# fLRR: Fast Low-Rank Representation Using Frobenius Norm

Haixian Zhang, Zhang Yi, and Xi Peng

Low Rank Representation (LRR) intends to find the representation with lowest-rank of a given data set, which can be formulated as a rank minimization problem. Since the rank operator is non-convex and discontinuous, most of the recent works use the nuclear norm as a convex relaxation. This letter theoretically shows that under some conditions, Frobenius-norm-based optimization problem has an unique solution that is also a solution of the original LRR optimization problem. In other words, it is feasible to apply Frobenius-norm as a surrogate of the nonconvex matrix rank function. This replacement will largely reduce the time-costs for obtaining the lowest-rank solution. Experimental results show that our method (i.e., fast Low Rank Representation, fLRR), performs well in terms of accuracy and computation speed in image clustering and motion segmentation compared with nuclear-norm-based LRR algorithm.

*Introduction:* Given a data set $X \in \mathcal{R}^{m\times n} (m < n)$ composed of column vectors, let $A$ be a data set composed of vectors with the same dimension as those in $X$. Both $X$ and $A$ can be considered as matrices. A linear representation of $X$ with respect to $A$ is a matrix $Z$ that satisfies the equation $X = AZ$. The data set $A$ is called a dictionary. In general, this linear matrix equation will have infinite solutions, and any solution can be considered to be a representation of $X$ associated with the dictionary $A$. To obtain an unique $Z$ and explore the latent structure of the given data set, various assumptions could be enforced over $Z$.

Liu et al. recently proposed Low Rank Representation (LRR) [1] by assuming that data are approximately sampled from an union of low-rank subspaces. Mathematically, LRR aims at solving

$$\min\ rank(Z) \quad \text{s.t. } X = AZ, \tag{1}$$

where $rank(Z)$ could be defined as the number of nonzero eigenvalues of the matrix $Z$. Clearly, (1) is non-convex and discontinuous, whose convex relaxation is as follows,

$$\min \|Z\|_* \quad \text{s.t. } X = AZ, \tag{2}$$

where $\|Z\|_*$ is the nuclear norm, which is a convex and continuous optimization problem.

Considering the possible corruptions, the objective function of LRR is

$$\min \|Z\|_* + \lambda\|E\|_p \quad \text{s.t. } X = AZ + E, \tag{3}$$

where $\|\cdot\|_p$ could be $\ell_1$-norm for describing sparse corruption or $\ell_{2,1}$-norm for characterizing sample-specified corruption.

The above nuclear-norm-based optimization problems are generally solved using Augmented Lagrange Multiplier algorithm (ALM) [2] which requires repeatedly performing Single Value Decomposition (SVD) over $Z$. Hence, this optimization program is inefficient.

Beyond the nuclear-norm, do other norms exist that can be used as a surrogates for rank-minimization problem in LRR? Can we develop a fast algorithm to calculate LRR? This letter addresses these problems by theoretically showing the equivalence between the solutions of a Frobenius-norm-based problem and the original LRR problem. And we further develop fast Low Rank Representation (fLRR) based on the theoretical results.

*Theoretical Analysis:* In the following analyses, Theorem 1 and Theorem 3 prove that Frobenius-norm-based problem is a surrogate of the rank-minimization problem of LRR in the case of clean data and corrupted ones, respectively. Theorem 2 shows that our Frobenius-norm-based method could produce a block-diagonal $Z$ under some conditions. This property is helpful to subspace clustering.

Let $A \in \mathcal{R}^{m\times n}$ be a matrix with rank $r$. The full SVD and skinny SVD of $A$ are $A = U\Sigma V^T$ and $A = U_r\Sigma_r V_r^T$, where $U$ and $V$ are two orthogonal matrices with the size of $m\times m$ and $n\times n$, respectively. In addition, $\Sigma$ is an $m\times n$ rectangular diagonal matrix, its diagonal elements are nonnegative real numbers. $\Sigma_r$ is a $r\times r$ diagonal matrix with singular values located on the diagonal in decreasing order, $U_r$ and $V_r$ consist of the first $r$ columns of $U$ and $V$, respectively. Clearly, $U_r$ and $V_r$ are column orthogonal matrices, i.e., $U_r^TU_r = I_r$, $V_r^TV_r = I_r$, where $I_r$ denotes the identity matrix with the size of $r\times r$. The pseudoinverse of $A$ is defined by $A^\dagger = V_r\Sigma_r^{-1}U_r^T$.

