Robust Low-Rank Subspace Segmentation with Semidefinite Guarantees

Manuscript Submitted to International Conference on Data Mining 2010
Yuzhao Ni, Ju Sun, Xiaotong Yuan, Shuicheng Yan, Loong-Fah Cheong

Abstract—Recently there is a line of research work proposing to employ Spectral Clustering (SC) to segment (group) high-dimensional structural data such as those (approximately) lying on subspaces or low-dimensional manifolds. By learning the affinity matrix in the form of sparse reconstruction, techniques proposed in this vein often considerably boost the performance in subspace settings where traditional SC can fail. Despite the success, there are fundamental problems that have been left unsolved: the spectrum property of the learned affinity matrix cannot be gauged in advance, and there is often one ugly symmetrization step that post-processes the affinity for SC input. Hence we advocate to enforce the symmetric semidefinite constraint explicitly during learning (Low-Rank Representation with Positive SemiDefinite constraint, or LRR-PSD), and show that factually it can be solved in an exquisite optimization scheme efficiently instead of general-purpose PSD solvers that usually scale up poorly. We provide rigorous mathematical derivations to show that, in its canonical form, LRR-PSD is equivalent to the recently proposed Low-Rank Representation (LRR) scheme [1], and hence offer theoretic and practical insights to both LRR-PSD and LRR, inviting future research. As per computational cost, our proposal is comparable to that of LRR, if not less. We validate our theoretic analysis and optimization scheme by experiments on both synthetic and real data sets.

Keywords—spectral clustering; affinity matrix learning; rank minimization; robust estimation; eigenvalue thresholding

I. INTRODUCTION

This paper deals with grouping or segmentation of high-dimensional data under subspace settings. The problem is formally defined as follows

Problem 1 (Subspace Segmentation). Given a set of sufficiently dense data vectors \( X = [x_1, \ldots, x_n] \), \( x_i \in \mathbb{R}^d \) representing a sample \( \forall i = 1, \ldots, n \). Suppose the data are drawn from a union of \( k \) subspaces \( \{S_i\}_{i=1}^k \) of unknown dimensions \( \{d_i\}_{i=1}^k \) respectively, segment all data vectors into their respective subspaces.

In this regard, the vast number of available clustering algorithms, ranging from the most basic k-means method to the most elegant spectral clustering (SC) method, can all be used towards solution. Nevertheless, there are strong reasons to believe that exploiting the very regularity associated with the data can enhance the clustering performance.

We choose SC as the basic framework for subspace segmentation. SC has been extensively researched (see [2] for a recent review) and employed for many applications (e.g. image segmentation [3] in computer vision). SC has remarkable ability to deal with highly complicated data structures which may easily fail simple clustering method such as k-means. The excellent performance of SC can be partially explained via its connection to the kernel method which has been extensively studied in machine learning, specially the recent unification of weighted kernel k-means and SC [4]. The implicit data transformation into higher-dimensional spaces is likely to make the clustering task easy for basic clustering algorithms.

Analogous to the freedom to choose the kernel function in kernel methods, SC is flexible enough to admit any similarity measures in the form of affinity matrices as its input. Despite the existence of research work on SC with general affinity matrices that are not semidefinite (see e.g., [5]), in practice the Gaussian kernel \( s(x_i, x_j) = \exp(-\|x_i - x_j\|^2/\sigma^2) \) and the linear kernel \( s(x_i, x_j) = x_i^\top x_j \) are evidently the most commonly employed. Use of these kernels naturally ensures about symmetry and positive semidefiniteness of the affinity matrix. When there are processing steps that cause asymmetry, e.g., construction of nearest-neighbors based affinity matrix, there is normally an additional symmetrization step involved before the subsequent eigen-analysis.

A. Prior Work

The most intuitive way to solve the subspace segmentation problem is perhaps robust model fitting. In this aspect, classic methods such as RANSAC [6] and Expectation Minimization [7] can be employed, based on some assumptions about the data distribution, and possibly also the parametric form (mixture of Gaussians).

Most customized algorithms on this problem, however, are contributed from researchers in computer vision, to solve the 3D multibody motion segmentation (MMS) problem (see e.g., [8] for the problem statement and review of existing algorithms). In this problem, geometric argument shows that trajectories of same rigid-body motion lie on a subspace of dimension at most 4. Hence MMS serves as a typical application of subspace segmentation. There are factorization based methods [9], algebraic methods exemplified by the Generalized Principal Component Analysis (GPCA) [10], and local subspace affinity (LSA) [11] to address MMS. They are all directly or indirectly linked to SC methods,
and can be considered as different ways to construct the affinity matrix for subsequent SC (for the former is the linear kernel, and the latter two kernels defined with local subspace distances).

Of special interest to the current investigation is the recent line of work on constructing the affinity matrix by sparsity-induced optimization. Cheng et al [12] (ℓ1 graph) and Elhamifar et al [13] (sparse subspace clustering, SSC) have independently proposed to use sparse reconstruction coefficients as similarity measures. To obtain the sparse coefficients, they reconstruct one sample using all the rest samples, while regularizing the coefficient vector by ℓ1 norm to promote sparsity. Hence the problem to solve boils down to Lasso (i.e., ℓ1-regularized least square problem, [14]), which has been well studied on theoretic and computational sides (ref e.g., [15]). Most recently Liu et al [1] has proposed to compute the reconstruction collectively, and regularize the rank of the affinity matrix) for capturing global data structures. This is made possible by employing the nuclear norm minimization as a surrogate, and they also provide a robust version to resist noise and outliers.

Nuclear norm minimization as a surrogate for rank minimization is a natural generalization of the trace heuristic used for positive semidefinite matrix in several fields, such as control theory [16]. The need for rank minimization has theoretically stemmed from the exploding research efforts in compressed sensing sparked by the seminal paper [17]. In fact, generalizing from vector sparsity to spectrum sparsity for matrices is natural. The practical driving forces come from applications such as collaborative filtering, sensor localization, to just name a few [18]. From the computational side, there are several cutting-edge customized algorithms for solving the otherwise large-scale PSD problem that is likely to plague most popular off-the-shelf PSD solvers. These algorithms include prominently singular value thresholding [19], accelerated proximal gradient (APG), and augmented Lagrange multiplier (ALM) methods (see [20] for a brief review).

