A PENALTY DECOMPOSITION METHOD FOR NUCLEAR
NORM MINIMIZATION WITH $l_1$ NORM FIDELITY TERM

DUO WANG
School of Mathematics and Statistics
Henan University of Science and Technology, Luoyang 471023, China

ZHENG-FEN JIN*
School of Mathematics and Statistics
Henan University of Science and Technology, Luoyang 471023, China
Henan Joint International Research Laboratory of Cyberspace Security Applications
Luoyang 471023, China

YOU LIN SHANG
School of Mathematics and Statistics,
Henan University of Science and Technology, Luoyang 471023, China

(Communicated by Hedy Attouch)

Abstract. It is well known that the nuclear norm minimization problems are
widely used in numerous fields such as machine learning, system recognition,
and image processing, etc. It has captured considerable attention and taken
good progress. Many researchers have made great contributions to the nu-
clear norm minimization problem with $l_2$ norm fidelity term, which is just
able to deal with those problems with Gaussian noise. In this paper, we pro-
pose an efficient penalty decomposition (PD) method to solve the nuclear norm
minimization problem with $l_1$ norm fidelity term. One subproblem admits a
closed-form solution due to its special structure, and another can also get a
closed-form solution by linearizing its quadratic term. The convergence re-
sults of the proposed method are given as well. Finally, the effectiveness and
efficiency of the proposed method are verified by some numerical experiments.

1. Introduction. The nuclear norm minimization problem with linear equality
constraints can be formulated as

$$
\min_{X \in \mathbb{R}^{m \times n}} \|X\|_*, \quad s.t. \quad \mathcal{A}(X) = b,
$$

where $X \in \mathbb{R}^{m \times n}$ is an unknown matrix, $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is a linear map and

$b \in \mathbb{R}^p$. In (1), $\|X\|_*$ is the nuclear norm of matrix $X$ [19], which is defined as

2000 Mathematics Subject Classification. Primary: 90C06, 90C25, 90C30.
Key words and phrases. Nuclear norm minimization, penalty decomposition method, $l_1$ norm
fidelity term, matrix completion.

This work of Y. Shang is supported by the National Natural Science Foundation of
China(No.11471102). This work of Z.-F Jin is supported by the National Natural Science
Foundation of China(No.61370220) and the Plan for Scientific Innovation Talent of Henan
Province(No.174200510011). This work of Duo Wang is supported by the National Natural Science
Foundation of China(No.11701150).

* Corresponding author: jinzhengfen@whu.edu.cn.

695
the sum of its positive singular values. Assume that the matrix $X$ has $r$ positive singular values of $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, then its nuclear norm is formulated as

$$\|X\|_* = \sum_{i=1}^{r} \sigma_i.$$  

In some sense, (1) is the tightest convex relaxation of the matrix rank minimization problem [19]

$$\min_{X \in \mathbb{R}^{m \times n}} \text{rank}(X), \quad \text{s.t.} \quad A(X) = b. \quad (2)$$  

The nuclear norm minimization problem (1) is convex, which is equivalent to a semidefinite programming (SDP) problem, and admits a lot of effective algorithms such as SeDuMi [25] and SDPT3 [28]. Unfortunately, both solvers are not valid for solving a higher dimensional problem. Subsequently, Cai et al presented a singular value thresholding (SVT) [8] algorithm overcoming the aboved disadvantage to solve the following problem

$$\min_{X \in \mathbb{R}^{m \times n}} \tau \|X\|_* + \frac{1}{2} \|X\|_F^2, \quad \text{s.t.} \quad A(X) = b, \quad (3)$$

where $\tau \geq 0$ is a given parameter, and $\|\cdot\|_F$ is the Frobenius norm. Obviously, the solution to (3) converges to the one of (1) as $\tau \to \infty$. Nevertheless, when the matrix is not very low rank, the efficiency is not remarkable. Afterwards, a fixed point continuation with approximate SVD (FPCA) [19] method of Ma et al is applied to solve

$$\min_{X \in \mathbb{R}^{m \times n}} \tau \|X\|_* + \frac{1}{2} \|A(X) - b\|_2^2. \quad (4)$$

Moreover, Toh and Yun introduced an accelerated proximal gradient (APG) [26] method for solving the Lagrangian version of (4). More recently, Yang and Yuan developed an alternating direction method (ADM) [31], where the closed-form solutions of some subproblems can be easily obtained by using a linearized technique. The other ADMM type algorithms can be seen in [29], [10], [17], etc.

