

A Smooth Approximation Algorithm of Rank-Regularized Optimization Problem and Its Applications

Bo Li¹(✉), Lianbao Jin¹, Chengcai Leng¹, Chunyuan Lu¹, and Jie Zhang²

¹ Nanchang Hangkong University, Nanchang 330063, Jiangxi, China
libo@nchu.edu.cn

² Dalian University of Technology, Dalian 116024, Liaoning, China

Abstract. In this paper, we propose a novel smooth approximation algorithm of rank-regularized optimization problem. Rank has been a popular candidate of regularization for image processing problems, especially for images with periodical textures. But the low-rank optimization is difficult to solve, because the rank is nonconvex and can't be formulated in closed form. The most popular methods is to adopt the nuclear norm as the approximation of rank, but the optimization of nuclear norm is also hard, it's time expensive as it needs computing the singular value decomposition at each iteration. In this paper, we propose a novel direct regularization method to solve the low-rank optimization. Contrast to the nuclear-norm approximation, a continuous approximation for rank regularization is proposed. The new method proposed in this paper is a 'direct' solver to the rank regularization, and it just need computing the singular value decomposition one time, so it's more efficient. We analyze the choosing criteria of parameters, and propose an adaptive algorithm based on Morozov discrepancy principle. Finally, numerical experiments have been done to show the efficiency of our algorithm and the performance on applications of image denoising.

Keywords: Rank regularization · Nuclear norm · Morozov discrepancy principle

1 Introduction

One of the most important task in computer vision is image denoising. The general idea is to regard a noisy image d as being obtained by corrupting a noiseless image m ; given a model for the noise corruption, the desired image m is a solution of the corresponding inverse problem.

$$d = m + n$$

Bo Li is supported by the NSFC China (61262050) and Natural Science Foundation of Jiangxi Province (20151BAB211006), Chengcai Leng is supported by NSFC China (61363049).

where n is the noise, in this paper we assume that the noise is normally distributed and additive. Although, the image denoising problem is the simplest possible inverse problem, it's still a difficult task for researchers due to the complicated structures and textures of natural images. Generally, the degraded imaging process can be described mathematically by the first class Fridman operator equation

$$(Af)(x, y) = \int \int \mathcal{K}(x - \xi, y - \eta)m(x, y)dxdy + noise = d(x, y)$$

where \mathcal{K} is usually a linear kernel function. For deblurring problem, \mathcal{K} is the point spread function(psf), and for denoising, it's just the identity operator. Since the the inverse problem is extremely ill-posed, most denoising procedures have to employ some sort of regularization. Generally, the image denoising problem can be solved under the frame of Tikhonov regularization.

$$\min_m \mathcal{J}(m) = \|\mathcal{K}m - d\|^2 + \alpha\Omega[m]$$

where $\Omega[m]$ is the stability functional for m . According to different types of problems, the $\Omega[m]$ can adopt different regularizations, such as Total Variation [1, 2], F-norm [3], L0-norm [4], and so on.

For images with low-rank structures, the rank regularization is a good choice. But the key problem is that the rank of an image matrix is difficult to be formulated in closed form. Recently, reference [4] showed that under surprisingly broad conditions, one can exactly recover the low-rank matrix A from $D = A + E$ with gross but sparse errors E by solving the following convex optimization problem:

$$\arg \min_A \|D - A\|_F^2 + \|A\|_*$$

where $\|\cdot\|_*$ represents the nuclear norm of a matrix (the sum of its singular values). In [5], this optimization is dubbed Robust PCA (RPCA), because it enables one to correctly recover underlying low-rank structure in the data, even in the presence of gross errors or outlying observations. This optimization can be easily recast as a semidefinite program and solved by an off-the-shelf interior point solver (e.g., [6–8]). However, although interior point methods offer superior convergence rates, the complexity of computing the step direction is $O(m^6)$. So they do not scale well with the size of the matrix.

