Article

Non-Convex Sparse and Low-Rank Based Robust Subspace Segmentation for Data Mining

Wenlong Cheng [1,2] (ID), **Mingbo Zhao [1,*], Naixue Xiong [3]** (ID) **and Kwok Tai Chui [2]**

[1] School of Information Science & Technology, Donghua University, Shanghai 200051, China;
    cheng.python@gmail.com
[2] Department of Electronic Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong, China;
    ktchui3-c@my.cityu.edu.hk
[3] School of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA;
    xiongnaixue@gmail.com
* Correspondence: mbzhao4@gmail.com; Tel.: +86-131-0684-8616

Abstract: Parsimony, including sparsity and low-rank, has shown great importance for data mining in social networks, particularly in tasks such as segmentation and recognition. Traditionally, such modeling approaches rely on an iterative algorithm that minimizes an objective function with convex l_1-norm or nuclear norm constraints. However, the obtained results by convex optimization are usually suboptimal to solutions of original sparse or low-rank problems. In this paper, a novel robust subspace segmentation algorithm has been proposed by integrating l_p-norm and Schatten p-norm constraints. Our so-obtained affinity graph can better capture local geometrical structure and the global information of the data. As a consequence, our algorithm is more generative, discriminative and robust. An efficient linearized alternating direction method is derived to realize our model. Extensive segmentation experiments are conducted on public datasets. The proposed algorithm is revealed to be more effective and robust compared to five existing algorithms.

Keywords: subspace segmentation; low-rank representation; non-convex; LADMAP

1. Introduction

High dimensionality research for data mining is an essential topic in modern imaging applications, such as social networks and the Internet of Things (IoT). It is worth noting that data of high dimension is often supposed to reside in several subspaces of lower dimension. For instance, facial images with various lightning conditions and expressions lie in a union of a nine-dimensional linear subspace [1]. Moreover, moving motions in videos [2] and hand-written digits [3] can also be approximated by multiple low-dimensional subspaces. Inspiringly, these characteristics enable effective segmentation, recognition, and classification to be carried out. The problem of subspace segmentation [4] is formulated as determining the number of subspaces and partitions the data according to the intrinsic structure.

Many subspace segmentation algorithms have emerged in the past decades. Some of these methods are algebraic or statistical. Among the algebraic methods, generalized principal component analysis (GPCA) [5] is mostly widely used. GPCA characterizes the data subspace with the gradient of a polynomial, and segmentation is obtained by fitting the data with polynomials. However, the performance drops quickly in the presences of noise, and polynomial fitting computation is time consuming. As to the statistical algorithms, including random sample consensus (RANSAC) [6], factorization-based methods [7,8], and probabilistic principal component analysis (PPCA) [9], the estimation of exact subspace models exquisitely changed the performance of this type of methods. Recently, spectral-type methods [10] like sparse representation (SR) [7,11,12], low-rank representation

(LRR) [13–15], and the extensions base on SR or LRR [16–22] have attracted much attention, as they are robust to noise [11–14], with strong theoretical backgrounds [16] and are easy to implement. Once an affinity graph is learned from SR or LRR, segmentation results can be obtained by means of spectral clustering. Therefore, building an affinity graph that accurately captures relevant data structures is a key point for SR- and LRR-based spectral-type models.

A good affinity graph should preserve the local geometrical structure, as well as the global information [11]. Recently, Elhamifar et al. [23] presented the sparse subspace clustering (SSC) based on an l_1-graph. The l_1-graph is constructed by using SR coefficients deduced in the l_1-norm minimization to represent the relationships among samples. Attributed to SR, the l_1-graph is sparse, capable of finding neighborhood according to the data and robust to noisy data. Inspired by l_1-graph, various SR-graphs have been proposed [24]. Wang et al. [25] provided l_2-Graph based subspace clustering to eliminate errors from various types of projection spaces. Peng et al. [26] proposed a unified framework for representation-based subspace clustering methods to cluster both the out-of-sample and the large-scale data. Peng et al. [27] introduced principal coefficients embedding to automatically identify the number of features, as well as learn the underlying subspace in the presence of Gaussian noise. Wang et al. [28] proposed a race lasso-based regularizer for multi-view data while keeping individual views well encapsulated. However, SR-based methods seek the sparsest representation coefficients of each data point individually, without a global structure regularizer. This drawback alters the robustness of these methods in the presence of outliers [15], when the data is not "clean" enough. To account for the underlying global information, Liu et al. [15] presented a two-step algorithm which firstly computes low-rank representation coefficients for data and then uses the coefficient matrix to build an affinity graph (LRR-graph). The LRR-graph jointly represents all the data by solving a nuclear norm optimization problem and, thus, becomes a better option at capturing global information. Numerous LRR-graph-based methods have also been proposed for spectral clustering. For example, Wang et al. [28] proposed a multi-graph Laplacian-regularized LRR to characterize the non-linear spectral graph structure from each view. Both SSC and LRR are based on convex relaxations of the initial problems, SSC uses l_1-norm to approximate the number of non-zero elements, while LRR applies nuclear norm to compute the number of non-vanishing singular values. These convex relaxations yield solutions that deviate far from solutions to the original problems. Hence, it is desirable to rather use non-convex surrogates, without causing any significant increase in computational complexity.

A large number of studies on non-convex surrogates for l_0-norm problem have been addressed recently. Xu et al. [29] introduced the $l_{1/2}$-norm in noisy signal recovery with an efficient iterative half-thresholding algorithm. Similarly, Zhang et al. [30] proposed the l_p-norm minimization with a generalized iterative shrinkage and thresholding method (GIST). Their study shows l_p-norm based model is more effective on image denoising and image deburring. Numerous researchers also considered the other non-convex surrogate functions, such as homotopic L_0-minimization [31], smoothly clipped absolute deviation (SCAD) [32], multi-stage convex relaxation [17], logarithm [33], half-quadratic regularization (HQR) [34], exponential-type penalty (ETP) [35], and minimax concave penalty (MCP) [36]. Recent years have also witnessed the progress in non-convex rank minimization functions. Mohan et al. [37] developed an efficient IRLS-p to minimize rank function, and improved the recovery performance for matrix completion. Xu et al. [18] introduced the $S_{1/2}$-norm with an efficient ADMM solver for video background modeling. Kong et al. [38] proposed a Schatten p-norm constrained model to recover noisy data. Another popular non-convex rank minimization is the truncated nuclear norm [39]. They all compete with the state-of-the-art algorithms to some extent.