In many applications, the measurement $b$ in (1) may be contaminated by a small amount of noise. In consideration of noise, quite a few researchers have presented two typical noisy models with $l_2$ norm fidelity term, which can deal with the Gaussian noisy problem effectively. One is the inequality constrained nuclear norm minimization problem

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* \quad \text{s.t.} \quad \|A(X) - b\|_2 \leq \delta, \quad (5)$$

where $\delta \geq 0$ is a parameter reflecting the level of noise. Obviously, model (5) reduces to (1) when $\delta = 0$.

The other one is the nuclear norm regularized least square problem

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* + \frac{\gamma}{2} \|A(X) - b\|_2^2, \quad (6)$$

where $\gamma \geq 0$ is a regularization parameter. When the $\gamma$ and $\delta$ are chosen appropriately, the model (5) is equivalent to (6).

Unfortunately, it was demonstrated that $l_2$ norm fidelity term is less valid for non-Gaussian additive noise because it may amplify the influence of noise [21]. As we know, the recovering low rank matrix problems are widely used in the domains of system identification [19], [22], [12], machine learning [14], [13], [30] and video denoising [15]. In addition, the impulsive noise, Gaussian noise and their mixture widely exist in the corrupted images and videos [1], [20]. Therefore, it is necessary
to find a substitutable formulation to overcome the disadvantage of the $l_2$ norm fidelity term. The literatures [2], [6], [9], [23] have demonstrated that the $l_1$ norm fidelity term is particularly efficient for solving non-Gaussian additive noise such as impulsive noise. Thus, in this paper, we consider the nuclear norm minimization problem with $l_1$ norm fidelity term, namely

$$
\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* + \frac{1}{\nu} \|A(X) - b\|_1,
$$

where the observation $b \in \mathbb{R}^p$ might contain some noise and $\nu > 0$ is a penalty parameter. It is well known that when the penalty parameter $\nu$ is less than a certain threshold, the $l_1$ fidelity term makes problem (7) into an exact penalty function problem and reduces it to problem (1). The model (7) can solve the problems not only with impulsive noise but also Gaussian noise or their mixture efficiently. Given the two non-smooth terms in the objective function, we transform the convex and non-smooth objective function into variable separated form by introducing an auxiliary variable. Here we introduce an auxiliary variable $r$, the problem (7) is converted equivalently into

$$
\min_{X,r} \|X\|_* + \frac{1}{\nu} \|r\|_1, \quad s.t. \quad A(X) - b = r, X \in \mathbb{R}^{m \times n}, r \in \mathbb{R}^p.
$$

Then we construct an efficient algorithm to solve the problem (8) according to its separable structure.

Recently, Lu and Zhang [18] introduced a penalty decomposition (PD) method for solving the general rank minimization problem and tested its performance by the matrix completion and nearest low-rank correlation matrix problems. Inspired by the well performance of their method, in this paper, we extend it to solve the problem (8) and compare it with the well known algorithm FPCA. Since both of the subproblems admit closed-form solutions and the block coordinate descent (BCD) method or the Gauss-Seidel method performs well if each subproblem can be solved efficiently [18], we apply the BCD method to minimize each penalty subproblem.

Precisely the convergence of the BCD method was studied in [18], [4], [5], [27]. If the objective function $f$ is strictly convex differentiable, then the sequence generated by this method converges to a stationary point of $f$ [5]. More recently, in [4], the global sublinear rate of convergence of the block coordinate gradient projection algorithm for smooth convex programming problems has been established. In nonconvex and nonsmooth setting, Attouch et al. [3] studied the convergence properties of the proximal-forward-backward (PFB) applied for minimizing a nonconvex and nonconvex function $f$. Moreover, based on the KL property of $f$, Bolte et al. [7] established the global convergence results for a proximal alternating linearized minimization (PALM) algorithm for nonconvex and nonsmooth minimization problems. The problem (8) is convex and nonsmooth, so the convergence of our proposed algorithm can be easily obtained according to the literature [7]. Finally, the efficiency of the proposed algorithm is demonstrated by some numerical experiments.

The paper proceeds as follow. In Section 2, we will apply the PD for solving the equality constrained problem (8) and give the convergence result. The numerical results are presented in Section 3. Finally, we conclude the paper with some remarks in Section 4.