B. Our Investigation and Contributions

We advocate to learn an affinity matrix that makes a valid kernel directly, i.e., being symmetric positive semidefinite. This is one critical problem the previous sparse-reconstruction and global low-rank minimization approaches has bypassed. Without consideration in this aspect, the empirical behaviors of the learnt affinity matrices are poorly justified. We will focus on the global framework proposed in [1] as the global conditions are easier to gauge and additional constraints on the learnt affinity matrix can be put in directly.

We constrain the affinity matrix to be symmetric directly (LRR-PSD). Surprisingly, during analysis of the connection with the canonical form (LRR) proposed in [1], we find out the two formulations are exactly equivalent, and moreover we can accurately characterize the spectrum of the optimal solution. In addition, we successfully establish the uniqueness of the solution in that form, and hence correct one critical error reported in [1] stating that the optimal solutions are not unique.

More interestingly, we show that our advocated formulation (LRR-PSD) also admits a simple solution as reported in [1], but at a lower computational cost. As a nontrivial byproduct, we also provide a rigorous but elementary proof to nuclear-norm regularized simple least square problem with a positive semidefinite constraint, which complements the elegant closed-form solution to the general form [19].

To sum up, we highlight our contributions in two aspects: 1) we provide a rigorous proof of the equivalence between LRR and LRR-PSD, and establish the uniqueness of the optimal solution. In addition, we offer a sensible characterization of the spectrum for the optimal solution; and 2) we show that LRR-PSD can also be efficiently solved in a scheme similar to that of LRR-PSD but with notable difference at a possibly lower cost.

II. ROBUST LOW-RANK SUBSPACE SEGMENTATION WITH SEMIDEFINITE GUARANTEES

We will first set forth the notation used throughout the paper, and introduce necessary analytic tools. The canonical optimization framework of learning a low-rank affinity matrix for subspace segmentation/clustering will be presented next, and the equivalence between LRR-PSD and LRR will be formally established. Taking on the analysis, we briefly discuss about the spectrum in the robust versions of LRR-PSD and LRR, and touches on other noise assumptions. We will then proceed to present the optimization algorithm to tackle LRR-PSD under noisy settings.

A. Notation and Preliminaries

1) Summary of Notations: We will use bold capital and bold lowercase for matrices and vectors respectively, such as X and b, and use normal letters for scalars and entries of matrices and vectors, e.g., λ, X_{ij} (the (i, j)^{th} entry of matrix X). We will consider real vector and matrix spaces exclusively in this paper and use \mathbb{R}^n or \mathbb{R}^{m \times n} alike to denote the real spaces of proper dimensionality or geometry.

We are interested in five norms of a matrix X \in \mathbb{R}^{m \times n}. The first three are functions of singular values \{\sigma_i\} and they are: 1) the operator norm or induced 2-norm denoted by \|X\|_2, which is essentially the largest singular value \sigma_{\max}; 2) the Frobenius norm, defined as \|X\|_F = \left(\sum_{i=1}^d \sigma_i^2\right)^{1/2}; and 3) the nuclear norm, or sum of the singular values \|X\|_* = \sum_{i=1}^d \sigma_i, assuming d = \min(m, n). The additional two include: 4) the matrix \ell_1 norm \|X\|_1 which generalizes the vector \ell_1 norm to the concatenation of matrix columns; and 5) the group norm \|X\|_{2,1}, which sums up the \ell_2 norms of columns. Besides, the Euclidean inner product between matrices is \langle X, Y \rangle = \text{trace}(X^\top Y). This also
induces an alternative calculation of the Frobenius norm, \( \|X\|_F = \sqrt{\text{trace}(X^\top X)} \).

We will denote the spectrum (the set of eigenvalues) of a square matrix by \( \lambda(N) \), for \( N \in \mathbb{R}^{n \times n} \) (similarly the collection of singular values for a general matrix \( \sigma(X) \)). We denote the set of all \( n \times n \) real symmetric matrix by \( S^n \), and the corresponding semidefinite cone as \( S^n_+ \). The convention definiteness is not defined for asymmetric matrices.

\[ \|X\|_p \leq \|X\| \leq \|X\|_1 \leq \|X\|_2 \leq \|X\|_\infty \] 

where \( \|X\| \) is unitarily invariant if \( \|UXV\| = \|X\| \) for all unitary matrices \( U \) and \( V \) (i.e. \( U^{-1} = U^\top, V^{-1} = V^\top \)) of compatible dimension.

Interestingly common encountered unitarily invariant norms are all functions of the singular values, and lie within two general families: 1) Schatten-\( p \) norms, arising from applying the \( p \)-norm to the vector of singular values, \( \|X\|_p = \left( \sum_{i=1}^d \sigma_i^p \right)^{1/p} \); and 2) Ky-Fan \( k \) norms, representing partial ordered sums of largest singular values, \( \|X\|_{KFk} = \sum_{i=1}^k \sigma_i \), assuming \( k \leq d \) and \( \sigma_1 \geq \ldots \geq \sigma_d \) for \( d = \min(m,n) \). Of our interest here is that the nuclear norm and Frobenius norm are Schatten-1 norm and Schatten-2 norms, respectively. This fact will be critical for several places of our later argument.

Next we will state one crucial fact about the duality between the nuclear norm and the operator norm. For any norm \( \| \cdot \| \), its dual norm \( \| \cdot \|_\| \) is defined via the variational characterization

\[ \|X\|_\| = \sup_{Y} \{ \langle Y, X \rangle \mid \|Y\| \leq 1 \}, \tag{1} \]

where \( \|Y\| \leq 1 \) can always be taken as equality for the supermum to achieve, since the inner product \( \langle Y, X \rangle \) is homogeneous w.r.t. \( Y \). Then we have a formal statement about the duality

**Lemma 3 (\[22\], Proposition 2.1). The dual norm of the operator norm \( \| \cdot \|_2 \) in \( \mathbb{R}^{m \times n} \) is the nuclear norm \( \| \cdot \|_* \).**