Combining the non-convex l_p-norm regularizer with the Schatten p-norm ($0 < p \leq 1$), in this study, we propose a robust method named non-convex sparse and low-rank-based robust subspace segmentation (l_pSpSS). Our l_p-norm error function can better predict errors, which further improves the robustness of subspace segmentation. Meanwhile, the Schatten p-norm-regularized objective function shows a better ability to approximate the rank of coefficient matrix compared with the nuclear norm. Our method can provide a more accurate description of the global information and

better measurement of data redundancy. Thus, our new objective is to solve joint the l_p-norm and Schatten p-norm ($0 < p \le 1$) minimization together. When $p \rightarrow 0$, our proposed l_pSpSS turns to be more robust and effective than SR- and LRR-based subspace segmentation algorithms. In addition, we enforce non-negative constraint to the reconstruction coefficients, which aids interpretability and allows better solutions in numerous application areas such as text mining, computer vision, and bioinformatics. Traditionally, an alternating direction method (ADM) [40] can solve this optimization problem efficiently. However, to increase the speed and scalability of the algorithm, we choose an efficient solver commonly named the linearized alternating direction method with adaptive penalty (LADMAP) [19]. As it is based on fewer auxiliary parameters and without an inverse of its matrix, it is more efficient than ADM. Numerical experimental results verify our proposed method, which consistently obtains better segmentation results.

The rest of this paper is structured as follows: In Section 2, the notations, as well as the overview of SSC and LRR, will be presented. Section 3 is dedicated to introducing our novel non-convex sparse and low-rank based robust subspace segmentation. Section 4 conducts multiple numerical experiments to examine the effectiveness and robustness of l_pSpSS. Section 5 concludes this work.

2. Background

This section is divided into three parts. First, the notation and definition are illustrated. The background of two algorithms, SSC and LRR, will be discussed in Sections 2.2 and 2.3, respectively.

2.1. Notations and Definitions ∈

Suppose $X = [x_1, \cdots, x_N]$ is an image matrix consists of N sufficiently dense data points $\{x_i \in R^m\}_{i=1}^N$, $\{S_i\}_{i=1}^n$ is an arrangement of n subspaces. Let $\{x_i\}$ be drawn from n subspaces $\{S_1 \cup S_2 \cdots \cup S_n\}$ of lower dimension. Given X, the goal of subspace segmentation is to partition the data points into the underlying low-dimensional subspaces.

The l_p-norm ($0 < p < \infty$) of vector $x \in R^{n \times 1}$ can be expressed as $\|x\|_p = \left(\sum_1^n \|x_i\|^p \right)^{1/p}$, in which x_i is the i-th element. Therefore, the p-norm of $x \in R^{n \times 1}$ to the power p can be expressed as $\|x\|_p^p = \sum_1^n \|x_i\|^p$.

The Schatten p-norm of a matrix $x \in R^{n \times m}$ is expressed as:

$$\|X\|_{S_p} = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^p \right)^{1/p} \tag{1}$$

in which $0 < p \le 1$, and σ_i is the i-th largest singular value. Thus, it can be deduced that:

$$\|X\|_{S_p}^p = \left(\sum_{i=1}^{\min(m,n)} \sigma_i^p \right) \tag{2}$$

The Schatten 1-norm is just nuclear norm $|X|_*$, while the Schatten 0-norm is the approximation of the rank of X. Compared with $|X|_*$, $|X|_{S_p}$ is a better approximation of the rank of X.

2.2. Sparse Subspace Clustering

Recently, SSC [16] has grabbed considerable attention. The hypothesis states that data are drawn from several subspaces of lower dimension, and can be sparsely self-expressed. More formally, SSC aims to solve the following program:

$$\min_Z \sum_i \|Z_i\|_0 \quad s.t. X = XZ \quad and \quad diag(Z) = 0 \tag{3}$$

where $Z_i = [z_{i1}, z_{i2}, \cdots, z_{iN}]^T \in R^N$ are the reconstruction coefficients, and $\|z\|_0$ refers to the number of nonzero values. As it is difficult to solve this non-convex objective, a convex l_1 minimization problem is proposed by solving the following program:

$$\min_Z \|Z\|_1 s.t. X = XZ \quad and \quad diag(Z) = 0 \tag{4}$$

The minimization problem in Equation (4) can be concluded using the alternating direction method of multipliers (ADMM) [19]. Afterwards, the coefficient matrix Z can be utilized to construct the affinity matrix as $W = |Z| + |Z^T|$. Finally, W is performed via spectral clustering and the segmentation result is drawn. While SSC works well in practice, the model is invalid when the obtained similarity graph is poorly connected (we refer readers to Soltanolkotabi et al. [41] for very recent results in this direction).

2.3. Low-Rank Representation-Based Subspace Segmentation

The difference between LRR and SSC is that, LRR seeks the lowest rank representation Z but not the sparsest representation. LRR is based on the assumption that for observed data $X \in R^{d \times N}$ drawn from n low-dimensional subspaces, and the rank of coefficient matrix $r = rank(Z) = \sum_{i=1}^n z_i$ is assumed to be much smaller than $\min\{d, N\}$. The LRR is formulated as:

$$\min_z rank(Z) \quad s.t. X = XZ \tag{5}$$

As the rank function minimization is non-convex, Equation (5) can be reformulated as the following convex minimization problem:

$$\min_z \|Z\|_* \quad s.t. X = XZ \tag{6}$$

in which $\|Z\|_*$ is the nuclear norm, which yields a good approximation to the matrix rank of Z. Singular value threshold (SVD) can be used to efficiently solve Equation (6), when there is no error present in X.

