

FAST FIRST-ORDER METHODS FOR STABLE PRINCIPAL COMPONENT PURSUIT[¶]

N. S. AYBAT [†], D. GOLDFARB [‡], AND G. IYENGAR [§]

Abstract. The stable principal component pursuit (SPCP) problem is a non-smooth convex optimization problem, the solution of which has been shown both in theory and in practice to enable one to recover the low rank and sparse components of a matrix whose elements have been corrupted by Gaussian noise. In this paper, we first show how several existing fast first-order methods can be applied to this problem very efficiently. Specifically, we show that the subproblems that arise when applying optimal gradient methods of Nesterov, alternating linearization methods and alternating direction augmented Lagrangian methods to the SPCP problem either have closed-form solutions or have solutions that can be obtained with very modest effort. Later, we develop a new first order algorithm, NSA, based on partial variable splitting. All but one of the methods analyzed require at least one of the non-smooth terms in the objective function to be smoothed and obtain an ϵ -optimal solution to the SPCP problem in $O(1/\epsilon)$ iterations. NSA, which works directly with the fully non-smooth objective function, is proved to be convergent under mild conditions on the sequence of parameters it uses. Our preliminary computational tests show that the latter method, NSA, although its complexity is not known, is the fastest among the four algorithms described and substantially outperforms ASALM, the only existing method for the SPCP problem. To best of our knowledge, an algorithm for the SPCP problem that has $O(1/\epsilon)$ iteration complexity and has a per iteration complexity equal to that of a singular value decomposition is given for the first time.

1. Introduction. In [2, 12], it was shown that when the data matrix $D \in \mathbb{R}^{m \times n}$ is of the form $D = X^0 + S^0$, where X^0 is a low-rank matrix, i.e. $\text{rank}(X^0) \ll \min\{m, n\}$, and S^0 is a sparse matrix, i.e. $\|S^0\|_0 \ll mn$ ($\|\cdot\|_0$ counts the number of nonzero elements of its argument), one can recover the low-rank and sparse components of D by solving the *principal component pursuit* problem

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* + \xi \|D - X\|_1, \quad (1.1)$$

where $\xi = \frac{1}{\sqrt{\max\{m, n\}}}$.

For $X \in \mathbb{R}^{m \times n}$, $\|X\|_*$ denotes the nuclear norm of X , which is equal to the sum of its singular values, $\|X\|_1 := \sum_{i=1}^m \sum_{j=1}^n |X_{ij}|$, $\|X\|_\infty := \max\{|X_{ij}| : 1 \leq i \leq m, 1 \leq j \leq n\}$ and $\|X\|_2 := \sigma_{\max}(X)$, where $\sigma_{\max}(X)$ is the maximum singular value of X .

To be more precise, let $X^0 \in \mathbb{R}^{m \times n}$ with $\text{rank}(X^0) = r$ and let $X^0 = U\Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$ denote the singular value decomposition (SVD) of X^0 . Suppose that for some $\mu > 0$, U and V satisfy

$$\max_i \|U^T e_i\|_2^2 \leq \frac{\mu r}{m}, \quad \max_i \|V^T e_i\|_2^2 \leq \frac{\mu r}{n}, \quad \|UV^T\|_\infty \leq \sqrt{\frac{\mu r}{mn}}, \quad (1.2)$$

where e_i denotes the i -th unit vector.

THEOREM 1.1. [2] Suppose $D = X^0 + S^0$, where $X^0 \in \mathbb{R}^{m \times n}$ with $m < n$ satisfies (1.2) for some $\mu > 0$, and the support set of S^0 is uniformly distributed. Then there are constants c , ρ_r , ρ_s such that with probability of at least $1 - cn^{-10}$, the principal component pursuit problem (1.1) exactly recovers X^0 and S^0 provided that

$$\text{rank}(X^0) \leq \rho_r m \mu^{-1} (\log(n))^{-2} \quad \text{and} \quad \|S^0\|_0 \leq \rho_s mn. \quad (1.3)$$

In [13], it is shown that the recovery is still possible even when the data matrix, D , is corrupted with a dense error matrix, ζ^0 such that $\|\zeta^0\|_F \leq \delta$, by solving the *stable principal component pursuit* (SPCP) problem

$$(P) : \min_{X, S \in \mathbb{R}^{m \times n}} \{\|X\|_* + \xi \|S\|_1 : \|X + S - D\|_F \leq \delta\}. \quad (1.4)$$

Specifically, the following theorem is proved in [13].