In fact, the duality taken together with the characterization of dual norms implies \( \langle Y, X \rangle \leq \|Y\|_2 \|X\|_* \), which has been used extensively in the analysis of nuclear norm problems. Our last piece of review touches on the core of the problem, i.e., how rank minimization problems (NP-Hard) could be (conditionally) solved via nuclear norm minimization formulation which is convex. This myth lies with the concept of convex envelope, which means the tightest convex pointwise approximation to a function (tightest convex lower bound). Formally, for any (possibly nonconvex, e.g., the rank function currently under investigation) function \( f : C \rightarrow \mathbb{R} \), where \( C \) denotes a given convex set, the convex envelope of \( f \) is the largest convex function \( g \) such that \( g(x) \leq f(x) \) for all \( x \in C \). The following lemma relates the rank function to the nuclear norm via convex envelope

**Lemma 4 ([10], Theorem 1, pp.54 and Sec.5.1.5 for proof). The convex envelope of rank \( (X) \) on the set \( \{X \in \mathbb{R}^{m \times n} \mid \|X\|_2 \leq 1 \} \) is the nuclear norm \( \|X\|_* \).**

This lemma justifies the heuristic to use the nuclear norm as a surrogate of the rank, and much of recent work, e.g. low-rank matrix completion [18] and Robust Principal Component Analysis (RPCA) [23], proves theoretically under mild conditions, the optimization can be exactly equivalent. We will borrow heavily the idea of this surrogate, and build on the theoretical underpinnings to develop our formulation and analysis of subspace segmentation.

**B. Subspace Segmentation with Clean Data – An Amazing Equivalence**

To tackle the subspace segmentation problem, Liu et al [1] have proposed to learn the affinity matrix for SC via solving the following rank minimization problem

\[ \min \text{ rank}(Z), \text{ s.t. } X = XZ. \tag{2} \]

As a convex surrogate, the rank objective is replaced by the nuclear norm, and hence the formulation

\[ \text{(LRR) } \min \|Z\|_*, \text{ s.t. } X = XZ. \tag{3} \]

instead, we advocate to solve the problem incorporating the semidefinite constraint directly to produce a valid kernel directly as argued above

\[ \text{(LRR-PSD) } \min \|Z\|_*, \text{ s.t. } X = XZ, Z \succeq 0. \tag{4} \]

Liu et al [1] has established one important characterization about solution(s) to LRR for \( X \in \mathbb{R}^{d \times n} \) and \( Z \in \mathbb{R}^{n \times n} \).

**Theorem 5 ([1], Theorem 3.1). Assume the data sampling is sufficient such that \( n_1 > \text{rank}(X_i) = d_i \) and the data have been ordered by group. If the subspace are independent then there exists an optimal solution \( Z^* \) to problem LRR that is block-diagonal:**

\[ Z^*_n = \begin{bmatrix} Z_1^* & 0 & \cdots & 0 \\ 0 & Z_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_k^* \end{bmatrix} \]
Lemma 7 and validate several critical technical results preceding for a beautiful interplay between classic matrix computation and theoretical dimensions. Furthermore, reproduce the sketch here for completeness.

Theorem 6 (LRR-PSD/LRR Equivalence). The LRR problem and LRR-PSD problem is exactly equivalent, i.e. the identical optimal solution to each problem must be symmetric semidefinite. Moreover, 1) LRR and LRR-PSD both have unique but identical minimizers; 2) The spectrum of the minimizer $Z^*$ is as specified in Theorem [7].

Proof: Theorem [11] will show that LRR has a unique minimizer that is symmetric positive semidefinite, which naturally satisfies the constraint for LRR-PSD. The equivalence and uniqueness for both follow.

The development of the above results will testify the beautiful interplay between classic matrix computation and properties of nuclear norms we reviewed above. We present and validate several critical technical results preceding formal presentation of our proof.

Lemma 7 (24, Lemma 7.1.2 on Real Matrices). If $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{p \times p}$, and $M \in \mathbb{R}^{n \times p}$ satisfy

$$AM = MB, \quad \text{rank}(M) = p,$$

then there exists an orthogonal $Q \in \mathbb{R}^{n \times n}$ such that

$$Q^T AQ = T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

for $T \in \mathbb{R}^{n \times n}$, $T_{11} \in \mathbb{R}^{p \times p}$ and $T_{12}, 0$ and $T_{22}$ of compatible dimensions. Furthermore, $\lambda(T_{11}) = \lambda(A) \cap \lambda(B)$. \[\square\]

Since the proof is critical for subsequent arguments, we reproduce the sketch here for completeness.

Proof: Let

$$M = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, Q \in \mathbb{R}^{n \times n}, R_1 \in \mathbb{R}^{p \times p}$$

be a QR factorization of $X$. By substituting this into Eq. 5, rearranging and substituting at

$$\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

where

$$Q^T AQ = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

with $T_{11} \in \mathbb{R}^{p \times p}$ and others of compatible dimension. Since $R_1$ is nonsingular, $T_{21}R_1 = 0$ implying $T_{21} = 0$.

Moreover, $T_{11}R_1 = R_1B \iff T_{11} = R_1BR_1^{-1}$, suggesting $T_{11}$ and $B$ are similar and hence $\lambda(T_{11}) = \lambda(B)$. Lemma 7.1.1 [24] dictates that $\lambda(A) = \lambda(T) = \lambda(T_{11}) \cup \lambda(T_{22})$, which leads to the conclusion. \[\square\]

By induction, one can easily establish the following famous decomposition of Schur.

Lemma 8 (Real Schur Decomposition, 24 Theorem 7.4.1). If $X \in \mathbb{R}^{n \times n}$, then there exists an orthogonal $Q \in \mathbb{R}^{n \times n}$ such that

$$Q^T XQ = \begin{bmatrix} R_{11} & R_{12} & \ldots & R_{1m} \\ 0 & R_{22} & \ldots & R_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & R_{mm} \end{bmatrix},$$

where each $R_{ii}$ is either a $1 \times 1$ or $2 \times 2$ matrix having complex conjugate eigenvalues of $X$.

It is clear, that when $X$ only has real eigenvalues one can find an orthogonal $Q$ such that $Q^T XQ$ turns out to be an upper triangular matrix containing $\lambda(X)$ as its diagonal, or $Q^T XQ = \text{Diag}(\lambda) + N$, for $N$ being a strictly upper triangular matrix. It is evident irrespective of the matrix $Q$, we always have

$$\|N\|_F^2 = \|X\|_F^2 - \sum_{i=1}^n \lambda_i^2, \quad (10)$$

and we are ready to present the next useful property saying that $\|N\|_F^2$ is an indicator for symmetry.

