SIViP (2017) 11:1271–1278
DOI 10.1007/s11760-017-1084-9

ORIGINAL PAPER

Augmented Lagrangian alternating direction method for low-rank minimization via non-convex approximation

Yongyong Chen[1] · Yongli Wang[1] · Mingqiang Li[2] · Guoping He[3]

Received: 22 June 2016 / Revised: 18 October 2016 / Accepted: 20 March 2017 / Published online: 6 April 2017

Abstract This paper concerns the low-rank minimization problems which consist of finding a matrix of minimum rank subject to linear constraints. Many existing approaches, which used the nuclear norm as a convex surrogate of the rank function, usually result in a suboptimal solution. To seek a tighter rank approximation, we develop a non-convex surrogate to approximate the rank function based on the Laplace function. An iterative algorithm based on the augmented Lagrangian multipliers method is developed. Empirical studies for practical applications including robust principal component analysis and low-rank representation demonstrate that our proposed algorithm outperforms many other state-of-the-art convex and non-convex methods developed recently in the literature.

Keywords Low-rank minimization · Non-convex approximation · Iterative algorithm · Difference of convex programming

1 Introduction

In this paper, we consider the low-rank minimization model

$$\min_{L,S} \operatorname{rank}(L) + \lambda\|S\|_0 \quad \text{s.t.} \quad AL + S = D, \tag{1}$$

✉ Yongli Wang
wangyongli@sdkd.net.cn

1 College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

2 School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

3 Shandong Academy of Science, Jinan 250014, China

where $\lambda > 0$ is a weight parameter which balances the contribution of the rank versus the sparsity. $D \in R^{m\times n}$ is the data matrix (without loss of generality, we assume $m \leq n$). As we know, both the rank function and the l_0 norm are non-convex and discontinuous. The convex relaxation replacing the rank function and the l_0 norm, respectively, with the nuclear norm and l_1 norm can be converted into

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1 \quad \text{s.t.} \quad AL + S = D, \tag{2}$$

where $\|L\|_*$ is defined as the sum of the singular values of matrix L and $\|S\|_1 = \sum_{ij} |S_{ij}|$. The following are some interesting examples arising in many applications.

Example 1 Robust Principal Component Analysis (RPCA) The RPCA problem seeks to recover a low-rank matrix L plus a sparse matrix S from the corrupted data matrix $D \in R^{m\times n}$, which can be formulated as

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1 \quad \text{s.t.} \quad L + S = D. \tag{3}$$

Problem (3) can be written as (2) by defining $A = I$, where I denotes the identity matrix. This formulation has received broad attention in many applications, such as image processing, computer vision, web data ranking and bioinformatics [1–3]. Specifically, [1] has proved that when the rank of L and the sparsity of S satisfy some mild conditions, L and S can be exactly recovered with a high probability using (3). Unfortunately, when the large errors fasten on a number of columns of S, the convex relaxation (3) will fail. To overcome this dilemma, [4] has proposed an outlier pursuit

$$\min_{L,S} \|L\|_* + \lambda\|S\|_{2,1} \quad \text{s.t.} \quad L + S = D, \tag{4}$$

where $\|S\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m} S_{ij}^2}$. Under some mild conditions, outlier pursuit (4) can exactly recover the column support and exactly identify outliers.

Example 2 Low-Rank Representation (LRR)

Given the data matrix $A = [a_1, a_2, \ldots, a_n] \in R^{m\times n}$, each column of which is a sample. LRR aims to seek a low-rank representation matrix $L \in R^{n\times n}$ showing mutual similarity of the samples, or say LRR uses the data matrix A itself as the dictionary, i.e., $AL = A$. Mathematically, LRR can be formulated as

$$\min_{L} \|L\|_* \quad \text{s.t.} \quad AL = A. \tag{5}$$

In real applications, the data matrix A is often noisy or even grossly corrupted. In the case of data contaminated by outliers, LRR solves the convex optimization problem [5]

$$\min_{L,S} \|L\|_* + \lambda\|S\|_{2,1} \quad \text{s.t.} \quad AL + S = A. \tag{6}$$

Problem (6) cannot be completely conformed to (2) by defining $D = A$, but there are still many similarities.

Low-rank minimization can be applied in transform-invariant low-rank textures [6], matrix completion [7,8], low-rank subspace clustering [9] and so on. The literature [10] first gave a comprehensive overview to the concept of low-rank modeling and summarized the models and algorithms for low-rank matrix recovery. Furthermore, low-rank minimization has been investigated with application to hyper-spectral data recovery [11], accelerated dynamic magnetic resonance imaging [12] and unveiling traffic anomalies [13].