2. Algorithm. In this section, we will construct our PD method for solving the problem (8) and give the convergence result of the proposed method.
2.1. Penalty decomposition method for problem (8). In this subsection, we mainly focus on solving the problem (8) via PD method. Given a parameter $\mu > 0$, the corresponding quadratic penalty function of (8) is

$$P_\mu(X, r) = \|X\|_* + \frac{1}{2\nu} \|r\|_1 + \frac{\mu}{2} \|A(X) - b - r\|_2^2,$$

(9)

where $\mu$ is a penalty parameter. Because of the strong convexity, the solution of penalty problem (9) is unique. In addition, as $\mu \to \infty$, the solution of problem (9) converges to the one of (8). Now we apply the block coordinate decent (BCD) method to minimize the penalty problem (9), i.e.,

$$\min_{X, r} \{P_\mu(X, r) : X \in \mathbb{R}^{m \times n}, r \in \mathbb{R}^p\}.$$

(10)

For simplicity, we use $l$ and $k$ represent the inner iteration index and the outer iteration index, respectively.

Given $\{(X^k_l, r^k_l)\}$, the main steps are to solve the both penalty subproblems, which can be described as follows:

$$X^k_{l+1} = \arg\min_X P_\mu(X, r^k_l), \, X \in \mathbb{R}^{m \times n},$$

(11)

$$r^k_{l+1} = \arg\min_r P_\mu(X^k_{l+1}, r), \, r \in \mathbb{R}^p.$$

(12)

Firstly, for fixed $\{(X^k_l, r^k_l)\}$, the minimizer $X^k_{l+1}$ of (9) with respect to $X$ is given by

$$X^k_{l+1} = \arg\min_X \|X\|_* + \frac{\mu}{2} \|A(X) - b - r^k_l\|_2^2, \, X \in \mathbb{R}^{m \times n}.$$

(13)

Obviously, its analytic solution is not clear due to the linear mapping $A$. Thus we linearize the last quadratic term to get a closed-form solution of the above subproblem. Let $G(X) = \frac{1}{2} \|A(X) - b - r^k_l\|_2^2$, it is clear to see that $G(X)$ is a quadratic function. Then let $g^k$ be the gradient of $G(X)$ at $X^k_l$, which can be represented as

$$g^k = A^*(A(X^k_l) - b - r^k_l),$$

then we can obtain that

$$G(X) \approx G(X^k_l) + \langle g^k, X - X^k_l \rangle + \frac{1}{2\tau} \|X - X^k_l\|_F^2 = G(X^k_l) + \frac{1}{2\tau} \|X - X^k_l + \tau g^k\|_F^2 - \frac{\tau}{2} (g^k)^2.$$

(14)

Hence, the X-subproblem (11) can be approximately solved by the following model

$$X^k_{l+1} = \arg\min_X \{\|X\|_* + \frac{\mu}{2\tau} \|X - (X^k_l - \tau g^k)\|_F^2\}, \, X \in \mathbb{R}^{m \times n}.$$

(15)

To get the exact solution of (15), let $Y = X^k_l - \tau g^k$, then the solution of X-subproblem can be reformulated as

$$X^k_{l+1} = \arg\min_X \{\tau \|X\|_* + \frac{\mu}{2} \|X - Y\|_F^2\}, \, X \in \mathbb{R}^{m \times n}.$$

(16)

According to the theorem in [8], [19], we need to compute the singular value decomposition (SVD) of matrix $Y \in \mathbb{R}^{m \times n}$ of rank $r$ firstly, i.e.,

$$Y = U\Sigma V^T, \Sigma = \text{diag}\{\sigma_i\}_{1 \leq i \leq r},$$

where $U$ and $V$ are unitary matrices, and $\sigma_i$ are sorted eigenvalues of $Y$.
where $U$ and $V$ are $m \times r$ and $r \times n$ matrices respectively with orthonormal columns, and the $\sigma_i$ is the $i$–th positive singular value of matrix $Y$. Then the $X^k_{t+1}$ can be written directly as

$$X^k_{t+1} = D_{r/\mu}(Y),$$

(17)

where $D_{r/\mu}(Y) = U(0 - \tau/\mu I)^{+} V^T$, and $(t)^{+} := \max\{t, 0\}$.

Secondly, for fixed $X^k_{t+1}$, we can obtain $r^k_{t+1}$ by minimizing (9) with respect to $r$, that is

$$r^{k+1} = \arg\min_{r \in \mathbb{R}^p}\{\|X^k_{t+1}\|_1 + \frac{1}{\nu} \|r\|_1 + \frac{\mu}{2} \|A(X^k_{t+1}) - b - r\|_2^2\}$$

$$= \arg\min_{r \in \mathbb{R}^p}\{\frac{1}{\nu} \|r\|_1 + \frac{\mu}{2} \|A(X^k_{t+1}) - b - r\|_2^2\}$$

$$= \arg\min_{r \in \mathbb{R}^p}\{\frac{1}{\nu\mu} \|r\|_1 + \frac{1}{2} \|r - (A(X^k_{t+1}) - b)\|_2^2\}$$

$$= \text{Shrink}(A(X^k_{t+1}) - b, \frac{1}{\nu\mu}),$$

(18)

where $\text{Shrink}(.,.)$ is the well known vector shrinkage [19].