In recent years, the research for more scalable algorithms for high dimensional convex optimization problems has prompted a return to first-order methods. One striking example of this is the current popularity of iterative thresholding algorithms for l^1 -norm minimization problems arising in compressed sensing [9–11]. Similar iterative thresholding techniques [12–15] can be applied to the problem of recovering a low-rank matrix from an incomplete (but clean) subset of its entries. This optimization is closely related to the RPCA problem, and the algorithm and convergence proof extend quite naturally to RPCA. However, the iterative thresholding scheme proposed in [4] exhibits extremely slow convergence: solving

one instance requires about 10^4 iterations, each of which has the same cost as one singular value decomposition.

In this paper, we attempt to solve the rank regularization problem in a new direction. Contrast to the nuclear-norm approximation introduced in [4], we propose a continuous approximation for rank regularization, and analyze the regularity of the algorithm. Compared with the nuclear norm approximation, the new method proposed in this paper is a ‘direct’ solver to the rank regularization, and it’s a smooth optimization problem with fruitful theory and computation foundation. Finally, the denoising experiments based on the new regularization are proposed for images with periodical textures.

The paper is organized as follows. In Sect. 2 we describe the proposed smooth approximation algorithm in detail, and the criteria of automatic parameter choosing is discussed theoretically; the numerical experiments results will be shown in Sect. 3 and conclusions in Sect. 4.

2 Proposed Method

In this paper, we consider the rank regularization optimization problem, which can be described as following model:

$$\min_u \mathcal{J}(u) = \|u - u_0\|^2 + \text{rank}(u) \quad (1)$$

As the rank of an image matrix is nonconvex and difficult to be formulated in closed form, so the optimization problem (1) is difficult to be solved directly. The nuclear norm is a relaxation of the rank, and it can be proved that under some conditions this relaxation is a good convex approximation.

$$\min_u \mathcal{J}(u) = \|u - u_0\|^2 + \alpha \|u\|_*$$

Many optimization algorithms for this problem have been researched, such as the Accelerated Proximal Gradient (APG) [16–18], the Augmented Lagrange Multiplier (ALM) Method [19], and Iterative Singular Value Thresholding (ISVT) [12]. One common shortcoming of these methods is that they have to calculate the singular value decomposition at each iteration, which increase the complexity of computation heavily.

In this paper, we attempt to solve the rank regularization problem in a new direction. Despite using the nuclear-norm as relaxation, we propose a ‘direct’ solver to the rank regularization. The main contribution of this algorithm includes: firstly, it’s a smooth optimization problem, easily to be solved, and we propose an automatic parameter choosing criteria based on the Morozov discrepancy principle; secondly, compared with traditional methods, it just needs calculating singular value decomposition one time, so it’s more efficient.

From the basis of matrix theory, the rank of a matrix equals to the number of non-zero singular value terms. For a matrix A , $\text{rank}(A) = \#\{\sigma \neq 0\}$, where σ is the vector composed by the singular value of matrix A . From this point of view, to compute the rank of a matrix is a process of counting non-zeros. It’s a

generalization of L_0 norm to two dimensional space. But 'the counting process' is also difficult to be formed in the closed form and can't be optimized directly. In this paper, we propose a continuous function to describe the process of "counting non-zeros".

Firstly, for a matrix A , the operation of singular value decomposition should be calculated, $A = S \cdot V \cdot D$, where S, D are the unitary matrix, and V is a diagonal matrix composed by the singular value of A . Then, we construct a characteristic function $p(\sigma(i))$ for each singular value $\sigma(i)$, and the function $p(\sigma_i)$ should have the following property: $p(\sigma_i) = \begin{cases} 0, & \text{if } \sigma_i = 0; \\ 1, & \text{else.} \end{cases}$ In this paper, we choose the gaussian function as the approximation of characteristic function, with only one parameter s as the variance.

$$p(\sigma) = 1 - \exp\left(-\frac{\sigma^2}{s^2}\right), \tag{2}$$

From Fig. 1, we can see that when $s \rightarrow 0$, the function $p(\cdot)$ is a good approximation of characteristic function. So the rank can be approximately formulated as the sum of the characteristic functions of each singular value, $rank(A) = \sum_i p(\sigma(i))$. With the proposed relaxation of rank, the denoising model (1) can be formulated as following

$$\min_{\sigma} \mathcal{J}(\sigma) = \|S \cdot V \cdot D - u_0\|^2 + \lambda \sum_i p(\sigma(i)) \tag{3}$$

where $V = diag\{\sigma\}$, and the final denoised image $u = S \cdot V(\sigma) \cdot D$.

