. For any matrices M and N of consistent sizes, we have \( P_U P_I(M) = P_T P_U(M) \) and \( P_V^L P_I(N) = P_T P_V^L (N) \).

**Lemma 5** If a matrix H satisfies \( \| H \|_{2,\infty} \leq 1 \) and is support on I, then \( \| H \| \leq \sqrt{|I|} \).

**Proof** This lemma is adapted from Xu et al. (2010). We present a proof here for completeness.

\[
\| H \| = \| H^T \| = \max_{\| x \|_2 \leq 1} \| H^T x \|_2 = \max_{\| x \|_2 \leq 1} \| x^T H \|_2 = \max_{\| x \|_2 \leq 1} \sqrt{\sum_{i \in I} (x^T [H]_i)^2} \leq \sqrt{\sum_{i \in I} 1} = \sqrt{|I|}.
\]
Lemma 6 For any two column-orthonomal matrices $U$ and $V$ of consistent sizes, we have $\|UV^T\|_{2,\infty} = \max_i \|V^T e_i\|_2$.

Lemma 7 For any matrices $M$ and $N$ of consistent sizes, we have

$$\|MN\|_{2,\infty} \leq \|M\|\|N\|_{2,\infty},$$

$$|\langle M, N \rangle| \leq \|M\|_{2,\infty}\|N\|_{2,1}$$

Proof We have

$$\|MN\|_{2,\infty} = \max_i \|MN e_i\|_2 = \max_i \|M[N]_i\|_2 \leq \max_i \|M\|\|N]_i\|_2 = \|M\| \max_i \|N\|_2 = \|M\|\|N\|_{2,\infty}.$$

$$|\langle M, N \rangle| = \sum_i \|M]_i^T[N\|_i \leq \sum_i \|M]_i^T[N\|_i \leq \sum_i \|M\|_i\|N\|_2 \leq \sum_i \|M\|_i\|N\|_2 = \|M\|_{2,\infty}\|N\|_{2,1}.$$

The exactness described in (7) seems “mysterious”. Actually, they can be “seamlessly” achieved by imposing two additional constraints in (8), as shown in the following theorem.

Theorem 2 Let the pair $(Z', C')$ satisfy $X = XZ' + C'$. Denote the SVD of $Z'$ as $U'S'V'^T$, and the column support of $C'$ as $I'$. If $P_{V_0}(Z') = Z'$ and $P_{I_0}(C') = C'$, then $U'U'^T = V_0V_0^T$ and $I' = I_0$.

Remark 3 The above theorem implies that the exactness described in (7) is equivalent to two linear constraints: $P_{V_0}(Z^*) = Z^*$ and $P_{I_0}(C^*) = C^*$. As will be seen, this can largely facilitates the proof of Theorem 4.

Proof To prove $U'U'^T = V_0V_0^T$, we only need to prove that rank$(Z') \geq r_0$, as $P_{V_0}(Z') = Z'$ implies that $U'$ is a subspace of $V_0$. Notice that $P_{I_0}(X) = X_0$. Then we have

$$X_0 = P_{I_0}(X) = P_{I_0}(XZ' + C') = P_{I_0}(XZ').$$

So, $r_0 = \text{rank}(X_0) = \text{rank}(X \cdot P_{I_0}(Z')) \leq \text{rank}(P_{I_0}(Z')) \leq \text{rank}(Z')$.

To ensure $I' = I_0$, we only need to prove that $I_0 \cap I'^c = \emptyset$, since $P_{I_0}(C') = C'$ has produced $I' \subseteq I_0$. Via some computations, we have that

$$P_{I_0}(X_0) = 0 \Rightarrow U_0 \Sigma_0 P_{I_0}(V_0^T) = 0 \Rightarrow P_{I_0}(V_0^T) = 0 \Rightarrow V_0 P_{I_0}(V_0^T) = 0.$$

(8)
Also, we have

\[ V_0 \in P_{I_0}^c \Rightarrow V_0^T = V_0^T V_X V_X^T \]
\[ \Rightarrow V_0 V_0^T = V_0 V_0^T V_X V_X^T, \]  \hspace{1cm} (9)

which simply leads to \( V_0 V_0^T V_X P_{I_0}(V_X^T) = V_0 P_{I_0}(V_0^T). \) Recalling (9), we further have

\[ V_0 P_{I_0}(V_0^T) = 0 \Rightarrow V_0 V_0^T V_X P_{I_0}(V_X^T) = V_0 P_{I_0}(V_0^T) = 0 \]
\[ \Rightarrow V_0 V_0^T V_X P_{I_0 \cap I_0^c}(V_X^T) = 0, \]  \hspace{1cm} (10)

where the last equality holds because \( I_0 \cap I_0^c \subseteq I_0. \) Also, note that \( I_0 \cap I_0^c \subseteq I_0^c. \) Then we have the following:

\[ X = XZ' + C' \Rightarrow P_{I_0 \cap I_0^c}(X) = X P_{I_0 \cap I_0^c}(Z') \]
\[ \Rightarrow U_X \Sigma_X P_{I_0 \cap I_0^c}(V_X^T) = U_X \Sigma_X V_X^T P_{I_0 \cap I_0^c}(Z') \]
\[ \Rightarrow P_{I_0 \cap I_0^c}(V_X^T) = V_X^T P_{I_0 \cap I_0^c}(Z') \]
\[ \Rightarrow V_X P_{I_0 \cap I_0^c}(V_X^T) = V_X V_X^T P_{I_0 \cap I_0^c}(Z') \]
\[ \Rightarrow V_0 V_0^T P_{I_0 \cap I_0^c}(V_X^T) = V_0 V_0^T V_X V_X^T P_{I_0 \cap I_0^c}(Z') \]  \hspace{1cm} (11)

Recalling (9) and (10), then we have

\[ V_0 V_0^T V_X P_{I_0 \cap I_0^c}(V_X^T) = 0 \Rightarrow V_0 V_0^T V_X V_X^T P_{I_0 \cap I_0^c}(Z') = 0 \]
\[ \Rightarrow V_0 V_0^T P_{I_0 \cap I_0^c}(Z') = 0 \]
\[ \Rightarrow P_{I_0 \cap I_0^c}(Z') = 0, \]

where the last equality is from the conclusion of \( Z' = V_0 V_0^T Z'. \) By \( X = X_0 + C_0, \)

\[ P_{I_0 \cap I_0^c}(C_0) = P_{I_0 \cap I_0^c}(X - X_0) = P_{I_0 \cap I_0^c}(X). \]

Notice that \( P_{I_0 \cap I_0^c}(X) = X P_{I_0 \cap I_0^c}(Z'). \) Then by (11), we have

\[ P_{I_0 \cap I_0^c}(C_0) = 0, \] and so \( I_0 \cap I_0^c = \emptyset. \)

4.3 Dual Conditions

To prove that LRR can exactly recover the row space and column support, Theorem 2 suggests us to prove that the pair \((Z', C')\) is a solution to (2), and every solution to (2) also satisfies the two constraints in Theorem 2. To this end, we write down the optimal conditions of (2), resulting in the dual conditions for ensuring the exactness of LRR.

At first, we define two operators that are closely related to the subgradient of \( \|C'\|_{2,1} \) and \( \|Z'\|_*. \)
Definition 4  1. Let $(Z', C')$ satisfy $X = XZ' + C'$, $\mathcal{P}_{I_0}(Z') = Z'$ and $\mathcal{P}_{I_0}(C') = C'$. We define the following: 

$$
\mathcal{B}(C') \triangleq \{H|\mathcal{P}_{I_0}(H) = 0; \forall i \in I_0: [H]_i = \frac{[C']_i}{\|C'\|_2}\}.
$$

Observe that $\mathcal{B}(C')$ is a column-normalized version of $C'$.

2. Let the SVD of $Z'$ as $U'\Sigma'V'^{\top}$, we further define the operator $\mathcal{P}_{T(Z')}$ as

$$
\mathcal{P}_{T(Z')}() \triangleq \mathcal{P}_{U'}() + \mathcal{P}_{V'}() - \mathcal{P}_{U'}\mathcal{P}_{V'}() = \mathcal{P}_{U'}() + \mathcal{P}_{V'}() - \mathcal{P}_{U'}^L\mathcal{P}_{V'}().
$$

Next, we present and prove the dual conditions for exactly recovering the row space and column support of $X_0$ and $C_0$, respectively.