If $(X^k_{t+1},r^k_{t+1})$ satisfies $\|A(X^k_{t+1}) - b - r^k_{t+1}\|_2 \leq \epsilon$, then let $(X^k, Y^k) = (X^k_{t+1}, r^k_{t+1})$, and start the outer loops.

In light of all the analysis above, we give the following framework of the PD method for solving problem (8). Here we note it as $PD - l_1$.

**Initialization:** Input $X^0$, $r^0$. Choose an accuracy parameter $\epsilon \geq 0$. Given constants $\mu_0 > 0$, $\tau > 0$, $\nu > 0$, $\sigma > 1$. Set $k = 0$.

**Step 1.** Set $l = 0$ and find an approximate solution $(X^k, r^k)$ by the following steps (1)-(4):

1. Compute $X^k_{t+1}$ via (17) with fixed $X^k$ and $r^k$.
2. Compute $r^{k+1}_{t+1}$ via (18) with fixed $X^k_{t+1}$.
3. Set $(X^{k+1}, r^{k+1}) := (X^k_{t+1}, r^{k+1}_{t+1})$. If $(X^k, r^k)$ satisfies $\|A(X^k) - b - r^k\|_2 \leq \epsilon$, then go to **Step 2**.
4. set $l := l + 1$ and go to step (1).

**Step 2.** Set $\mu_{k+1} := \sigma \mu_k$.

**Step 3.** Set $k := k + 1$ and go to **Step 1**.

**Remark 1.** The above algorithm is inspired by the PD method in literature [18]. Differently, we extend the PD method to solve the model (8), which can solve the problems not only with impulsive noise but also Gaussian noise or their mixture efficiently. One subproblem admits a closed-form solution due to its special structure, and another can also get a closed-form solution by linearizing its quadratic term.

**Remark 2.** For the above algorithm, the terminated condition can also be the relative error between the original matrix $M$ and the optimal solution $X^*$ produced by the $PD - l_1$, that is

$$\text{RelErr} = \frac{\|X^* - M\|_F}{\|M\|_F} \leq \epsilon$$

for some $\epsilon > 0$. 

2.2. Convergence analysis. In this subsection, we firstly give the convergence of the inner iterations of $PD - l_1$. Since the quadratic penalty function of $(8)$ is convex and nonsmooth, it can be seen as a special case of the Problem $(M)$ [7] with $f(X) = \|X\|_*, g(r) = \frac{1}{p}\|r\|_1$, $H(X, r) = \frac{2}{p}\|A(X) - b - r\|_2^2$. Obviously, the objective function satisfies the KL property. Moreover, both functions of $X \rightarrow H(X, r)$ (for fixed $r$) and $r \rightarrow H(X, r)$ (for fixed $X$) are $C^{1, 1}$. Hence, from Theorem 1 in [7], we give the following convergence theorem for inner iterations of $PD - l_1$ without proof.

**Theorem 2.1.** Let the sequence $(X_l, r_l)$ be generated by the above BCD method which is assumed to be bounded. Then,

(i) The Sequence $(X_l, r_l)$ has finite length, that is

$$\sum_{l=1}^{\infty} \|X_{l+1} - X_l\|_F + \|r_{l+1} - r_l\|_F < \infty.$$ 

(ii) The Sequence $(X_l, r_l)$ converges to an optimal solution $(X^*, r^*)$ of the problem $(8)$.

Next, by the theory of the penalty function method [11], [24] and convergence results for the PD method in [18], we describe the convergence property of the proposed algorithm. The following convergence result shows that the proposed method can solve the problem (7) effectively.

**Theorem 2.2.** Let $\epsilon = 0$ in $PD - l_1$. Suppose that the solution to the equality constrained problem $(8)$ exists and let the sequence $(X^k, r^k)$ be generated by the above PD method. Then, for any accumulation point $(X^*, r^*)$ of the sequence $\{(X^k, r^k)\}$, $(X^*, r^*)$ is an optimal solution of $(8)$.