When the data X is noisy, an extension of LRR is proposed as follows:

$$\min_{Z,E} \|Z\|_* + \lambda E_p \quad s.t. X = XZ + E \tag{7}$$

in which $\lambda \geq 0$ is the tradeoff parameter, trading off low rankness between reconstruction error. $E_p \in R^{d \times N}$ is the noise term with different regularization strategies, which depends on the property of E_p. When the noise term is Gaussian noise, $\|E\|_p = \|E\|_F^2$, in which $\| \cdot \|_F$ refers to the Frobenius norm. When the noise term are entry-wise corruptions, $\|E\|_p = \|E\|_1$, in which $\| \cdot \|_1$ refers to the l_1 norm. When the noise term are sample-specific corruption and outliers, $\|E\|_p = \|E\|_{2,1}$, in which $\|E\|_{2,1} = \sum_{i=1}^N \sqrt{\sum_{i=1}^d E_{ij}^2} = \sum_{i=1}^N \|E_j\|_2$ refers to the $l_{2,1}$ norm. Equation (7) can be solved by ADMM [19] to obtain the coefficient matrix Z. Afterwards, the coefficient matrix Z can be utilized for the construction of affinity matrix $W = |Z| + |Z^T|$. Finally, spectral clustering can be applied to W for segmentation results.

3. Non-Convex Sparse and Low-Rank Based Robust Subspace Segmentation

In this section, we first propose the non-convex sparse and low-rank-ased robust subspace segmentation model, in which we combine the l_p-norm with the Schatten p-norm together for clustering, and then use LADMAP to solve $l_p S_p SS$. Finally, we analyze of the time complexity of $l_p SpSS$.

3.1. Model of $l_p SpSS$

We consider the non-convex sparse and low-rank-based subspace segmentation for data contaminated by noise and corruption. Notice that the nuclear norm is replaced by the Schatten

p-norm, when p is smaller than 1, the underlying global information can be captured more effectively. Additionally, the l_p norm ($0 < p \leq 1$) of the coefficient matrix is also introduced as an error function, in order to harvest stronger robustness to noise [42]. It has been demonstrated in some recent research [43] that the Schatten p-norm is more powerful than the nuclear norm in matrix completion, and the recovery performance of the l_p-norm is also superior to the convex l_1-norm [36]. Our l_pSpSS will surely be more effective than the convex methods.

We begin by considering the relaxed low-rank subspace segmentation problem, which is equivalent to:

$$\min_{Z,E} \|Z\|_{S_p}^p + \beta \|Z\|_p^p + \lambda \|E\|_{2,1}$$
$$s.t.\, X = XZ + E, Z \geq 0 \tag{8}$$

In which, the first term $\|Z\|_p^p$ is the l_p-norm, which improves the integration of the local geometrical structure. Meanwhile, the second term $\|Z\|_{S_p}^p$ is the Schatten p-norm, which can better approximate the rank of Z. Moreover, the third term reconstruction error $\|E\|_{2,1}$ is the $l_{2,1}$ norm, which can better characterize errors like corruption and outliers. β and λ are trade-off parameters. Regarding the widely-used non-negative constraint ($Z \geq 0$), which is to ensure direct use of the reconstruction coefficients in the affinity graph construction.

3.2. Solution to l_pSpSS

3.2.1. Brief Description of LADMAP

We adopt LADMAP [19] to solve the objective function (Equation (8)) constrained by the l_p-norm norm and Schatten-p regularizers. An auxiliary variable W is introduced and the optimization problem becomes separable. Thus, Equation (8) is rewritten as:

$$\min_{Z,E} \|Z\|_{S_p}^p + \beta \|Z\|_p^p + \lambda \|E\|_{2,1}$$
$$s.t.\, X = XZ + E, Z = W, W \geq 0 \tag{9}$$

To remove two linear constraints in Equation (9), we introduce two Lagrange multipliers Y_1 and Y_2, hence, the optimization problem is defined using the following Lagrangian function:

$$
\begin{aligned}
L(Z,W,E,Y_1,Y_2,u) &= \|Z\|_{S_p}^p + \beta \|W\|_p^p + \lambda \|E\|_{2,1} + \langle Y_1, X - XZ - E \rangle + \langle Y_2, Z - W \rangle + \tfrac{\mu}{2}(\|X - XZ - E\|_F^2) + (\|Z - W\|_F^2) \\
&= \|Z\|_{S_p}^p + \beta \|W\|_p^p + \lambda \|E\|_{2,1} + q(Z,W,E,Y_1,Y_2,\mu) - \tfrac{1}{2\mu}(\|Y_1\|_F^2 + \|Y_2\|_F^2)
\end{aligned} \tag{10}
$$

where $q(Z,W,E,Y_1,Y_2,\mu) = \tfrac{\mu}{2}\left(\|X - XZ - E + Y_1/\mu\|_F^2 + \|Z - W + Y_2/\mu\|_F^2 \right)$, Y_1 and Y_2 are the Lagrange multipliers, and $\mu \geq 0$ is a trade-off parameter. We solve Equation (10) by minimizing L to update each variable with the other variables fixed. The updating schemes at each iteration can be designed as follows:

$$W(k+1) = \arg\min_{W \geq 0} \beta \|W\|_p^p + \frac{\mu(k)}{2} \|Z(k+1) - W + Y_2(k)/\mu(k)\|_F^2 \tag{11}$$

$$
\begin{aligned}
Z(k+1) &= \arg\min_Z \|Z\|_{S_p}^p + \langle \nabla_Z q(Z(k), W(k), E(k), Y_1(k), Y_2(k), \mu(k), Z - Z(k) \rangle + \tfrac{\mu(k)\theta}{2} \|Z - Z(k)\|_F^2 \\
&= \arg\min_Z \|Z\|_{S_p}^p + \tfrac{\theta\mu(k)}{2} \Big| Z - Z(k) + [-X^T(X - XZ(k) - E(k) + \tfrac{Y_1(k)}{\mu(k)}) + (Z(k) - W(k) + \tfrac{Y_2(k)}{\mu(k)})]/\theta \Big|_F^2
\end{aligned} \tag{12}
$$

$$E(k+1) = \arg\min_E \lambda \|E\|_{2,1} + \tfrac{u(k)}{2} \|X - XZ(k+1) + \tfrac{Y_1(k)}{u(k)} - E\|_F^2 = \Omega_{\lambda u(k)^{-1}}(X - XZ(k+1) + \tfrac{Y_1(k)}{\mu(k)}) \tag{13}$$

In Equation (12), $\nabla_Z q$ is the partial differential of q with respect to Z, $\theta = \|X\|_F^2$. In particular, the detailed procedures of LADMAP are shown in Algorithm 1. The first, Equation (11), and the second, Equation (12), are solved using the following subsections. The last convex problem (Equation (13)) can solved by the $l_{2,1}$-norm minimization operator [15].