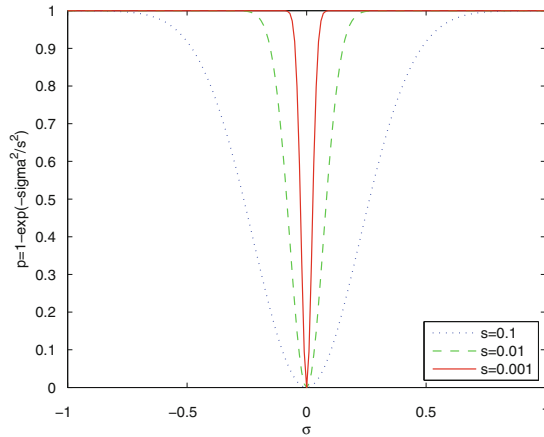


Fig. 1. The curve of characteristic function $p(\sigma) = 1 - \exp\left(-\frac{\sigma^2}{s^2}\right)$ with different choosing parameter $s = 0.1, 0.01, 0.001$.

Compared with the nuclear norm, the model (3) is a new relaxation to the rank regularization. And we can see that it's a smooth optimization problem, especially, it just need compute the singular value decomposition just one time.

From the property of unitary of matrix S and D , the model (3) can be reformulated as

$$\min_{\sigma} \|V - S^T u_0 D^T\|^2 + \lambda \sum_i p(\sigma_i)$$

In the discrete representation, it can be described as

$$\min_{\sigma} \sum_{i,j} (\sigma_i - (S^T u_0 D^T)_{ij})^2 + \lambda \sum_i p(\sigma_i) \tag{4}$$

By computing the derivative with respect to σ_i , the final problem is to solve the following smooth equation system, which can be efficiently solved by Newton methods.

$$\sigma_i + \frac{2\lambda}{s^2} \sigma_i \exp(-\frac{\sigma_i^2}{s^2}) = \sum_j (S^T u_0)_{ij} * d_{ij} \tag{5}$$

The detail process is shown in Algorithm 1.

Algorithm 1

- 1 Initialization, $[S_0, V_0, D_0] = SVD(u_0)$, λ, s
 - 2 Iteration
 - (1) $\sigma_{new} = \arg \min_{\sigma} |S * V * D - u_0|_2^2 + \lambda \sum_i p(\sigma_i)$; solved by (5)
 - (2) update $u_{new} = S * V(\sigma_{new}) * D$;
 - (3) if $|u_{new} - u_0| < tolerance$, stop;
 - (4) let $u_0 = u_{new}$, and goto (1).
-

In Algorithm 1, there are two parameters λ and s need to be initialized, How to choose these two parameters is an important and hard work. In the following, we will discuss the choosing criteria of λ and s .

The choose of parameter λ is difficult for regularization problem, because if the weighting parameter λ is too small, the algorithm will be unstable, and if λ is too large, the regularity will be too strong and the resulting algorithm will not accurately optimize the originally problem. In this paper, we propose an adaptive choosing criteria of parameter λ based on the Morozov discrepancy principle.

