Brief Papers

Revisiting $L_{2,1}$-Norm Robustness With Vector Outlier Regularization

Bo Jiang$^\circledast$ and Chris Ding

Abstract—In many real-world applications, data usually contain outliers. One popular approach is to use the $L_{2,1}$-norm function as a robust loss/error function. However, the robustness of the $L_{2,1}$-norm function is not well understood so far. In this brief, we propose a new vector outlier regularization (VOR) framework to understand and analyze the robustness of the $L_{2,1}$-norm function. Our VOR function defines a data point to be the outlier if it is outside a threshold with respect to a theoretical prediction, and regularizes it, i.e., pull it back to the threshold line. Thus, in the VOR function, how far an outlier lies away from its theoretical predicted value does not affect the final regularization and analysis results. One important aspect of the VOR function is that it has an equivalent continuous formulation, based on which we can prove that the $L_{2,1}$-norm function is the limiting case of the proposed VOR function. Based on this theoretical result, we thus provide a new and intuitive explanation for the robustness property of the $L_{2,1}$-norm function. As an example, we use the VOR function to matrix factorization and propose a VOR principal component analysis (PCA) (VORPCA). We show some benefits of VORPCA on data reconstruction and clustering tasks.

Index Terms—$L_{2,1}$-norm, outlier regularization, principal component analysis (PCA).

I. INTRODUCTION

Real-world image data sets often contain noises and errors. Traditionally, this is often handled using principal component analysis (PCA), linear discriminant analysis (LDA), and many other dimension-reduction methods [1]–[4]. Among the dimension-reduction methods, PCA is one of most widely used linear algorithms [2]. For Gaussian-type noises, these methods are very effective. However, sometimes the noises are large, such as outliers, corrupted, occluded images, different illuminations, and shading conditions. For these kinds of noises or outliers, PCA-type dimension-reduction methods usually break down. To address this problem, some robust dimension-reduction or subspace-extraction methods are developed [5]–[12]. One popular way is to use some matrix norms such as $L_{1}$-norm [13] and $L_{2,1}$-norm [14]–[17] to develop robust formulations. $L_{1}$-norm-based approaches [10], [13], [18]–[21] usually demonstrate the good effects for recovering true signals from large corruption, while $L_{2,1}$-norm approaches [14]–[17], [22]–[24] often show the robustness property in dealing with the outliers. In addition, $L_{2,1}$-norm defines the distance by using the $L_{2,1}$-norm function. In many real applications, data usually have Gaussian distribution, and thus, the $L_{2}$-norm function is more suitable for data distance/metric measurement used in many learning applications, such as data reconstruction [14], [25], regression [16], and graph matching [17].

However, the robustness property of the $L_{1}$ and $L_{2,1}$-norm functions is not well understood so far. In this brief, we focus on the $L_{2,1}$-norm function and aim to introduce a novel vector outlier regularization (VOR) function to understand and analyze the robustness of the $L_{2,1}$-norm function. Our VOR function defines a data point to be the outlier if it is outside a threshold with respect to a theoretical prediction and regularizes it, i.e., pull it back to the threshold line. Although the VOR function is a discrete function, it has an equivalent continuous representation. Based on this continuous representation, we can prove that $L_{2,1}$-norm function is the limiting case of the proposed VOR function. This provides a new intuitive way to analyze the robustness property of the $L_{2,1}$-norm function. As an example, we use the VOR function to matrix factorization and propose a novel VORPCA. VORPCA can be regarded as a balanced model between the standard PCA and $R_{1}$-PCA [14], [15] and degenerated to $R_{1}$-PCA at the small tolerance limit. Using VORPCA, we provide a new intuitive analysis and interpretation for the robustness of the $L_{2,1}$-norm-based $R_{1}$-PCA model. From data reconstruction, VORPCA can preserve small high-rank components, and thus retain fine details in data reconstruction, which is beneficial to data learning problems, such as classification, clustering, and so on. One limitation of VORPCA is nonconvex, and only the local optimal solution can be obtained.

