

RESEARCH

Open Access



Proximal iteratively reweighted algorithm for low-rank matrix recovery

Chao-Qun Ma and Yi-Shuai Ren* 

*Correspondence:
renyishuai1989@126.com
Business School, Hunan University,
Changsha, 410082, China

Abstract

This paper proposes a proximal iteratively reweighted algorithm to recover a low-rank matrix based on the weighted fixed point method. The weighted singular value thresholding problem gains a closed form solution because of the special properties of nonconvex surrogate functions. Besides, this study also has shown that the proximal iteratively reweighted algorithm lessens the objective function value monotonically, and any limit point is a stationary point theoretically.

Keywords: compressed sensing; matrix rank minimization; reweighted nuclear norm minimization; Schatten- p quasi-norm minimization

1 Introduction

The low-rank matrix recovery problem has been a research hotspot recently [1, 2], and it has a range of applications in many fields such as signal or image processing [3, 4], subspace segmentation [5], collaborative filtering [6], and system identification [7]. Matrix rank minimization under affine equality constraints is generally formulated as follows:

$$\min_X \text{rank}(X) \quad \text{s.t.} \quad A(X) = b, \quad (1.1)$$

where the linear map $A : R^{m \times n} \rightarrow R^P$ and the vector b are known.

Unfortunately, solving the above rank minimization problem (1.1) directly is an NP-hard problem [8], thus this problem is computationally infeasible. Therefore, the convex relations of these methods have been proposed and studied in the literature. For example, Recht *et al.* [8] proposed a nuclear norm minimization method for the matrix reconstruction. The tightest convex relaxation of problem (1.1) is the following nuclear norm minimization problem:

$$\min_X \|X\|_{\star} \quad \text{s.t.} \quad A(X) = b, \quad (1.2)$$

where $\|X\|_{\star} = \sum_{i=1}^r \sigma_i(X)$ is the sum of all the singular values of $X \in R^{m \times n}$ with $\text{rank}(X) = r$ (without loss of generality, $n \leq m$). It has been shown that problem (1.2) shares common solutions with problem (1.1) under some sufficient conditions (see, *e.g.*, [8, 9]).

However, the exact recovery of the low-rank matrix requires more measurements via nuclear norm minimization. Recently, some experimental observations and theoretical

guarantees have shown the superiority of ℓ_p quasi-norm minimization to ℓ_1 minimization in compressive sampling [10]. Therefore, the ℓ_p quasi-norm minimization [11–13] was introduced instead of the nuclear norm minimization in order to give a better approximation to the original problem (1.1). Therefore, the ℓ_p quasi-norm minimization can be formulated as

$$\min_X \|X\|_p^p \quad \text{s.t.} \quad A(X) = b, \tag{1.3}$$

where $\|X\|_p = (\sum_{i=1}^r \sigma_i^p(X))^{1/p}$ for some $p \in (0, 1)$.

However, in practice, the observed data in the low-rank matrix recovery problem may be contaminated with noise, namely $b = AX + e$, where e contains measurement errors dominated by certain normal distribution. In order to recover the low-rank matrix robustly, problem (1.3) can be modified to

$$\min_X \|X\|_p^p \quad \text{s.t.} \quad \|A(X) - b\|_2 \leq \varepsilon, \tag{1.4}$$

where $\|\cdot\|_2$ is the ℓ_2 norm of vector and $\varepsilon \geq \|e\|_2$ is some constant.

Under some conditions, problems (1.3) and (1.4) can be rewritten as the following unconstrained model:

$$\min_X \tau \|X\|_p^p + \frac{1}{2} \|A(X) - b\|_2^2, \tag{1.5}$$

where $\tau > 0$ is a given parameter. Since the above problem (1.5) is nonconvex and NP-hard, thus the researchers throughout the world proposed and analyzed some iterative reweighted algorithms [13–15]. The key idea of the iterative reweighted technique is to solve a convex problem with a given weight at each iteration and update the weight at every turn.