As the final optimal σ is dependent on the choosing of parameter λ , so the problem (3) should be reformulated as

$$\min_{\sigma} J(\sigma, \lambda) = \|S \cdot \text{diag}(\sigma_i^\lambda) \cdot D - u_{0,\delta}\|^2 + \alpha \sum_i p(\sigma_i^\lambda) \tag{6}$$

There are two variables σ and λ . For a fixed λ , the optimization for σ is simple to be solved by the Euler equation:

$$\sigma_i^\lambda + \lambda p(\sigma_i^\lambda)' = b, \tag{7}$$

And for fixed σ , the parameter λ can be chosen by the Morozov discrepancy principle.

Proposition 1. *The criteria of choosing parameter λ Definite functional $\phi(\lambda)$,*

$$\phi(\lambda) = \|S \cdot \text{diag}(\sigma^\lambda) \cdot D - u_{0,\delta}\|^2 - \delta^2 \tag{8}$$

according to the Morozov discrepancy principle, the parameter λ should obey $\frac{d\phi(\lambda)}{d\lambda} = 0$, which follows the Newton iteration

$$\lambda_{k+1} = \lambda_k - \frac{\phi(\lambda_k)}{\phi'(\lambda_k)} \tag{9}$$

according to (7), $\phi'_{\lambda_k} = \sigma_i \frac{d\sigma_i}{d\lambda} = -\sigma_i(1 + \lambda_k p(\sigma_i^\lambda))^{-1} p(\sigma_i^\lambda)'$.

We can see that the function $p(\sigma)$ is not a convex function, especially when the parameter $s \rightarrow 0$, it's difficult to obtain the global optimal solution. But for bigger s , this is a smooth optimization problem, and the global solution can be surely solved. So in this paper, we choose a descent sequence $s, s_0 > s_1 > \dots > s_n$.

Algorithm 2. Algorithm with adaption of parameter λ and s

1 Initialization, $[S_0, V_0, D_0] = SVD(u_0), \lambda_0, s_0$

2 Iteration

(1) $\sigma_{new} = \arg \min_{\sigma} |S * V(\sigma^\lambda) * D - u_0|_2^2 + \lambda \sum_i f(\sigma_i^\lambda)$; solved by (5)

(2) update λ according to the Morozov discrepancy principle by (9)

(3) update $u_{new} = S * V(\sigma_{new}) * D$;

(4) if $|u_{new} - u_0| < tolerance$, stop;

(5) let $u_0 = u_{new}, s = 1/2 * s_0$ and goto (1).

3 Numerical Experiments

In this section, we report our experimental results with the proposed algorithm in the previous section. We present experiments on both synthetic data and real images. All the experiments are conducted and timed on the same PC with an Intel Core i5 2.50 GHz CPU that has 2 cores and 16 GB memory, running Windows 7 and Matlab (Version 7.10).

We compared the proposed method with some of the most popular low rank matrix completion algorithm, including the Accelerated Proximal Gradient algorithm (APG), the Iterative Singular Value Thresholding algorithm (ISVT) [11] and the Augmented Lagrange Multiplier (ALM).

3.1 Synthetic Data

We generate the synthetic matrix $X \in R^{250 \times 250}$ with rank equal to 25, as shown in Fig. 2(a), it's a block-diagonal matrix. Then the gaussian noise with level 0.35 is added, as shown in Fig. 2(b). For the compared algorithms, we use the

default parameters in their released codes. The results were shown in Fig. 2(c-f), and the PSNR values, estimated rank and the costing time were listed in the Table 1. According to the experimental results, we can learn that our algorithm can achieve the exact estimation of rank and higher PSNR value while having the smallest time complexity.