Proposition 9. For any square matrix $X \in \mathbb{R}^{n \times n} \text{ having only real eigenvalues, } X \text{ is symmetric if and only if } \|X\|_F^2 - \sum_{i=1}^n \lambda_i^2 = 0.$

Proof: $(\implies)$ For symmetric matrix $X$, $\sigma_i(X) = |\lambda_i(X)|$, we have $\|X\|_F^2 - \sum_{i=1}^n \lambda_i^2 = \sum_{i=1}^n \sigma_i^2 - \sum_{i=1}^n \lambda_i^2 = 0.$

$(\impliedby)$ As discussed, we can always find an orthogonal $Q$ such that $Q^T XQ = \text{Diag}(\lambda) + N$, for $N$ being strict diagonal. Hence we obtain $\|X\|_F^2 = \sum_{i=1}^n \lambda_i^2 + \|N\|_F^2$, together with $\|X\|_F^2 = \sum_{i=1}^n \lambda_i^2$ indicating $\|N\|_F^2 = 0$, or $N = 0_{n \times n}$. Symmetry follows in that $X^T = (Q \text{Diag}(\lambda) Q^T)^T = Q \text{Diag}(\lambda) Q^T = X$, which concludes the verification. \[\square\]

The next lemma deals with an important inequality of nuclear norms on vertically-partitioned or horizontally-partitioned matrices.

Lemma 10 (25). Adaptation of Theorem 4.4, pp 33-34. Let $A \in \mathbb{R}^{m \times n}$ be partitioned in the form

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \text{ (rep. } A = [A_1 \ A_2]) \quad (11)$$

and the sorted singular values of $A$ be $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \geq 0$ and those of $A_1$ be $\tau_1 \geq \tau_2 \geq \cdots \tau_d \geq 0$ for...
Proof: The original proof in [25] has shown that \( \sigma_i \geq \tau_i \), \( \forall i = 1, \ldots, d \). This is because \( \sigma_i^2 \) and \( \tau_i^2 \), \( \forall i = 1, \ldots, d \) are eigenvalues of \( A^\top A \) (rep. \( AA^\top \)) and \( A_i^\top A \) (rep. \( A_i A_i^\top \)), respectively, and \( A_i^\top A = A_i A_i^\top + A_i A_i^\top \) (rep. \( AA^\top = AA^\top + AA^\top \)). Theorem 3.14 [25] dictates that \( \sigma_i^2 \geq \tau_i^2 + \lambda_i^2 \), \( \forall i = 1, \ldots, d \), where \( \lambda_i^2 \) is the smallest eigenvalue of \( A_i^\top A_i \) (rep. \( A_i A_i^\top \)). It follows \( \sum_{i=1}^d \sigma_i \geq \sum_{i=1}^d \tau_i \), and hence we have \( \|A\|_* \geq \|A_i\|_* \).

We next show stronger results saying that the inequality is strict unless \( A_2 = 0 \).

\( \implies \) Since \( \sigma_i \geq \tau_i \), \( \forall i = 1, \ldots, d \), requiring \( \|A\|_* = \|A_1\|_* \) or \( \sum_{i=1}^d \sigma_i = \sum_{i=1}^d \tau_i \) amounts to imposing \( \sigma_i = \tau_i \), \( \forall i = 1, \ldots, d \), which suggests \( \sum_{i=1}^d \sigma_i^2 = \sum_{i=1}^d \tau_i^2 \), or trace \( A^\top A = \text{trace} (A_i A_i^\top) \) (rep. trace \( AA^\top ) = \text{trace} (A_i A_i^\top) \), identically \( \|A\|_F^2 = \|A_i\|_F^2 \). But we also have from the above argument trace \( A^\top A = \text{trace} (A_i A_i^\top) + \text{trace} (A_i A_i^\top) \) (rep. trace \( AA^\top ) = \text{trace} (A_i A_i^\top) + \text{trace} (A_i A_i^\top) \), or \( \|A\|_F^2 = \|A_i\|_F^2 + \|A_i\|_F^2 \). Taking them together we obtain \( \|A_2\|_F^2 = 0 \), implying \( A_2 = 0 \).

\( \iff \) Simple substitution verifies the equality and also completes the proof.

Theorem 11. Optimization problem LRR has a unique minimizer \( Z^* \). Moreover \( Z^* \) is semidefinite, with the only positive eigenvalue 1 of multiplicity \( r = \text{rank}(X) \) and the rest all zeros.

Proof: We first show the claim about the spectrum of \( Z^* \), and then proceed to prove the uniqueness, and finally the symmetry. It will be manifest the tight coupling of these three aspects.

(Spectrum analysis) By \( XZ = X \), we have \( Z^\top X^\top = X^\top \). Taking \( r \) independent columns from \( X^\top \) (i.e., \( r \) independent rows from \( X \)) and organize as a submatrix \( M \) of \( X^\top \), we obtain \( Z^\top M = MI \). By Lemma 7 and its proof, we have a QR factorization of \( M \) and one similarity transform of \( Z^* \), as respectively

\[
M = [U \quad U_u^\perp] \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix},
\]

\[
[U \quad U_u^\perp]^\top Z^\top [U \quad U_u^\perp] = T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}.
\]

where \( U_u^\perp \) spans the complementary dual subspace of \( U \). Further we have \( \lambda(T_{11}) = \lambda(Z^*) \cap \lambda(I) \), indicating that \( T_{11} \) has the only eigenvalue of 1 with multiplicity determined by the dimension of \( T_{11} = U_u^\perp Z^\top U \), or \( \text{rank}(M) \), which is also \( r = \text{rank}(X) \).

We continue to show that towards minimal \( \|Z^*\|_* \), \( \lambda(T_{22}) \) is 0. By the unitary invariance property of nuclear norm, \( \min : \|Z\|_* \iff \min : \|T\|_* \). Noticing that

\[
\begin{bmatrix} T_{12} \\ T_{22} \end{bmatrix} = \begin{bmatrix} U & U_u^\perp \end{bmatrix} Z^\top \begin{bmatrix} U & U_u^\perp \end{bmatrix} = Q \begin{bmatrix} Z^\top U \end{bmatrix},
\]

we have the freedom to choose the values since \( U_u^\perp \) is not dictated by data. To obtain the minimizer, we have to make \( T_{12} = T_{22} = 0 \) in that

\[
\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \geq \begin{bmatrix} T_{11} \\ 0 \end{bmatrix} \iff \|T_{11}\|_* = \|T_{11}\|_*,
\]

where the inequality have followed Lemma 10 and equality is obtained only when \( T_{12} = T_{22} = 0 \). By Lemma 7.1.1 [24], \( \lambda(T) = \lambda(T_{11}) \cup \lambda(T_{22}) \) with 1 of multiplicity \( r \) and 0 of multiplicity \( n - r \), and hence \( \lambda(Z^*) \). We can carry this result to \( Z^* \) thanks to the fact that \( \lambda(Z) = \lambda(Z^\top) \) (ref. §1.4.1. [26]).