Different from previous studies, based on the weighted fixed point method, this paper puts forward a proximal iteratively reweighted algorithm to recover a low-rank matrix. Due to the special properties of nonconvex surrogate functions, the algorithm iteratively has a closed form solution to solve a weighted singular value thresholding problem. Also, in theory, this study has proved that the proximal iteratively reweighted algorithm decreases the objective function value monotonically, and any limit point is a stationary point.

The remainder of this paper is organized as follows. Section 2 introduces some notations and preliminary lemmas, and Section 3 describes the main results. The conclusion is followed in Section 4.

2 Preliminaries

Recently, Lai *et al.* [13] considered the following unconstrained problem:

$$\min_X F(X) = \tau \operatorname{tr}((X^T X + \varepsilon I)^{p/2}) + \frac{1}{2} \|A(X) - b\|_2^2, \tag{2.1}$$

where I is the $n \times n$ identity matrix and $\varepsilon > 0$ is a smoothing parameter. By the definition in [13], we have

$$\text{tr}((X^T X + \varepsilon I)^{p/2}) = \sum_{i=1}^n (\sigma_i(X)^2 + \varepsilon)^{p/2}. \tag{2.2}$$

Lemma 2.1 ([16]) *Let $\varphi(X) = \psi \circ \sigma(X) = \sum_{i=1}^n (|\sigma_i(X)| + \varepsilon)^p$, where the function $\varphi : \mathbb{R}^{m \times n} \rightarrow [-\infty, +\infty]$ with $n \leq m$ is orthogonally invariant; $\psi : \mathbb{R}^{m \times n} \rightarrow [-\infty, +\infty]$ is an absolutely symmetric function and $p \in (0, 1)$, then $\varphi = \psi \circ \sigma$ is subdifferentiable at matrix $X \in \mathbb{R}^{m \times n}$ and*

$$\partial\varphi(X) = pU \text{Diag} \left\{ \frac{c_i}{(\sigma_i(X) + \varepsilon)^{1-p}} : i \in \Omega \right\} V^T$$

with $X = U\Sigma V^T$ being the SVD of X , and

$$c_i = \begin{cases} 1, & \sigma_i(X) > 0, \\ [-1, 1], & \sigma_i(X) = 0, \end{cases}$$

is a constant depending only on the value of $\sigma_i(X)$ for each $i \in \Omega$.

From Lemma 2.1, let $m = n$ and the matrix Y be a semidefinite matrix, then $Y = Y^T$ and the subdifferentiable of the function

$$\varphi(Y) = \sum_{i=1}^n (|\sigma_i(Y)| + \varepsilon)^{p/2} = \text{tr}((Y + \varepsilon I)^{p/2}) \tag{2.3}$$

is

$$\partial\varphi(Y) = \frac{p}{2} U_1 \text{Diag} \left\{ \frac{c_i}{(\sigma_i(Y) + \varepsilon)^{1-\frac{p}{2}}} : i \in \Omega \right\} U_1^T,$$

with $Y = U_1 \Sigma_1 U_1^T$ being the SVD of Y , and $\Omega = \{1, 2, \dots, n\}$.

From (2.3), it is easier to know exactly that $\varphi(Y)$ is concave, thus $-\varphi(Y)$ is convex. Besides, a vector Y^* is said to be a subgradient of a convex function f at a point Y if $f(z) \geq f(Y) + \langle Y^*, Y - z \rangle$, for any Z . Therefore, based on the definition of subgradient of the convex function, we have

$$-\varphi(Y) \geq -\varphi(Y_k) + \langle -G_k, Y - Y_k \rangle, \tag{2.4}$$

where $-G_k$ is the subgradient of $-\varphi(Y)$ at Y_k , i.e., $-G_k \in \partial(-\varphi(Y_k))$. The inequality of (2.4) is equivalent to

$$\varphi(Y) \leq \varphi(Y_k) + \langle G_k, Y - Y_k \rangle. \tag{2.5}$$

Then $\varphi(Y_k) + \langle G_k, Y - Y_k \rangle$ is used as a surrogate function of $\varphi(Y)$.