There is no denying that the convex relaxation has made great success in theoretical research and practical applications. However, there are still some shortages and deficiencies in the convex relaxation. On the one hand, the nuclear norm, which adds all singular values together, is essentially the l_1 norm of the singular values. This implies that large singular values are penalized more heavily than small ones. On the other hand, most theoretical analysis are based on a strong assumption that the underlying matrix must satisfy incoherence property, which may not be guaranteed in practical scenarios. Moreover, the convex relaxation has poor convergence rate as the matrix dimensions grow. From what has been discussed above, the nuclear norm may not be a good approximation of the matrix rank in some cases [14,15]. Recently, solving low-rank minimization problems has attracted broad attention by using non-convex proxy instead of nuclear norm. Specifically, the popular non-convex low- rank regularizers include l_p norm $(0 < p < 1)$ [16], capped norms [14], truncated nuclear norm regularization [15] and log-determinant function [17]. Motivated by the literature [18] which proposed a l_0 norm minimization model with different smooth approximations in seismic exploration, we develop a non-convex Laplace function to achieve a more accurate approximation to the rank of matrix than the nuclear norm. Furthermore, we propose an efficient iterative algorithm called NALM to solve the non-convex optimization problem.

Our contribution

In this paper, we propose a novel non-convex approximation of the rank function called Laplace norm, which is different from the nuclear norm. Our motivation stems from the compressive sensing which is concerned with the recovery of a sparse vector variable in some transform domain. Compared with traditional nuclear norm minimization problem, our model based on the Laplace norm is a non-convex problem. Thus, we present an iterative algorithm based on the augmented Lagrangian multipliers method to solve the non-convex optimization problem. Furthermore, we apply difference of convex (DC) programming [19] to solve the resulting subproblem which is a combination of concave and convex functions. Our empirical studies with applications to background extraction in surveillance video, face image shadow removal and subspace clustering validate that our proposed algorithm outperforms many other state-of-the-art convex and non-convex methods and Laplace norm is a more accurate and robust approximation to the rank function.

Organization

The rest of this paper is organized as follows. In Sect. 2, we propose our non-convex rank formulation for the low-rank minimization problem and provide a brief description of the augmented Lagrangian multipliers method (ALM). Sections 3 and 4 describe the procedure of NALM to solve the robust principal component analysis and low-rank representation problem, respectively. Experimental results are presented in Sect. 5. Finally, we make some conclusions and discussions in Sect. 6.

2 Proposed algorithm

In this section, we develop a novel non-convex matrix rank approximation and give an overview to the augmented Lagrangian multipliers method.

2.1 Laplace norm

We define the new norm of matrix L based on the Laplace function as

$$\operatorname{rank}(L) \approx \|L\|_\gamma = \sum_{i=1}^{m}\left(1 - \mathrm{e}^{-\sigma_i(L)/\gamma}\right), \quad \gamma > 0, \tag{7}$$

which is called Laplace norm. Certainly, Laplace norm is pseudo-norm but has nice properties.

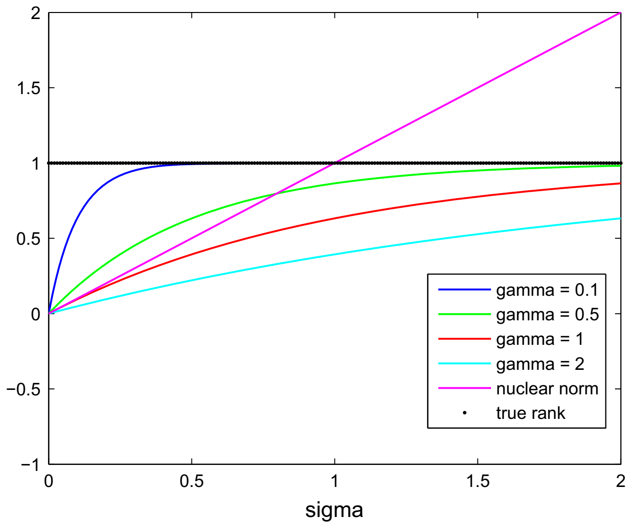

Fig. 1 Approximation of the rank function using Laplace norm, nuclear norm and true rank with an increasing value of σ

Proposition 1 *The Laplace norm has the following properties.*

(1) $\lim_{\gamma \to 0} \|L\|_\gamma = \text{rank}(L)$.
(2) $1 - e^{-\sigma_i(L)/\gamma} = 0$ *when* $\sigma_i(L) = 0$.[1]
(3) $\|L\|_\gamma$ *is unitarily invariant, that is,* $\|L\|_\gamma = \|ULV\|_\gamma$ *for any orthonormal* $U \in R^{m\times m}$ *and* $V \in R^{n\times n}$.
(4) *positive definiteness:* $\|L\|_\gamma \geq 0$ *for any* $L \in R^{m\times n}$ *and* $\|L\|_\gamma = 0$ *if and only if* $L = 0$.

Figure 1 plots the approximation of the rank function using Laplace norm, nuclear norm and true rank with an increasing value of σ. It illustrates that the smaller the γ is, the more accurate the Laplace norm approximation would be. In our empirical evaluations, the low-rank matrix L can be recovered more accurately than the nuclear norm by choosing proper γ.