Given a matrix $M \in \mathcal{R}^{m\times n}$, the Frobenius norm of $M$ is defined by $\|M\|_F = \sqrt{trace\,(M^TM)} = \sqrt{\sum_{i=1}^{\min\{m,n\}}\sigma_i^2}$, where $\sigma_i$ is a singular value of $M$. Clearly, $\|M\|_F = 0$ if and only if $M = 0$.

*Lemma 1: Suppose $P$ is a column orthogonal matrix, i.e., $P^TP = I$. Then, $\|PM\|_F = \|M\|_F$.*

*Lemma 2: For the matrices $M$ and $N$ with same number of columns, it holds that*

$$\left\| \begin{bmatrix} M \\ N \end{bmatrix} \right\|_F^2 = \|M\|_F^2 + \|N\|_F^2. \tag{4}$$

The proofs of the above two lemmas are trivial.

*Theorem 1: Suppose that $X \in span\{A\}$, the Frobenius norm minimization problem*

$$\min \|Z\|_F \quad \text{s.t. } X = AZ, \tag{5}$$

*has an unique solution $Z^* = A^\dagger X$ which is also the lowest-rank solution of LRR in terms of (1).*

*Proof:* Let the full and skinny SVDs of $A$ be $A = U\Sigma V^T$ and $A = U_r\Sigma_rV_r^T$, respectively. Then, the pseudoinverse of $A$ is $A^\dagger = V_r\Sigma_r^{-1}U_r^T$. Defining $V_c$ by $V^T = \begin{bmatrix} V_r^T \\ V_c^T \end{bmatrix}$ and $V_c^TV_r = 0$. Moreover, it can be easily checked that $Z^*$ satisfies $X = AZ^*$ owing to $X \in span\{A\}$.

To prove that $Z^*$ is the unique solution of the optimization problem (5), two steps are required. First, we will prove that, for any solution $Z$ of $X = AZ$, it must hold that $\|Z\|_F \geq \|Z^*\|_F$. Using Lemma 1, we have

$$\begin{aligned} \|Z\|_F &= \left\| \begin{bmatrix} V_r^T \\ V_c^T \end{bmatrix} [Z^* + (Z - Z^*)] \right\|_F \\ &= \left\| \begin{bmatrix} V_r^TZ^* + V_r^T(Z - Z^*) \\ V_c^TZ^* + V_c^T(Z - Z^*) \end{bmatrix} \right\|_F. \end{aligned}$$

As $A\,(Z - Z^*) = 0$, i.e., $U_r\Sigma_rV_r^T\,(Z - Z^*) = 0$, it follows that $V_r^T(Z - Z^*) = 0$. Denote $B = \Sigma_r^{-1}U_r^TX$, then $Z^* = V_rB$. Because $V_c^TV_r = 0$, we have $V_c^TZ^* = V_c^TV_rB = 0$. Then,

$$\|Z\|_F = \left\| \begin{bmatrix} B \\ V_c^T(Z - Z^*) \end{bmatrix} \right\|_F. \tag{6}$$

By Lemma 2,

$$\|Z\|_F^2 = \|B\|_F^2 + \|V_c^T(Z - Z^*)\|_F^2, \tag{7}$$

then, $\|Z\|_F \geq \|B\|_F$.

By Lemma 1,

$$\|B\|_F = \|V_rB\|_F = \|Z^*\|_F, \tag{8}$$

thus, $\|Z\|_F \geq \|Z^*\|_F$ for any solution $Z$ of $X = AZ$.

In the second step, we will prove that if there exists another solution $Z$ of (5), $Z = Z^*$ must hold. Clearly, $Z$ is a solution of (5) which implies that $X = AZ$ and $\|Z\|_F = \|Z^*\|_F$. From (7) and (8),

$$\|Z\|_F^2 = \|Z^*\|_F^2 + \|V_c^T\,(Z - Z^*)\,\|_F^2. \tag{9}$$

Since $\|Z\|_F = \|Z^*\|_F$, it must hold that $\|V_c^T\,(Z - Z^*)\,\|_F = 0$, and so $V_c^T\,(Z - Z^*) = 0$. Together with $V_r^T\,(Z - Z^*) = 0$, this gives $V^T\,(Z - Z^*) = 0$. Because $V$ is an orthogonal matrix, it must hold that $Z = Z^*$. The above proves that $Z^*$ is the unique solution of the optimization problem (5).