Algorithm 1. LADMAP for solving Equation (9).

Input: Data X, tradeoff variables $\beta > 0$, $\lambda > 0$
Initialize: $Z(0) = W(0) = E(0)$, $Y_1(0) = Y_1(0) = 0$, $\mu(0) = 0.1$, $\mu_{max} = 10^7$, $\rho_0 = 1.1$, $\varepsilon_1 = 10^{-7}$, $\varepsilon_2 = 10^{-6}$, $k = 0$,
$\theta = \|X\|_F^2$, and the number of maximum iteration *MaxIter* = 1000.
While not converged and $k \leq$ *MaxIter* **do**

(1) Compute $W(k+1)$ by solving Equation (11) with $Z(k+1)$, $E(k+1)$ are fixed;
(2) Compute $Z(k+1)$ by solving Equation (12) with $W(k+1)$, $E(k+1)$ are fixed;
(3) Compute $E(k+1)$ by solving Equation (13) with $W(k+1)$, $Z(k+1)$ are fixed;
(4) Compute the multipliers $Y_1(k+1)$ and $Y_2(k+1)$ as follows:

$$\begin{cases} Y_1(k+1) = Y_1(k) + \mu(k)(X - XZ(k+1) - E(k+1)) \\ \quad Y_2(k+1) = Y_2(k) + \mu(k)(Z(k+1) - W(k+1)) \end{cases}$$

(5) Compute $\mu(k+1)$ as follows:

$$\mu(k+1) = \min(\mu_{\max}, \rho\mu(k)), \text{ in which}$$

$$\rho = \begin{cases} \rho_0, if \mu(k)\Omega/\|X\|_F \leq \varepsilon_2 \\ 1, otherwise \end{cases}$$

(6) Converges when

$$\begin{cases} \|X - XZ(k) - E(k)\|_F/\|X\|_F \leq \varepsilon_1 or \\ \quad \mu(k)\Omega/\|X\|_F \leq \varepsilon_2 \end{cases}$$

In which $\Omega = \max(\sqrt{\theta}\|Z(k) - Z(k+1)\|_F, \|W(k) - W(k+1)\|_F, \|E(k) - E(k+1)\|_F)$
(7) Increment of iteration: $k = k + 1$.

End while
Output: The optimal solution W, Z and E.

3.2.2. Solving the Non-Convex l_p-Norm Minimization Subproblem (Equation (11))

For each element in $\{X_{ij} | (i, j) \in \Omega\}$, we can decouple Equation (11) into a simplified formula:

$$\min \frac{1}{2}(x - y)^2 + \lambda|x|^p \tag{14}$$

Recently, Zhang et al. solved this l_p-norm optimization problem via the proposed GIST [30]. For l_p-norm minimization, the thresholding function is:

$$\tau_p(\lambda) = (2\lambda(1 - p))^{1/(2-p)} + \lambda p(2\lambda(1 - p))^{(p-1)/(2-p)} \tag{15}$$

Meanwhile, the generalized soft-thresholding function is:

$$T_p(y; \lambda) = \begin{cases} 0, & if |y| \leq \tau_p(\lambda) \\ \text{sgn}(y)S_p(|y|; \lambda), & if |y| > \tau_p(\lambda) \end{cases} \tag{16}$$

The corresponding thresholding rule in the generalized soft-thresholding function is $T_p^{GST}(y; \lambda) = 0$ when $|y| \leq \tau_p^{GST}(\lambda)$, and corresponding shrinkage rule is $T_p^{GST}(y; \lambda) = \text{sgn}(y)S_p^{GST}(|y|; \lambda)$ when $|y| > \tau_p^{GST}(\lambda)$.

3.2.3. Solving the Non-Convex Schatten p-Norm Minimization Subproblem (Equation (12))

We can reformulate Equation (12) as the following simplified notation:

$$\min_X \frac{1}{2}\|X - A\|_F^2 + \lambda X_{s_p}^p \tag{17}$$

After applying SVD on *X*, *X* is decomposed into summation of *r* rank-*p* matrices $X = U\Delta V^T$. Here, *U* is the left singular vector, Δ is the non-zero singular diagonal matrix, and *V* is the right singular vector. The *i*-th singular value δ_i is solved by:

$$\min_{\delta_i \geq 0} \frac{1}{2}(\delta_i - \sigma_i)^2 + \lambda \delta_i^p \tag{18}$$

Equation (16) can be used to solve Equation (18) again. For *p* = 1, we can obtain the same solution with nuclear norm minimization [44].

3.3. Convergence and Computational Complexity Analysis

Although Algorithm 1 is described in three major alternative steps for solving *W*, *Z*, and *E*, we can actually combine steps for *Z* and *E* easily into one larger block step by simultaneously solving for (*Z*, *E*). Thus, the convergence conclusion of two variables LADMAP in [45] can be applied to our case. Finally, the convergence of the algorithm is ensured.

Suppose the size of *X* are $d \times n$, *k* is number of total iterations, and *r* is the lowest rank for *X*. The major time consumption of Algorithm 1 is mainly determined by Step 2, as it involves time-consuming SVDs. In Step 1, each component of $\nabla_{Z_k} q$ can be computed in $O(rn^2)$ by using the skinny SVD to update *W*. In Step 2, the complexity of the SVD to update *Z* is approximately $O(d^2n)$. In Step 3, the computation complexity of $l_{2,1}$ minimization operator is about $O(dn)$. The total complexity is, thus, $O(krn^2 + kd^2n)$. Since $r \leq min(d, n)$, the time cost is, at most, $O(krn^2)$.

3.4. Affinity Graph for Subspace Segmentation

Once Equation (9) was solved by LADMAP, we can obtain the optimal coefficient matrix Z^*. Since every sample is reconstructed by its neighbors, *Z** naturally characterizes the relationships among samples. Such information is a good indicator of similarity among samples, we use the reconstruction coefficients to build the affinity graph. The non-convex l_p-norm ensures that each sample only connects to few samples. As a result, the weights of the graph tend to be sparse. While the non-convex Schatten *p*-norm ensures samples lying in the same subspace are highly correlated and tend to be assigned into the same cluster, *Z** is theoretically able to capture the global information, and the graph weights are constrained with non-negativity, as they reflect similarities between data points.