2.2 Proposed non-convex formulation

With the above-mentioned Laplace norm, we consider the general framework as follows

$$\min_{L,S} \|L\|_\gamma + \lambda\|S\|_r \quad \text{s.t.} \quad AL + S = D. \tag{8}$$

where r depends on various problems. For example, when the low- rank minimization model (8) is used to solve LRR, $r = 2, 1$.

[1] The contribution of zero singular values in Laplace norm is the same as the true rank function and the nuclear norm.

2.3 The augmented Lagrangian multipliers method

In Ref. [20], the augmented Lagrangian multipliers method is introduced for solving the following constrained optimization problem with linear equality constraints

$$\min_{Z} f(Z) \quad \text{s.t.} \quad MZ - B = 0, \tag{9}$$

where $f : R^{p\times q} \to R$. The augmented Lagrangian function for this problem is defined as

$$L(Z, \Lambda, \mu) = f(Z) + \langle \Lambda, MZ - B\rangle + \frac{\mu}{2}\|MZ - B\|_F^2. \tag{10}$$

where $\mu > 0$ is called the penalty parameter and Λ is the Lagrange multipliers. $\langle\cdot,\cdot\rangle$ denotes the standard inner product in a finite-dimensional Euclidean space and $\|\cdot\|_F$ is the Frobenius norm of matrix variables. ALM first updates Z by minimizing $L(Z, \Lambda, \mu)$ with Λ being fixed and updates Λ with Z fixed at its latest value, until some convergence criteria are satisfied. The details of ALM are summarized in Algorithm 1.

Algorithm 1 ALM:General framework of augmented Lagrangian multipliers method

Require: Choose $\mu_0 > 0$ and Λ_0, Set $k = 0$
1: **while** not converged **do**
2: Update Z: $Z_{k+1} = \arg\min_Z L(Z, \Lambda_k, \mu_k)$
3: Update Λ: $\Lambda_{k+1} = \Lambda_k + \beta * \mu(MZ_{k+1} - B)$;
4: Update μ_k to μ_{k+1}.
5: **end while**
Ensure: Z_k

Recently, it has been shown that ALM is very efficient and extensible for many large-scale programming problems arising in statistics and machine learning [21,22], provided that the resulting subproblems are sufficiently simple to have closed-form solutions. In Sects. 3 and 4, we will focus on the applications of RPCA and LRR using ALM. When the resulting subproblems do not have closed-form solutions, we apply difference of convex (DC) programming such that the new subproblems can easily derive the closed-form solutions.

3 Application to robust principal component analysis

Using Laplace norm to replace the nuclear norm, (4) can be formulated as

$$\min_{L,S} \|L\|_\gamma + \lambda\|S\|_{2,1} \quad \text{s.t.} \quad L + S = D. \tag{11}$$

This problem can be viewed as a special case of the non-convex low- rank minimization model by defining $A = I$.

Denoting $Z = [L^T, S^T]^T$, $B = D$, $M = [I, I]$ and $f(Z) = \|L\|_\gamma + \lambda\|S\|_{2,1}$, the constrained RPCA (11) then can be rewritten as (9). The augmented Lagrangian function of (11) is given by

$$L(L, S, \Lambda, \mu) = \|L\|_\gamma + \lambda\|S\|_{2,1} + \langle \Lambda, L + S - D\rangle + \frac{\mu}{2}\|L + S - D\|_F^2$$

Thus, step 2 of Algorithm 1 can be written as

$$(L_{k+1}, S_{k+1}) = \arg\min_{L,S} \|L\|_\gamma + \lambda\|S\|_{2,1} + \frac{\mu}{2}\|L + S - D + \Lambda_k/\mu\|_F^2. \tag{12}$$

It is extremely difficult and trivial to solve the subproblem (12) concerning with matrix variables L and S simultaneously. Following the idea of nonlinear block Gauss–Seidel (NLBGS) technique, we can update L and S alternatively by keeping the other fixed at its latest value and then update the Lagrange multiplier, i.e.,

$$S_{k+1} = \arg\min_S \lambda\|S\|_{2,1} + \frac{\mu}{2}\|L_k + S - D + \Lambda_k/\mu\|_F^2; \tag{13}$$

$$L_{k+1} = \arg\min_L \|L\|_\gamma + \frac{\mu}{2}\|L + S_{k+1} - D + \Lambda_k/\mu\|_F^2; \tag{14}$$

$$\Lambda_{k+1} = \Lambda_k + \beta\mu(L_{k+1} + S_{k+1} - D), \quad \beta \in \left(0, \frac{\sqrt{5}+1}{2}\right) \tag{15}$$

where β is a step length. β aims to update the penalty parameter μ in each iteration, which is recommended in [7,22,23].

As to (13), we can obtain its closed-form solutions by the following lemma.