**Proof.** Let $(\bar{X}, \bar{r})$ be an optimal solution of problem $(8)$, that is

$$\|\bar{X}\|_* + \frac{1}{\nu}\|\bar{r}\|_1 \leq \|X\|_* + \frac{1}{\nu}\|r\|_1, \text{ for all } (X, r) \text{ with } A(X) - b = r.$$ 

Since $\{(X^k, r^k)\}$ minimizes $P_{\mu_k}(X, r)$ for each $k$, we have that

$$P_{\mu_k}(X^k, r^k; \mu_k) \leq P_{\mu_k}(\bar{X}, \bar{r}; \mu_k),$$

which leads to the inequality

$$\|X^k\|_* + \frac{1}{\nu}\|r^k\|_1 + \frac{\mu_k}{2}\|A(X^k) - b - r^k\|_2^2 \leq \|\bar{X}\|_* + \frac{1}{\nu}\|\bar{r}\|_1 + \frac{\mu_k}{2}\|A(\bar{X}) - b - \bar{r}\|_2^2$$

$$= \|\bar{X}\|_* + \frac{1}{\nu}\|\bar{r}\|_1.$$  \hspace{1cm} (19)

By rearranging this expression, we obtain

$$\|A(X^k) - b - r^k\|_2^2 \leq \frac{2}{\mu_k}(\|\bar{X}\|_* + \frac{1}{\nu}\|\bar{r}\|_1 - \|X^k\|_* - \frac{1}{\nu}\|r^k\|_1).$$  \hspace{1cm} (20)

Suppose that $(X^*, r^*)$ is an accumulation point of $\{(X^k, r^k)\}$, so that there is an infinite subsequence $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}, k \to \infty} (X^k, r^k) = (X^*, r^*).$$
3. **Numerical experiments.** In this section, we carry out some numerical experiments to illustrate that the model (7) with \( l_1 \)-fidelity can solve different noise cases and perform better than the models with \( l_2 \)-fidelity. Firstly, we test the PD – \( l_1 \) method against the state-of-art algorithm FPCA for solving a series of matrix completion problems with different noise cases. Then, we test a gray image to graphically illustrate the practical performance of PD – \( l_1 \). Finally, we apply the proposed PD – \( l_1 \) to solve the problem (7) with different experimental settings.

All experiments are performed under Windows 7 premium and MATLAB 2016a running on a Lenovo laptop with an Intel core CPU at 2.5 GHz and 4 GB memory.

In all performance assessments, we first generate a real low rank matrix \( M = M_L M_R^T \), where matrix \( M_L \in \mathbb{R}^{m \times r} \) and \( M_R \in \mathbb{R}^{n \times r} \) are generated with independently identically distributed Gaussian entries. We produce \( M \) via the Matlab script "\( \text{randn}(m,r) \times \text{randn}(r,n) \)". Obviously, the rank of \( M \) is \( r \). We use \( sr = p/mn \) to denote the sampling ratio, where \( p \) is the number of measurements. Additionally, the number of degree of freedom for a real-valued rank \( r \) matrix is defined by \( dr = r(m + n - r) \). In all tests, the quadratic penalty parameter \( \mu \) is chosen as \( \mu = 2.5/\min(m,n) \) by experience [31]. Furthermore, we set the approximate parameter \( \tau = 1.0 \) for matrix completion problems. In the noisy cases, the measurement \( b \) is produced by \( b = A(X) + \omega_G + \omega_I \), where \( \omega_G \) is the additive Gaussian noise of zero mean and standard deviation \( \sigma \), which can be generated by \( \omega_G = \sigma \times \text{randn}(p, 1) \), and \( \omega_I \) is the impulsive noise, which can be set to \( \pm 1 \) at \( N \) random positions of \( b \), where \( N = m \times n \times q \), and \( q = 0.01, 0.001, 0.005 \) can be chosen respectively. For each test, we use \( X^* \) to represent the optimal solution produced by our algorithm, and use relative error (RelErr) to measure the quality of \( X^* \) for original matrix \( M \), i.e.,

\[
\text{RelErr} = \frac{\|X^* - M\|_F}{\|M\|_F},
\]

We say that \( M \) is recovered successfully by \( X^* \) if the corresponding RelErr is less than \( 10^{-3} \), which has been used in [19], [8], [16].