After obtaining the coefficient matrix Z^*, the reconstruction coefficients of each sample are normalized and thresholded to zero. Therefore, the obtained normalized sparse $\hat{Z}^*$ can be used to compute the affinity graph $W = (\hat{Z}^* + (\hat{Z}^*)^T)/2$. Finally, *W* carries out spectral clustering to obtain the segmentation results. Our proposed non-convex sparse and low-rank based subspace segmentation is outlined in Algorithm 2.

Algorithm 2. $l_p SpSS$.

Input: Matrix $X = [x_1, x_2, \cdots, x_n] \in R^{d \times n}$, tradeoff parameters β, λ

1. Normalize each sample x_i to obtain $\hat{X} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n]$.

2. Solve the non-convex sparse and low-rank constrained program by Algorithm 1:

   $$\min \|Z\|_{S_p}^p + \beta \|Z\|_p^p + \lambda \|E\|_2^p,$$
   $$s.t. X = XZ + E, Z \geq 0$$

   and get (*Z**, *E**).

3. Normalize coefficient matrix Z^*, and threshold small values by θ to obtain $\hat{Z}^*$.

4. Set weights for affinity graph by $W = (\hat{Z}^* + ((\hat{Z}^*)^T)/2$.

5. The data is segmented by spectral clustering.

Output: The segmentation results.

---

4. Experimental Evaluation and Discussion

In the following, we will discuss the performance of our proposed l_pSpSS model. Firstly, the experimental setting is detailed in Section 4.1. From Section 4.2, Section 4.3, Section 4.4, we will test the segmentation performance of l_pSpSS on CMU-PIE, COIL20, and USPS. In Section 4.5, we will examine the robustness of l_pSpSS to block occlusions and pixel corruptions. Finally, the discussion of experimental results will be given in Section 4.6.

4.1. Experimental Settings

The proposed l_pSpSS approach will be evaluated on realistic images and compared with five related works. We use four publicly-available datasets, including CMU-PIE [46], COIL20 [47], USPS [48], and Extended Yale B [49]. Among them, datasets [46] and [49] contain face images with various poses/illuminations/facial expressions, COIL20 consists of different general objects, and USPS includes handwritten digit images. Our proposed l_pSpSS will be compared with five segmentation methods, including PCA, SSC [16], LRR [15], and NNLRS [50], while K-means serves as a baseline for comparison.

We adopt the same experimental settings as Zhuang's work [50]. For the compared methods, a grid search strategy is used for selecting model parameters, and the optimal segmentation is achieved by tuning the parameters carefully. As to our l_pSpSS, there are two regularized parameters, β and λ, affecting its performance. We take a stepwise selection strategy to search the best parameters. For example, we search the possible candidate interval λ may exit, with β fixed, and alternatively search $\lambda's$ most possible candidate interval, with β fixed. Finally, the best values are found in a two-dimensional candidate space of (β, λ).

To quantitatively and effectively measure the segmentation performance, two quantity metrics, namely accuracy (AC) and normalized mutual information (NMI) [51], are used in our experiments. All the experiments are implemented by MATLAB, on a MacBook Pro with a 2.6 GHz Intel Core i7 CPU and 16 GB memory.

4.2. Segmentation Results on CMU-PIE Database

In this experiment, we compare l_pSpSS with the other five methods on the CMU-PIE facial images dataset. It includes 41,368 pictures of 68 persons, acquired with various postures and lighting scenarios. The resolution of each image is $32 \times 32 = 1024$ pixels. Typical examples of CMU-PIE are shown in Figure 1. For each given cluster number $K = 4,...,68$ in the whole dataset, the segmentation results with different K were averaged on the twenty tests. The averaged segmentation performance of proposed and existing algorithms on the CMU-PIE dataset [46] are reported in Table 1.

Figure 1. Typical examples of the CMU-PIE dataset.

Table 1. Segmentation results of proposed and existing algorithms on the CMU-PIE dataset.

Clusters K	AC (%)						NMI (%)					
	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS
4	48.5	52.4	100	100	100	100	64.1	66.8	100	100	100	100
12	41.9	47.3	81.5	89.5	79	91.1	63.4	62.9	84.6	95.5	96.7	96.8
20	38.8	36.7	80.6	81.3	88.3	92.2	62.3	58.8	85.9	90.8	94.3	98.5
28	35.7	34.9	78.2	77.4	87.9	91.9	61.7	61.1	86.4	89.9	94	96.9
36	34.3	34.7	77.1	68.8	78.7	87.6	60.5	60.6	85.7	82.3	93.6	96.5
44	33.8	33.7	75.1	71.7	81.3	84.2	59.1	62.6	84.7	84.2	93.8	95.7
52	33.1	33.7	69.9	71.1	75.4	88.2	58.1	61.6	85.3	84.9	93.1	96
60	33	33.2	68.1	65.6	79.6	84.7	52.7	53.6	84.9	80.2	93.5	95.4
68	31	32.8	66.7	65.1	86.2	88.1	46.8	46.7	85.5	79.3	87.9	96.6
Average	36.7	37.7	74.7	73.8	84.1	89.8	58.7	59.4	85.4	87.4	93.4	96.9

We can see that our proposed l_pSpSS achieves the best segmentation AC and NMI on CMU-PIE dataset, which proves the effectiveness of our l_pSpSS. For example, the average segmentation accuracy of NNLRS and l_pSpSS are 84.1% and 89.8%, respectively. l_pSpSS improves the segmentation accuracy by 5.7% compared with NNLRS (the second best algorithm). The improvement of l_pSpSS indicates the importance of the non-convex SR and LRR affinity graph.

4.3. Segmentation Results on COIL20 Database

When it comes to the evaluation using second dataset COIL20 [49], the proposed l_pSpSS is compared with five existing algorithms. This dataset contains 1440 images of 20 objects, with 72 different views. The resolution of each picture is 32 × 32 = 1024 pixels. Typical examples of COIL20 are shown in Figure 2. For each given cluster number K = 2,...,20 in the whole dataset, the segmentation results with different K were averaged on the twenty tests. The averaged segmentation performances of the proposed and existing algorithms on the COIL20 dataset [47] are reported in Table 2.

Figure 2. Typical examples of the COIL20 dataset.

Table 2. Segmentation results of proposed and existing algorithms on the COIL20 dataset.