Lemma 1 [24,25] *Given a matrix $W_k = D - L_k - \Lambda_k/\mu$, then the subproblem*

$$\min_S \frac{1}{2}\|S - W_k\|_F^2 + \frac{\lambda}{\mu}\|S\|_{2,1} \tag{16}$$

has a closed-form solution S_{k+1} and the jth column of S_{k+1} is

$$S_{k+1}(:, j) = \begin{cases} \frac{\|W_k(:,j)\|_2 - \frac{\lambda}{\mu}}{\|W_k(:,j)\|_2} W_k(:, j), & if\ \frac{\lambda}{\mu} < \|W_k(:, j)\|_2 \\ 0, & otherwise. \end{cases} \tag{17}$$

As described earlier, the subproblem (14) is non-convex due to the concave Laplace norm. The intrinsic structure of the objective function in (14), which is a combination of concave and convex functions, motivates us to apply the difference of convex (DC) programming. We follow the similar linearization technique of the concave term in each iteration. The gradient of Laplace norm is given in the following lemma.

Lemma 2

$$\partial\|L\|_\gamma = \left\{U\mathrm{diag}(l)V^T : l_i = \mathrm{e}^{-\sigma_i(L)/\gamma}/\gamma\right\} \tag{18}$$

where the columns of U and V are the left and right singular vectors of L, respectively.

According to Lemma 2, at the $(k+1)$th iteration the subproblem (14) can be reformulated as

$$L_{k+1} = \arg\min_L \ \langle\partial\|L_k\|_\gamma, L\rangle + \frac{\mu}{2}\|L + S_{k+1} - D + \Lambda_k/\mu\|_F^2 \tag{19}$$

Computing the derivative of (19) with respect to L and setting the derivative to 0, we get

$$\partial\|L_k\|_\gamma + \mu(L + S_{k+1} - D + \Lambda_k/\mu) = 0. \tag{20}$$

It is easy to prove that the solution of (20) can be obtained by

$$L_{k+1} = D - S_{k+1} - (\Lambda_k + \partial\|L_k\|_\gamma)/\mu. \tag{21}$$

Now, a pseudo-code of the iterative scheme based on the ALM approach for (11) is as follows in Algorithm 2.

Algorithm 2 NALM1:Augmented Lagrangian Alternating Direction Method for non-convex RPCA

Require: Given data matrix $D \in R^{m\times n}$, choose $\mu > 0$, L_0, Λ_0 and $tol > 0$, Set $k = 0$.
1: **while** not converged **do**
2: Update S_{k+1} using (17);
3: Compute the gradient $\partial\|L_k\|_\gamma$ using (18);
4: Update L_{k+1} using (21);
5: Update Λ_{k+1} using (15).
6: **end while**
Ensure: L_k, S_k

4 Application to low-rank representation

Using the Laplace norm as the approximation of the rank function, the low-rank representation (6) can be written as

$$\min_{L,S} \|L\|_\gamma + \lambda\|S\|_{2,1} \quad \text{s.t.} \quad AL + S = A. \tag{22}$$

As discussed in [17,26,27], the observed data matrix A may be corrupted by impulsive noise which is spare but large and Gaussian noise which is small but dense under realistic situations. Then, the data matrix can be formulated as

$A = AL + S + E$, where S, E stand for the impulsive noise matrix and the Gaussian noise matrix, respectively. Taking both kinds of noise into consideration, we propose to optimize the following formula

$$\min_{L,S,E} \|L\|_\gamma + \lambda_1\|S\|_{2,1} + \lambda_2\|E\|_F^2 \quad \text{s.t.} \\ AL + S + E = A. \tag{23}$$

By introducing an auxiliary variable $X = L$, the optimization problem (23) can be transformed as

$$\min_{X,L,S} \|L\|_\gamma + \lambda_1\|S\|_{2,1} + \lambda_2\|A - AX - S\|_F^2 \quad \text{s.t.} \\ X = L. \tag{24}$$

The augmented Lagrangian function of (24) is

$$\begin{aligned} L(X, L, S, \Lambda, \mu) = \|L\|_\gamma &+ \lambda_1\|S\|_{2,1} \\ &+\lambda_2\|A - AX - S\|_F^2 + \langle \Lambda, L - X\rangle \\ &+\frac{\mu}{2}\|L - X\|_F^2. \end{aligned} \tag{25}$$

Thus, the iterative scheme of ALM for (24) can be summarized as follows

$$\begin{aligned} X_{k+1} &= \arg\min_X \lambda_2\|A - AX - S_k\|_F^2 + \frac{\mu}{2}\|L_k - X \\ &\quad +\Lambda_k/\mu\|_F^2; \end{aligned} \tag{26}$$

$$S_{k+1} = \arg\min_S \frac{\lambda_1}{\lambda_2}\|S\|_{2,1} + \|S - A + AX_{k+1}\|_F^2; \tag{27}$$

$$L_{k+1} = \arg\min_L \|L\|_\gamma + \frac{\mu}{2}\|L - X_{k+1} + \Lambda_k/\mu\|_F^2; \tag{28}$$

$$\Lambda_{k+1} = \Lambda_k + \beta\mu(L_{k+1} - X_{k+1}), \quad \beta \in \left(0, \frac{\sqrt{5}+1}{2}\right). \tag{29}$$