II. WHAT IS THE $L_{2,1}$-NORM ROBUSTNESS

In this section, we introduce the robustness property of the $L_{2,1}$-norm function. We show it from the matrix factorization problem. Formally, let $X = (x_{1}, \ldots, x_{n}) \in \mathbb{R}^{p \times n}$ be the observed $n$ data points in the feature vector space. Let $U \in \mathbb{R}^{p \times k}$, $k < p$ and $V = (v_{1}, \ldots, v_{n}) \in \mathbb{R}^{k \times n}$, and we consider two types of low-rank matrix factorization problems [2], [14]

$$\min_{U, V} \quad E_{2}(U, V) = \sum_{i=1}^{n} \|x_{i} - Uv_{i}\|_{2}^{2} = \|X - UV\|_{F}^{2}$$

$$\min_{U, V} \quad E_{21}(U, V) = \sum_{i=1}^{n} \|x_{i} - Uv_{i}\|_{2,1}$$

where $\cdot_{2,1}$ is the $L_{2,1}$-norm of the vector. The traditional intuitive understanding of $L_{2,1}$ robustness is follows.

Suppose $x_{i}$ is an outlier, then the residual $r_{i} = \|x_{i} - Uv_{i}\|_{2,1}$ is larger than the residuals of other vector data points. In $E_{2}(U, V)$, due to the squaring, $r_{i}^{2}$ would be much larger than other squared residuals, and thus easily dominates the objective function. In $E_{21}(U, V)$, the error for each data point is $r_{i} = \|x_{i} - Uv_{i}\|_{2,1}$, which is not squared, and thus diminishes the undue influence of those outliers and thus makes the learning more robust or stable.

However, this understanding of $L_{2,1}$ robustness is questionable. Although the large errors due to outliers are not squared in $L_{2,1}$, they are still large, and thus, one would expect they would still significantly influence the cost function and, therefore, the final results. In other words, the $L_{2,1}$ function is insensitive with respect to the outlyingness of the outliers: as long as a data point $x_{i}$ is an outlier, how far away $x_{i}$ lies does not affect the final results.
The VOR function is due to a process of VOR. The L1 in the two data sets (left- and right-hand-side panels) are different. In fact, the outliers do not seem to influence the errors due to the outliers (in the 2-D toy data). Two images of each person are selected and corrupted to generate the outlier images. More details are given in Section VI. Note that 1-norm robustness on the AT&T face data. In each, the errors due to the outliers (in the 2-D toy data) are much larger than those from nonoutliers. Fig. 2 shows the reconstruction results on the AT&T face data. Two images of each person are selected and corrupted to generate the outlier images. More details are given in Section VI. Note that 2,1 results perform more robustly than the L2 results when the outliers exist.

In the remaining of this brief, we show that this 2,1 robustness property is due to a process of VOR. The 2,1 function minimization is the limiting case of VOR. In the following, we first propose the general formulation of the VOR function, followed by a continuous representation of the VOR function. Then, we apply VOR in the matrix factorization form, based on which we provide a new insight and explanation on 2,1-norm robustness.

### III. VOR Function

In this section, we first propose the general formulation of our VOR function, followed by a continuous representation of the VOR function.

#### A. Continuous Representation

The VOR function [see (3)] is discrete. Interestingly, it also has an equivalent continuous formulation. Formally, we have Proposition 1.

Let \( X = (x_1 \ldots x_n) \in \mathbb{R}^{p \times n} \) be the observed input data in the feature vector space. Let \( F = (f_1 \ldots f_n) \in \mathbb{R}^{p \times n} \) be the corresponding theoretical model prediction. We define a vector data point \( x_i \) to be significantly distorted or highly corrupted if the difference \( \|x_i - f_i\| \) between the observed measured data \( x_i \) and the theoretical prediction \( f_i \) is larger than a tolerance limit \( \delta \). We wish to correct these highly corrupted data points. One intuitive and effective way is to move them toward the prediction manifold, but keep them at the boundary (tolerance limit). We achieve this purpose by defining the following function:

\[
\tilde{x}_i = \begin{cases} 
  x_i & \text{if } \|x_i - f_i\| \leq \delta \\
  f_i + \delta \frac{x_i - f_i}{\|x_i - f_i\|} & \text{if } \|x_i - f_i\| > \delta
\end{cases}
\]

where \( \delta > 0 \) is a threshold (or tolerance) parameter and \( \|v\| \) is the Euclidean norm of vector \( v \). We call it VOR function. Fig. 3 shows an illustration of the function of (3). Red points \( x_1 \cdots x_n \) are the input data. Their theoretical predictions \( f_1 \cdots f_n \) are on the plane (the subspace). Top and bottom planes indicate the threshold planes of tolerant limits: data points \( x_i \) outside the threshold plane are considered outliers; they are pulled back to the threshold plane \( x_i \) by “outlier regularization.” Data points within the tolerant limits remain unchanged.