Next, we prove that $Z^*$ is also a solution of the LRR optimization problem (1). Clearly, for any solution $Z$ of $X = AZ$, it holds that $rank(Z) \geq rank(AZ) = rank(X)$. On the other hand, $rank(Z^*) = rank(A^\dagger X) \leq rank(X)$. Thus, $rank(Z^*) = rank(X)$. This shows that $Z^*$ is the lowest-rank solution of the LRR optimization problem (1). The proof is complete. ■

In the following, Theorem 2 will show that the optimal $Z$ of (5) will be block-diagonal if the data are sampled from a set of independent subspaces $\{S_1, S_2, \cdots, S_k\}$, where the dimensionality of $S_i$ is $r_i$ and $i = \{1, 2, \cdots, k\}$. Note that, $\{S_1, S_2, \cdots, S_k\}$ are independent if and only if $S_i \bigcap \sum_{j\neq i} S_j = \{0\}$. Suppose that $X = [X_1, X_2, \cdots, X_k]$ and $A = [A_1, A_2, \cdots, A_k]$, where $A_i$ and $X_i$ contain $m_i$ and $n_i$ data points belonging to $S_i$, respectively. Moreover, $S_i = span\{A_i\}$ and $rank(A_i) = r_i$. By using the assumption that the data are drawn from a set of independent subspaces, we can easily prove Theorem 2.

*Theorem 2: Suppose that $X_i \in S_i$ and the subspaces $\{S_i | i = 1, \cdots, k\}$ are independent. Then, the unique solution $Z^*$ to the optimization problem (5) must be block-diagonal, i.e.,*

$$Z^* = \begin{bmatrix} Z_{11}^* & 0 & \cdots & 0 \\ 0 & Z_{22}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{kk}^* \end{bmatrix} \tag{10}$$

*where $Z_{ii}^*$ is an $m_i \times n_i$ matrix.*

Due to possible corruptions in $X$, it is hard to satisfy that $X$ is strictly sampled from the union of subspaces. To address this problem, we relax the equation $X = AZ$ to the inequality $\|X - AZ\|_F < \epsilon$, and the objective function of fLRR in the case of noisy data is as follows:

$$\min \|Z\|_F \quad \text{s.t.} \ \|X - AZ\|_F < \epsilon, \tag{11}$$

where $\epsilon > 0$ denotes the error tolerance, and the above constraint optimization problem (11) is equivalent to the unconstraint optimization problem:

$$E(Z) = \frac{\lambda}{2}\|Z\|_F^2 + \frac{1}{2}\|X - AZ\|_F^2 \tag{12}$$

where $\lambda > 0$ is a parameter.

*Theorem 3: The optimization problem (12) has a global solution:*

$$Z^\dagger = \left(\lambda I + A^T A\right)^{-1} A^T X, \tag{13}$$

*with*

$$rank(X) - rank\left(X - AZ^\dagger\right) \le rank\left(Z^\dagger\right) \le rank(X). \tag{14}$$

*Proof:* Let $\frac{\partial E(Z)}{\partial Z} = 0$, we have $Z^\dagger = \left(\lambda I + A^T A\right)^{-1} A^T X$. Because $rank(Z^\dagger) = rank\left(\left(\lambda I + A^T A\right)^{-1} A^T X\right) \le rank(X)$, and

$$\begin{aligned} rank(X) &= rank\left(X - AZ^\dagger + AZ^\dagger\right) \\ &\le rank\left(X - AZ^\dagger\right) + rank\left(Z^\dagger\right), \end{aligned}$$

it follows that $rank(X) - rank\left(X - AZ^\dagger\right) \le rank\left(Z^\dagger\right) \le rank(X)$.

Let the full SVD of $A$ be $A = U\Sigma V^T$ and skinny SVD of A be $A = U_r \Sigma_r V_r^T$, where $U = [U_r \ U_c]$ and $V^T = \begin{bmatrix} V_r^T \\ V_c^T \end{bmatrix}$ are orthogonal matrices. From (13), we have

$$\begin{aligned} Z^\dagger &= V\left(\lambda I + \Sigma^T \Sigma\right)^{-1} \Sigma^T U^T X && (15) \\ &= V_r S U_r^T X && (16) \end{aligned}$$

where $S$ is a diagonal matrix with the diagonal element $s_i = \frac{\sigma_i}{\sigma_i^2 + \lambda}$. If $\lambda \to 0$, then, $S \to \Sigma_r^{-1}$, and thus, $Z^\dagger \to V_r \Sigma_r^{-1} U_r^T X = A^\dagger X = Z^*$, which is consistent with the solution obtained in Theorem 1. This shows that the closed form solution $Z^\dagger$ presented in Theorem 3 is an approximation of the LRR in the case of corrupted data. This completes the proof. ■

Based on the above analyses, we develop fast Low Rank Representation (fLRR) for subspace clustering as follows:

1 For a given data set $X$, calculate its coefficient matrix by $Z^* = \left(\lambda I + X^T X\right)^{-1} X^T X$;
2 Cluster $X$ into $k$ subspaces by performing spectral clustering approach over $|Z^*| + |Z^*|^T$.