The solution of (26) is given by

$$\begin{aligned} X_{k+1} = &\left(2\lambda_2 A^T A + \mu I\right)^{-1} \left(2\lambda_2 A^T A - 2\lambda_2 A^T S_k \right. \\ &\left. +\mu L_k + \Lambda_k\right). \end{aligned} \tag{30}$$

We use the similar techniques as described in Sect. 3 to solve the subproblems (27) and (28). According to Lemma 1, the closed-form solution S_{k+1} can be gained by

$$S_{k+1}(:, j) = \begin{cases} \frac{\|W_k(:,j)\|_2 - \frac{\lambda_1}{2\lambda_2}}{\|W_k(:,j)\|_2} W_k(:, j), & if\ \frac{\lambda_1}{2\lambda_2} < \|W_k(:, j)\|_2 \\ 0, & otherwise, \end{cases} \tag{31}$$

where $W_k = A - AX_{k+1}$.

The DC programming is employed to solve the subproblem (28) whose gradient can be obtained by Lemma 2. Then, the original non-convex subproblem (28) can be solved by a series of convex problems, i.e.,

$$\begin{aligned} L_{k+1} = \arg\min_L &\langle \partial\|L_k\|_\gamma, L\rangle \\ &+\frac{\mu}{2}\|L - X_{k+1} + \Lambda_k/\mu\|_F^2. \end{aligned} \tag{32}$$

More specifically, the closed-form solution of (32) is obtainable by

$$L_{k+1} = X_{k+1} - (\Lambda_k + \partial\|L_k\|_\gamma)/\mu. \tag{33}$$

Based on the aforementioned analysis, the procedure of applying ALM to solve (23) is outlined in Algorithm 3.

Algorithm 3 NALM2:Augmented Lagrangian Alternating Direction Method for non-convex LRR

Require: Given data matrix $A \in R^{m\times n}$, choose $\mu > 0$, L_0,S_0, Λ_0 and $tol > 0$, Set $k = 0$
1: **while** not converged **do**
2: Update X_{k+1} using (30);
3: Update S_{k+1} using (31);
4: Compute the gradient $\partial\|L_k\|_\gamma$ using (18);
5: Update L_{k+1} using (33);
6: Update Λ_{k+1} using (29).
7: **end while**
Ensure: L_k

5 Experimental results

In this section, we present several experiments that validate our proposed algorithm and demonstrate its performance in some practical applications including background extraction in surveillance video, face image shadow removal and subspace clustering. The first two applications belong to RPCA, while the last application pertains to LRR. All the experiments were performed on a Lenovo laptop with an Intel Core i3-3240T 2.90GHz CPU that has 4 cores and 4GB of memory, running with Windows 8 and MATLAB (R2013a).

5.1 Experimental settings and implementation details

Our numerical experiments can be divided into two broad categories.

(1) In Sects. 5.2 and 5.3, we will compare our proposed NALM1 with the state-of-the-art robust principal component analysis algorithms: IALM [22] which focused on solving (3) by using the nuclear norm, GoDec [28] which developed the non-convex RPCA formulation by using iterative hard thresholding and NcRPCA [29] which consisted of alternating between projecting appropriate residuals onto low-rank and sparse matrices.

(2) In Sect. 5.4, we will compare our NALM2 with the state-of-the-art subspace clustering algorithms: SSC [30], LRSC [9], LSA [31], SCC [32] and LRR [5].

In our experiments, unless otherwise specified we follow the default values for parameters used in the solvers. In Sects. 5.2 and 5.3, the weight parameter λ in (3) was suggest to $\lambda = 1/\sqrt{\max(m,n)}$ in [1,2]. However, we find that λ with small magnitude is proper to our algorithm and we set $\lambda = 10^{-4}$. For the penalty parameter, μ is also chosen to set an increasing sequence of values, which is similar to IALM. The step length β is set $\beta = 1.618$ and the initial value of μ is equal to 10^{-3}, 0.5, respectively. We set $\gamma = 10^{-2}$, which is crucial to the Laplace norm since it determines the efficiency of the low-rank approximation. In Sect. 5.4, $\lambda, \beta, \mu, \gamma$ are set $\lambda_1 = 0.1, \lambda_2 = 5, \beta = 1.618, \mu_0 = 1, \gamma = 5 \times 10^{-2}$, respectively.