[†]IEOR Department, Columbia University. Email: nsa2106@columbia.edu.

[‡]IEOR Department, Columbia University. Email: goldfarb@columbia.edu.

[§]IEOR Department, Columbia University. Email: gi10@columbia.edu.

[¶]Research partially supported by ONR grant N000140310514, NSF Grant DMS 10-16571 and DOE Grant DE-FG02-08-25856.

THEOREM 1.2. [13] Suppose $D = X^0 + S^0 + \zeta^0$, where $X^0 \in \mathbb{R}^{m \times n}$ with $m < n$ satisfies (1.2) for some $\mu > 0$, and the support set of S^0 is uniformly distributed. If X^0 and S^0 satisfy (1.3), then for any ζ^0 such that $\|\zeta^0\|_F \leq \delta$ the solution, (X^*, S^*) , to the stable principal component pursuit problem (1.4) satisfies $\|X^* - X^0\|_F^2 + \|S^* - S^0\|_F^2 \leq Cmn\delta^2$ for some constant C with high probability.

Principal component pursuit and stable principal component pursuit both have applications in video surveillance and face recognition. For existing algorithmic approaches to solving principal component pursuit see [2, 3, 6, 7, 13] and references therein. In this paper, we develop four different fast first-order algorithms to solve the SPCP problem (P) . The first two algorithms are direct applications of Nesterov's optimal algorithm [9] and the proximal gradient method of Tseng [11], which is inspired by both FISTA and Nesterov's infinite memory algorithms that are introduced in [1] and [9], respectively. In this paper it is shown that both algorithms can compute an ϵ -optimal, feasible solution to (P) in $\mathcal{O}(1/\epsilon)$ iterations. The third and fourth algorithms apply an alternating direction augmented Lagrangian approach to an equivalent problem obtained by partial variable splitting. The third algorithm can compute an ϵ -optimal, feasible solution to the problem in $\mathcal{O}(1/\epsilon^2)$ iterations, which can be easily improved to $\mathcal{O}(1/\epsilon)$ complexity. Given $\epsilon > 0$, all first three algorithms use suitably smooth versions of at least one of the norms in the objective function. The fourth algorithm (NSA) works directly with the original non-smooth objective function and can be shown to converge to an optimal solution of (P) , provided that a mild condition on the increasing sequence of penalty multipliers holds. To best of our knowledge, an algorithm for the SPCP problem that has $\mathcal{O}(1/\epsilon)$ iteration complexity and has a per iteration complexity equal to that of a singular value decomposition is given for the first time.

The only algorithm that we know of that has been designed to solve the SPCP problem (P) is the algorithm ASALM [10]. The results of our numerical experiments comparing NSA algorithm with ASALM has shown that NSA is faster and also more robust to changes in problem parameters.

2. Proximal Gradient Algorithm with Smooth Objective Function. In this section we show that Nesterov's optimal algorithm [8, 9] for simple sets is efficient for solving (P) .

For fixed parameters $\mu > 0$ and $\nu > 0$, define the smooth $C^{1,1}$ functions $f_\mu(\cdot)$ and $g_\nu(\cdot)$ as follows

$$f_\mu(X) = \max_{U \in \mathbb{R}^{m \times n} : \|U\|_2 \leq 1} \langle X, U \rangle - \frac{\mu}{2} \|U\|_F^2, \quad (2.1)$$

$$g_\nu(S) = \max_{W \in \mathbb{R}^{m \times n} : \|W\|_\infty \leq 1} \langle S, W \rangle - \frac{\nu}{2} \|W\|_F^2. \quad (2.2)$$

Clearly, $f_\mu(\cdot)$ and $g_\nu(\cdot)$ closely approximate the non-smooth functions $f(X) := \|X\|_*$ and $g(S) := \|S\|_1$, respectively. Also let $\chi := \{(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} : \|X + S - D\|_F \leq \delta\}$ and $L = \frac{1}{\mu} + \frac{1}{\nu}$, where $\frac{1}{\mu}$ and $\frac{1}{\nu}$ are the Lipschitz constants for the gradients of $f_\mu(\cdot)$ and $g_\nu(\cdot)$, respectively. Then Nesterov's optimal algorithm [8, 9] for simple sets applied to the problem:

$$\min_{X, S \in \mathbb{R}^{m \times n}} \{f_\mu(X) + \xi g_\nu(S) : (X, S) \in \chi\}, \quad (2.3)$$

is given by **Algorithm 1**.