*Baselines:* To verify our claim that fLRR could obtain the lowest-rank representation based on Frobenius-norm-based problem. We carried out some experiments using two facial data sets (AR [3] and Extended Yale B (ExYaleB) [4]) and one motion database (Hopkins155 raw data [5]) in comparison to the existing well-known nuclear-norm-based LRR method (LRR) [1]. For fair comparison, we perform the same spectral clustering algorithm [6] on the affinity matrices achieved by fLRR and LRR, respectively. In all the tests, we tuned the parameters of fLRR and LRR and reported their best results on each database.

The used AR database contains 1400 clean faces randomly selected from 50 males and 50 females, and each image is resized from $165 \times 120$ to $55 \times 40$ (1/9). ExYaleB contains 2214 frontal-face images of 38 subjects, and each image is resized from $192 \times 168$ to $48 \times 42$ (1/16).

*Results:* Table 1 reports the Accuracy and Normalized Mutual Information (NMI) of LRR and fLRR on two facial data sets. Clearly, fLRR has very competitive clustering quality compared with that of LRR. More importantly, fLRR is obviously faster than LRR.

**Table 1:** Performance on the data sets of ExYaleB and AR. The values in parentheses denote the tuned values.

| Database | Algorithms | Accuracy (%) | NMI (%) | Time(s) |
|---|---|---|---|---|
| ExYaleB | LRR (1) | 66.86 | 73.46 | 5130.50 |
| | fLRR (0.1) | 71.00 | 76.31 | 133.12 |
| AR | LRR (6) | 80.86 | 91.18 | 973.84 |
| | fLRR (0.1) | 80.64 | 90.90 | 15.08 |
| **Speed up**: 14.50 – 64.58 times | | | | |

Next, we used the Hopkins 155 motion database [5] to examine the motion segmentation performance of fLRR. The data set has 156 video sequences, each one contains $39 \sim 550$ data points in two or three motions. Each motion is associated with a subspace. Table 2 shows the clustering error rates of the methods.

**Table 2:** Segmentation errors (%) achieved by LRR and fLRR over Hopkins155 raw data.

| Sequence | Algorithms | mean | max | median | std. |
|---|---|---|---|---|---|
| 2 motion | LRR (4) | 3.65 | 41.37 | 0.21 | 8.71 |
| | fLRR(0.008) | 2.68 | 38.65 | 0 | 6.92 |
| 3 motion | LRR (2) | 7.41 | 38.29 | 2.56 | 10.38 |
| | fLRR(0.004) | 6.52 | 33.94 | 2.56 | 8.57 |
| All | LRR (2.4) | 4.83 | 41.17 | 0.51 | 9.34 |
| | fLRR(0.008) | 3.90 | 38.65 | 0.29 | 8.26 |

*Conclusions:* Nuclear-norm-based optimization has been extensively used in these years as a surrogate of the low rank optimization problem. The nuclear-norm-based problems are often solved by using the ALM method which is complicated and inefficient for large scale setting. Differ from the existing works, this letter theoretically and experimentally showed that the Frobenius-norm can be used as a replacement of the rank function in the LRR optimization problem. Nevertheless, there is an unsolved problem in the selection of the parameter $\lambda$ in the closed form solution. More researches on this problem are required.

Haixian Zhang, Zhang Yi, and Xi Peng (*College of Computer Science, Sichuan University, Chengdu 610065, CHINA.*)

E-mail: zhangyi@scu.edu.cn.

## References

1 Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., and Ma, Y., 'Robust Recovery of Subspace Structures by Low-Rank Representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**, (1), pp. 171-184.
2 Yang, J. and Zhang, Y., 'Alternating Direction Algorithms for L1-Problems in Compressive Sensing', *SIAM Journal on Scientific Computing*, 2011, **33**, (1), pp. 250-278.
3 Martinez, A.M., The AR Face Database', CVC Technical Report, (1998)
4 Georghiades, A.S., Belhumeur, P.N., and Kriegman, D.J., 'From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**, (6), pp. 643-660.
5 Tron, R. and Vidal, R., 'A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms', *IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2007)
6 Ng, A.Y., Jordan, M.I., and Weiss, Y., 'On Spectral Clustering: Analysis and an Algorithm', *Advances in Neural Information Processing Systems*, (2002).