3 Main results

Let $Y = X^T X$, then $Y = V \Sigma^2 V^T$ can be obtained, where $X = U \Sigma V^T$ with $U \in \mathbb{R}^{m \times n}$, $V \in \mathbb{R}^{n \times n}$, and $\Sigma = \text{Diag}\{\sigma_i(X)\} \in \mathbb{R}^{n \times n}$, then $\sigma_i(Y) = (\sigma_i(X))^2$. From (2.2), (2.3), and (2.5),

$$\text{tr}((X^T X + \varepsilon I)^{p/2}) \leq \text{tr}((X_k^T X_k + \varepsilon I)^{p/2}) + \langle W_k, X^T X - X_k^T X_k \rangle \tag{3.1}$$

can be obtained, where $W_k \in \frac{\rho}{2} V \text{Diag}\{\frac{c_i}{((\sigma_i(X_k))^2 + \varepsilon)^{1-\frac{\rho}{2}}} : i \in \Omega\} V^T$.

In order to introduce the following lemma, the definitions of Lipschitz continuous of a function and the norm $\|\cdot\|_F$ are given, namely a function is Lipschitz continuous with constant L if, for any x, y , $|f(x) - f(y)| \leq L\|x - y\|$; and the $\|\cdot\|_F$ of a matrix X is defined as $\|X\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$.

Lemma 3.1 ([17]) *Let $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ be a continuously differentiable function with Lipschitz continuous gradient and the Lipschitz constant $L(f)$. Then, for any $L \geq L(f)$,*

$$f(X) \leq f(Y) + \langle \nabla f(Y), X - Y \rangle + \frac{L}{2} \|X - Y\|_F^2, \quad \forall X, Y \in \mathbb{R}^{m \times n}. \tag{3.2}$$

Now let $f(X) = \frac{1}{2} \|A(X) - b\|_2^2$, thus the Lipschitz constant of the gradient $\nabla f(X) = A \star (A(X) - b)$ is $L(f) = \lambda(A \star A)$, where $\lambda(A \star A)$ is the maximum eigenvalue of $A \star A$.

By using (2.1), (2.3), (3.1), and (3.2), we update X_{k+1} by minimizing the sum of these two surrogate functions

$$\begin{aligned} X_{k+1} &= \arg \min \varphi(X_k^T X_k) + \langle W_k, X^T X - X_k^T X_k \rangle + f(X_k) + \langle \nabla f(X_k), X - X_k \rangle \\ &\quad + \frac{L(f)}{2} \|X - X_k\|_F^2 \\ &= \arg \min \tau \langle W_k, X^T X \rangle + \frac{\rho}{2} \left\| X - \left(X_k - \frac{1}{\rho} \nabla f(X_k) \right) \right\|_F^2, \end{aligned} \tag{3.3}$$

where $\rho \geq \frac{L(f)}{2}$.

Lemma 3.2 *If the function $g(X) = \langle Q, X^T X \rangle$ with $X \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{n \times n}$, then the gradient of $g(X)$ is $\nabla g(X) = 2XQ$.*

Proof Consider the auxiliary function $\theta : \mathbb{R} \rightarrow \mathbb{R}$, given by $\theta(t) = g(X + tY)$, for any arbitrary matrix $Y \in \mathbb{R}^{m \times n}$. From the basic calculus, it can be known that $\theta'(0) = \langle \nabla g(X), Y \rangle$. By the definition of the derivative of function, it follows that

$$\begin{aligned} \theta'(0) &= \lim_{t \rightarrow 0} \frac{\theta(t) - \theta(0)}{t} = \lim_{t \rightarrow 0} \frac{g(X + tY) - g(X)}{t} \\ &= \lim_{t \rightarrow 0} \frac{\langle Q, (X + tY)^T (X + tY) \rangle - \langle Q, X^T X \rangle}{t} \\ &= \lim_{t \rightarrow 0} \frac{t \langle Q, X^T Y \rangle + t \langle Q, Y^T X \rangle + t^2 \langle Q, Y^T Y \rangle}{t} \\ &= \langle Q, X^T Y \rangle + \langle Q, Y^T X \rangle = \text{tr}(Q^T X^T Y) + \text{tr}(Y^T X Q) = \langle 2XQ, Y \rangle \end{aligned} \tag{3.4}$$

thus the gradient of $g(X)$ is $\nabla g(X) = 2XQ$. □

Based on the above analysis, this paper proposes the following algorithm.