Because of the simple form of the set χ , it is easy to ensure that all iterates (Y_k^x, Y_k^s) , (Z_k^x, Z_k^s) and (X_{k+1}, S_{k+1}) lie in χ . Hence, **Algorithm 1** enjoys the full convergence rate of $\mathcal{O}(L/k^2)$ of the Nesterov's method. Thus, setting $\mu = \Omega(\epsilon)$ and $\nu = \Omega(\epsilon)$, **Algorithm 1** computes an ϵ -optimal and feasible solution to problem (P) in $k^* = \mathcal{O}(1/\epsilon)$ iterations. The iterates (Y_k^x, Y_k^s) and (Z_k^x, Z_k^s) that need to be computed at each iteration of **Algorithm 1** are solutions to an optimization problem of the form:

$$(P_s) : \min_{X, S \in \mathbb{R}^{m \times n}} \left\{ \frac{L}{2} \left(\|X - \tilde{X}\|_F^2 + \|S - \tilde{S}\|_F^2 \right) + \langle Q_x, X \rangle + \langle Q_s, S \rangle : (X, S) \in \chi \right\}. \quad (2.4)$$

The following lemma shows that the solution to problems of the form (P_s) can be computed efficiently.

Algorithm 1 SMOOTH PROXIMAL GRADIENT(X_0, S_0)

```

1: input:  $X_0 \in \mathbb{R}^{m \times n}$ ,  $S_0 \in \mathbb{R}^{m \times n}$ 
2:  $k \leftarrow 0$ 
3: while  $k \leq k^*$  do
4:   Compute  $\nabla f_\mu(X_k)$  and  $\nabla g_\nu(S_k)$ 
5:    $(Y_k^x, Y_k^s) \leftarrow \operatorname{argmin}_{X,S} \{ \langle \nabla f_\mu(X_k), X \rangle + \langle \nabla g_\nu(S_k), S \rangle + \frac{L}{2} (\|X - X_k\|_F^2 + \|S - S_k\|_F^2) : (X, S) \in \chi \}$ 
6:    $\Gamma_k(X, S) := \sum_{i=0}^k \frac{i+1}{2} \{ \langle \nabla f_\mu(X_i), X \rangle + \langle \nabla g_\nu(S_i), S \rangle \}$ 
7:    $(Z_k^x, Z_k^s) \leftarrow \operatorname{argmin}_{X,S} \{ \Gamma_k(X, S) + \frac{L}{2} (\|X - X_0\|_F^2 + \|S - S_0\|_F^2) : (X, S) \in \chi \}$ 
8:    $(X_{k+1}, S_{k+1}) \leftarrow \left( \frac{k+1}{k+3} \right) (Y_k^x, Y_k^s) + \left( \frac{2}{k+3} \right) (Z_k^x, Z_k^s)$ 
9:    $k \leftarrow k + 1$ 
10: end while
11: return  $(X_{k^*}, S_{k^*})$ 

```

LEMMA 2.1. *The optimal solution (X^*, S^*) to problem (P_s) can be written in closed form as follows. When $\delta > 0$,*

$$X^* = \left(\frac{\theta^*}{L + 2\theta^*} \right) (D - q_s(\tilde{S})) + \left(\frac{L + \theta^*}{L + 2\theta^*} \right) q_x(\tilde{X}), \quad (2.5)$$

$$S^* = \left(\frac{\theta^*}{L + 2\theta^*} \right) (D - q_x(\tilde{X})) + \left(\frac{L + \theta^*}{L + 2\theta^*} \right) q_s(\tilde{S}), \quad (2.6)$$

where $q_x(X) := X - \frac{1}{L} Q_x$, $q_s(S) := S - \frac{1}{L} Q_s$ and

$$\theta^* = \max \left\{ 0, \frac{L}{2} \left(\frac{\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F}{\delta} - 1 \right) \right\}. \quad (2.7)$$