In Sects. 5.2 and 5.3, the stopping criterion we used to terminate the involved algorithms is the following relative error rule

$$\text{RelErr} = \frac{\|L_k + S_k - D\|_F}{\|D\|_F} \leq \text{tol},$$

where tol > 0 is a predefined tolerance and is set tol $= 10^{-3}$. To evaluate both effectiveness and efficiency of involved algorithms, we use the relative error (denoted by RelErr), the rank of L (denoted by Rank(L)), the CPU time in seconds (abbreviated as Time(s)) and the number of iterations (abbreviated as Iter). In Sect. 5.4, following [9], we use the subspace clustering error defined as

$$\text{clustering error} = \frac{\#\text{ of misclassified points}}{\text{total }\#\text{ of points}} \tag{34}$$

to evaluate the performance of involved algorithms.

5.2 Background extraction in surveillance video

A key application of RPCA is background extraction in surveillance video [1,33], which aims to detect moving objects from a stationary background. In this experiment, a benchmark data set escalator which is downloaded from http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html is used for RPCA. The data set contains a color (three channels) video with 3417 frames of 130 × 160 size. Our experiments employ both grayscale and color videos, which are different from the experiments in [1,10,28] and similar to the experiments in [34]. As is common that by stacking each frame as a column vector in the lexicographic order, we get a data matrix D whose columns are consistent with the sequence of the video frames. Due to memory limitation, in the first example which uses a sequence of 100 grayscale frames from escalator, the data matrix D is in $R^{20800\times 100}$. However, the data matrix D is in $R^{20800\times 300}$ in the second example which employs a sequence of 100 color frames from escalator.

Table 1 Comparison results of NALM1, IALM, GoDec and NcRPCA with the grayscale and color videos

Algorithms	Escalator/escalator (c)			
	RelErr	Rank(L)	Time(s)	Iter
NALM1	4.88e−6/4.93e−6	1/1	0.85/3.27	1/1
IALM	9.89e−4/6.79e−4	40/141	7.67/30.24	17/19
GoDec	9.43e−4/7.99e−4	1/3	48.65/115.64	100/100
NcRPCA	9.89e−4/6.79e−4	1/3	8.53/23.19	45/45

As stated in [35], the matrix of aligned images will have low rank, ideally rank one, so we set the desired rank in GoDec and NcRPCA for the grayscale video to be one, however, and to be three for the color video. The experimental results are depicted in Table 1. In Table 1, Escalator(c) denotes the tested color videos and the notation $ae-b$ means $a \times 10^{-b}$. From Table 1, we can observe that NALM1 is at least ten times faster and has higher accuracy than the other involved algorithms. Figure 2a, b depicts a visual comparison of background extraction using different algorithms. As can be seen, the non-convex algorithms including GoDec, NcRPCA and NALM1 result in better visual effect than the convex algorithm IALM, since the three moving escalators are removed to the foreground. Although our NALM1 is slightly worse in vision than NcRPCA, it is much faster than GoDec, NcRPCA and IALM. In addition, compared with IALM which results in L with rank 40, NALM1, GoDec and NcRPCA can obtain the desired rank-one matrix L using grayscale videos.

5.3 Face image shadow removal

Another important application of RPCA in [1] is removing artifacts such as shadows, specularities and saturations from face images. In this experiment, we use face images from the Extended Yale B database [36] which has 38 subjects; each subject has 64 images of size 192 × 168. So the data matrix D is in $R^{32256\times 64}$ for each subject.

As illustrated in Fig. 2c, our proposed algorithm removes the artifacts including local defects as the sparse component, while the performance of other algorithms suffers in the presence of some artifacts. Notice that the non-convex algorithms are superior to IALM. Table 2 states a quantitative comparison among the four algorithms. Similar to the conclusion in Sect. 5.2, NALM1 is at least ten times faster and has higher accuracy than the competing algorithms.