Algorithm 1 (Proximal iteratively reweighted algorithm to solve problem (2.1))

- 1: *Input:* $\rho \geq \frac{L(f)}{2}$, where $L(f)$ is the Lipschitz constant of $f(x)$.
- 2: *Initialization:* $k = 0, W_k$.
- 3: Update X^{k+1} by solving the following problem:

$$X_{k+1} = \arg \min \tau \langle X^T X, W_k \rangle + \frac{\rho}{2} \left\| X - \left(X_k - \frac{1}{\rho} A \star (A(X_k) - b) \right) \right\|_F^2.$$

- 4: Update the weight W_{k+1} by

$$W_{k+1} \in -\partial(-\varphi(Y_{k+1})), \quad \text{where } Y_{k+1} = X_k^T X_k.$$

- 5: Output low-rank matrix X^k .

Theorem 3.3 *Let $\rho \geq \frac{L(f)}{2}$, where $L(f)$ is the Lipschitz constant of $f(x)$. The sequence $\{X_k\}$ generated in Algorithm 1 satisfies*

(1)

$$F(X_k) - F(X_{k+1}) \geq \left(\rho - \frac{L(f)}{2} \right) \|X_k - X_{k+1}\|_F^2.$$

(2) *The sequence $\{X_k\}$ is bounded.*

(3) $\sum_{k=1}^\infty \|X_k - X_{k+1}\|_F^2 < \frac{2\theta}{2\rho - L(f)}$. *In particular, $\lim_{k \rightarrow \infty} \|X_k - X_{k+1}\|_F = 0$.*

Proof Since X_{k+1} is the globally optimal solution of problem (3.3), and the zero matrix is contained in the subgradient with respect to X . That is, there exists a matrix $X_{k+1} \in$ such that

$$2\tau X_{k+1} W_k + \nabla f(X_k) + \rho(X_{k+1} - X_k) = 0. \tag{3.5}$$

By using the above equality of (3.4) and (3.5), we get

$$2\tau \langle X_{k+1} W_k, X_{k+1} - X_k \rangle + \langle \nabla f(X_k), X_{k+1} - X_k \rangle + \rho \|X_{k+1} - X_k\|_F^2 = 0. \tag{3.6}$$

Since the function $\langle W_k, X^T X \rangle$ is a convex function on X , thus

$$\langle W_k, X_k^T X_k \rangle - \langle W_k, X_{k+1}^T X_{k+1} \rangle \geq 2 \langle X_{k+1} W_k, X_k - X_{k+1} \rangle,$$

and the above equality also can be rewritten as

$$\langle W_k, X_k^T X_k - X_{k+1}^T X_{k+1} \rangle \geq 2 \langle X_{k+1} W_k, X_k - X_{k+1} \rangle. \tag{3.7}$$

Then it follows from (3.6) and (3.7) that

$$\tau \langle W_k, X_k^T X_k - X_{k+1}^T X_{k+1} \rangle \geq -\langle \nabla f(X_k), X_k - X_{k+1} \rangle + \rho \|X_{k+1} - X_k\|_F^2. \tag{3.8}$$

Let $f(X) = \frac{1}{2} \|A(X) - b\|_2^2$, and according to Lemma 3.1,

$$f(X_k) - f(X_{k+1}) \geq \langle \nabla f(X_k), X_k - X_{k+1} \rangle - \frac{L(f)}{2} \|X_k - X_{k+1}\|_F^2 \tag{3.9}$$

can be obtained. Since the function $\text{tr}((X^T X + \varepsilon I)^{p/2})$ is concave, and just like (3.1), then it can be obtained

$$\text{tr}((X_k^T X_k + \varepsilon I)^{p/2}) - \text{tr}((X_{k+1}^T X_{k+1} + \varepsilon I)^{p/2}) \geq \langle W_k, X_k^T X_k - X_{k+1}^T X_{k+1} \rangle. \tag{3.10}$$