Theorem 3 Let $(Z', C')$ satisfy $X = XZ' + C'$, $\mathcal{P}_{I_0}(Z') = Z'$ and $\mathcal{P}_{I_0}(C') = C'$. Then $(Z', C')$ is an optimal solution to (2) if there exists a matrix $Q$ that satisfies

(a) $\mathcal{P}_{T(Z')}(X^TQ) = U'\Sigma'V'^{\top}$,

(b) $\|\mathcal{P}_{T(Z')}^*(X^TQ)\| < 1$,

(c) $\mathcal{P}_{I_0}(Q) = \lambda \mathcal{B}(C')$,

(d) $\|\mathcal{P}_{I_0}(Q)\|_{2,\infty} < \lambda$.

Further, if $\mathcal{P}_{I_0} \cap \mathcal{P}_{V'} = \{0\}$, then any optimal solution to (2) will have the exact row space and column support.

Proof By standard convexity arguments (Rockafellar, 1970), a feasible pair $(Z', C')$ is an optimal solution to (2) if there exists $Q'$ such that

$$
Q' \in \partial\|Z'\|_* \quad \text{and} \quad Q' \in \lambda X^T\partial\|C'\|_{2,1}.
$$

Note that (a) and (b) imply that $X^TQ \in \partial\|Z'\|_*$. Furthermore, letting $T'$ be the column support of $C'$, then by Theorem 2 we have $T' = I_0$. Therefore (c) and (d) imply that $Q \in \lambda \partial\|C'\|_{2,1}$, and so $X^TQ \in \lambda X^T\partial\|C'\|_{2,1}$. Thus, $(Z', C')$ is an optimal solution to (2).

Notice that the LRR problem (2) may have multiple solutions. For any fixed $\Delta \neq 0$, assume that $(Z' + \Delta_1, C' - \Delta)$ is also optimal. Then by $X = X(Z' + \Delta_1) + (C' - \Delta) = XZ' + C'$, we have

$$
\Delta = X\Delta_1.
$$

By the well-known duality between operator norm and nuclear norm, there exists $W_0$ that satisfies $\|W_0\| = 1$ and $\langle W_0, \mathcal{P}_{T(Z')}(\Delta_1) \rangle = \|\mathcal{P}_{T(Z')}(\Delta_1)\|_*$. Let $W = \mathcal{P}_{T(Z')}(W_0)$, then we have that $\|W\| \leq 1$, $\langle W, \mathcal{P}_{T(Z')}(\Delta_1) \rangle = \|\mathcal{P}_{T(Z')}(\Delta_1)\|_*$ and $\mathcal{P}_{T(Z)}(W) = 0$. Let $F$ be such that

$$
[F]_i = \begin{cases} 
-\frac{[\Delta]_i}{\|\Delta\|_2}, & \text{if } i \notin I_0 \text{ and } [\Delta]_i \neq 0, \\
0, & \text{otherwise}.
\end{cases}
$$

13
Then $\mathcal{P}_{T(Z')}^2(X^TQ) + W$ is a subgradient of $\|Z'\|_*$, and $\mathcal{P}_{I_0}(Q)/\lambda + F$ is a subgradient of $\|C'\|_{2,1}$. By the convexity of nuclear norm and $\ell_{2,1}$ norm, we have

$$\begin{align*}
\|Z' + \Delta_1\|_* + \lambda\|C' - \Delta\|_{2,1} &\geq \|L'\|_* + \lambda\|C'\|_{2,1} + \langle \mathcal{P}_{T(Z')}^2(X^TQ) + W, \Delta_1 \rangle - \lambda(\mathcal{P}_{I_0}(Q)/\lambda + F, \Delta) \\
&= \|L'\|_* + \lambda\|C'\|_{2,1} + \langle \mathcal{P}_{T(Z')}^2((I_1)\|\| + \lambda\|P_{I_0}(\Delta)\|_{2,1} + \langle \mathcal{P}_{T(Z')}^2(X^TQ), \Delta_1 \rangle - \langle \mathcal{P}_{I_0}(Q), \Delta \rangle.
\end{align*}$$

Notice that

$$\begin{align*}
\langle \mathcal{P}_{T(Z')}^2(X^TQ), \Delta_1 \rangle - \langle \mathcal{P}_{I_0}(Q), \Delta \rangle &= \langle X^TQ - \mathcal{P}_{T(Z')}^2(X^TQ), \Delta_1 \rangle - \langle Q - \mathcal{P}_{I_0}(Q), \Delta \rangle \\
&= \langle - \mathcal{P}_{T(Z')}^2(X^TQ), \Delta_1 \rangle + \langle \mathcal{P}_{I_0}(Q), \Delta \rangle + \langle Q, X \Delta_1 - \Delta \rangle \\
&= \langle - \mathcal{P}_{T(Z')}^2(X^TQ), \Delta_1 \rangle + \langle \mathcal{P}_{I_0}(Q), \Delta \rangle \\
&\geq -\|\mathcal{P}_{T(Z')}^2(X^TQ)\|\|\mathcal{P}_{T(Z')}^2(\Delta_1)\|_* - \|\mathcal{P}_{I_0}(Q)\|_2,\infty\|\mathcal{P}_{I_0}(\Delta)\|_{2,1}.
\end{align*}$$

where the last inequality is from Lemma [4] and the well-known conclusion that $\|\langle MN \rangle \| \leq \|M\|\|N\|_*$ holds for any matrices $M$ and $N$.

The above deductions have proven that

$$\begin{align*}
\|Z' + \Delta_1\|_* + \lambda\|C' - \Delta\|_{2,1} &\geq \|L'\|_* + \lambda\|C'\|_{2,1} + (1 - \|\mathcal{P}_{T(Z')}^2(X^TQ)\|)\|\mathcal{P}_{T(Z')}^2(\Delta_1)\|_* \\
&\quad + (\lambda - \|\mathcal{P}_{I_0}(Q)\|_2,\infty)\|\mathcal{P}_{I_0}(\Delta)\|_{2,1}.
\end{align*}$$

However, since both $(Z', C')$ and $(Z' + \Delta_1, C' - \Delta)$ are optimal to (2), we must have

$$\|Z' + \Delta_1\|_* + \lambda\|C' - \Delta\|_{2,1} = \|L'\|_* + \lambda\|C'\|_{2,1},$$

and so

$$(1 - \|\mathcal{P}_{T(Z')}^2(X^TQ)\|)\|\mathcal{P}_{T(Z')}^2(\Delta_1)\|_* + (\lambda - \|\mathcal{P}_{I_0}(Q)\|_2,\infty)\|\mathcal{P}_{I_0}(\Delta)\|_{2,1} \leq 0.$$ Recalling the conditions (b) and (d), then we have

$$\|\mathcal{P}_{T(Z')}^2(\Delta_1)\|_* = \|\mathcal{P}_{I_0}(\Delta)\|_{2,1} = 0,$$

i.e., $\mathcal{P}_{T(Z')}^2(\Delta_1) = \Delta_1$ and $\mathcal{P}_{I_0}(\Delta) = \Delta$. By Lemma [4],

$$Z' \in \mathcal{P}_{V_X}^L, Z' + \Delta_1 \in \mathcal{P}_{V_X}^L \quad \text{and so} \quad \Delta_1 \in \mathcal{P}_{V_X}^L.$$

Also, notice that $\Delta = X\Delta_1$. Thus, we have

\begin{align*}
\mathcal{P}_{I_0}(\Delta) &= 0 &\Rightarrow& & X\mathcal{P}_{I_0}(\Delta_1) &= 0 \\
&\Rightarrow& & V_X^T\mathcal{P}_{I_0}(\Delta_1) &= 0 \\
&\Rightarrow& & \mathcal{P}_{V_X}^L\mathcal{P}_{I_0}(\Delta_1) &= 0 \\
&\Rightarrow& & \mathcal{P}_{I_0}(\mathcal{P}_{V_X}^L(\Delta_1)) &= 0 \\
&\Rightarrow& & \mathcal{P}_{I_0}(\Delta_1) &= 0,
\end{align*}
which implies that $\mathcal{P}_{I_0}(\Delta_1) = \Delta_1$. Furthermore, we have

$$
\begin{align*}
\mathcal{P}_{I_0}(\Delta_1) &= \Delta_1 = \mathcal{P}_{T(Z')}\mathcal{U}'(\Delta_1) + \mathcal{P}_{V'}\mathcal{P}_{U'}(\Delta_1) \\
&= \mathcal{P}_{U'}(\mathcal{P}_{I_0}(\Delta_1)) + \mathcal{P}_{V'}\mathcal{P}_{U'}(\Delta_1) \\
&= \mathcal{P}_{I_0}\mathcal{P}_{U'}(\Delta_1) + \mathcal{P}_{V'}\mathcal{P}_{U'}(\Delta_1) \\
\Rightarrow \mathcal{P}_{I_0}\mathcal{P}_{U'}(\Delta_1) &= \mathcal{P}_{V'}\mathcal{P}_{U'}(\Delta_1).
\end{align*}
$$

Since $\mathcal{P}_{I_0}\mathcal{P}_{U'}(\Delta_1) = \mathcal{P}_{U'}(\Delta_1)$, the above result implies that $\mathcal{P}_{U'}(\Delta_1) \in \mathcal{P}_{I_0} \cap \mathcal{P}_{V'}$.