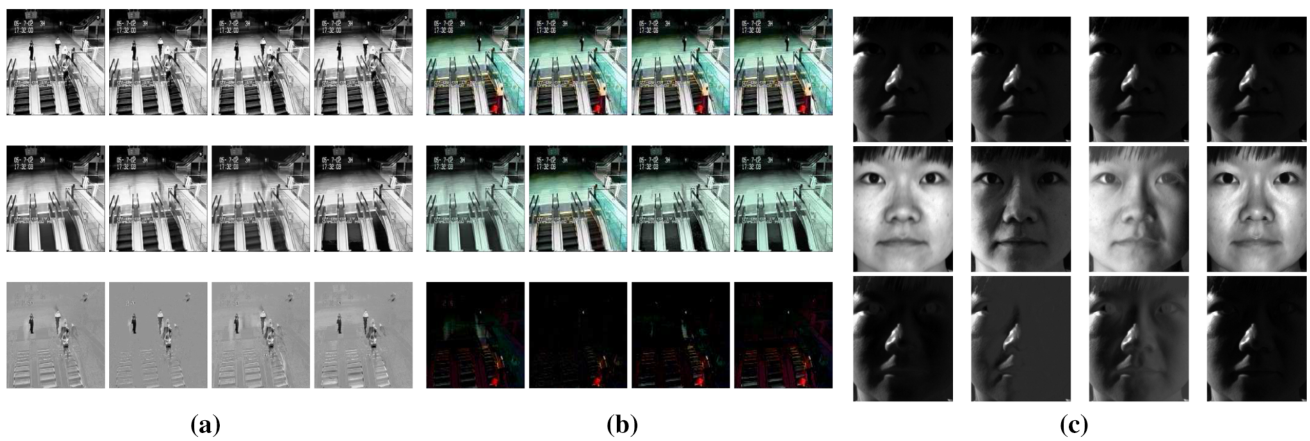

Fig. 2 Using RPCA for background extraction (**a**, **b**) and face image shadow removal (**c**). The *rows* from *top* to *bottom* correspond to the original video or the original face images, the low-rank components and sparse components, respectively. The *columns* from *left* to *right* in (**a**–**c**) are implemented by NALM1, IALM, GoDec and NcRPCA, respectively

Table 2 Comparison results of NALM1, IALM, GoDec and NcRPCA with the Extended Yale B database

Algorithms	Extended Yale B database			
	RelErr	Rank(L)	Time(s)	Iter
NALM1	3.07e−6	1	0.46	1
IALM	6.40e−4	32	5.87	17
GoDec	2.67e−3	1	53.56	100
NcRPCA	6.40e−4	1	10.41	49

Table 3 Clustering error rates (%) of different algorithms on Extended Yale B database

Method	LSA	SCC	LRR	LRSC	SSC	NALM2
2 Objects						
Mean	32.80	16.62	9.52	5.32	1.86	**1.55**
Median	47.66	7.82	5.47	4.69	**0.00**	0.078
3 Objects						
Mean	52.29	38.16	19.52	8.47	3.10	**2.33**
Median	50.00	39.06	14.58	7.81	**1.04**	1.56
5 Objects						
Mean	58.02	58.90	34.16	12.24	4.31	**3.15**
Median	56.87	59.38	35.00	11.25	2.50	**2.25**
8 Objects						
Mean	59.19	66.11	41.19	23.72	5.85	**3.90**
Median	58.59	64.65	43.75	28.03	4.49	**3.32**
10 Objects						
Mean	60.42	73.02	38.85	30.36	10.94	**4.11**
Median	57.5	75.78	41.09	28.75	5.63	**2.97**

The lowest clustering error rates are highlighted in bold

5.4 Clustering and classification

After obtaining the low-rank representation matrix L from (23), we follow the idea of [5] to construct the affinity matrix that is used to perform spectral clustering in a way similar to [37]. Following the previous work [9], we resize all images from the Extended Yale B database to 48×42 pixels and vectorize each image as a data point to construct the data matrix A. Inspired by [30], we divide the the 38 objects into four groups: subjects 1–10, 11–20, 21–30, 31–38. For each of the first three groups, all choices of $n \in \{2, 3, 5, 8, 10\}$ are considered, and for the last group, all choices of $n \in \{2, 3, 5, 8\}$ objects are considered. Finally, we implement the subspace clustering algorithm on each set of n subjects.

The clustering error rates defined as (34) of NALM2 as well as the LSA, SCC, LRR, LRSC and SSC whose results are cited from [30] on Extended Yale B database are given in Table 3. As can be seen from Table 3, our proposed non-convex algorithm NALM2 has superior performance than other competing algorithms in terms of mean accuracy for all cases. Unfortunately, NALM2 is inferior but comparable than SSC in cases with two or three objects, however, superior on other cases. Moreover, NALM2 based on the non-convex regularizer is more stable than LRR [5] using the convex nuclear norm whose error rates increase thoroughly as the number of subjects increases.

6 Conclusions and discussion

This paper focuses on studying a non-convex approach to the low-rank minimization problem, which involves extracting a low-rank structure, by a novel approximation of the rank function called Laplace norm. Our proposed formulation can give more accurate approximation to the rank function than the nuclear norm. This is due to the fact that unlike traditional nuclear norm, which adds all singular values together, Laplace norm can give the balanced penalization for different singular values. In addition, we present a simple and efficient iterative algorithm based on the augmented Lagrangian multipliers method to solve the non-convex optimization problem. Numerical experimental results demonstrate that our proposed algorithm outperforms many other state-of-the-art convex and non-convex approaches.

On the other hand, we hope that our current idea will motivate to extend the proposed Laplace norm in the low-rank matrix completion occurring in image inpainting and recommender systems.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 11241005) and Research Project Supported by Shanxi Scholarship Council of China (2015-093).