(Uniqueness of \( Z^* \)) Suppose a perturbed version \( Z' = Z^* + H \) is also a minimizer. So \( XZ' = X(Z^* + H) = X = XZ^* \), suggesting \( XH = 0 \) or \( H \) is in \( \text{Null}(X) \) (which is complement to the row space). We have

\[
H^\top X^\top = 0 \iff H^\top [U \quad U_u^\perp] \begin{bmatrix} R \\ 0 \end{bmatrix} = 0
\]

\[
\iff H^\top U R = 0 \iff H^\top U = 0,
\]

where the last equality holds because \( R \) is nonsingular. If \( H \neq 0 \), we have

\[
[U \quad U_u^\perp]^\top Z^\top [U \quad U_u^\perp] = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} U & U_u^\perp \end{bmatrix} \begin{bmatrix} U & U_u^\perp \end{bmatrix} = \begin{bmatrix} T_{11} & U^\top H^\top U_u^\perp \\ 0 & (U_u^\perp)^\top H^\top U_u^\perp \end{bmatrix},
\]

where we have substituted the analytic values of \( Z^* \) and its corresponding \( T_{ij}, \forall i,j = \{1, 2\} \) as discussed above and the fact \( H^\top U = 0 \). Since \( H^\top U_u^\perp \neq 0 \) (otherwise together with \( H^\top U = 0 \) we would obtain \( H = 0 \)), employing the inequality in Lemma 10 again we see that

\[
\begin{bmatrix} T_{11} & U^\top H^\top U_u^\perp \\ 0 & (U_u^\perp)^\top H^\top U_u^\perp \end{bmatrix} \iff \|T_{11}\|_* = \|Z\|_*.
\]

In other words, the objective is strictly increased unless \( H^\top = 0 \) or \( H = 0 \), which establishes the uniqueness.

(Symmetry of \( Z^* \)) We will need a formal characterization of the operator norm \( \|Z\|_* \) (the largest singular value) for any matrix \( Z \in \mathbb{R}^{n \times n} \) as follows (see e.g., [27], §3.1.1 and its proof, we only consider square matrices here although the definition is general)

\[
\|Z\|_2 = \sup \{ \|Zv\| : \|v\| = 1, v \in \mathbb{R}^n \},
\]

where we slightly abuse the general notation \( \| \cdot \| \) and use it for the \( \ell_2 \) vector norm temporarily. It is easy to check that \( \|Z^*\|_2 = \|Z^*\|_2 = 1 \) as follows. Recall independent
columns $M$ of $X^T$ defines an orthonormal basis $U$ for the column space of $X^T$, as in (12), and $[U^T\ U^{1\perp}]$ constitutes an orthogonal basis for $\mathbb{R}^n$, with $M \in \mathbb{R}^{n \times r}$. Hence any vector $v \in \mathbb{R}^n$ can be represented as $v = Uv_U + U^{1\perp}v_{U^{1\perp}}$, where $v_U$ and $v_{U^{1\perp}}$ denote the component vectors corresponding to $U$ and $U^T$, respectively. Analysis in (13) and the facts $T_{12} = T_{22} = 0$ and $Q^T$ being orthogonal imply $Z^*U^{1\perp} = 0$. Hence we have

$$
\|Z^*v\|^2 = \text{trace}\left((v^T Z^*Z^*v)^2\right)
= \text{trace}\left([\text{Uv}_U + U^{1\perp}v_{U^{1\perp}}]^T Z^*Z^* (Uv_U + U^{1\perp}v_{U^{1\perp}})\right)
= \text{trace}\left([Uv_U]^T Z^*Z^* (Uv_U)\right)
= \|Z^*v_U\|^2 = \|Uv_U\|^2 = \|v_U\|^2,
$$

(19)

where the second last equality follows because $Z^*U^{1\perp} = 0$. Therefore, we have that $Z^*$ has at most $r$ nonvanishing singular values, denoted as $1 = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. Recall that we have also established $Z^*$ has $r$ nonvanishing eigenvalues of value 1, $\lambda_1 = \lambda_2 = \cdots = \lambda_r$. We have $\|Z^*\|^2_F = \sum_{i=1}^r \sigma_i^2 \leq \sum_{i=1}^r \lambda_i^2$. However, as indicated in proof of Proposition 9 for $Z^*$ having real eigenvalues, $\|Z^*\|^2_F \geq \sum_{i=1}^r \lambda_i^2$. Taking them together, we obtain $\|Z^*\|^2_F = \sum_{i=1}^r \lambda_i^2$ and hence $\sigma_1 = \sigma_2 = \cdots = \sigma_r = 1$. By Proposition 9 we have verified that $Z^*$ is symmetric.

The nonnegative spectrum established before and symmetry together dictate that $Z^* \succeq 0$, and hence concludes the proof.

C. Robust Subspace Segmentation with Data Containing Outliers and Noises

To account for noises and outliers, explicit distortion terms can be introduced into the objective and constraint. Hence we obtain the robust version of LRR-PSD and LRR respectively as follows

$$
\min_{Z,E} \|Z\|_* + \lambda \|E\|_\ell, \text{ s.t. } X = XZ + E, Z \succeq 0,
$$

(20)

$$
\min_{Z,E} \|Z\|_* + \lambda \|E\|_\ell, \text{ s.t. } XZ + E = X.
$$

(21)

We have used $\| \cdot \|_\ell$ to mean generic norms. We caution that we cannot in general expect these two versions to be equivalent despite the provable equivalence of LRR-PSD and LRR. Remarkably, the problem has changed much due to the extra variable. Nevertheless, it is still possible to partially gauge the behaviors of the solutions as follows.