By the assumption of $\mathcal{P}_{I_0} \cap \mathcal{P}_{V'} = \{0\}$, we have $\mathcal{P}_{U'}(\Delta_1) = 0$. Recalling Theorem 2, we have that $\mathcal{P}_{U'} = \mathcal{P}_{U_0}^L$, and so $\Delta_1 \in \mathcal{P}_{U_0}^L$. Thus, the solution $(Z' + \Delta_1, C' - \Delta)$ also satisfies $X = X(Z' + \Delta_1) + (C' - \Delta)$, $\mathcal{P}_{U_0}^L(Z' + \Delta_1) = Z' + \Delta_1$ and $\mathcal{P}_{I_0}(C' - \Delta) = C' - \Delta$. Recalling Theorem 2 again, it can be concluded that the solution $(Z' + \Delta_1, C' - \Delta)$ also exactly recovers the row space and column support, i.e., all possible solutions to (2) equally produce the exact recovery. 

### 4.4 Obtaining Dual Certificates

In this section, we complete the proof of Theorem 1 by constructing a matrix $Q$ that satisfies the conditions in Theorem 3, and proving $\mathcal{P}_{I_0} \cap \mathcal{P}_{V'} = \{0\}$ as well. This is done by considering an alternate optimization problem, often called the “oracle problem”. The oracle problem arises by imposing the equivalent conditions as additional constraints in (2):

**Oracle Problem:**

$$
\begin{align*}
\min_{Z,C} \|Z\|_* + \lambda\|C\|_{2,1} \\
X &= XZ + C, \mathcal{P}_{V_0}^L(Z) = Z; \mathcal{P}_{I_0}(C) = C.
\end{align*}
$$

(12)

Note that the above problem is always feasible, as $(V_0V_0^T, C_0)$ is a feasible solution. Thus, an optimal solution, denoted as $(\hat{Z}, \hat{C})$, exists. Observe that because of the two additional constraints, $(\hat{Z}, \hat{C})$ satisfies (7). Therefore, to show Theorem 1 holds, it suffices to show that $(\hat{Z}, \hat{C})$ is the optimal solution to LRR. With this perspective, we next show that $(\hat{Z}, \hat{C})$ is an optimal solution to (2), and obtain the dual certificates by the optimal conditions of (12).

In the rest of the paper, we need to use the following two notations: $\hat{U}\hat{\Sigma}\hat{V}^T$ is the SVD of $\hat{Z}$, and $\hat{Z}$ is the column support of $\hat{C}$.

**Lemma 8** There exists an orthonormal matrix $\hat{V}$ such that

$$
\hat{V}\hat{V}^T = \hat{\Sigma}\hat{V}^T.
$$

In addition,

$$
\mathcal{P}_T(\cdot) \triangleq \mathcal{P}_U(\cdot) + \mathcal{P}_V(\cdot) - \mathcal{P}_U\mathcal{P}_V(\cdot)
$$

$$
= \mathcal{P}_{U_0}^L(\cdot) + \mathcal{P}_V(\cdot) - \mathcal{P}_{U_0}^L\mathcal{P}_V(\cdot).
$$
Lemma 9 Let $\hat{H} = B(\hat{C})$, then we have
$$V_0P_{Z_0}(\hat{V}^T) = \lambda P_{I_0}^L(X^T\hat{H}).$$

Proof Notice that the Lagrange dual function of the oracle problem \([12]\) is
$$\mathcal{L}(Z,C,Y,Y_1,Y_2) = \|Z\|_* + \lambda\|C\|_{2,1} + \langle Y,X - XZ - C \rangle + \langle Y_1, P_{V_0}^L(Z) - Z \rangle + \langle Y_2, P_{Z_0}(C) - C \rangle,$$
where $Y, Y_1$ and $Y_2$ are Lagrange multipliers. Since $(\hat{Z}, \hat{C})$ is a solution to problem \([12]\), we have
$$0 \in \partial \mathcal{L}_Z(\hat{Z}, \hat{C}, Y, Y_1, Y_2) \quad \text{and} \quad 0 \in \partial \mathcal{L}_C(\hat{Z}, \hat{C}, Y, Y_1, Y_2).$$
Hence, there exists $\hat{W}, \hat{H}$ and $\hat{L}$ such that
$$p_f(\hat{W}) = 0, \|\hat{W}\| \leq 1, V_0\hat{V}^T + \hat{W} \in \partial \|\hat{Z}\|_*,$$
$$\hat{H} = B(\hat{C}), \mathcal{P}_{Z_0}(\hat{L}) = 0, \|\hat{L}\|_{2,\infty} \leq 1, \hat{H} + \hat{L} \in \partial \|\hat{C}\|_{2,1},$$
$$V_0\hat{V}^T + \hat{W} - X^TY - P_{V_0}^L(Y_1) = 0,$$
$$\lambda(\hat{H} + \hat{L}) - Y - P_{Z_0}(Y_2) = 0.$$ 
Let $A = \hat{W} - Y_1$ and $B = \lambda\hat{L} - Y_2$, then the last two equations above imply that
$$V_0\hat{V}^T + P_{V_0}^L(A) = \lambda X^T\hat{H} + P_{Z_0}(X^TB). \quad (13)$$
Furthermore, we have
$$P_{V_0}^L P_{Z_0}(V_0\hat{V}^T + P_{V_0}^L(A)) = P_{V_0}^L P_{Z_0}(V_0\hat{V}^T) + P_{V_0}^L P_{Z_0} P_{V_0}^L(A)$$
$$= V_0P_{Z_0}(\hat{V}^T) + P_{V_0}^L P_{Z_0} P_{Z_0}(A)$$
$$= V_0P_{Z_0}(\hat{V}^T). \quad (14)$$
Similarly, we have
$$P_{V_0}^L P_{Z_0}(\lambda X^T\hat{H} + P_{Z_0}(X^TB)) = P_{V_0}^L P_{Z_0}(\lambda X^T\hat{H}) + P_{V_0}^L P_{Z_0} P_{Z_0}(X^TB)$$
$$= P_{V_0}^L P_{Z_0}(\lambda X^T\hat{H}) + \lambda P_{V_0}^L(X^TP_{Z_0}(\hat{H}))$$
$$= \lambda P_{V_0}^L(X^T\hat{H}). \quad (15)$$
Combing \([13]\), \([14]\) and \([15]\) together, we have
$$V_0P_{Z_0}(\hat{V}^T) = \lambda P_{V_0}^L(X^T\hat{H}).$$

Before constructing a matrix $Q$ that satisfies the conditions in Theorem \([16]\), we shall prove that $P_{Z_0} \cap \mathcal{P}_V = \{0\}$ can be satisfied by choosing appropriate parameter $\lambda$. 

\[16\]
**Definition 5** Recalling the definition of $\tilde{V}$, define matrix $G$ as
\[
G \triangleq \mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T.
\]
Then we have
\[
G = \sum_{i \in \mathcal{I}_0} [\tilde{V}^T]_i ([\tilde{V}^T]_i)^T \preceq \sum_{i} [\tilde{V}^T]_i ([\tilde{V}^T]_i)^T = \tilde{V}^T \tilde{V} = I,
\]
where $\preceq$ is the generalized inequality induced by the positive semi-definite cone. Hence, $\|G\| \leq 1$.

The following lemma states that $\|G\|$ can be far away from 1 by choosing appropriate $\lambda$.

**Lemma 10** Let $\psi = \|G\|$, then $\psi \leq \lambda^2\|X\|^2\gamma n$.

**Proof** Notice that
\[
\psi = \|\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T\| = \|V_0\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T V_0^T\|
\]
\[
= \|(V_0\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T))(V_0\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T\|.
\]
By Lemma 9 we have
\[
\psi = \|\lambda\mathcal{P}_{V_0}^L(X^T \hat{H})(\lambda\mathcal{P}_{V_0}^L(X^T \hat{H}))^T\|
\]
\[
= \lambda^2\|\mathcal{P}_{V_0}^L(X^T \hat{H})(\mathcal{P}_{V_0}^L(X^T \hat{H}))^T\|
\]
\[
\leq \lambda^2\|\mathcal{P}_{V_0}^L(X^T \hat{H})\|\|\mathcal{P}_{V_0}^L(X^T \hat{H})^T\|
\]
\[
\leq \lambda^2\|X\|^2|\mathcal{I}_0| = \lambda^2\|X\|^2\gamma n,
\]
where $\|\hat{H}\|^2 \leq |\mathcal{I}_0| = \gamma n$ is due to Lemma 5.