Fig. 3. Illustration of the function of (3). Red points \( x_1 \cdots x_n \) are the input data. Their theoretical predictions \( f_1 \cdots f_n \) are on the plane (the subspace). Top and bottom planes indicate the threshold planes of tolerant limits: data points \( x_i \) outside the threshold plane are considered outliers; they are pulled back to the threshold plane \( x_i \) by “outlier regularization.” Data points within the tolerant limits remain unchanged.

Let \( X = (x_1 \ldots x_n) \in \mathbb{R}^{p \times n} \) be the observed input data in the feature vector space. Let \( F = (f_1 \ldots f_n) \in \mathbb{R}^{p \times n} \) be the corresponding theoretical model prediction. We define a vector data point \( x_i \) to be significantly distorted or highly corrupted if the difference \( \|x_i - f_i\| \) between the observed measured data \( x_i \) and the theoretical prediction \( f_i \) is larger than a tolerance limit \( \delta \). We wish to correct these highly corrupted data points. One intuitive and effective way is to move them toward the prediction manifold, but keep them at the boundary (tolerance limit). We achieve this purpose by defining the following function:

\[
\tilde{x}_i = \begin{cases} 
  x_i & \text{if } \|x_i - f_i\| \leq \delta \\
  f_i + \delta \frac{x_i - f_i}{\|x_i - f_i\|} & \text{if } \|x_i - f_i\| > \delta
\end{cases}
\]
Proposition 1: \( \tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \in \mathbb{R}^{P \times n} \) of (3) is the optimal solution to the following problem, that is:

\[
\tilde{X} = \arg \min_Z \|X - Z\|_{2,1} + \frac{1}{2\delta} \|Z - F\|_F^2.
\] (4)

Proof: Due to the definition of \( L_{2,1} \)-norm, problem (4) can be decoupled into \( n \) separate independent subproblems

\[
\tilde{x}_i = \arg \min_{x_i} \|x_i - z_i\| + \frac{1}{2\delta} \|z_i - f_i\|^2.
\] (5)

We now prove that the solution of (5) is given by (3). Setting \( u = z_i - x_i \), (5) can be written as

\[
\min_{u} \delta \|u\| + \frac{1}{2} \|u - (f_i - x_i)\|^2.
\] (6)

In the Appendix, we prove that the solution of (6) is given by

\[
u^* = \max \left(1 - \delta \frac{\left\| f_i - x_i \right\|}{\left\| f_i - x_i \right\|}, 0 \right) (f_i - x_i).
\] (7)

Thus, for (5), \( \tilde{x}_i = z_i^* = u^* + x_i \). If \( \|f_i - x_i\| \leq \delta \), \( u^* = 0 \); thus, \( z_i^* = x_i \), which is the same as the definition of (3).

If \( \|f_i - x_i\| > \delta \), we have

\[
u^* = \left(1 - \frac{\delta}{\left\| f_i - x_i \right\|}\right) (f_i - x_i) = f_i - x_i - \delta \frac{x_i - f_i}{\|f_i - x_i\|}.
\]

Thus \( z_i^* = u^* + x_i = f_i + \delta \frac{x_i - f_i}{\|f_i - x_i\|} \) which is the same as (3). This completes the proof. \( \square \)

IV. VOR IN MATRIX FACTORIZATION

Here, we apply the VOR function to matrix factorization and propose our VORPCA. Let \( X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{P \times n} \) be the observed input data. Let \( F = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^{P \times n} \) be the corresponding theoretical prediction. Here, we set the theoretical prediction model to be rank- \( k \) approximation of the same as PCA, i.e., \( F = UV \), where \( U \in \mathbb{R}^{P \times k}, V = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^{k \times n} \) and \( \text{rank}(UV) \leq k \).

Using the continuous representation [see (4)], our VORPCA is formulated as

\[
\min_{Z, U, V} \|X - Z\|_{2,1} + \frac{1}{2\delta} \|Z - UV\|_F^2.
\] (8)

Illustration: Let \( \tilde{X} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \) be the optimal solution \( Z \) of problem (8). In Fig. 4, we show the results of VORPCA on a simple 2-D data set. The original data \( \{x_i\} \) are shown as black dots. Reconstructed data \( \{\tilde{x}_i\} \) are shown as red circles (nonoutliers) and blue squares (outliers). Red line indicates the prediction correct subspace (standard PCA on the reconstructed data \( \{\tilde{x}_i\} \)), while green lines show the boundary (tolerance limit). Outliers are brought back to the correct subspace by VORPCA at several \( \delta \) values but kept at the boundary (tolerance limit) at convergence.