3.1. **Test on matrix completion problems.** In this subsection we present some numerical results to illustrate the feasibility and efficiency of the proposed algorithm.
For our purpose, we test our proposed method by using the nuclear norm matrix completion problem with $l_1$-fidelity, which can be formulated as

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* + \frac{1}{\nu} \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}|.$$  \hfill (22)

In order to assess the performance of $PD - l_1$, we test it against the state-of-art algorithm FPCA used to solve the nuclear norm matrix completion problem with $l_2$-fidelity:

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* + \frac{\mu}{2} \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}|^2.$$  \hfill (23)

Firstly, we test the two problems of (22) and (23) with three noisy cases: impulsive noise only, Gaussian and impulsive noise, and Gaussian noise only. We set the parameter $\text{maxit} = 150$ to observe their performance more clearly. The test results are displayed in Figure 1. Observing the Figure 1, it is clear to see that the model (22) with $l_1$-fidelity can be dramatically better than (23) whenever the observation data may contain Gaussian noise, impulsive noise or their mixtures. Particularly, it is clear that $PD - l_1$ can recover the matrix successfully via solving the model (22), but the model (23) is inefficient for the data with impulsive noise. Moreover, the accuracy of solution for model (22) can attain $10^{-4}$ within 150 iterations, which is a little higher than model (23).

Secondly, we test the performance of PD method for the model (22) and (23) with different $sr$ and $r$. Numerical results can be seen in the Figure 2. In the first row of Figure 2, we set $m = n = 500$ and $r = 10$ with different $sr$. It can be seen that both models work successfully for the matrix completion problem with noiseless. Moreover, it is clear that the $PD - l_1$ for model (22) converged faster than the model (23) except for the case $sr = 0.3$ and can get an approximate exact solution, which is less than $10^{-10}$. Observing the second row of Figure 2, we can see that both models perform well with $r$ increasing. The model (22) can get a higher accuracy as $10^{-10}$ but the model (23) is not able to attain a high accuracy, which just attains a relatively middle accuracy. Moreover, from the Iter point of view, model (22) clearly saves much number of iterations. Taking everything together, it can be concluded that the model (22) is a better choice if a higher accuracy of solution is needed.

To end this subsection, we test the $PD - l_1$ method for recovering a real gray image to illustrate its superiority. Numerical results can be seen in the Figure 3. In this experiment, we test a real $512 \times 512$ gray image (a), which has been widely used in many simulations. Firstly, we get a low rank image (b) by applying matrix SVD. Then we randomly select 70% samples from the low rank image and add some different kinds of noise to obtain the corrupted images (c) and (e). Finally, we get the recovered images (d) and (f) via solving the matrix completion model (22) by using the proposed $PD - l_1$. Furthermore, the relative error is $\text{RelErr} = 4.760e - 3$, $4.827e - 3$ respectively. Comparing (d) with (b), it is clear to see that the proposed $PD - l_1$ is efficient for recovering the corrupted images with impulsive noise. In addition, (f) shows that the $PD - l_1$ recovers the corrupted images with Gaussian noise and impulsive noise successfully. This test illustrates that the proposed $PD - l_1$ performs well on real image restoration.
3.2. Nuclear norm minimization with $l_1$-fidelity term. In this subsection, we report the efficiency of the $PD - l_1$ for solving the problem (7) with different experimental settings. Firstly, we test $PD - l_1$ method for solving problem (7) with impulsive noise. In this test, we choose $m = n$ from 100 to 1000 and set $sr = 0.5$, $r = 5$, $q = 0.01$. The test results are given in Table 1. From Table 1, we can see that the accuracy of the solution is $10^{-4}$. In other words, the problem is solved successfully. The limited numerical results demonstrate that the proposed $PD - l_1$ algorithm can effectively solve the nuclear norm minimization problem with impulsive noise.

Secondly, we test the performance of $PD - l_1$ method for solving problem (7) with different $sr$. The numerical results can be seen in Figure 4. From the Figure 4, we can see that as the higher the sampling ratio is, the higher accuracy can be obtained.

Finally, in order to further reveal the efficiency of the $PD - l_1$ method, we test $PD - l_1$ to solve the problem (7) with different noises. The test results are given in Table 2. From Table 2, we can see that all experimental examples are solved...
Figure 2. Numerical comparisons of model (22) and (23) with noiseless (m=n=500). First row: \( r = 10 \), \( sr \) is 0.3, 0.5, 0.7 respectively. Second row: \( sr = 0.5 \), \( r \) is 5, 15, 30 respectively.

successfully, especially for the impulsive noise and mixture noise. So base on the above analysis, we can conclude that the \( PD - l_1 \) method is efficient and promising for the problem (7).