References

1. Candès, E.J., Li, X., Ma, Y., et al.: Robust principal component analysis? J. ACM **58**(3), 11 (2011)
2. Wright, J., Ganesh, A., Rao, S., et al.: Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. In: 23rd Annual Conference on Advances in Neural Information Processing Systems 22, Vancouver, 7–10 Dec 2009 (2009)
3. De la Torre, F., Black, M.J.: Robust principal component analysis for computer vision. ICCV **1**, 362–369 (2001)
4. Xu, H., Caramanis, C., Sanghavi, S.: Robust PCA via outlier pursuit. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems, NIPS'10, pp. 2496–2504, Curran Associates Inc, Vancouver, 6–9 Dec 2010 (2010)
5. Liu, G., Lin, Z., Yan, S., et al.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
6. Zhang, Z., Liang, X., Ganesh, A., et al.: TILT: transform invariant low-rank textures. Int. J. Comput. Vis. **99**(1), 1–24 (2010)
7. Chen, C., He, B., Yuan, X.: Matrix completion via an alternating direction method. IMA J. Numer. Anal. **32**(1), 227–245 (2012)
8. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. Found. Comput. Math. **9**(6), 717–772 (2009)
9. Vidal, R., Favaro, P.: Low rank subspace clustering (LRSC). Pattern Recognit. Lett. **43**, 47–61 (2014)
10. Zhou, X., Yang, C., Zhao, H., et al.: Low rank modeling and its applications in image analysis. ACM Comput. Surv. **47**(2), 36 (2015)
11. Xu, Y., Yin, W., Wen, Z., et al.: An alternating direction algorithm for matrix completion with nonnegative factors. Front. Math. China **7**(2), 365–384 (2012)
12. Otazo, R., Candès, E.J., Sodickson, K.: Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magn. Reson. Med. **73**(3), 1125–1136 (2015)
13. Mardani, M., Mateos, G., Giannakis, G.: Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies. IEEE Trans. Inf. Theory **59**(8), 5186–5205 (2013)
14. Sun, Q., Xiang, S., Ye, J.: Robust principal component analysis via capped norms. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 311–319 (2013)
15. Hu, Y., Zhang, D., Ye, J., et al.: Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans. Pattern Anal. Mach. Intell. **35**(9), 2117–2130 (2013)
16. Marjanovic, G., Solo, V.: On optimization and matrix completion. IEEE Trans. Signal Process. **60**(11), 5714–5724 (2012)
17. Peng, C., Kang, Z., Li, H., et al.: Subspace clustering using log-determinant rank approximation. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 925–934 (2015)
18. Wang, Y., Liu, P., Li, Z., et al.: Data regularization using Gaussian beams decomposition and sparse norms. J. Inverse Ill Posed Probl. **21**(1), 1–23 (2013)
19. Tao, P., An, L.: Convex analysis approach to dc programming: theory, algorithms and applications. Acta Math. Vietnam. **22**(1), 289–355 (1997)
20. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic press, Cambridge (1982)
21. Boyd, S., Parikh, N., Chu, E., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
22. Lin, Z., Chen, M., Ma, Y.: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055 (2010)
23. Bazaraa, M., Sherali, H., Shetty, C.: Nonlinear Programming: Theory and Algorithms. Wiley, Hoboken (1993)
24. Huang, J., Nie, F., Huang, H., et al.: Robust manifold nonnegative matrix factorization. ACM Trans. Knowl. Discov. Data (TKDD) **8**(3), 11 (2014)
25. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B **68**(1), 49–67 (2006)
26. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. Optim. **20**(4), 1956–1982 (2010)
27. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observation. SIAM J. Optim. **21**(1), 57–81 (2011)
28. Zhou, T., Tao, D.: Godec: Randomized low-rank and sparse matrix decomposition in noisy case. In: 28th International Conference on Machine Learning (ICML), pp. 33–40, Bellevue, 28 June–2 July 2011 (2011)
29. Netrapalli, P., Niranjan, U.N., Sanghavi, S., et al.: Non-convex robust PCA. In: Ghahraman, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 1107–1115. Curran Associates, Inc. (2014)
30. Elhamifar, E., Vidal, R.: Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **35**(11), 2765–2781 (2013)
31. Yan, J., Pollefeys, M.: A General Framework for Motion Segmentation: Independent, Articulated, Rigid, Non-rigid, Degenerate and Non-degenerate. ECCV 2006. Springer, Berlin (2006)
32. Chen, G., Lerman, G.: Spectral curvature clustering (SCC). Int. J. Comput. Vis. **81**(3), 317–330 (2009)
33. Li, L., Huang, W., Gu, I., et al.: Statistical modeling of complex backgrounds for foreground object detection. IEEE Trans. Image Process. **13**(11), 1459–1472 (2004)
34. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. **141**(1–2), 349–382 (2013)
35. Peng, Y., Ganesh, A., Wright, J., et al.: RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2233–2246 (2012)
36. Lee, K.C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 684–698 (2005)
37. Agarwal, P.K., Mustafa, N.H.: k-Means projective clustering. In: Proceedings of the Twenty-Third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 155–165 (2004)