The above lemma bounds $\psi$ far way from 1. In particular, for $\lambda \leq \frac{3}{\sqrt{\|X\|\sqrt{\gamma}}}$, we have $\psi \leq \frac{1}{4}$. So we can assume that $\psi < 1$ in sequel.

**Lemma 11** If $\psi < 1$, then $\mathcal{P}_{\mathcal{V}} \cap \mathcal{P}_{\mathcal{I}_0} = \mathcal{P}_{\mathcal{V}} \cap \mathcal{P}_{I_0} = \{0\}$.

**Proof** Let $M \in \mathcal{P}_{\mathcal{V}} \cap \mathcal{P}_{\mathcal{I}_0}$, then we have
\[
\|M\|^2 = \|MM^T\| = \|\mathcal{P}_{\mathcal{I}_0}(M)(\mathcal{P}_{\mathcal{I}_0}(M))^T\| = \|\mathcal{P}_{\mathcal{I}_0}(M\tilde{V}\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(M\tilde{V}\tilde{V}^T))^T\|
\]
\[
= \|M\|\|\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T \tilde{V}^T M^T\|
\]
\[
\leq \|M\|^2\|\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T \tilde{V}^T\| = \|M\|^2\|\mathcal{P}_{\mathcal{I}_0}(\tilde{V}^T)(\mathcal{P}_{\mathcal{I}_0}(\tilde{V}))^T\| = \|M\|^2\psi
\]
\[
\leq \|M\|^2.
\]
Since $\psi < 1$, the last equality can hold only if $\|M\| = 0$, and hence $M = 0$. Also, note that $\mathcal{P}_{\mathcal{V}} = \mathcal{P}_{\mathcal{V}}$, which completes the proof.

The following lemma plays a key role in constructing $Q$ that satisfies the conditions in Theorem 8.
Lemma 12 If \( \psi < 1 \), then the operator \( \mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi \) is an injection from \( \mathcal{P}_\psi \) to \( \mathcal{P}_\psi \), and its inverse operator is \( \mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi \mathcal{P}_\psi^{-1} \).

Proof For any matrix \( M \) such that \( \| M \| = 1 \), we have

\[
\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi (M) = \mathcal{P}_\psi \mathcal{P}_{I_0} (M \tilde{V} \tilde{V}^T) = \mathcal{P}_\psi (M \tilde{V} \tilde{V}^T) = \tilde{M} \tilde{V} \mathcal{P}_{I_0} (\tilde{V}^T) \tilde{V}^T = \tilde{M} \tilde{V} \mathcal{P}_{I_0} (\tilde{V}^T) \tilde{V}^T = \tilde{M} \mathcal{P}_{I_0} (\tilde{V}^T) \tilde{V}^T = \tilde{M} \mathcal{P}_{I_0} (\tilde{V}^T) \tilde{V}^T,
\]

which leads to \( \| \mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi \| \leq \| G \| = \psi \). Since \( \psi < 1 \), \( 1 + \sum_{i=1}^\infty (\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi)^i \) is well defined, and has a spectral norm not larger than \( 1/(1 - \psi) \).

Note that

\[
\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi = \mathcal{P}_\psi (I - \mathcal{P}_{I_0}) \mathcal{P}_\psi = \mathcal{P}_\psi (I - \mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi),
\]

thus for any \( M \in \mathcal{P}_\psi \) the following holds

\[
\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi (I + \sum_{i=1}^\infty (\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi)^i)(M) = \mathcal{P}_\psi (I - \mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi)(I + \sum_{i=1}^\infty (\mathcal{P}_\psi \mathcal{P}_{I_0} \mathcal{P}_\psi)^i)(M) = \mathcal{P}_\psi (M) = M.
\]

Lemma 13 We have

\[
\| \mathcal{P}_{I_0} (\tilde{V}^T) \|_{2, \infty} \leq \sqrt{\frac{\mu_{I_0}}{(1 - \gamma)n}}.
\]

Proof Notice that \( X = X \tilde{Z} + \tilde{C} \) and \( \mathcal{P}_{I_0} (X) = X_0 = \mathcal{P}_{I_0} (X_0) \). Then we have

\[
X = X \tilde{Z} + \tilde{C} \Rightarrow \mathcal{P}_{I_0} (X_0) = X \mathcal{P}_{I_0} (\tilde{Z}) \Rightarrow \tilde{V}_0^T = \mathcal{P}_{I_0} (\tilde{V}_0^T) = \Sigma_0^{-1} U_0 \tilde{X} \tilde{\Sigma} \mathcal{P}_{I_0} (\tilde{V}^T),
\]

which implies that the rows of \( \mathcal{P}_{I_0} (\tilde{V}^T) \) span the rows of \( \tilde{V}_0^T \). However, the rank of \( \mathcal{P}_{I_0} (\tilde{V}^T) \) is at most \( r_0 \) (this is because the rank of both \( \tilde{U} \) and \( \tilde{V} \) is \( r_0 \)). Thus, it can be concluded that \( \mathcal{P}_{I_0} (\tilde{V}^T) \) is of full row rank. At the same time, we have

\[
0 \preceq \mathcal{P}_{I_0} (\tilde{V}^T) (\mathcal{P}_{I_0} (\tilde{V}^T))^T \preceq I.
\]

So, there exists a symmetric, invertible matrix \( Y \in \mathbb{R}^{r_0 \times r_0} \) such that

\[
\| Y \| \leq 1 \quad \text{and} \quad Y^2 = \mathcal{P}_{I_0} (\tilde{V}^T) (\mathcal{P}_{I_0} (\tilde{V}^T))^T.
\]
This in turn implies that $Y^{-1}P_{I_0}(\hat{V}^T)$ has orthonomal rows. Since $P_{I_0}(V_0^T) = V_0^T$ is also row orthonomal, it can be concluded that there exists a row orthonomal matrix $R$ such that

$$Y^{-1}P_{I_0}(\hat{V}^T) = RP_{I_0}(V_0^T).$$

Then we have

$$\|P_{I_0}(\hat{V}^T)\|_{2,\infty} = \|YRP_{I_0}(V_0^T)\|_{2,\infty} \leq \|YP_{I_0}(V_0^T)\|_{2,\infty} \leq \|P_{I_0}(V_0^T)\|_{2,\infty} \leq \sqrt{\frac{\mu r_0}{(1-\gamma)n}},$$

where the last inequality is from the definition of $\mu$.

By the definition of $\bar{V}$, we further have

$$\|P_{I_0}(\bar{V}^T)\|_{2,\infty} = \|P_{I_0}(V_0^T \bar{U} \bar{V}^T)\|_{2,\infty} = \|V_0^T \bar{U} \bar{P}_{I_0}(\hat{V}^T)\|_{2,\infty} \leq \|P_{I_0}(\hat{V}^T)\|_{2,\infty} \leq \sqrt{\frac{\mu r_0}{(1-\gamma)n}}.$$

Now we define $Q_1$ and $Q_2$ used to construct the matrix $Q$ that satisfies the conditions in Theorem 3.

**Definition 6** Define $Q_1$ and $Q_2$ as follows:

$$Q_1 \triangleq \lambda P_{V_0}(X^T \hat{H}) = V_0P_{I_0}(\hat{V}^T),$$

$$Q_2 \triangleq \lambda P_{V_0}P_{I_0}P_{\bar{V}}(I + \sum_{i=1}^{\infty}(P_{\bar{V}}P_{I_0}P_{\bar{V}})^i)P_{\bar{V}}(X^T \hat{H})$$

$$= \lambda P_{I_0}P_{\bar{V}}(I + \sum_{i=1}^{\infty}(P_{\bar{V}}P_{I_0}P_{\bar{V}})^i)P_{\bar{V}}P_{V_0}(X^T \hat{H}),$$

where the equalities are due to Lemma 9 and Lemma 10.

The following Theorem almost finishes the proof of Theorem 1.