Suppose an optimal $E^*$ is somehow achieved (i.e., we assume it is fixed), we are then only concerned with

$$
\min_{Z,E} \|Z\|_* \text{ s.t. } XZ + E^* = X, (Z \succeq 0).
$$

(22)

Since columns of $E^*$ must be in the column space of $X$, we assume $E^* = X\delta E$. Then we obtain from the equality constraint $X(Z + \delta E) = X$. By employing similar process in the proof of Theorem 11 one can easily verify that

$$
\begin{bmatrix}
U & U^{1\perp}
\end{bmatrix}^T (Z^* + \delta E)^T \begin{bmatrix}
U & U^{1\perp}
\end{bmatrix}
= \begin{bmatrix}
T_{11}^2 & T_{12}^2 & T_{13}^{E^*} & T_{14}^{E^*} & T_{15}^{E^*} & T_{16}^{E^*}
\end{bmatrix}
$$

(23)

where the notation is consistent with the proof in Theorem 11. So towards minimizing $\|Z^*\|_*$, we can always have $T_{11}^* = 0$, or $T_{12}^* = T_{22}^* = 0$, for any $\delta E^T$. So the rest of spectrum of $Z^*$ is determined by $T_{11}^*$, and we have that $Z^*$ can have at most $r$ nonvanishing eigenvalues, where $r = \text{rank}(X)$. Note that $(T_{11}^{E^*} + T_{11}^{E^*})$ has $r$ eigenvalues of 1, so spectrum of $T_{11}^{E^*}$ will be perturbation of that since the norm of $T_{11}^{E^*}$ is in general small. This is also confirmed by our numerical experiments in IV-C.

Moreover, we have intentionally left the norm for $E$ unspecified since it apparently depends on the noise model we assume. The use of $\|E\|_{2,1}$ assumes the noise is sample-specific. In practice, however, a more natural assumption is uniformly random, i.e., each dimension of every data sample has the same chance of getting corrupted. In this case, the simple $\|E\|_1$ will suffice. We demonstrate via experiments IV-C and show that indeed $\|\cdot\|_1$ is more robust in that case.

The above comments about spectrum properties and noise model selection apply to both settings.

D. Solving Robust LRR-PSD via Eigenvalue Thresholding

The equivalence of LRR-PSD and LRR does not readily translate to the respective robust versions, and hence we need to figure out ways of solving the robust LRR-PSD. Due to the strong connection between these two problems, however, we will still try to employ the Augmented Lagrange Multiplier (ALM) method (see e.g., [20]) to tackle this as in [1].

We first convert the problem into its equivalent form as

$$
\min_{Z,E,J} \|J\|_* + \lambda \|E\|_\ell, \text{ s.t. } X = XZ + E, Z = J, Z \succeq 0,
$$

(24)

where we have used $\|E\|_\ell$ to mean generic norms. Forming the partial ALM problem, we have

$$
\min_{Z,E,J \geq 0, Y_1, Y_2} \|J\|_* + \lambda \|E\|_\ell
+ \langle Y_1, X - XZ - E \rangle + \langle Y_2, Z - J \rangle
+ \frac{\mu}{2} \|X - XZ - E\|^2_F + \frac{\mu}{2} \|Z - J\|^2_F.
$$

(25)
We can then follow the inexact ALM routine\(^5\) to update \(Z, E, J, Y_1, Y_2\) alternately. While fixing others, how to update \(E\) depends on the norm \(|\cdot||\cdot\|\). There are bunch of norms facilitates closed-form solution, such as the \(|\cdot||\cdot|_2\), as we discussed and \(\sum\sum\) (see e.g., \(\text{[20]}\)).

Updating \(J(J \geq 0)\) is the major obstacle to clean up. To be specific, we will be facing problem of this form to update \(J\)

\[
M^* = \arg\min_M \frac{1}{\mu} \|M\|_s + \frac{1}{2} \|M - G\|_F^2, \text{ s.t. } M \succeq 0, \tag{26}
\]

where \(G\) may or may not be symmetric. We will next show in Theorem\(^6\) that symmetric \(G\) facilitates a closed-form solution, and generalize this in Theorem\(^7\) which basically states that asymmetric \(G\) also leads to a closed-form solution. Moreover, the major computational cost lies with eigen-decomposition of a symmetric square matrix, as against singular value decomposition of a square matrix of the same size in solving the counterpart in robust\(\text{LRR}\).

**Lemma 12 (\text{[21]}, Lemma 3.2).** For any block partitioned matrix \(X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\), this inequality holds

\[
\|X\|_s \geq \left\| \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \right\|_s = \|A\|_s + \|D\|_s. \tag{27}
\]

Similar inequality also holds for the square of Frobenius norm \(\|\cdot\|_F^2\).

**Theorem 13.** For any symmetric matrix \(S \in S^n\), the unique closed form solution to the optimization problem

\[
M^* = \arg\min_M \frac{1}{\mu} \|M\|_s + \frac{1}{2} \|M - S\|_F^2, \text{ s.t. } M \succeq 0, \tag{28}
\]

takes the form

\[
M^* = Q \text{ Diag} \left(\max(\lambda - 1/\mu, 0)\right) Q^T, \tag{29}
\]

whereby \(S = QAQ^T\), for \(A = \text{Diag}(\lambda)\), is the spectrum(eigen-) decomposition of \(S\) and \(\lambda\) should be understood element-wise.\(^6\)

**Proof:** Observing that the objective is strictly convex over a convex set, we asserts there exists a unique minimizer. The remaining task to single out the minimizer. Symmetric \(S\) admits a spectrum factorization \(S = QAQ^T\), where \(Q^{-1} = Q^T\). We set \(M = Q^T MQ\), and hence the optimization in Eq.\(^{28}\) can be cast into

\[
\tilde{M}^* = \arg\min_M \frac{1}{\mu} \|\tilde{M}\|_s + \frac{1}{2} \|\tilde{M} - A\|_F^2, \text{ s.t. } \tilde{M} \succeq 0. \tag{30}
\]

By the unitary invariance property of the Frobenius norm and the nuclear norm, and the fact that \(M \succeq 0 \iff Q^T MQ \succeq 0\) with unitary (orthogonal) \(Q\), we assert these two optimization problems are exactly equivalent (in the sense that \(M\) and \(\tilde{M}\) can be recovered from each other deterministically).