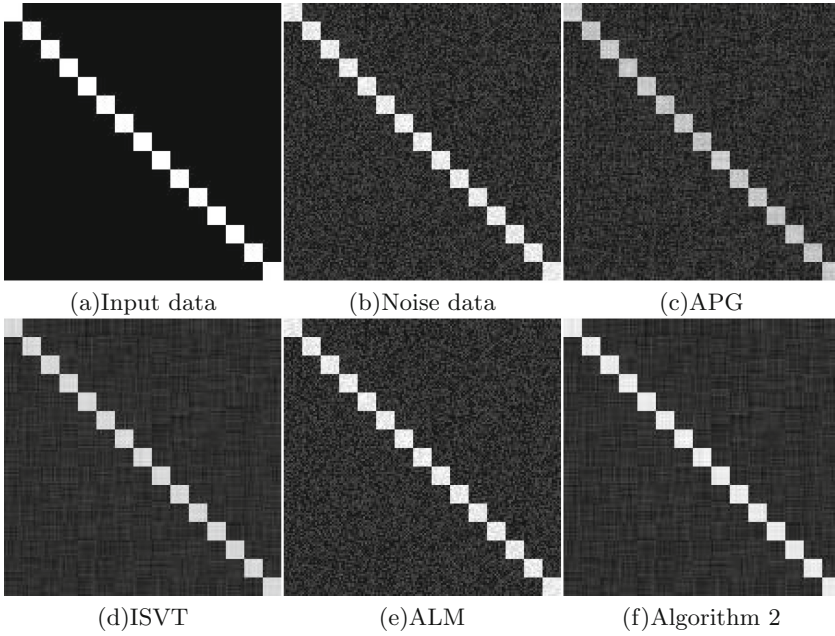


Fig. 2. Experimental results for synthetic data.

Table 1. PSNR, estimated rank and costing time of different algorithms

	APG	ALM	ISVT	Algorithm 2
Building image	32.5175	34.0694	35.7420	37.3545
Fabric image	81	55	27	25

3.2 Real Image Denoising

As general natural images may not have the low rank property, so we choose a certain type of natural images which have periodical textures, such as images of buildings, fabric, and so on (as shown in Figs. 3 and 4). The PSNR values for two test images are shown in Fig. 5.

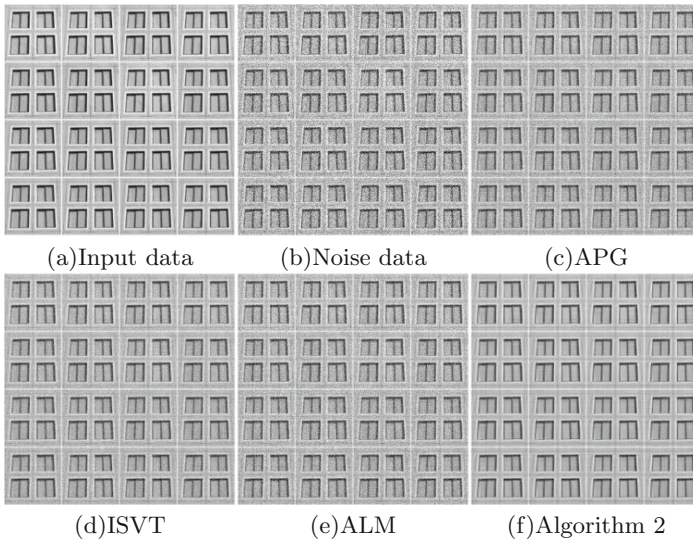


Fig. 3. The image denoising results based on different algorithms.

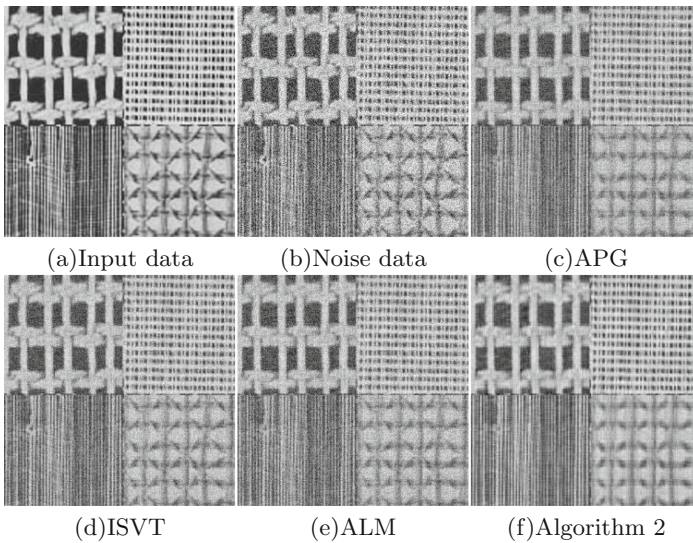


Fig. 4. The image denoising results based on different algorithms.