Clusters K	AC (%)						NMI (%)					
	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS
2	88.3	88.1	96.2	90.3	98.3	98.7	80.4	81.4	88.7	90.8	92.2	92.6
4	84.7	84	81.5	88.8	96.1	98	77.5	78.7	84.6	87.3	87.7	91.2
6	74.5	83.4	80.6	83.9	94.6	95.1	72.3	73.8	85.9	86.2	89.1	90.6
8	73.8	71.1	78.2	77	86.9	92.5	75.3	74.8	86.4	86.8	90.1	88.5
10	71.2	69.4	77.1	74.9	87.2	89.1	74.1	74.8	85.7	85.9	89.5	91.9
12	68.8	68.5	75.1	70.3	86.8	87.9	75.4	75.6	84.7	86.8	90	89.5
14	65.2	66.3	69.9	66.5	84.9	86.6	74.1	75	85.3	85.6	88.6	92.1
16	66.4	67.3	68.1	67	85.5	87.7	74.8	74.6	84.9	84.1	88.9	91.7
18	63.5	65.8	67.1	65.8	83.1	84.6	74.9	74.7	84.4	84.6	88.8	91.1
20	62.8	64.3	66.7	64	78.8	81.7	75.8	74.1	85.5	86.4	90.1	91
Average	71.9	72.8	76.1	74.9	88.2	90.2	75.5	75.8	85.6	86.5	89.5	91

Experimental results on COIL20 indicates that our l_pSpSS outperforms the other five existing algorithms. For example, the average AC and NMI of l_pSpSS are 90.2% and 91.0%, which are higher than for NNLRS (the second best algorithm) by 2.0% and 1.5%, respectively. Especially, when the cluster number is large, the superiority of l_pSpSS is very obvious.

4.4. Segmentation Results on USPS Handwritten Digit Dataset

In this experiment, we compare l_pSpSS with the other five methods on the USPS handwritten digit dataset. It contains 9298 images of 10 classes, with a variety of orientations. The resolution of each picture is $16 \times 16 = 256$ pixels. Figure 3 shows typical sample images in the USPS dataset. For each given cluster number $K = 2,...,10$, the segmentation results with different K were averaged on the twenty tests. The averaged segmentation performances of proposed and existing algorithms on the USPS handwritten digit dataset are reported in Table 3.

Figure 3. Typical examples of USPS dataset.

Table 3 shows that our proposed l_pSpSS still obtains the best segmentation performance. This result demonstrates that a non-convex sparse and low-rank graph is better to model complex related data than traditional SR and LRR based graphs. Experimental results have demonstrated that our proposed l_pSpSS model can not only represent the global information, but also preserves the local geometrical structures in the data by incorporating the non-convex l_p-norm regularizer.

Table 3. Segmentation results of proposed and existing algorithms USPS handwritten digit dataset.

Clusters K	AC (%)						NMI (%)					
	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS
2	94.1	94.3	94.2	94.6	96.6	98.3	71.9	72.2	81.4	73.8	79.3	82.9
3	88.1	88.8	89.3	89.3	94.7	96.4	71.1	71.4	79.8	75.9	80.3	83.4
4	82.2	79.2	83.3	84	90.1	91.2	67.1	68.2	79.4	72.3	77.3	80.3
5	79.1	78.2	79.1	80.8	87.3	88.3	65	66.7	77.9	70.6	79.2	81.1
6	77.4	74.3	75.2	75.1	88.3	90.2	65.1	66.3	76.3	73.6	75.7	79.1
7	74.8	73.3	74.2	75.6	82.7	84.3	62.7	63.2	81.4	69.8	74.5	75
8	71.5	71.8	74.3	76.3	80.6	82.4	61.3	63.4	79.8	68.9	73.2	75.1
9	68.7	69.2	75.3	75	79.4	80.6	59.9	60.2	79.4	67.3	72.7	74.7
10	65.4	63.3	74.2	74.3	75.1	77.4	59.4	60.7	77.3	66.6	71.6	72.6
Average	77.9	76.9	79.9	80.6	86.1	87.7	64.8	65.9	79.2	71	76	78.2

4.5. Segmentation Results on Dataset with Block Occlusions and Pixel Corruptions

Finally, we evaluate the robustness of each model on the more challenging Extended Yale B face dataset. It has 38×64 facial images, with various lighting scenarios. To reduce the resources and

budget, the resolution of each picture is downsized to 96×84. This dataset more challenging for subspace segmentation, as 50% of the samples with hard shadows or specularities. Figure 4 shows typical sample images of Extended Yale B.
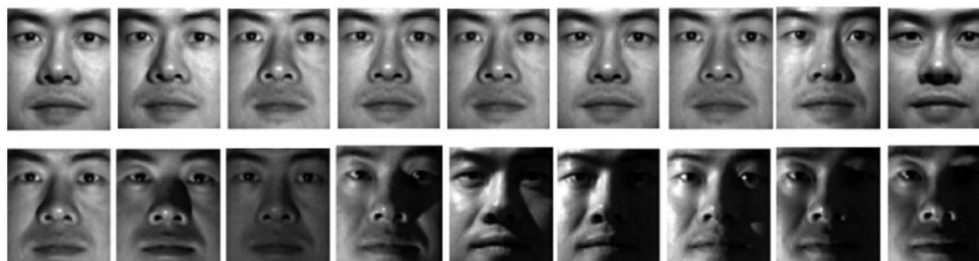

Figure 4. Typical sample images of Extended Yale B dataset.

We select 1134 images from first 18 individuals to evaluate the different methods. Two types of corruptions are introduced into this experiment. For Type I block occlusions, different block sizes (from 5×5 to 20×20) are added to randomly selected locations of the images. For Type II random pixel corruption, randomly-chosen pixels on each images are substituted with equally-distributed random values. The proportion of corrupted pixels per image is from 0 to 20%. Some examples from corrupted Extended Yale B face images are shown in Figure 5. The averaged segmentation AC of proposed and existing algorithms with multiple block occlusions and pixel corruptions are tabulated in Table 4 as well as Table 5.

Figure 5. Typical sample images of the corrupted dataset.

Table 4. Segmentation results of proposed and existing algorithms on Extended Yale B dataset with multiple block occlusions.