Now, combining (3.8), (3.9), and (3.10), we get

$$\begin{aligned} F(X_k) - F(X_{k+1}) &= \tau \text{tr}((X_k^T X_k + \varepsilon I)^{p/2}) + f(X_k) - \tau \text{tr}((X_{k+1}^T X_{k+1} + \varepsilon I)^{p/2}) - f(X_{k+1}) \\ &\geq \left(\rho - \frac{L(f)}{2} \right) \|X_{k+1} - X_k\|_F^2 \geq 0. \end{aligned}$$

Thus, $F(X_k)$ is monotonically decreasing. Given the facts of all inequalities above for $k \geq 1$, it can be obtained

$$F(X_1) - F(X_{k+1}) \geq \left(\rho - \frac{L(f)}{2} \right) \sum_{i=1}^k \|X_{i+1} - X_i\|_F^2, \tag{3.11}$$

and from (3.11) it follows that

$$\left(\rho - \frac{L(f)}{2} \right) \sum_{i=1}^k \|X_{i+1} - X_i\|_F^2 \leq F(X_1) < +\infty. \tag{3.12}$$

Then, for $k \rightarrow \infty$, (3.12) implies that

$$\lim_{k \rightarrow \infty} \|X_{k+1} - X_k\|_F = 0.$$

Since the objective function $F(X)$ in problem (2.1) is nonnegative and satisfies

$$F(X) \rightarrow \infty, \quad \text{as } \|X\|_F \rightarrow \infty,$$

then $X_k \in \{X : 0 \leq F(X) \leq F(X_1)\}$ and the sequence $\{X_k\}$ is bounded.

Therefore, the proof has been completed. □

Theorem 3.4 *Let $\{X_k\}$ be the sequence generated in Algorithm 1. Then any accumulation point of $\{X_k\}$ is a stationary point X^\star of the problem. Moreover, for $k = 1, 2, \dots, N$, there always exists*

$$\min_{1 \leq k \leq N} \|X_{k+1} - X_k\|_F^2 \leq \frac{F(X_1) - F(X^\star)}{n(\rho - \frac{L(f)}{2})}.$$

Proof Since the sequence $\{X_k\}$ generated in Algorithm 1 is bounded, there exist an accumulation point X^\star and a subsequence $\{X_{k_j}\}$ such that $\lim_{j \rightarrow \infty} X_{k_j} = X^\star$. Assume that X_{k_j}

is the solution of problem (3.3), it can be obtained

$$2\tau X_{kj+1} W_{kj} + \nabla f(X_{kj}) + \rho(X_{kj+1} - X_{kj}) = 0.$$

Let $j \rightarrow \infty$, according to Theorem 3.3, $\lim_{j \rightarrow \infty} \|X_{kj+1} - X_{kj}\|_F = 0$ can be obtained. Hence, there exists the matrix

$$W^\star = \frac{\rho}{2} V_2 \text{Diag} \left\{ \frac{1}{((\sigma_i(X^\star))^2 + \varepsilon)^{1-\frac{\rho}{2}}} \right\} V_2^T = \frac{\rho}{2} ((X^\star)^T X^\star + \varepsilon I)^{\rho/2-1},$$

where $X^\star = U_2 \Sigma_2 V_2^T$ with $U_2 \in \mathbb{R}^{m \times n}$, $V_2 \in \mathbb{R}^{n \times n}$, and $\Sigma = \text{Diag} \left\{ \frac{1}{((\sigma_i(X^\star))^2 + \varepsilon)^{1-\frac{\rho}{2}}} \right\}$.

By the above analysis, it can be known that

$$\tau \rho X^\star ((X^\star)^T X^\star + \varepsilon I)^{\rho/2-1} + \nabla f(X^\star) = 0,$$

then X^\star is a stationary point of problem (2.1).