Next we argue that a minimizer \(\tilde{M}^*\) must be a diagonal matrix. Let \(f(M) = 1/\mu \|M\|_s + 1/2 \|M - \Lambda\|_F^2\). In fact, for a non-diagonal matrix \(\tilde{M}_0\), we can always restrict it to diagonal elements to get \(\tilde{M}_0\) such that \(f(\tilde{M}_0) < f(\tilde{M}_0)\) by Lemma.\(^1\) and the fact \(\Lambda\) being diagonal. The strict inequality holds since restriction from a non-diagonal matrix to its diagonal elements results in strict decrease in square of the Frobenius norm. So assuming \(\tilde{M} = \text{Diag}(\xi_1, \cdots, \xi_n)\) and \(\Lambda = \text{Diag}(\lambda_1, \cdots, \lambda_n)\), the problem reduces to a quadratic program w.r.t. \(\{\xi_i\}_{i=1}^n\)

\[
\{\xi_i\}_{i=1}^n = \arg\min_{\{\xi_i\}_{i=1}^n} \frac{1}{\mu} \sum_{i=1}^n \xi_i + \frac{1}{2} \sum_{i=1}^n \|\xi_i - \lambda_i\|^2, \text{ s.t. } \xi_i \geq 0, \forall i. \tag{31}
\]

The programming is obviously separable and simple manipulation suggests the unique closed form solution \(\xi_i^* = \max(\lambda_i - 1/\mu, 0)\), which concludes the proof.\(\blacksquare\)

**Remark 14.** Note that uniqueness of the solution may not be directly translated from Eq.\(^{30}\) to \(\text{[28]}\) since one may argue \(Q\) is not unique in general. There are three causes to the ambiguity: 1) general sign reversal ambiguity of eigenvectors, 2) freedom with eigenvectors corresponding to the zero eigenvalues, and 3) freedom with eigenvectors corresponding to eigenvalues with multiplicity greater than 1. Noticing that \(M^* = \sum_{i=1}^n \max(\lambda_i - 1/\mu, 0) q_i q_i^T\), the sign ambiguity and problems caused by zero-valued eigenvalues are readily removed in view of form of the summant \(\max(\lambda_i - 1/\mu, 0) q_i q_i^T\). For the last problem, assume one repeated eigenvalue \(\lambda_i\) has one set of its eigenvectors arranged column-wise in \(V = [v_1, \cdots, v_k]\), which essentially spans a \(k\)-dimensional subspace (and acts as the basis). So this part of contribution to \(M^*\) can be written as \(\max(\lambda_i - 1/\mu, 0) q_i q_i^T\). Realizing that generating a new set of eigenvectors via linear combination can be accounted for by a rotation to the original basis vectors, namely \(V = VR_{k \times k}\) for \(R^T R = I\) in that subspace, we have \(\lambda_i V V^T = \lambda_i V R (R^T)^T = \lambda_i V V^T\). Hence the sum is not altered by any cause.

In fact, building on Theorem\(^13\) we can proceed to devise a more general result on any real square matrix as follows\(^6\)

**Theorem 15.** For any square matrix \(P \in \mathbb{R}^{n \times n}\), the unique closed form solution to the optimization problem

\[
M^* = \arg\min_M \frac{1}{\mu} \|M\|_s + \frac{1}{2} \|M - P\|_F^2, \text{ s.t. } M \succeq 0, \tag{32}
\]

\(^5\)Toh and Yun\(^{28}\) have shed some light on the results (ref. Remark 3 in their paper) but lack a detailed development and theoretic proof, and our proof is derived independent of their work. Moreover, solution to the general case in next theorem extends this result.

\(^6\)Moreover, using the similar arguments, plus Lemma\(^{10}\) we are able to produce a nonconstructive proof to the well known results about singular value thresholding\(^{19}\) without any use of subgradient. We will not pursue this direction as it is out of the scope of this paper.
Hence we reach an equivalent formation of the original
objective function can be cast in its equivalent form
where we observe have that

\[ M^* = \text{arg min}_M \frac{1}{\mu} \| M \|_s + \frac{1}{2} \| M - \bar{P} \|_F^2, \quad \text{s.t.} \ M \succeq 0, \]  \tag{34}

with \( \bar{P} = (P + P^\top)/2 \). Solution to Eq. (34) readily follows
from Theorem 13.

The above two theorems (Theorem 13 and 15) have enabled a fast solution to updating \( J \). Moreover, since they ensure the symmetry of output \( J \) irrespective of the symmetry of \( G \), we can be assured the alternation optimization process converges to a solution of \( J \) that satisfies the constraint \( J \succeq 0 \).

III. COMPLEXITY ANALYSIS AND SCALABILITY

For solving the ALM problems corresponding to robust
LRR-PSD and robust LRR, the main computational cost
per iteration comes from either eigen-decomposition of
a symmetric matrix or SVD of a square matrix of the same
size. In numerical linear algebra, computing a stable
SVD of matrix \( X \in \mathbb{R}^{n \times n} \) is to convert it to a symmetric
eigen-decomposition problem on an augmented matrix

\[
\tilde{X} = \begin{bmatrix} 0 & X^\top \\ X & 0 \end{bmatrix}.
\]

Hence from SVD to eigen-decomposition of comparable
size, we can expect a constant factor of speedup that
depends on the matrix dimension. Figure 1 provides
benchmark results on computational times of SVD and eigen-
decomposition on matrices of sizes ranging from very small
up to 5000. Tested solvers include these provided in Matlab
and these in PROPACK. It is evident that for matrices
of the same size, eigen-decomposition is essentially faster in
both solver package. We will stick to the built-in functions
in Matlab as we find in practice PROPACK is sometimes
unstable when solving full problems (it is specialized in
solving large and sparse matrices).

IV. EXPERIMENTS

In this section, we systematically verify both the theoretic
analysis provided before, and related claims.

A. Experiment Setups

We use two data sets throughout our experiments.

Toy Data (TD). Following setting in [1], 5 independent subspaces
\( \{S_i\}_{i=1}^{5} \subset \mathbb{R}^{100} \) are constructed, whose bases \( \{U_i\}_{i=1}^{5} \) are generated by
\( U_{i+1} = TU_i, 1 \leq i \leq 4 \), where \( T \) represents a
random rotation and \( U_1 \) a random orthogonal matrix of dimension
100 × 4. So each subspace has a dimension of 4. 20 data vectors
are sampled from each subspace by \( X_i = U_iQ_i, 1 \leq i \leq 5 \) with
\( Q_i \) being a \( 4 \times 20 \) iid zero mean unit variance Gaussian matrix
\( \mathcal{N}(0, 1) \). Collection of this clean \( X \) should have rank 20.