V. REVISITING \( R_1 \)-PCA ROBUSTNESS WITH VORPCA MODEL

In the VORPCA model [see (8)], when the tolerance \( \delta \to 0 \), the second reconstruction term is weighted with an infinite weight. Thus, \( Z = UV \) and the VORPCA problem becomes \( L_{2,1} \)-norm-based PCA (\( R_1 \)-PCA) [14], [15]

\[
\min_{U, V} \|X - UV\|_{2,1}.
\] (9)

On the other hand, when \( \delta \to \infty \), the first term is weighted with an infinite weight. Thus, \( X = Z \) and the VORPCA problem becomes standard PCA [2]

\[
\min_{U, V} \|X - UV\|_F^2.
\] (10)

Formally, we have Proposition 2.

Proposition 2: When \( \delta \to 0 \), VORPCA becomes \( R_1 \)-PCA. When \( \delta \to \infty \), VORPCA becomes PCA.

Proof: In order to prove it, we use the quadratic-penalty function [26] method. Given an equality-constrained problem

\[
\min_x f(x) \text{ s.t. } c_j(x) = 0.
\] (11)

The quadratic penalty function \( G(x; \delta) \) is defined as

\[
G(x; \delta) = f(x) + \frac{1}{2\delta} \sum_i c_i^2(x)
\] (12)

where \( \delta > 0 \) is the penalty parameter. As discussed in [26], when \( \delta \to 0 \), the constraint violation is penalized with increasing severity, and the optimal solution of the penalty function \( G(x; \delta) \) is identical to the constrained problem [see (11)].

First, we show that VORPCA [see (8)] is a quadratic-penalty function for the equality-constrained problem

\[
\min_{Z, U, V} \|X - Z\|_{2,1} \text{ s.t. } Z = UV.
\] (13)

As discussed earlier, when \( \delta \to 0 \), the constraint violation is penalized enough, and the optimal solution of VORPCA is identical to that of the constrained problem [see (13)], which is the same as \( R_1 \)-PCA [see (9)].

Second, we can also rewrite VORPCA as

\[
\min_{Z, U, V} 2\delta \|X - Z\|_{2,1} + \|Z - UV\|_F^2.
\] (14)

Similarly, it is also a quadratic-penalty function for the equality-constrained problem as

\[
\min_{Z, U, V} \|Z - UV\|_F^2 \text{ s.t. } X = Z.
\] (15)
When \( \delta \to \infty \), the constraint is penalized with increasing severity, and the optimal solution of VORPCA is identical to (15), which is the same as the standard PCA.

From Proposition 2, we can note that VORPCA can be regarded as a kind of balanced model between the PCA and the \( R_1 \)-PCA. As demonstrated in Fig. 5, when \( \delta \) is large enough, the reconstruction results of VORPCA are almost identical to the standard PCA (L2-norm). When \( \delta \) is small enough, the reconstruction results of VORPCA are almost identical to \( R_1 \)-PCA (L21PCA). That is, \( R_1 \)-PCA is the small tolerance limit of VORPCA and PCA is the large tolerance limit of VORPCA.

In Section II, we have shown that \( R_1 \)-PCA performs robustly to the outliers. The fact that \( R_1 \)-PCA is the small tolerance limit of VORPCA offers some explanations on the robustness property of the \( L_{2,1} \)-norm function. That is, at small \( \delta \), most data points become outliers and are regularized using the VOR function, i.e., pulled toward the theoretical prediction. Obviously, true outliers do not affect the final results, and furthermore, the outlyingness of the true outliers does not matter either. This provides a new kind of explanation that how robustness is performed in the \( R_1 \)-PCA model, i.e., the outliers are corrected based on the VOR function.

VI. ALGORITHM AND APPLICATIONS

A. Computational Algorithm

The above VORPCA problem can be efficiently computed by using the following algorithm.

Step 0: Initialize \( U, V \) by solving PCA on \( X \).

Repeat Step 1 and Step 2 until convergence.

Step 1: Fixing \( U, V \), we compute the optimal \( Z \) by solving (4).

Step 2: Fixing \( Z \), the optimizations for \( U \) and \( V \) are obtained by solving PCA on \( Z \).

The convergence of the algorithm is guaranteed, because each update has a closed-form solution, which decreases the objective function in each iteration. In practical, the convergence condition is that the maximum element changes of \( U \) and \( V \) between two consecutive iterations are less than \( \epsilon \) or the maximum number of iterations reaches maximum iteration.