Moreover, by using (3.11), for $k = 1, 2, \dots, N$, it can be obtained

$$\begin{aligned} F(X_1) - F(X_{N+1}) &\geq \left(\rho - \frac{L(f)}{2} \right) \sum_{k=1}^N \|X_{k+1} - X_k\|_F^2 \\ &\geq N \left(\rho - \frac{L(f)}{2} \right) \min_{1 \leq k \leq N} \|X_{k+1} - X_k\|_F^2. \end{aligned}$$

Thus

$$\begin{aligned} \min_{1 \leq k \leq N} \|X_{k+1} - X_k\|_F^2 &\leq \frac{F(X_1) - F(X_{N+1})}{n(\rho - \frac{L(f)}{2})} \\ &\leq \frac{F(X_1) - F(X^\star)}{n(\rho - \frac{L(f)}{2})} \end{aligned}$$

can be obtained, which completes the proof. □

4 Conclusion

A proximal iteratively reweighted algorithm based on the weighted fixed point method for recovering a low-rank matrix problem has been presented in this paper. Due to the special properties of the nonconvex surrogate function, the algorithm in this study iteratively has a closed form solution to solving a weighted singular value thresholding problem. Finally, it has been proved that the algorithm can decrease the objective function value monotonically and any limit point is a stationary point.

Acknowledgements

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (No. 71431008).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

The authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 November 2017 Accepted: 28 December 2017 Published online: 08 January 2018

References

1. Lu, C, Lin, Z, Yan, S: Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization. *IEEE Trans. Image Process.* **24**, 646-654 (2015)
2. Fornasier, M, Rauhut, H, Ward, R: Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. Optim.* **24**, 1614-1640 (2011)
3. Akcay, H: Subspace-based spectrum estimation infrequency-domain by regularized nuclear norm minimization. *Signal Process.* **99**, 69-85 (2014)
4. Takahashi, T, Konishi, K, Furukawa, T: Rank minimization approach to image inpainting using null space based alternating optimization. In: *Proceedings of IEEE International Conference on Image Processing*, pp. 1717-1720 (2012)
5. Liu, G, Lin, Z, Yu, Y: Robust subspace segmentation by low-rank representation. In: *Proceedings of the 27th International Conference on Machine Learning*, pp. 663-670 (2010)
6. Candes, EJ, Recht, B: Exact matrix completion via convex optimization. *Found. Comput. Math.* **9**(6), 717-772 (2009)
7. Liu, Z, Vandenberghe, L: Interior-point method for nuclear norm approximation with application to system identification. *SIAM J. Matrix Anal. Appl.* **31**(3), 1235-1256 (2009)
8. Recht, B, Fazel, M, Parrilo, P: Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. *SIAM Rev.* **52**, 471-501 (2010)
9. Recht, B, Xu, W, Hassibi, B: Null space conditions and thresholds for rank minimization. *Math. Program.* **127**, 175-202 (2011)
10. Baraniuk, R: Compressive sensing. *IEEE Signal Process. Mag.* **24**(4), 118-124 (2007)
11. Zhang, M, Haihuang, Z, Zhang, Y: Restricted p -isometry properties of nonconvex matrix recovery. *IEEE Trans. Inf. Theory* **59**(7), 4316-4323 (2013)
12. Liu, L, Huang, W, Chen, DR: Exact minimum rank approximation via Schatten p -norm minimization. *J. Comput. Appl. Math.* **267**, 218-227 (2014)
13. Lai, M, Xu, Y, Yin, W: Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM J. Numer. Anal.* **51**(2), 927-957 (2014)
14. Konishi, K, Uruma, K, Takahashi, T, Furukawa, T: Iterative partial matrix shrinkage algorithm for matrix rank minimization. *Signal Process.* **100**, 124-131 (2014)
15. Mohan, K, Fazel, M: Iterative reweighted algorithms for matrix rank minimization. *J. Mach. Learn. Res.* **13**(1), 3441-3473 (2012)
16. Li, YF, Zhang, YJZ, Huang, ZH: A reweighted nuclear norm minimization algorithm for low rank matrix recovery. *J. Comput. Appl. Math.* **263**, 338-350 (2014)
17. Toh, K-C, Yun, S: An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pac. J. Optim.* **6**, 615-640 (2010)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