When $\delta = 0$,

$$X^* = \frac{1}{2} (D - q_s(\tilde{S})) + \frac{1}{2} q_x(\tilde{X}) \text{ and } S^* = \frac{1}{2} (D - q_x(\tilde{X})) + \frac{1}{2} q_s(\tilde{S}). \quad (2.8)$$

Proof. Suppose that $\delta > 0$. Writing the constraint in problem (P_s) , $(X, S) \in \chi$, as

$$\frac{1}{2} \|X + S - D\|_F^2 \leq \frac{\delta^2}{2}, \quad (2.9)$$

the Lagrangian function for (2.4) is given as

$$\mathcal{L}(X, S; \theta) = \frac{L}{2} (\|X - \tilde{X}\|_F^2 + \|S - \tilde{S}\|_F^2) + \langle Q_x, X - \tilde{X} \rangle + \langle Q_s, S - \tilde{S} \rangle + \frac{\theta}{2} (\|X + S - D\|_F^2 - \delta^2).$$

Therefore, the optimal solution (X^*, S^*) and optimal Lagrangian multiplier $\theta^* \in \mathbb{R}$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions:

- i. $\|X^* + S^* - D\|_F \leq \delta$,
- ii. $\theta^* \geq 0$,
- iii. $\theta^* (\|X^* + S^* - D\|_F - \delta) = 0$,
- iv. $L(X^* - \tilde{X}) + \theta^*(X^* + S^* - D) + Q_x = 0$,
- v. $L(S^* - \tilde{S}) + \theta^*(X^* + S^* - D) + Q_s = 0$.

Conditions iv and v imply that (X^*, S^*) satisfy (2.5) and (2.6), from which it follows that

$$X^* + S^* - D = \left(\frac{L}{L + 2\theta^*} \right) (q_x(\tilde{X}) + q_s(\tilde{S}) - D). \quad (2.10)$$

Case 1: $\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F \leq \delta$. Setting $X^* = q_x(\tilde{X})$, $S^* = q_s(\tilde{S})$ and $\theta^* = 0$, clearly satisfies (2.5), (2.6) and conditions i (from (2.10)), ii and iii. Thus, this choice of variables satisfies all the five KKT conditions.

Case 2: $\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F > \delta$. Set $\theta^* = \frac{L}{2} \left(\frac{\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F}{\delta} - 1 \right)$. Since $\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F > \delta$, $\theta^* > 0$; hence, ii is satisfied. Moreover, for this value of θ^* , it follows from (2.10) that $\|X^* + S^* - D\|_F = \delta$. Thus, KKT conditions i and iii are satisfied.

Therefore, setting X^* and S^* according to (2.5) and (2.6), respectively; and setting

$$\theta^* = \max \left\{ 0, \frac{L}{2} \left(\frac{\|q_x(\tilde{X}) + q_s(\tilde{S}) - D\|_F}{\delta} - 1 \right) \right\},$$

satisfies all the five KKT conditions.

Now, suppose that $\delta = 0$. Since $S^* = D - X^*$, problem (P_s) can be written as

$$\min_{X \in \mathbb{R}^{m \times n}} \|X - \tilde{X} + \frac{Q_x}{L}\|_F^2 + \|D - X - \tilde{S} + \frac{Q_s}{L}\|_F^2,$$

which is also equivalent to the problem: $\min_{X \in \mathbb{R}^{m \times n}} \|X - q_x(\tilde{X})\|_F^2 + \|X - (D - q_s(\tilde{S}))\|_F^2$. Then (2.8) trivially follows from first-order optimality conditions for this problem and the fact that $S^* = D - X^*$. $\square$

3. Proximal Gradient Algorithm with Partially Smooth Objective Function. In this section we show how the proximal gradient algorithm, Algorithm 3 in [11], can be applied to the problem

$$\min_{X, S \in \mathbb{R}^{m \times n}} \{f_\mu(X) + \xi \|S\|_1 : (X, S) \in \chi\}, \quad (3.1)$$

where $f_\mu(\cdot)$ is the smooth function defined in (2.1) such that $\nabla f_\mu(\cdot)$ is Lipschitz continuous with constant $L_\mu = \frac{1}{\mu}$. This algorithm is given in **Algorithm 2**.