B. Image Reconstruction

One main advantage of VORPCA is that the reconstructed data \( Z \) do not shrink the magnitude of the data. To show this, we introduce a baseline method, trace-norm-based \( L_{2,1} \)-PCA (we call it TrL21PCA) [24] as

\[
\min_{Z} \| X - Z \|_{2,1} + \beta \| Z \|_{tr}.
\]  

(16)

Here, \( \beta \) is a positive weighting parameter. The trace norm (also called nuclear norm) \( \| Z \|_{tr} \) is the sum of singular values of \( Z \). The trace norm is a surrogate of rank (\( Z \)), with the purpose to achieve low-rank [9], [27]. One advantage of trace norm is that it is convex.

To help illustrate the main points, we run both VORPCA and TrL21PCA on the occluded images from the AT&T face data set. Fig. 6 shows the results. Here, we observe that: 1) both VORPCA and TrL21PCA reconstruction are robust with respect to large occlusion errors and 2) finer details of individual images are mostly suppressed in TrL21PCA, but are partially retained in VORPCA. Fig. 7 shows the singular values of computed \( Z \). Here, one can see that the singular values of the TrL21PCA-reconstructed data are downshifted (evenly suppressed) for all terms. From \( k = 40 \) and up, all singular values are zero. In contrast, the singular values of VORPCA-reconstructed data remain close to the original data for ranks \( k = 1 \) to 40, but reduce significantly beyond these ranks. They remain nonzero for all higher ranks. Small but nonzero higher ranks help retain certain fine details in the reconstructed images. A concise measure of noise removal can be defined as follows. Let \( \hat{X}_0 \) be the original nonoccluded images representing the true signals. Let \( E \) be the occlusion, i.e., \( X = X_0 + E \) is the input data. Let \( Z \) be computed from the TrL21PCA and VORPCA models. Then, we define the noise-free residual as \( \| Z - \hat{X}_0 \|_F \). Fig. 8 shows the residual results. We can see that VORPCA can usually return a lower residual value than TrL21PCA. This is consistent with Fig. 3.

Fig. 6. Reconstruction results from TrL21PCA and VORPCA on the AT&T face data. In each panel, for images of one person, the top line indicates the original occluded images, middle line indicates the reconstruction from TrL21PCA, and bottom line indicates the reconstruction from VORPCA. Finer details of individual images are suppressed in TrL21PCA, but partially retained in VORPCA.

Fig. 7. Singular values of solution \( Z \) from TrL21PCA and VORPCA on the AT&T face data. Left: entire scale. Right: in small vertical scale such that small singular values are more clear. The presence of small high-rank components (with nonzero singular values) in VORPCA helps retain fine details in VORPCA reconstruction.

1) In VORPCA, the data rank is only suppressed on higher rank terms. The important lower ranks \( 1 \leq k \leq K \) are not suppressed, but instead protected. This can be seen from Fig. 7 (\( K = 40 \)), where lower rank singular values remain nearly identical to the singular values of the input data.

2) In VORPCA, higher rank components do not appear directly in cost function. They are suppressed, but not completely eliminated, as shown in Fig. 7. Small but nonzero higher
its robustness. We provide the theoretical analysis and continuous formulation of VORPCA to demonstrate the robustness of the $R_1$-PCA model.

**APPENDIX**

We prove that the solution of (6) is given by (7). For simplification, we ignore subscript $i$ in (6) and write it as

$$\min_{\mathbf{u}} \delta \| \mathbf{u} \| + \frac{1}{2} \| \mathbf{u} - \mathbf{a} \|^2. \quad (17)$$

*Proof:* It is clear that, given the magnitude of the vector $\mathbf{u}$, the direction of $\mathbf{u}$ must be in the same direction of the vector $\mathbf{a}$ in order to minimize the second term. Thus, the direction of $\mathbf{u}$ must be in the same direction of the vector $\mathbf{a}$, i.e., we must have $\mathbf{u} = \rho \mathbf{a}$, where $\rho \geq 0$ is a scalar.