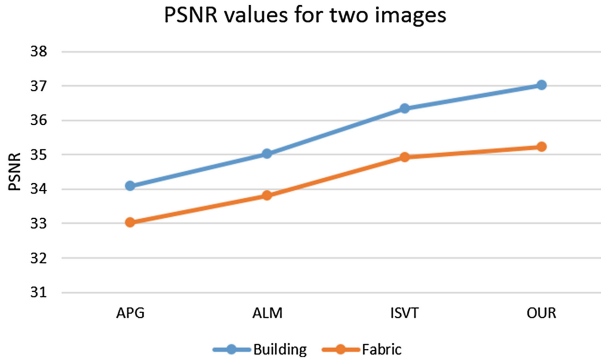


Fig. 5. The PSNR values of different algorithms for building and fabric image

4 Conclusions

In this paper, we propose a novel rank-regularized optimization algorithm. The new method proposed in this paper is a ‘direct’ solver to the rank regularization, it’s a smooth optimization and it just need computing the singular value decomposition one time. An automatic parameter choosing criteria is proposed based on the Morozov discrepancy principle. Finally, the experimental results show that the proposed algorithm can achieve the better estimation of rank and the higher PSNR value while having the smallest time complexity.

References

1. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
2. Rudin, L., Osher, S.: Total variation based image restoration with free local constraints. *IEEE Int. Conf. Image Process.* **1**, 31–35 (1994)
3. Zhang, H., Yi, Z., Peng, X.: fLRR: fast low-rank representation using Frobenius-norm. *Electron. Lett.* **50**(13), 936–938 (2014)
4. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. *ACM Trans. Graph. (TOG)* **31**(6), 139 (2012)
5. Wright, J., Ganesh, A., Rao, S., Ma, Y.: Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization. *Adv. Neural Inf. Process. Syst.* **22**, 2080–2088 (2009)
6. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming (web page and software), June 2009. <http://stanford.edu/boyd/cvx>
7. Chandrasekharan, V., Sanghavi, S., Parillo, P., Wilsky, A.: Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21**(2), 572–596 (2011)
8. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
9. Cai, J.-F., Osher, S., Shen, Z.: Linearized Bregman iterations for compressed sensing. *Math. Comp.* **78**, 1515–1536 (2009)

10. Hale, E.T., Yin, W., Zhang, Y.: Fixed-point continuation for L1-minimization: methodology and convergence. *SIAM J. Optim.* **19**(3), 1107–1130 (2008)
11. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for L1-minimization with applications to compressed sensing. *SIAM J. Imaging Sci.* **1**, 143–168 (2008)
12. Cai, J.-F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20**(4), 1956–1982 (2010)
13. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52**(3), 471–501 (2010)
14. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. *Found. Comput. Math.* **9**(6), 717–772 (2009)
15. Ganesh, A., et al.: Fast algorithms for recovering a corrupted low-rank matrix. In: 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE (2009)
16. Lin, Z., Ganesh, A., Wright, J., Wu, L., Chen, M., Ma, Y.: Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. Technical report, UILU-ENG-09-2214, UIUC (2009)
17. Hu, Y., Wei, Z., Yuan, G.: Inexact accelerated proximal gradient algorithms for matrix $l_{2,1}$ -norm minimization problem in multi-task feature learning. *Stat. Optim. Inf. Comput.* **2**(4), 352–367 (2014)
18. Goldfarb, D., Qin, Z.: Robust low-rank tensor recovery: models and algorithms. *SIAM J. Matrix Anal. Appl.* **35**(1), 225–253 (2014)
19. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. In: *Advances in neural information processing systems*, pp. 612–620 (2011)