**Theorem 4** Let the SVD of the dictionary matrix $X$ as $U_X \Sigma_X V_X^T$. Assume $\psi < 1$. Let

$$Q \triangleq U_X \Sigma_X^{-1}V_X^T(V_0 \bar{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2).$$

If

$$\frac{\gamma}{1-\gamma} < \frac{\beta^2(1-\psi)^2}{(3-\psi+\beta)^2 \mu r_0},$$

and

$$\frac{(1-\psi)\sqrt{\mu r_0}}{\|X\| \sqrt{n(\beta(1-\psi) - (1+\beta) \sqrt{\frac{1}{1-\gamma} \mu r_0})}} < \lambda < \frac{1-\psi}{\|X\| \sqrt{n(2-\psi)}},$$

then $Q$ satisfies the conditions in Theorem 3 i.e., it is the dual certificate.
Proof By Lemma[11] it is concluded that $\psi < 1$ can ensure that $\mathcal{P}_\psi \cap \mathcal{P}_{I_0} = \{0\}$. Hence it is sufficient to show that $Q$ simultaneously satisfies

(S1) $\mathcal{P}_U(X^T Q) = \bar{U} \bar{V}^T,$
(S2) $\mathcal{P}_\psi(X^T Q) = \bar{U} \bar{V}^T,$
(S3) $\mathcal{P}_{I_0}(Q) = \lambda \bar{H},$
(S4) $\|\mathcal{P}_T(X^T Q)\| < 1,$
(S5) $\|\mathcal{P}_{I_0^0}(Q)\|_{2,\infty} < \lambda.$

We prove that each of these five conditions holds, in S1-S5. Then in S6, we show that the condition on $\lambda$ is not vacuous, i.e., the lower bound is strictly less than the upper bound.

First of all, we shall simplify the formula of $X^T Q$ that will be used several times in the following process. Recalling the setting [3] that assumes $\mathcal{P}_{V_0}(V_0) = V_0$, we have that $\mathcal{P}_{V_0}(Q_1) = Q_1$ and

$$
\mathcal{P}_V^L(Q_2) = \lambda \mathcal{P}_{I_0} \mathcal{P}_V(I + \sum_{i=1}^{\infty} (\mathcal{P}_V \mathcal{P}_{I_0} \mathcal{P}_V)^i) \mathcal{P}_V \mathcal{P}_{V_0} \mathcal{P}_V^L(X^T \bar{H})
$$

Further, we have

$$
X^T Q = V_X V_X^T (V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - Q_2) = \mathcal{P}_{V_X}^L(V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - Q_2)
$$

S1: Note that $\mathcal{P}_{V_0}^L(Q_1) = \lambda \mathcal{P}_{V_0}^L(X^T \bar{H})$ and $\mathcal{P}_{V_0}^L(Q_2) = 0$. Thus we have

$$
\mathcal{P}_U(X^T Q) = \mathcal{P}_U(V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - Q_2)
$$

$$
= \mathcal{P}_{V_0}^L(V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - Q_2)
$$

$$
= V_0 \bar{V}^T + \lambda X^T \bar{H} - \mathcal{P}_{V_0}^L(Q_1) - \mathcal{P}_{V_0}^L(Q_2)
$$

$$
= V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - \mathcal{P}_{V_0}^L(Q_2)
$$

$$
= V_0 \bar{V}^T + \lambda X^T \bar{H} - Q_1 - Q_2.
$$
S2: First note that

\[ P_V(Q_2) = \lambda P_V P_{I_0} P_V (I + \sum_{i=1}^{\infty} (P_V P_{I_0} P_V)^i) P_V P_{V_0}^L (X^T \hat{H}) \]

\[ = \lambda P_V P_{V_0}^L (X^T \hat{H}), \]

which is from that the operator \( P_V P_{I_0} P_V \) is an injection from \( P_V \) to \( P_V \), and its inverse is given by \( I + \sum_{i=1}^{\infty} (P_V P_{I_0} P_V)^i \).

Thus we have

\[ P_V(X^T Q) = P_V (V_0 \tilde{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2) \]

\[ = P_V (V_0 \tilde{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2) \]

\[ = V_0 \tilde{V}^T + \lambda \tilde{V}^T (X^T \hat{H}) - \lambda \tilde{V}^T P_{V_0}^L (X^T \hat{H}) - P_V (Q_2) \]

\[ = V_0 \tilde{V}^T + \lambda \tilde{V}^T P_{V_0}^L (X^T \hat{H}) - P_V (Q_2) \]

\[ = V_0 \tilde{V}^T = \hat{U} \tilde{V}^T. \]

S3: We have

\[ P_{I_0}(Q) = P_{I_0}(U_X \Sigma_X^{-1} V_X^T (V_0 \tilde{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2)) \]

\[ = U_X \Sigma_X^{-1} V_X^T V_0 P_{I_0} (\tilde{V}^T) + \lambda U_X U_X^T P_{I_0} (\hat{H}) - U_X \Sigma_X^{-1} V_X^T P_{I_0} (Q_1) \]

\[ = U_X \Sigma_X^{-1} V_X^T V_0 P_{I_0} (\tilde{V}^T) + \lambda U_X U_X^T \hat{H} - U_X \Sigma_X^{-1} V_X^T P_{I_0} (Q_1) \]

\[ = U_X \Sigma_X^{-1} V_X^T V_0 P_{I_0} (\tilde{V}^T) + \lambda U_X U_X^T \hat{H} - U_X \Sigma_X^{-1} V_X^T V_0 P_{I_0} (\tilde{V}^T) \]

\[ = \lambda U_X U_X^T \hat{H} = \lambda \tilde{P}_{V_X} (\hat{H}). \]

By \( \hat{C} = X (I - \hat{Z}) \), we have that \( \hat{C} \in \mathcal{P}_{U_X} \) and so

\[ \hat{H} = B(\hat{C}) \in \mathcal{P}_{U_X}, \]

which finishes the proof of \( P_{I_0}(Q) = \lambda \hat{H} \).

S4: Since \( P_{T_\perp} (V_0 \tilde{V}^T) = P_{T_\perp} (Q_1) = 0 \), we have

\[ P_{T_\perp}(X^T Q) = P_{T_\perp} (V_0 \tilde{V}^T + \lambda X^T \hat{H} - Q_1 - Q_2) \]

\[ = \lambda P_{T_\perp} P_{V_0}^L (X^T \hat{H}) - \lambda P_{T_\perp} P_{V_0}^L P_{V_0} (I + \sum_{i=1}^{\infty} (P_V P_{I_0} P_V)^i) P_V (X^T \hat{H}). \]

First, it can be calculated that

\[ \| \tilde{P}_{V_0}^L (X^T \hat{H}) \| \leq \| X^T \hat{H} \| \leq \| X \| \| \hat{H} \| \leq \| X \| \sqrt{n}, \]

where \( \| \hat{H} \| \leq \sqrt{n} \) is due to Lemma 5.
Second, we have the following

\[
\|P_{V_0} P_{\hat{V}} P_{\hat{V}} (I + \sum_{i=1}^{\infty} (P_{\hat{V}} P_{I_0} P_{\hat{V}}) P_{\hat{V}} (X^T \hat{H})) \| \\
\leq \|P_{I_0} P_{\hat{V}} (I + \sum_{i=1}^{\infty} (P_{\hat{V}} P_{I_0} P_{\hat{V}}) P_{\hat{V}} (X^T \hat{H})) \| \\
\leq \|(I + \sum_{i=1}^{\infty} (P_{\hat{V}} P_{I_0} P_{\hat{V}}) P_{\hat{V}} (X^T \hat{H})) \| \\
\leq \frac{1}{1 - \psi} \|P_{\hat{V}} (X^T \hat{H})\| \\
\leq \frac{\|X\| \sqrt{\gamma n}}{1 - \psi}.
\]

Thus we have that

\[
\|P_{\hat{V}} (X^T Q)\| < 1 \iff \lambda (\|X\| \sqrt{\gamma n} + \|X\| \sqrt{\gamma n} / (1 - \psi)) < 1 \\
\iff \lambda < \frac{1 - \psi}{\|X\| \sqrt{\gamma n} (2 - \psi)}.
\]

**S5:** Note that \( P_{I_0} (X^T \hat{H}) = P_{I_0} (Q_1) = 0 \). So we only need to bound the rest two parts.

By Lemma 7, we have

\[
\|P_{I_0} (U_X \Sigma_X^{-1} V_X^T V_0 \hat{V})\|_{2,\infty} = \|U_X \Sigma_X^{-1} V_X^T V_0 P_{I_0} (\hat{V})\|_{2,\infty} \\
\leq \|U_X \Sigma_X^{-1} V_X^T V_0\| \|P_{I_0} (\hat{V})\|_{2,\infty} \\
= \|\Sigma_X^{-1} V_X^T V_0\| \|P_{I_0} (\hat{V})\|_{2,\infty} \\
\leq \frac{1}{\beta \|X\|} \|P_{I_0} (\hat{V})\|_{2,\infty} \\
\leq \frac{1}{\beta \|X\|} \sqrt{\frac{\mu \nu_0}{(1 - \gamma)n}}, \tag{16}
\]

where \( \|\Sigma_X^{-1} V_X^T V_0\| \leq \frac{1}{\beta \|X\|} \) is due the definition of \( \beta \), and the last inequality is due to Lemma 13.