Substituting to (17), we need to minimize

$$f(\rho) = \delta \| \mathbf{a} \| \rho + (1/2) \| \mathbf{a} \|^2 (\rho - 1)^2$$

subject to $\rho \geq 0$. The derivative of $f(\rho)$ with respect to variable $\rho$ is

$$f'(\rho) = \delta \| \mathbf{a} \| + (\rho - 1) \| \mathbf{a} \|^2.$$

The optimal $\rho$ satisfies the Karush–Kuhn–Tucker (KKT) complementarity slackness condition [26], that is

$$0 = pf'(\rho) = \rho(\delta \| \mathbf{a} \| + (\rho - 1) \| \mathbf{a} \|^2),$$

which is $\rho = \delta \| \mathbf{a} \| + (\rho - 1) \| \mathbf{a} \|^2 = 0$. The solution is $\rho^* = \max(1 - \delta / \| \mathbf{a} \|, 0)$. This gives $\mathbf{u}^* = \rho^* \mathbf{a}$. Replacing $\mathbf{a} = \mathbf{f}_i - \mathbf{x}_i$ gives (7).

**REFERENCES**

[1] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[2] R. O. Duda, P. E. Hart, and D. G. Stork, Eds., *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.

[3] X. He, D. Cai, S. Yan, and H.-J. Zhang, “Neighborhood preserving embedding,” in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1208–1213.

[4] S. T. Roweis, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[5] F. De la Torre and M. J. Black, “A framework for robust subspace learning,” *Int. J. Comput. Vis.*, vol. 54, no. 1, pp. 117–142, Aug. 2003.

[6] H. Aanaes, R. Fisker, K. Astrom, and J. M. Carstensen, “Robust factorization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1215–1225, Sep. 2002.

[7] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.

[8] S. Ma, D. Goldfarb, and L. Chen, “Fixed point and Bregman iterative methods for matrix rank minimization,” *Math. Program.*, vol. 128, nos. 1–2, pp. 321–353, Jun. 2011.

[9] M. Fazel, “Matrix rank minimization with applications,” Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 2002.

[10] C. Ding and B. Jiang, “L1-norm error function robustness and outlier regularization,” *IEEE Comput. Sci. Conf. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2005, pp. 739–746.

[11] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas, “Robust PCA via alternating iteratively reweighted low-rank matrix factorization,” in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3383–3387.

[12] Z. Lin, C. Xu, and H. Zha, “Robust matrix factorization by majorization minimization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 208–220, Jan. 2018.
[14] C. Ding, D. Zhou, X. He, and H. Zha, “R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization,” in Proc. 23rd Int. Conf. Mach. Learn., 2006, pp. 281–288.
[15] N. Kwak, “Principal component analysis based on L1-norm maximization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 9, pp. 1672–1680, Sep. 2008.
[16] C.-X. Ren, D.-Q. Dai, and H. Yan, “Robust classification using f2,1-norm based regression model,” Pattern Recognit., vol. 45, no. 7, pp. 2708–2718, Jul. 2012.
[17] Y.-F. Yu, G. Xu, M. Jiang, H. Zhu, D.-Q. Dai, and H. Yan, “Joint transformation learning via the L2,1-norm metric for robust graph matching,” IEEE Trans. Cybern., to be published.
[18] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. A. Pados, “Efficient L1-norm principal-component analysis via bit flipping,” IEEE Trans. Signal Process., vol. 65, no. 16, pp. 4252–4264, Aug. 2017.
[19] Y. Liu and D. A. Pados, “Conformity evaluation of data samples by L1-norm principal-component analysis,” Proc. SPIE. Congress Sens. VII, Diverse Modalities Big Data Anal., vol. 10658, May 2018, Art. no. 1065809.
[20] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, “Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization,” in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 2080–2088.
[21] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, “Sparse and low-rank matrix decompositions,” IFAC Proc. Volumes, vol. 42, no. 10, pp. 1493–1498, 2009.
[22] G. Liu, Z. Lin, and Y. Yu, “Robust subspace segmentation by low-rank representation,” in Proc. ICML, 2010, pp. 663–670.
[23] P. Favaro, R. Vidal, and A. Ravichandran, “A closed form solution to robust subspace estimation and clustering,” in Proc. CVPR, Jun. 2011, pp. 1801–1807.
[24] H. Xu, C. Caramanis, and S. Sanghavi, “Robust PCA via outlier pursuit,” IEEE Trans. Inf. Theory, vol. 58, no. 5, pp. 3047–3064, May 2012.
[25] B. Jiang, C. Ding, B. Luo, and J. Tang, “Graph-Laplacian PCA: Closed-form solution and robustness,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 3492–3498.
[26] J. Nocedal and S. Wright, Numerical Optimization. Berlin, Germany: Springer, 2006.
[27] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” SIAM Rev., vol. 52, no. 3, pp. 471–501, Jan. 2010.
[28] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia object image library (coil-20),” Dept. Comput. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-006-96, 1996.
[29] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 888–905, Aug. 2000.