Algorithm 2 PARTIALLY SMOOTH PROXIMAL GRADIENT(X_0, S_0)

```

1: input:  $X_0 \in \mathbb{R}^{m \times n}$ ,  $S_0 \in \mathbb{R}^{m \times n}$ 
2:  $(Z_0^x, Z_0^s) \leftarrow (X_0, S_0)$ ,  $k \leftarrow 0$ 
3: while  $k \leq k^*$  do
4:    $(Y_k^x, Y_k^s) \leftarrow \left(\frac{k}{k+2}\right)(X_k, S_k) + \left(\frac{2}{k+2}\right)(Z_k^x, Z_k^s)$ 
5:   Compute  $\nabla f_\mu(Y_k^x)$ 
6:    $(Z_{k+1}^x, Z_{k+1}^s) \leftarrow \operatorname{argmin}_{X, S} \left\{ \sum_{i=0}^k \frac{i+1}{2} \{\xi \|S\|_1 + \langle \nabla f_\mu(Y_i^x), X \rangle\} + \frac{L_\mu}{2} \|X - X_0\|_F^2 : (X, S) \in \chi \right\}$ 
7:    $(X_{k+1}, S_{k+1}) \leftarrow \left(\frac{k}{k+2}\right)(X_k, S_k) + \left(\frac{2}{k+2}\right)(Z_{k+1}^x, Z_{k+1}^s)$ 
8:    $k \leftarrow k + 1$ 
9: end while
10: return  $(X_{k^*}, S_{k^*})$ 

```

Mimicking the proof in [11], it is easy to show that **Algorithm 2**, which uses the prox function $\frac{1}{2}\|X - X_0\|_F^2$, converges to the optimal solution of (3.1). Given $(X_0, S_0) \in \chi$, e.g. $X_0 = \mathbf{0}$ and $S_0 = \bar{D}$, the current algorithm keeps all iterates in χ as in **Algorithm 1**, and hence it enjoys the full convergence rate of $\mathcal{O}(L/k^2)$. Thus, setting $\mu = \Omega(\epsilon)$, **Algorithm 2** computes an ϵ -optimal, feasible solution of problem (P) in $k^* = \mathcal{O}(1/\epsilon)$ iterations.

The only thing left to be shown is that the optimization subproblems in **Algorithm 2** can be solved efficiently. The subproblem that has to be solved at each iteration to compute (Z_{k+1}^x, Z_{k+1}^s) has the form:

$$(P_{ns}) : \min \left\{ \xi \|S\|_1 + \langle Q, X - \tilde{X} \rangle + \frac{\rho}{2} \|X - \tilde{X}\|_F^2 : (X, S) \in \chi \right\}, \quad (3.2)$$

for some $\rho > 0$. Lemma 3.1 shows that these computations can be done efficiently.

LEMMA 3.1. *The optimal solution (X^*, S^*) to problem (P_{ns}) can be written in closed form as follows. When $\delta > 0$,*

$$S^* = \text{sign} \left(D - q(\tilde{X}) \right) \odot \max \left\{ |D - q(\tilde{X})| - \xi \frac{(\rho + \theta^*)}{\rho \theta^*} E, \mathbf{0} \right\}, \quad (3.3)$$

$$X^* = \frac{\theta^*}{\rho + \theta^*} (D - S^*) + \frac{\rho}{\rho + \theta^*} q(\tilde{X}), \quad (3.4)$$

where $q(\tilde{X}) := \tilde{X} - \frac{1}{\rho} Q$, E and $\mathbf{0} \in \mathbb{R}^{m \times n}$ are matrices with all components equal to ones and zeros, respectively, and $\odot$ denotes the componentwise multiplication operator. $\theta^* = 0$ if $\|D - q(\tilde{X})\|_F \leq \delta$; otherwise, θ^* is the unique positive solution of the nonlinear equation $\phi(\theta) = \delta$, where

$$\phi(\theta) := \left\| \min \left\{ \frac{\xi}{\theta} E, \frac{\rho}{\rho + \theta} |D - q(\tilde{X})| \right\} \right\|_F. \quad (3.5)$$

Moreover, θ^* can be efficiently computed in $\mathcal{O}(mn \log(mn))$ time.