We expand \( Q_2 \) for convenience:

\[
Q_2 = \lambda P_{I_0} P_{\hat{V}} (I + \sum_{i=1}^{\infty} (P_{\hat{V}} P_{I_0} P_{V}) P_{\hat{V}} P_{V_0} (X^T \hat{H})) \\
= \lambda (I - V_0 V_0^T) (X^T \hat{H}) \tilde{V} \hat{V}^T (I + \sum_{i=1}^{\infty} \tilde{V} G_i \tilde{V}^T) \tilde{V} P_{I_0} (\hat{V}).
\]
Write \( Q_2 = \lambda (\hat{Q}_2 - \tilde{Q}_2) \), with

\[
\hat{Q}_2 \triangleq X^T \hat{H} \tilde{V} \tilde{V}^T (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \tilde{V} P_{I_0} (\tilde{V}^T),
\]

\[
\tilde{Q}_2 \triangleq V_0 V_0^T X^T \hat{H} \tilde{V} \tilde{V}^T (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \tilde{V} P_{I_0} (\tilde{V}^T).
\]

Then we have

\[
\| P_{I_0} (U_X \Sigma_X^{-1} V_X^T Q_2) \|_{2,\infty} = \| U_X \Sigma_X^{-1} V_X^T P_{I_0} (Q_2) \|_{2,\infty}
\]

\[
= \| U_X U_X^T \hat{H} \tilde{V} \tilde{V}^T (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \tilde{V} P_{I_0} (\tilde{V}^T) \|_{2,\infty}
\]

\[
\leq \| \hat{H} \| \| \tilde{V} \| \| V \| \| P_{I_0} (\tilde{V}^T) \|_{2,\infty}
\]

\[
\leq \| \hat{H} \| \| \tilde{V} \| \| (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \| \| P_{I_0} (\tilde{V}^T) \|_{2,\infty}
\]

\[
\leq \sqrt{\gamma n} \frac{1}{1 - \psi} \sqrt{\mu r_0} \frac{(1 - \gamma)}{n}
\]

\[
= \frac{1}{1 - \psi} \sqrt{\gamma} \mu r_0,
\]

(17)

and

\[
\| P_{I_0} (U_X \Sigma_X^{-1} V_X^T \hat{Q}_2) \|_{2,\infty} = \| U_X \Sigma_X^{-1} V_X^T P_{I_0} (\hat{Q}_2) \|_{2,\infty}
\]

\[
= \| U_X \Sigma_X^{-1} V_X^T V_0 V_0^T X^T \hat{H} \tilde{V} \tilde{V}^T (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \tilde{V} P_{I_0} (\tilde{V}^T) \|_{2,\infty}
\]

\[
\leq \| \Sigma_X^{-1} V_X^T V_0 \| \| V_0 \| X \| \| \tilde{V} \| \| \tilde{V} \| \| (I + \sum_{i=1}^{\infty} \tilde{V} G^i \tilde{V}^T) \| \| P_{I_0} (\tilde{V}^T) \|_{2,\infty}
\]

\[
\leq \frac{1}{\beta\| X \|} \| X \| \sqrt{\gamma n} \frac{1}{1 - \psi} \sqrt{\mu r_0} \frac{(1 - \gamma)}{n}
\]

\[
= \frac{1}{\beta(1 - \psi)} \sqrt{\gamma} \mu r_0.
\]

(18)
Combing (16), (17) and (18) together, we have

\[
\|\mathcal{P}_{Q_0}(Q)\|_{2,\infty} \leq \|\mathcal{P}_{Q_0}(U_X \Sigma_X^{-1} V_X V_T)\|_{2,\infty} + \lambda \|\mathcal{P}_{Q_0}(U_X \Sigma_X^{-1} V_X \tilde{Q}_2)\|_{2,\infty} \\
\leq \frac{1}{\beta} \|X\| \sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \lambda \frac{(1+\beta)\sqrt{\gamma}}{\beta(1-\psi)} \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} + \lambda \frac{\gamma}{\beta(1-\psi)} \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} \\
= \frac{1}{\beta} \|X\| \sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \lambda \frac{(1+\beta)}{\beta(1-\psi)} \sqrt{\frac{\gamma}{1-\gamma} \mu r_0}.
\]

Hence,

\[
\|\mathcal{P}_{Q_0}(Q)\|_{2,\infty} < \lambda \\
\iff \frac{1}{\beta} \|X\| \sqrt{\frac{\mu r_0}{(1-\gamma)n}} + \lambda \frac{(1+\beta)}{\beta(1-\psi)} \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} < \lambda \\
\iff \frac{1}{\beta} \|X\| \sqrt{\frac{\mu r_0}{(1-\gamma)n}} < \lambda \left(1 - \frac{1+\beta}{\beta(1-\psi)} \sqrt{\frac{\gamma}{1-\gamma} \mu r_0}\right) \\
\iff \lambda > \frac{1-\psi}{\|X\| \sqrt{n(\beta(1-\psi) - (1+\beta) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0})}},
\]

as long as \(\beta(1-\psi) - (1+\beta) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} > 0\), which is proven in the following step.

**S6:** We have shown that each of the 5 conditions hold. Finally, we show that the bounds on \(\lambda\) can be satisfied. But this amounts to a condition on the outlier fraction \(\gamma\). Indeed, we have

\[
\frac{(1-\psi)}{\|X\| \sqrt{n(\beta(1-\psi) - (1+\beta) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0})}} < \frac{1-\psi}{\|X\| \sqrt{n(2-\psi) \sqrt{\gamma}}}
\]

\[
\iff (2-\psi) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} < \beta(1-\psi) - (1+\beta) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0}
\]

\[
\iff \frac{\gamma}{1-\gamma} < \frac{\beta^2(1-\psi)^2}{(3-\psi + \beta)^2 \mu r_0},
\]

which can be satisfied, since the right hand side does not depend on \(\gamma\). Moreover, this condition also ensures \(\beta(1-\psi) - (1+\beta) \sqrt{\frac{\gamma}{1-\gamma} \mu r_0} > 0\).

We have thus shown that if \(\psi < 1\) and \(\lambda\) is within the given bounds, we can construct a dual certificate. From here, the following lemma immediately establishes our main result, Theorem 1.
Lemma 14 Let $\gamma^*$ be such that
\[
\frac{\gamma^*}{1 - \gamma^*} = \frac{324\beta^2}{49(11 + 4\beta)^2\mu r_0},
\]
then LRR, with $\lambda = \frac{3}{\|X\|\sqrt{\gamma^*}n}$, strictly succeeds as long as $\gamma \leq \gamma^*$.

Proof First note that
\[
\frac{324\beta^2}{49(11 + 4\beta)^2\mu r_0} = \frac{36\beta^2(1 - \frac{1}{\gamma})^2}{49(3 - \frac{1}{\gamma} + \beta)^2\mu r_0}.
\]

Lemma 10 implies that as long as $\gamma \leq \gamma^*$ we have the following:
\[
\psi \leq \lambda^2\|X\|^2\gamma n = \frac{9\gamma}{49\gamma^*} \leq \frac{9}{49} < \frac{1}{4}.
\]
Hence, we have
\[
\frac{\beta^2(1 - \psi)^2}{(3 - \psi + \beta)^2\mu r_0} > \frac{\beta^2(1 - \frac{1}{\gamma})^2}{(3 - \frac{1}{\gamma} + \beta)^2\mu r_0} \\
\Rightarrow \frac{\gamma^*}{1 - \gamma^*} < \frac{36\beta^2(1 - \psi)^2}{49(3 - \psi + \beta)^2\mu r_0} \\
\Rightarrow \mu r_0 < \frac{36\beta^2(1 - \psi)^2(1 - \gamma^*)}{49(3 - \psi + \beta)^2\gamma^*}.
\]

Note that $\frac{1 - \psi}{\sqrt{\gamma^*}(3 - \psi + \beta)\sqrt{\frac{\mu r_0}{1 - \gamma^*}}} \|X\|\sqrt{\gamma^*}$ as a function of $\sqrt{\frac{\mu r_0}{1 - \gamma^*}}$ is strictly increasing. Moreover, $\sqrt{\frac{\mu r_0}{1 - \gamma^*}} < \frac{\beta(1 - \psi)}{3 - \psi + \beta}$, and thus
\[
\frac{1 - \psi}{\sqrt{\gamma^*}(3 - \psi + \beta)\sqrt{\frac{\mu r_0}{1 - \gamma^*}} < \frac{\beta(1 - \psi)^2}{\sqrt{\gamma^*}(3 - \psi + \beta)\sqrt{\frac{\mu r_0}{1 - \gamma^*}}}
\]
where the last inequality holds because $\psi \geq 0$. 