Block Size	AC (%)						NMI (%)					
	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS
5×5	13.6	15.1	78.5	88.3	89.5	90.6	14.2	16.1	79.0	90.4	91.5	92.8
10×10	11.7	13.4	75.4	86.7	87.0	88.7	12.6	14.4	77.3	88.9	89.2	91.2
15×15	9.8	11.3	72.7	84.5	85.1	86.8	10.8	12.8	75.4	86.3	87.4	88.6
20×20	7.5	9.6	70.2	82.1	82.8	84.6	8.5	10.6	72.7	84.4	85.8	86.2

Table 5. Segmentation results of proposed and existing methods on Extended Yale B with pixel corruptions.

Corruption Rate	AC (%)						NMI (%)					
	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS	K-Means	PCA	SSC	LRR	NNLRS	l_pSpSS
0.05	11.2	13.1	68.8	83.9	86.9	88.3	14.2	17.1	74.3	89.9	90.8	92.3
0.1	7.6	9.4	64.5	78.3	81.4	83.6	11.6	13.4	68.3	87.4	87.6	89.1
0.15	5.8	8.8	62.4	72.2	76.2	79.2	7.8	9.8	65.4	82.3	83.9	84.5
0.2	3.5	5.6	60.7	68.4	72.8	75.2	4.5	6.6	62.7	77.4	78.8	79.2

Both experimental results show that our l_pSpSS achieves the best segmentation results again. Segmentation results suggest that proposed l_pSpSS is more robust than compared methods, especially when a significant portion of the realistic samples are corrupted.

4.6. Discussions

Our l_pSpSS outperforms five existing subspace segmentation algorithms. Especially, in the case of CMU-PIE face dataset, the improvement by l_pSpSS is largest. Our affinity graph can capture the local geometrical structure, as well as the global information of the data, hence, is both generative and discriminative.

l_pSpSS is more robust than the other compared methods, which can properly deal with multiple noises. Images in Extended Yale B dataset contain different errors, including block occlusions, pixel corruptions, illuminations, partition them is challenging. However, l_p-norm and Schatten p-norm are introduced for l_pSpSS affinity graph construction. Therefore, our model can better predict errors and is a better measurement for data redundancy.

The segmentation performance of SSC and LRR are almost the same. For example, the segmentation accuracy of SSC on CMU-PIE is 0.9% better than LRR, while the performance of LRR on USPS is 0.7% higher than SSC. The segmentation results heavily depends on the intrinsic structure of the testing dataset, it is difficult to determine which one is better.

The LRR-based algorithm is robust in handling noisy data. It aims at obtaining the low rankness of coefficient matrix, thus, the LRR-based methods can better model the global information. Furthermore, LRR can find similar clusters which measure the data redundancy, ensuring high quality and stability of the segmentation results. For data heavily contaminated with corruptions or outliers, the model can find lower ranks that will be more robust to noise.

However, SVD computation for Schatten p-norm minimization is performed in each iteration, which is very time consuming, and the best segmentation results are not achieved at the lowest value of p. Hence, we will be focused on the study of speeding up the Schatten p-norm solver and the selection of best p values in our future work.

5. Conclusions

This paper presents an accurate and robust for subspace segmentation, named l_pSpSS, by introducing the non-convex l_p-norm and Schatten p-norm minimization. Taking advantages from the original sparsity and low rankness of data of high dimension, both local geometrical structure and the global information of the data can be learnt. A linearized alternating direction method with adaptive penalty (LADMAP) is also introduced to search for optimal solutions. Numerous experiments on CMU-PIE, COIL20, USPS, and Extended Yale B verify the effectiveness and robustness of our l_pSpSS compared to five existing works.

Author Contributions: This paper is prepared by all authors. W.C., M.Z., N.X., and K.T.C. proposed the idea. W.C. designed the algorithm, performed the numerical experiments, and analyzed the experimental results. W.C., M.Z., N.X., and K.T.C. wrote and revised this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shi, X.; Yang, Y.; Guo, Z.; Lai, Z. Face recognition by sparse discriminant analysis via joint $l_{2,1}$-norm minimization. *Pattern Recogn.* **2014**, *47*, 2447–2453. [CrossRef]
2. Costeira, J.P.; Kanade, T. A multibody factorization method for independently moving objects. *Int. J. Comput. Vis.* **1998**, *29*, 159–179. [CrossRef]
3. Hastie, T.; Simard, P.Y. Metrics and models for handwritten character recognition. *Stat. Sci.* **1998**, *13*, 54–65. [CrossRef]
4. Parsons, L.; Haque, E.; Liu, H. Subspace clustering for high dimensional data: A review. *SIGKDD Explor.* **2004**, *6*, 90–105. [CrossRef]

5. Vidal, R.; Ma, Y.; Sastry, S. Generalized principal component analysis (GPCA). *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1945–1959. [CrossRef] [PubMed]

6. Fischler, M.A.; Bolles, R.C. Random sample consensus-A paradigm for model-fitting with applications to image-analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

7. Lu, Y.; Lai, Z.; Xu, Y.; Li, X.; Zhang, D.; Yuan, C. Non-negative Discriminant Matrix Factorization. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *45*, 4080–4091. [CrossRef]

8. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. In Proceedings of the International Conference on Intelligent Systems for Molecular Biology, Menlo Park, CA, USA, 14–17 August 1994; pp. 28–36.

9. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1999**, *61*, 611–622. [CrossRef]

10. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. In Proceedings of the Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–8 December 2006; pp. 849–856.

11. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [CrossRef] [PubMed]

12. Wright, J.; Ma, Y.; Mairal, J.; Sapiro, G.; Huang, T.S.; Yan, S.C. Sparse representation for computer vision and pattern recognition. *Proc. IEEE* **2010**, *98*, 1031–1044. [CrossRef]

13. Fang, X.; Xu, Y.; Li, X.; Lai, Z.; Wong, W.K. Robust semi-supervised subspace clustering via non-negative low-rank representation. *IEEE Trans. Cybern.* **2016**, *46*, 1828–1838. [CrossRef] [PubMed]

14. Peng, Y.; Ganesh, A.; Wright, J.; Xu, W.; Ma, Y. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2233–2246. [CrossRef] [PubMed]

15. Liu, G.; Lin, Z.; Yu, Y. Robust subspace segmentation by low-rank representation. In Proceedings of the International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 663–670.

16. Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781. [CrossRef] [PubMed]

17. Zhang, T. Analysis of multi-stage convex relaxation for sparse regularization. *J. Am. Stat. Assoc.* **2010**, *11*, 1081–1107. [CrossRef]

18. Rao, G.; Peng, Y.; Xu, Z. Robust sparse and low-rank matrix decomposition based on $S_{1/2}$ modeling. *Sci. China Inform. Sci.* **2013**, *43*, 733–748. [CrossRef]

19. Lin, Z.; Liu, R.; Su, Z. Linearized alternating direction method with adaptive penalty for low-rank representation. In Proceedings of the Annual Conference on Neural Information Processing Systems, Granada, Spain, 12–17 December 2011; pp. 612–620.

20. Zhuang, L.; Gao, H.; Lin, Z.; Ma, Y.; Zhang, X.; Yu, N. Non-negative low-rank and sparse graph for semi-supervised learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2328–2335.

21. Lu, C.; Tang, J.; Yan, S.; Lin, Z. Generalized non-convex non-smooth low-rank minimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4130–4137.

22. Peng, X.; Yu, Z.; Yi, Z.; Tang, H. Constructing the l_2-graph for robust subspace learning and subspace clustering. *IEEE Trans. Cybern.* **2017**, *47*, 1053–1066. [CrossRef] [PubMed]

23. Elhamifar, E.; Vidal, R. Sparse subspace clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2790–2797.

24. Tang, J.; Hong, R.; Yan, S.; Chua, T.-S.; Qi, G.-J.; Jain, R. Image annotation by *k*NN-sparse graph-based label propagation over noisily tagged web images. *IEEE Trans. Intell. Trans. Syst.* **2011**, *2*, 14. [CrossRef]

25. Wang, Y.; Zhang, W.; Wu, L.; Lin, X.; Fang, M.; Pan, S. In Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering. In Proceedings of the International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2153–2159.

26. Peng, X.; Tang, H.; Zhang, L.; Yi, Z.; Xiao, S. A unified framework for representation-based subspace clustering of out-of-sample and large-scale data. *IEEE Trans. Neural Netw. Learn Syst.* **2016**, *27*, 2499–2512. [CrossRef] [PubMed]

27. Peng, X.; Lu, J.; Yi, Z.; Yan, R. Automatic subspace learning via principal coefficients embedding. *IEEE Trans. Cybern.* **2016**, *PP*, 1–14. [CrossRef] [PubMed]

28. Wang, Y.; Lin, X.; Wu, L.; Zhang, W.; Zhang, Q.; Huang, X. Robust subspace clustering for multi-view data by exploiting correlation consensus. *IEEE Trans. Image Process.* **2015**, *24*, 3939–3949. [CrossRef] [PubMed]

29. Xu, Z.B.; Chang, X.Y.; Xu, F.M.; Zhang, H. $l_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw.* **2012**, *23*, 1013–1027. [CrossRef]

30. Zuo, W.M.; Meng, D.Y.; Zhang, L.; Feng, X.C.; Zhang, D. A generalized iterated shrinkage algorithm for non-convex sparse coding. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 217–224.

31. Trzasko, J.; Manduca, A. Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization. *IEEE Trans. Med. Image* **2009**, *28*, 106–121. [CrossRef] [PubMed]

32. Fan, J.Q.; Li, R.Z. Variable selection via non-concave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [CrossRef]

33. Friedman, J.H. Fast sparse regression and classification. *Int. J. Forecast.* **2012**, *28*, 722–738. [CrossRef]

34. Geman, D.; Yang, C. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *4*, 932–946. [CrossRef] [PubMed]

35. Gao, C.; Wang, N.; Yu, Q.; Zhang, Z. A feasible nonconvex relaxation approach to feature selection. In Proceedings of the Conference on Artificial Intelligence, San Francisco, CA, USA, 7–11 August 2011; pp. 356–361.

36. Zhang, C.H. Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **2010**, *38*, 894–942. [CrossRef]

37. Mohan, K.; Fazel, M. Iterative reweighted algorithms for matrix rank minimization. *J. Mach. Learn. Res.* **2012**, *13*, 3441–3473. [CrossRef]

38. Kong, D.; Zhang, M.; Ding, C. Minimal shrinkage for noisy data recovery using Schatten *p*-norm objective. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Prague, Czech Republic, 23–27 September 2013; pp. 177–193.

39. Hu, Y.; Zhang, D.; Ye, J.; Li, X.; He, X. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2117–2130. [CrossRef] [PubMed]

40. Shen, Y.; Wen, Z.; Zhang, Y. Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optim. Methods Softw.* **2014**, *29*, 239–263. [CrossRef]

41. Soltanolkotabi, M.; Candes, E.J. A geometric analysis of subspace clustering with outliers. *Ann. Stat.* **2012**, *40*, 2195–2238. [CrossRef]

42. Zhou, W.; Li, P.; Wang, X.; Li, F.; Liu, H.; Zhang, R.; Ma, T.; Liu, T.; Guo, D.; Yao, D. l_p norm spectral regression for feature extraction in outlier conditions. In Proceedings of the IEEE Conference on Digital Signal Processing, Singapore, 21–24 July 2015; pp. 535–538.

43. Nie, F.; Wang, H.; Cai, X.; Huang, H.; Ding, C. Robust matrix completion via joint Schatten *p*-norm and l_p-norm minimization. In Proceedings of the IEEE International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 566–574.

44. Cai, J.F.; Candes, E.J.; Shen, Z.W. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [CrossRef]

45. Favaro, P.; Vidal, R.; Ravichandran, A. A closed form solution to robust subspace estimation and clustering. In Proceedings of the International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1801–1807.

46. Baker, S.; Sim, T.; Bsat, M. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1615–1618. [CrossRef]

47. Geusebroek, J.-M.; Burghouts, G.J.; Smeulders, A.W. The Amsterdam library of object images. *Int. J. Comput. Vis.* **2005**, *61*, 103–112. [CrossRef]

48. Hull, J.J. A database for handwritten text recognition research. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 550–554. [CrossRef]

49. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]

50. Zhuang, L.; Gao, S.; Tang, J.; Wang, J.; Lin, Z.; Ma, Y.; Yu, N. Constructing a non-negative low-rank and sparse graph with data-adaptive features. *IEEE Trans. Image Process.* **2015**, *24*, 3717–3728. [CrossRef] [PubMed]
51. Cai, D.; He, X.F.; Han, J.W. Document clustering using locality preserving indexing. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 1624–1637. [CrossRef]