When $\delta = 0$,

$$S^* = \text{sign} \left(D - q(\tilde{X}) \right) \odot \max \left\{ |D - q(\tilde{X})| - \frac{\xi}{\rho} E, \mathbf{0} \right\} \quad \text{and} \quad X^* = D - S^*. \quad (3.6)$$

Proof. Suppose that $\delta > 0$. Let (X^*, S^*) be an optimal solution to problem (P_{ns}) and θ^* denote the optimal Lagrangian multiplier for the constraint $(X, S) \in \chi$ written as (2.9). Then the KKT optimality conditions for this problem are

- i. $Q + \rho(X^* - \tilde{X}) + \theta^*(X^* + S^* - D) = 0$,
- ii. $\xi G + \theta^*(X^* + S^* - D) = 0$ and $G \in \partial \|S^*\|_1$,
- iii. $\|X^* + S^* - D\|_F \leq \delta$,
- iv. $\theta^* \geq 0$,
- v. $\theta^* (\|X^* + S^* - D\|_F - \delta) = 0$.

From i and ii, we have

$$\begin{bmatrix} (\rho + \theta^*)I & \theta^*I \\ \theta^*I & \theta^*I \end{bmatrix} \begin{bmatrix} X^* \\ S^* \end{bmatrix} = \begin{bmatrix} \theta^*D + \rho q(\tilde{X}) \\ \theta^*D - \xi G \end{bmatrix}, \quad (3.7)$$

where $q(\tilde{X}) = \tilde{X} - \frac{1}{\rho} Q$. From (3.7) it follows that

$$\begin{bmatrix} (\rho + \theta^*)I & \theta^*I \\ 0 & \left(\frac{\rho \theta^*}{\rho + \theta^*} I \right) \end{bmatrix} \begin{bmatrix} X^* \\ S^* \end{bmatrix} = \begin{bmatrix} \theta^*D + \rho q(\tilde{X}) \\ \frac{\rho \theta^*}{\rho + \theta^*} (D - q(\tilde{X})) - \xi G \end{bmatrix}. \quad (3.8)$$

From the second equation in (3.8), we have

$$\xi \frac{(\rho + \theta^*)}{\rho \theta^*} G + S^* + q(\tilde{X}) - D = 0. \quad (3.9)$$

But (3.9) is precisely the first-order optimality conditions for the “shrinkage” problem

$$\min_{S \in \mathbb{R}^{m \times n}} \left\{ \xi \frac{(\rho + \theta^*)}{\rho \theta^*} \|S\|_1 + \frac{1}{2} \|S + q(\tilde{X}) - D\|_F^2 \right\}.$$

Thus, S^* is the optimal solution to the “shrinkage” problem and is given by (3.3). (3.4) follows from the first equation in (3.8), and it implies

$$X^* + S^* - D = \frac{\rho}{\rho + \theta^*} (S^* + q(\tilde{X}) - D). \quad (3.10)$$

Therefore,

$$\begin{aligned}
\|X^* + S^* - D\|_F &= \frac{\rho}{\rho + \theta^*} \|S^* + q(\tilde{X}) - D\|_F, \\
&= \frac{\rho}{\rho + \theta^*} \left\| \text{sign}(D - q(\tilde{X})) \odot \max \left\{ |D - q(\tilde{X})| - \xi \frac{(\rho + \theta^*)}{\rho \theta^*} E, \mathbf{0} \right\} - (D - q(\tilde{X})) \right\|_F, \\
&= \frac{\rho}{\rho + \theta^*} \left\| \max \left\{ |D - q(\tilde{X})| - \xi \frac{(\rho + \theta^*)}{\rho \theta^*} E, \mathbf{0} \right\} - |D - q(\tilde{X})| \right\|_F, \\
&= \frac{\rho}{\rho + \theta^*} \left\| \min \left\{ \xi \frac{(\rho + \theta^*)}{\rho \theta^*} E, |D - q(\tilde{X})| \right\} \right\|_F, \\
&= \left\| \min \left\{ \frac{\xi}{\theta^*} E, \frac{\rho}{\rho + \theta^*} |D - q(\tilde{X})| \right\} \right\|_F, \tag{3.11}
\end{aligned}$$

where the second equation uses (3.3). Now let $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be

$$\phi(\theta) := \left\| \min \left\{ \frac{\xi}{\theta} E, \frac{\rho}{\rho + \theta} |D - q(\tilde{X})| \right\} \right\|_F. \tag{3.12}$$

Case 1: $\|D - q(\tilde{X})\|_F \leq \delta$. $\theta^* = 0$, $S^* = \mathbf{0}$ and $X^* = q(\tilde{X})$ trivially satisfy all the KKT conditions.