Extended Yale B (EYB). Following setting in [1], 640 frontal face
images of 10 classes from the whole Yale B dataset are selected.
Each class contains about 64 images, and images are resized to
42 × 48. Raw pixel values are stacked into data vectors of dimension
2016 as features. This dataset is an example of heavily corrupted
data.

B. Equivalence of LRR-PSD and LRR

1) Spectrum Verification: Recall the key to establish the
equivalence of LRR-PSD and LRR lies with showing that the
eigenvalues and singular values of \( Z^* \) are identical, with 1 of
multiplicity equal to the data rank and the rest 0’s. In order
to verify this, we use TD without introducing any noise, and
hence the data matrix has rank 20. We simulate the clean
settings, i.e., LRR-PSD and LRR by gradually increasing the
regularization parameter \( \lambda \) of the robust versions [20] and
[21]. Intuitively for large enough \( \lambda \), the optimization tends to
put \( E = 0 \) and hence approaches the clean settings. Figure 2
presents the results along the regularization path (0.1 \( \sim \) 1).
It is evident during the passing to \( \lambda = 1 \), the eigenvalue
and singular value spectra match each other, and identically
produce 20 values of 1 and the rest all 0. This confirms
empirically the correctness of our theoretic analysis.

![Figure 1. Comparison of computation time for full SVD/eigen-decomposition. SVD and EIG are from Matlab built-in function (which essentially is wrapper for corresponding Lapack routines), and LANSVD, LANEIG from PROPACK.](image-url)
In pursuit of providing more insights into recent line of research work on sparse-reconstruction based affinity matrix learning for subspace segmentation, we have discovered an

2) Spectrum Perturbation Under Robust Setting: As we conjectured in Sec. II-C in most cases spectrum of the obtained affinity matrix from robust LRR-PSD or robust LRR will be perturbation of the ideal spectrum. Repeated experiments on many settings confirm about this, although we cannot offer a formal explanation to this yet. Here we only produce a visualization (figure 3) to show how things evolve under different noise level when we set $\lambda = 0.12$. The noise is added in sample-specific sense, as done in [1], i.e., some samples are chosen to be corrupted while others are kept clean. We do observe some breakdown cases when $\lambda$ is very small (not presented in the figure), which can be partially explained by that in that case the effect of nuclear norm regularization is weakened.

C. Selection of Noise Models

We have argued that the norm selection for the noise term $E$ should depend on the knowledge on noise patterns. We are going to compare the $\| \cdot \|_1$ noise model with the $\| \cdot \|_{2,1}$ noise model used in [1]. First we test on TD. Instead of adopting a sample-specific noise assumption, we assume that the corruptions are totally random w.r.t. data dimension and data sample which is more realistic. We add Gaussian noise with zero mean and variance $0.3\|X\|_F$, where $X$ is the whole data collection. Percentage of corruption is measured against the total number of entries in $X$. The evolution of SC performance against the percentage of corruption is presented in Figure 4. We can see the obvious better resistance against noises exhibited by the $\| \cdot \|_1$ form.

D. Speed Benchmark: LRR-PSD vs. LRR

We benchmark for the speed of robust LRR-PSD and robust LRR on EYB, and also present the clustering performance as compared to the conventional Gaussian kernel and linear kernel SC, which is obviously missing from [1]. Table II presents the accuracy obtained via various affinity matrices for SC, of which obviously LRR-PSD and LRR win out. To test the running time, we also include another set where each image in EYB is resized into $21 \times 24$ (Set 1). We denote the original setting Set 2, and use the first 20 classes of which each image resized into $42 \times 48$ to produce Set 3. We report running time (T), number of iterations (Iter), convergence tolerance (Tol) for each setting. Table III presents the results. Interestingly, LRR-DSP always converges with more iterations but less running time than that of LRR. The benefit of using eigen-decomposition in place of SVD is apparent.

V. SUMMARY

In pursuit of providing more insights into recent line of research work on sparse-reconstruction based affinity matrix learning for subspace segmentation, we have discovered an

Figure 2. Comparison of the eigen-spectrum (top) and singular value spectrum (bottom) for clean toy data (no artificial noises added) under the robust settings. Increasing the value of $\lambda$ in the robust settings, or effectively passing towards the clean formulation, the optimal $Z^*$ tends to produce 20 nonvanishing eigenvalues/singular values of 1. Left: by solving LRR. Right: by solving LRR-PSD. (Please refer to the color pdf and zoom in for better viewing effect.)

Figure 3. Evolution of the eigen-spectrum of the learnt affinity matrix under different noise levels. Left: solving by robust LRR; Right: solving by robust LRR-PSD. Surprisingly the spectra are always confined within $[0, 1]$ in this setting. (Please refer to the color pdf and zoom in for better viewing effect.)

Figure 4. Comparison of performance using different noise models. In essence we use $\|E\|_1$ and $\|E\|_{2,1}$ respectively in the objective. Instead of sample-specific noise, we assume random distributed noises, which is a more natural noise model. The $\ell_1$ version shows better resistance against noise.

<table>
<thead>
<tr>
<th></th>
<th>Gauss SC</th>
<th>Linear SC</th>
<th>SSC</th>
<th>LRR</th>
<th>LRR-PSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc.</td>
<td>24.84</td>
<td>30.16</td>
<td>37.66</td>
<td>59.53</td>
<td>60</td>
</tr>
</tbody>
</table>
important equivalence between the recently proposed LRR and our advocated version LRR-PSD in their canonical forms. This is a critical step towards understanding the behaviors of this family of algorithms. Moreover, we show that our advocated version, in its robust/denoising form, also facilitates a simple solution scheme that is as least as simple as the original optimization of LRR. Our experiments suggest in practice LRR-PSD is more likely to be flexible in solving large-scale problems.

ACKNOWLEDGEMENT

This work is partially supported by project grant NRF2007IDM-IDM002-069 on “Life Spaces” from the IDM Project Office, Media Development Authority of Singapore.

REFERENCES