25
Figure 3: The success rates obtained from 50 random trials. (a) When $\lambda = 0.2$, the success rates obtained under various settings of the outlier fraction $\gamma$. (b) When the outlier fraction is fixed to be $\gamma = 0.5$, plotting the success rate as a function of the parameter $\lambda$. In these experiments, the “success” is measured in terms of exact recovery, i.e., $U^*(U^*)^T = V_0V_0^T$ and $I^* = I_0$.

By $\psi < 1/4$, we also have

$$
\frac{1 - \psi}{\|X\| \sqrt{n(2 - \psi)}} \geq \frac{1 - \psi}{\|X\| \sqrt{\gamma^*n(2 - \psi)}} > \frac{1 - \frac{1}{4}}{\|X\| \sqrt{\gamma^*n(2 - \frac{1}{4})}} = \frac{3}{7\|X\| \sqrt{\gamma^*n}}.
$$

Hence, $\lambda = \frac{3}{7\|X\| \sqrt{\gamma^*n}}$ always satisfies the given bounds, as long as the outlier fraction $\gamma$ is not higher than $\gamma^*$.

5. Experiments

We present some numerical experiment results in this section. Our main goal is to validate the theoretical results obtained in previous section, not to verify the effectiveness of the LRR algorithm itself. For the latter, we refer the readers to the numerous works that use LRR to achieve state-of-the-art performances in applications including motion segmentation (Liu et al., 2010a; Liu and Yan, 2011; Favaro et al., 2011), image segmentation (Chen et al., 2011), saliency detection (Lang et al., 2011) and face recognition (Liu and Yan, 2011).

5.1 Results on Synthetic Data

Theorem 1 states that there exists a parameter $\lambda$ such that LRR can work well when the outlier fraction is not larger than a certain threshold. To explore this, we construct 5 pairwise disjoint subspaces $\{S_i\}_{i=1}^5$ whose bases $\{U_i\}_{i=1}^5 \in \mathbb{R}^{500}$ are computed by $U_{i+1} = TU_i, 1 \leq i \leq 4$, where $T$ is a random rotation and $U_1$ is a random orthonormal matrix of
Subspace Segmentation & Outlier Detection by Low-Rank Representation

Figure 4: Examples of the images in the Yale-Caltech dataset.

Dimension 500 × 5. Therefore, each subspace is a 5-dimensional subspace of \( \mathbb{R}^{500} \). From each subspace \( S_i \), we sample 40 data samples by \( X_i = U_i R_i, 1 \leq i \leq 5 \) where \( R_i \) is a 5 × 40 matrix with each entry uniformly distributed in \([-1, 1]\). We then construct the sample matrix \( X_0 \) as \( X_0 = [X_1, \ldots, X_5] \). Some outliers are randomly generated from zero mean Gaussian distribution with standard deviation \( s \), where \( s \) is set to be the average absolute value of the samples, to ensure that the samples and outliers have comparable magnitudes. Fixing all the other configurations, we change the number of outliers and the parameter \( \lambda \), and observe whether LRR succeeds or not. More precisely, we claim LRR succeeds if Equation (7) holds with a tolerance of 0.01%, i.e., \( \| U^* (U^*)^T - V_0 V_0^T \| < 10^{-4} \) (i.e., \( U^* (U^*)^T = V_0 V_0^T \)), and \( I^* = I_0 \) with \( I^* = \{ i : \| [C^*]_i \|_2 \geq 10^{-4} \| [X_i]_2 \| \} \). Figure 3(a) shows that LRR succeeds when \( \gamma \) is smaller than a threshold (0.6 in this example), with sharp phase-transition observed. In addition, Figure 3(b) illustrates that there exists a parameter range for obtaining exact recovery. These results are consistent with the statements in Theorem 1.

5.2 Results on Real Data

5.2.1 Datasets

To test LRR’s effectiveness in the presence of outliers and noise, we create a dataset, which we call “Yale-Caltech”, by combing Extended Yale Database B [Lee et al., 2005] and Caltech101 [Li et al., 2004]. For Extended Yale Database B, we remove the images pictured under extreme light conditions, i.e., we only use the images with view directions smaller than 45 degrees and light source directions smaller than 60 degrees, resulting in 1204 authentic samples drawn (approximately) from a union of 38 low-rank subspaces (each face class corresponds to a subspace). For Caltech101, we only select the classes containing no more than 40 images, resulting in 609 non-face outliers. Figure 4 shows some examples of this dataset.

5.2.2 Evaluation Metrics

Segmentation Accuracy (ACC): The segmentation results can be evaluated in a similar way as classification results. However, as segmentation does not provide label to each cluster, we postprocess the result to assign each cluster a label: given the ground truth classification results, the label of a cluster is the index of the ground truth class that contributes
Table 1: Segmentation accuracy (ACC) and AUC comparison on the Yale-Caltech dataset.

|            | PCA  | RPCA$_1$ | RPCA$_2,1$ | LRR  |
|------------|------|----------|------------|------|
| ACC (%)    | 77.15| 82.97    | 83.72      | 86.13|
| AUC        | 0.9653| 0.9819   | 0.9862     | 0.9927|

the maximum number of samples to the cluster. We then compute the segmentation accuracy (ACC) as the percentage of correctly classified samples.

**Areas Under Curve (AUC):** Recall that as shown in Theorem 1, the optimal solution $C^*$ is column sparse, and can be used to detect the outliers in data. In the noiseless case, this can be done by simply identifying all the nonzero columns of $C^*$. In the noisy cases, however, $C^*$ is only approximately column-sparse, and we have to threshold. That is, the $i$-th data vector of $X$ is considered to be outlier if and only if

$$
\|C^*_{i}\|_2 > \delta,
$$

where $\delta > 0$ is a parameter. To evaluate the effectiveness of outlier detection without choosing a parameter $\delta$, we consider the receiver operator characteristic (ROC) that is widely used to evaluate the performance of binary classifiers. The ROC curve is obtained by trying all possible thresholding values, and for each value, plotting the true positives rate on the Y-axis against the false positive rate value on the X-axis. We use the areas under the ROC curve, known as AUC, to evaluate the quality of outlier detection. Note that AUC score ranges between 0 and 1, and larger AUC score means more precise outlier detection.

5.2.3 Results

The goal of this test is to identify 609 non-face outliers and segment the rest 1204 face images into 38 clusters. The performance of segmentation and outlier detection is evaluated by ACC and AUC, respectively. While investigating segmentation performance, the affinity matrix is computed from all images, including both the face images and non-face outliers. Note here that the computation of ACC does not involve the outliers, as we need to clearly explore the segmentation aspect of LRR.

We resize all images into $20 \times 20$ pixels and form a data matrix $X$ of size $400 \times 1813$. Table 1 shows the results of standard PCA, RPCA$_1$ proposed in Candès et al. (2009), RPCA$_2,1$ proposed in Xu et al. (2010) and LRR. Table 1 shows that LRR achieves best performance among all methods, both for subspace segmentation and for outlier detection. We believe that the advantages of LRR, in terms of subspace segmentation, are mainly due to the fact that it directly targets on recovering the row space $V_0 V_0^T$, which is known to determine the correct segmentation. In contrast, PCA and RPCA methods are designed for recovering the column space $U_0 U_0^T$, which is designed for dimension reduction. In terms of outlier detection, the advantages of LRR are due to the fact that this dataset has a structure of
multiple subspaces, while PCA and RPCA methods are designed for the case where data come from a single subspace.

6. Conclusion

This paper studies the problem of subspace segmentation in the presence of outliers. We analyzed a convex formulation termed LRR, and showed that the optimal solution exactly recovers the row space of the authentic data and identifies the outliers. Since the row space determines the segmentation of data, LRR can perform subspace segmentation and outlier identification simultaneously.