Table 1. Numerical results of \( PD - l_1 \) for (7), m=n, sr=0.5, q=0.01.

| \((m, r)\) | \(p/dr\) | Iter | Time | RelErr |
|-----------|----------|------|------|--------|
| (100, 5)  | 5.13     | 392  | 2.84 | 9.905e-004 |
| (200, 5)  | 10.13    | 187  | 4.03 | 9.881e-004 |
| (300, 5)  | 15.13    | 130  | 6.51 | 9.916e-004 |
| (400, 5)  | 20.13    | 99   | 9.26 | 9.947e-004 |
| (500, 5)  | 25.13    | 91   | 14.44| 9.709e-004 |
| (600, 5)  | 30.13    | 78   | 19.31| 9.853e-004 |
| (700, 5)  | 35.13    | 85   | 36.56| 9.984e-004 |
| (800, 5)  | 40.13    | 88   | 54.63| 9.889e-004 |
| (900, 5)  | 45.13    | 60   | 51.40| 9.793e-004 |
| (1000, 5) | 50.13    | 62   | 72.48| 9.737e-004 |
Figure 3. (a): Original gray image (512 × 512); (b): Corresponding low rank images with $r = 30$; (c): Randomly masked image from rank 30 with $sr = 70\%$ and $q = 0.01$; (e): Randomly masked image from rank 30 with $sr = 70\%$ and $\sigma = 0.001$, $q = 0.01$; (d), (f): Recovered images by $PD - l_1$, whose “RelErr” are $4.760e - 3$, $4.827e - 3$ respectively.
Figure 4. Numerical results of $PD - l_1$ for problem (7) with impulsive noise: $m=n=500$, $r=10$, $sr=0.5, 0.7, 0.9$, $q=0.01$.

Table 2. Numerical results of $PD - l_1$ for (7), $m=n$, $sr=0.5$.

| $(m, r)$   | $p/dr$ | $q$  | $\sigma$ | Time  | RelErr     |
|------------|--------|------|-----------|-------|------------|
| (500, 10)  | 12.63  | 0.01 | 0.00      | 21.43 | 9.905e-004 |
| (500, 10)  | 12.63  | 0.01 | 0.001     | 21.65 | 9.822e-004 |
| (500, 10)  | 12.63  | 0.01 | 0.0001    | 21.77 | 9.781e-004 |
| (500, 10)  | 12.63  | 0.001| 0.0001    | 20.61 | 8.054e-004 |
| (500, 10)  | 12.63  | 0.005| 0.0001    | 20.53 | 8.659e-004 |
| (500, 10)  | 12.63  | 0.00 | 0.0001    | 2.15  | 7.722e-005 |

4. Concluding remarks. In this paper, we first proposed a $PD - l_1$ method for solving the nuclear norm minimization problem with $l_1$ norm fidelity term. The proposed method mainly minimized the quadratic penalty function of the problem (7) by alternatively solving the resulting penalty subproblems on $X$ and $r$. One subproblem admits a closed-form solution due to its special structure, and the other subproblem on $X$ can also get a closed-form solution by linearizing its quadratic term. Convergence analysis of the proposed method has been given. Then we presented some numerical experiments with different experiment settings to illustrate the performance of the proposed algorithm. Numerical results show that the $l_1$-fidelity model is better than the $l_2$ one when the observation data contains some impulsive noise or the mixture with Gaussian noise. Moreover, the recovery of some corrupted images have further illustrated that the $PD - l_1$ method is promising and practical. Here we devote to illustrating the efficiency of the $l_1$-fidelity model for the low rank minimization problem with non-Gaussian noise, so all the numerical
experiments use full SVD. In addition, methods based on augmented Lagrangian methods and ADMM can be applied as well to the problem (7). In the near future, we will improve our proposed method with some efficient acceleration techniques and study on other efficient algorithms for problem (7).

Acknowledgments. We thank editors for their valuable suggestions on the paper. We also thank the anonymous referees for their careful reading and very useful suggestions that improved the paper.

This work of Y. Shang is supported by the National Natural Science Foundation of China grant No.11471102. This work of Z.-F Jin is supported by the National Natural Science Foundation of China grant No.61370220 and the Plan for Scientific Innovation Talent of Henan Province Grant No.174200510011. This work of Duo Wang is supported by the National Natural Science Foundation of China grant No.11701150.

REFERENCES

[1] E. Abreu, M. Lightstone, S. K. Mitra and K. Arakawa, A new efficient approach for the removal of impulse noise from highly corrupted images, IEEE Transactions on Image Processing, 5 (1996), 1012–1025.
[2] S. Alliney, A property of the minimum vectors of a regularizing functional defined by means of the absolute norm, IEEE Transactions on Signal Processing, 45 (1997), 913–917.
[3] H. Attouch, J. Bolte and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods, Mathematical Programming, 137 (2013), 91–129.
[4] A. Beck and L. Tetruashvili, On the convergence of block coordinate descent type methods, SIAM Journal on Optimization, 23 (2013), 2037–2060.
[5] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Athena Scientific, Belmont, MA, 2014.
[6] P. Bloomfield and W. Steiger, Least Absolute Deviations, Theory, Applications, and Algorithms, Progress in Probability and Statistics, 6. Birkhäuser Boston, Inc., Boston, MA, 1983.
[7] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for non-convex and nonsmooth problems, Mathematical Programming, 146 (2014), 459–494.
[8] J. F. Cai, E. J. Candès and Z. Shen, A singular value thresholding algorithm for matrix completion, SIam Journal on Optimization, 20 (2008), 1956–1982.
[9] J. F. Cai, R. H. Chan and M. Nikolova, Two-phase approach for deblurring images corrupted by impulse plus gaussian noise, Inverse Problems and Imaging, 2 (2012), 187–204.
[10] C. Chen, B. He and X. Yuan, Matrix completion via an alternating direction method, Ima Journal of Numerical Analysis, 32 (2012), 227–245.
[11] Z. Dong and W. Zhu, An improvement of the penalty decomposition method for sparse approximation, Signal Processing, 113 (2015), 52–60.
[12] M. Fazel, H. Hindi and S. P. Boyd, Rank minimization and applications in system theory, in American Control Conference, 2004. Proceedings of the IEEE, 4 (2004).
[13] M. Fazel, H. Hindi and S. P. Boyd, A rank minimization heuristic with application to minimum order system approximation, in American Control Conference, 2001. Proceedings of the IEEE, 6 (2001).
[14] S. Gu, L. Zhang, W. Zuo and X. Feng, Weighted nuclear norm minimization with application to image denoising, in Computer Vision and Pattern Recognition, 2014, 2862–2869.
[15] H. Ji, C. Liu, Z. Shen and Y. Xu, Robust video denoising using low rank matrix completion, in Computer Vision and Pattern Recognition, 2010, 1791–1798.
[16] Z. F. Jin, Z. Wan, X. Zhao and Y. Xiao, A penalty decomposition method for rank minimization problem with affine constraints, Applied Mathematical Modelling, 39 (2015), 4859–4870.
[17] Z. F. Jin, Q. Wang and Z. Wan, Recovering low-rank matrices from corrupted observations via the linear conjugate gradient algorithm, Journal of Computational and Applied Mathematics, 256 (2014), 114–120.
[18] Z. Lu and Y. Zhang, Penalty decomposition methods for rank minimization, Optimization Methods and Software, 30 (2015), 531–558.
[19] S. Ma, D. Goldfarb and L. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, *Mathematical Programming*, **128** (2011), 321–353.

[20] Y. Marnissi, Y. Zheng, E. Chouzenoux and J. C. Pesquet, *A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise*, Research report, Laboratoire Informatique Gaspard Monge, 2016.

[21] C. A. Micchelli, L. Shen, Y. Xu and X. Zeng, Proximity algorithms for the l1/tv image denoising model, *Advances in Computational Mathematics*, **38** (2013), 401–426.

[22] K. Mohan and M. Fazel, Reweighted nuclear norm minimization with application to system identification, in *American Control Conference*, 2010, 2953–2959.

[23] M. Nikolova, A variational approach to remove outliers and impulse noise, *Journal of Mathematical Imaging and Vision*, **20** (2004), 99–120.

[24] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

[25] J. F. Sturm, Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones, *Optimization Methods and Software*, **11** (1999), 625–653.

[26] K. C. Toh and S. W. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, *Pacific Journal of Optimization*, **6** (2010), 615–640.

[27] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, *Journal of Optimization Theory and Applications*, **109** (2001), 475–494.

[28] R. H. Tütüncü, K. C. Toh and M. J. Todd, Solving semidefinite-quadratic-linear programs using SDPT3, *Mathematical Programming*, **95** (2003), 189–217.

[29] Y. H. Xiao and Z. F. Jin, An alternating direction method for linear-constrained matrix nuclear norm minimization, *Numerical Linear Algebra with Applications*, **19** (2012), 541–554.

[30] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang and Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39** (2016), 156–171.

[31] J. Yang and X. Yuan, Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization, *Mathematics of Computation*, **82** (2013), 301–329.

Received April 2018; revised February 2019.

E-mail address: 160310270300@stu.haust.edu.cn
E-mail address: jinzhengfen@whu.edu.cn
E-mail address: mathshang@sina.com