The analysis presented in this paper differs from previous work (e.g., Candès et al., 2009; Xu et al., 2010) largely due to the fact that the dictionary used in (2) is the data matrix $X$, as opposed to the (arguably easier) identity matrix $I$ used in Candès et al. (2009) and Xu et al. (2010). As a future direction, it is interesting to investigate whether the technique presented can be extended to general dictionary matrices other than $X$ or $I$.

Appendix A. Proofs

A.1 Proof of Lemma 2

Proof Suppose the SVD of $X_0$ is $U_0 \Sigma_0 V_0^T$, and the SVD of $C_0$ is $U_C \Sigma_C V_C^T$. Suppose $U_0^\perp$ and $U_C^\perp$ are the orthogonal complements of $U_0$ and $U_C$, respectively. By the independence between span ($C_0$) and span ($X_0$), $[U_0^\perp, U_C^\perp]$ spans the whole ambient space, and thus the following linear equation system has feasible solutions $Y_0$ and $Y_C$:

$$U_0^\perp(U_0^\perp)^T Y_0 + U_C^\perp(U_C^\perp)^T Y_C = I.$$ 

Let $Y = I - U_0^\perp(U_0^\perp)^T Y_0$, then it can be computed that

$$X_0^T Y = X_0^T \quad \text{and} \quad C_0^T Y = 0,$$

i.e., $X_0 = Y X_0$ and $Y C_0 = 0$ are feasible. By $P_{I_0}(X) = X_0$, $P_{I_0}(X) = C_0$, $P_{I_0}(X_0) = X_0$ and $P_{I_0}(X_0) = 0$, the following linear equation system has feasible solutions $Y$:

$$X_0 = Y X,$$

which simply leads to $V_0 \in P_{YX}^L$.

A.2 Proof of Lemma 3

Proof Suppose $U_X \Sigma_X V_X^T$ is the SVD of $X$, $U_0 \Sigma_0 V_0^T$ is the SVD of $X_0$, $U_C$ is the column space of $C_0$, and $U_C^\perp$ is the orthogonal complement of $U_C$. By $X = X_0 + C_0$, $(U_C^\perp)^T X = (U_C^\perp)^T X_0$ and thus

$$(U_C^\perp)^T U_X \Sigma_X V_X^T = (U_C^\perp)^T U_0 \Sigma_0 V_0^T,$$

from which it can be deduced that

$$(U_C^\perp)^T U_X = (U_C^\perp)^T U_0 \Sigma_0 (V_0^T V_X \Sigma_X^{-1}).$$
Since \( \text{span}(C_0) \) and \( \text{span}(X_0) \) are independent to each other, \((U_C^\perp)^T U_0 \) is of full column rank. Let the SVD of \((U_C^\perp)^T U_0 \) be \( U_1 \Sigma_1 V_1^T \), then we have

\[
V_0^T V_X \Sigma_X^{-1} = \Sigma_0^{-1} V_1 \Sigma_1^{-1} U_1^T (U_C^\perp)^T U_X.
\]

Hence,

\[
\|V_0^T V_X \Sigma_X^{-1}\| = \|\Sigma_0^{-1} V_1 \Sigma_1^{-1} U_1^T (U_C^\perp)^T U_X\| \leq \|\Sigma_0^{-1}\| \|\Sigma_1^{-1}\| \frac{1}{\sigma_{\min}(X_0) \sin(\theta)},
\]

where \( \|\Sigma_1^{-1}\| = \frac{1}{\sin(\theta)} \) is concluded from \cite{Knyazev2002}. By \( \|X\| \leq \|X_0\| + \|C_0\| \), we further have

\[
\beta = \frac{1}{\|\Sigma_X^{-1} V_X^T V_0\||X|} \geq \frac{\sigma_{\min}(X_0) \sin(\theta)}{\|X\|} \geq \frac{\sigma_{\min}(X_0) \sin(\theta)}{\|X_0\| + \|C_0\|} \frac{1}{\sin(\theta)} \frac{\bar{V}}{\text{cond}(X_0)} \frac{1}{(1 + \sqrt{\|C_0\|} / \|X_0\|)}.
\]

**Appendix B. List of Notations**

- **X**: The observed data matrix.
- **X_0**: The ground truth of the data matrix.
- **C_0**: The ground truth of the outliers.
- **\text{cond}(\cdot)**: The condition number of a matrix.
- **d**: The ambient data dimension, i.e., number of rows of \( X \).
- **n**: The number of data points, i.e., number of columns of \( X \).
- **I_0**: The indices of outliers, i.e., non-zero columns of \( C_0 \).
- **\gamma**: Fraction of outliers, which equals \(|I_0|/n\).
- **U_0, V_0**: The left and right singular vectors of \( X_0 \).
- **\mu**: Incoherence parameter of \( V_0 \).
- **\beta**: RWD parameter of the dictionary \( X \).
- **\hat{Z}, \hat{C}**: The optimal solution of the Oracle Problem.
- **\hat{U}, \hat{V}**: The left and right singular vectors of \( \hat{Z} \).
- **\bar{V}**: An auxiliary matrix defined in Lemma 8.
- **\mathcal{B}(\cdot)**: An operator defined in Definition 4.
- **\hat{H}**: An auxiliary matrix defined in Lemma 9 as \( \hat{H} = \mathcal{B}(\hat{C}) \).
- **\bar{G}**: An auxiliary matrix defined in Definition 5.
- **\phi**: Defined in Lemma 10 as \( \psi = \|\bar{G}\| \).
References

Emmanuel Candès and Yaniv Plan. Matrix completion with noise. In Proceeding of IEEE, volume 98, pages 925–936, 2010.

Emmanuel Candès and Benjamin Recht. Exact matrix completion via convex optimization. Foundations of Computational Mathematics, 9(6):717–772, 2009.

Emmanuel Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? Journal of the ACM, to appear, 2009.

Bing Chen, Guangcan Liu, Zhongyang Huang, and Shuicheng Yan. Multi-task low-rank affinities pursuit for image segmentation. In IEEE International Conference on Computer Vision, 2011.

Paulo Costeira and Takeo Kanade. A multibody factorization method for independently moving objects. International Journal on Computer Vision, 29(3):159–179, 1998.

J. Devlin, R. Gnanadesikan, and J. Kettenring. Robust estimation of dispersion matrices and principal components. Journal of the American Statistical Association, 76(374):354–362, 1981.

Yonina Eldar and Moshe Mishali. Robust recovery of signals from a structured union of subspaces. IEEE Transactions on Information Theory, 55(11):5302–5316, 2009.

Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pages 2790–2797, 2009.

Paolo Favaro, René Vidal, and Avinash Ravichandran. A closed form solution to robust subspace estimation and clustering. In IEEE Conference on Computer Vision and Pattern Recognition, 2011.

Maryam Fazel. Matrix rank minimization with applications. PhD thesis, 2002.

Martin Fischler and Robert Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6):381–395, 1981.

C. Gear. Multibody grouping from motion images. International Journal on Computer Vision, 29(2):133–150, 1998.

Amit Gruber and Yair Weiss. Multibody factorization with uncertainty and missing data using the EM algorithm. In IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 707–714, 2004.

Andrew V. Knyazev, Merico, and E. Argentati. Principal angles between subspaces in an A-based scalar product: Algorithms and perturbation estimates. SIAM J. Sci. Comput, 23:2009–2041, 2002.

Congyan Lang, Guangcan Liu, Jian Yu, and Shuicheng Yan. Saliency detection by multi-task sparsity pursuit. IEEE Transactions on Image Processing, 2011.
Kuang-Chih Lee, Jeffrey Ho, and David Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):684–698, 2005.

Fei-Fei Li, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *Workshop of CVPR*, pages 178–188, 2004.

Guangcan Liu and Shuicheng Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *IEEE International Conference on Computer Vision*, 2011.

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yi Ma, and Yong Yu. Robust recovery of subspace structures by low-rank representation. *Preprint*, 2010a.

Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *International Conference on Machine Learning (ICML)*, pages 663–670, 2010b.

Guangcan Liu, Zhouchen Lin, Yong Yu, and Xiaoou Tang. Unsupervised object segmentation with a hybrid graph model (HGM). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):910–924, 2010c.

Yi Ma, Harm Derksen, Wei Hong, and John Wright. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1546–1562, 2007.

Yi Ma, Allen Yang, Harm Derksen, and Robert Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, 2008.

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, 2010.

R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, USA, 1970.

Fernando Torre and Michael Black. Robust principal component analysis for computer vision. In *ICCV*, pages 362–369, 2001.

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

John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in Neural Information Processing Systems*, pages 2080–2088, 2009.

Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via outlier pursuit. In *Advances in Neural Information Processing Systems*, 2010.