Case 2: $\|D - q(\tilde{X})\|_F > \delta$. It is easy to show that $\phi(\cdot)$ is a strictly decreasing function of θ . Since $\phi(0) = \|D - q(\tilde{X})\|_F > \delta$ and $\lim_{\theta \rightarrow \infty} \phi(\theta) = 0$, there exists a unique $\theta^* > 0$ such that $\phi(\theta^*) = \delta$. Given θ^* , S^* and X^* can then be computed from equations (3.3) and (3.4), respectively. Moreover, since $\theta^* > 0$ and $\phi(\theta^*) = \delta$, (3.11) implies that X^* , S^* and θ^* satisfy the KKT conditions.

We now show that θ^* can be computed in $\mathcal{O}(mn \log(mn))$ time. Let $A := |D - q(\tilde{X})|$ and $0 \leq a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(mn)}$ be the mn elements of the matrix A sorted in increasing order, which can be done in $\mathcal{O}(mn \log(mn))$ time. Defining $a_{(0)} := 0$ and $a_{(mn+1)} := \infty$, we then have for all $j \in \{0, 1, \dots, mn\}$ that

$$\frac{\rho}{\rho + \theta} a_{(j)} \leq \frac{\xi}{\theta} \leq \frac{\rho}{\rho + \theta} a_{(j+1)} \Leftrightarrow \frac{1}{\xi} a_{(j)} - \frac{1}{\rho} \leq \frac{1}{\theta} \leq \frac{1}{\xi} a_{(j+1)} - \frac{1}{\rho}. \tag{3.13}$$

For all $\bar{k} < j \leq mn$ define θ_j such that $\frac{1}{\theta_j} = \frac{1}{\xi} a_{(j)} - \frac{1}{\rho}$ and let $\bar{k} := \max \left\{ j : \frac{1}{\theta_j} \leq 0, j \in \{0, 1, \dots, mn\} \right\}$. Then for all $\bar{k} < j \leq mn$

$$\phi(\theta_j) = \sqrt{\left(\frac{\rho}{\rho + \theta_j} \right)^2 \sum_{i=0}^j a_{(i)}^2 + (mn - j) \left(\frac{\xi}{\theta_j} \right)^2}. \tag{3.14}$$

Also define $\theta_{\bar{k}} := \infty$ and $\theta_{mn+1} := 0$ so that $\phi(\theta_{\bar{k}}) := 0$ and $\phi(\theta_{mn+1}) = \phi(0) = \|A\|_F > \delta$. Note that $\{\theta_j\}_{\{\bar{k} < j \leq mn\}}$ contains all the points at which $\phi(\theta)$ may not be differentiable for $\theta \geq 0$. Define $j^* := \max \{j : \phi(\theta_j) \leq \delta, \bar{k} \leq j \leq mn\}$. Then θ^* is the unique solution of the system

$$\sqrt{\left(\frac{\rho}{\rho + \theta} \right)^2 \sum_{i=0}^{j^*} a_{(i)}^2 + (mn - j^*) \left(\frac{\xi}{\theta} \right)^2} = \delta \text{ and } \theta > 0, \tag{3.15}$$

since $\phi(\theta)$ is continuous and strictly decreasing in θ for $\theta \geq 0$. Solving the equation in (3.15) requires finding the roots of a fourth-order polynomial (a.k.a. quartic function); therefore, one can compute $\theta^* > 0$ using the algebraic solutions of quartic equations (as shown by Lodovico Ferrari in 1540), which requires $\mathcal{O}(1)$ operations.

Note that if $\bar{k} = mn$, then θ^* is the solution of the equation

$$\sqrt{\left(\frac{\rho}{\rho + \theta^*} \right)^2 \sum_{i=1}^{mn} a_{(i)}^2} = \delta, \tag{3.16}$$

i.e. $\theta^* = \rho \left(\frac{\|A\|_F}{\delta} - 1 \right) = \rho \left(\frac{\|D - \tilde{X}\|_F}{\delta} - 1 \right)$. Hence, we have proved that problem (P_{ns}) can be solved efficiently.

Now, suppose that $\delta = 0$. Since $S^* = D - X^*$, problem (P_{ns}) can be written as

$$\min_{S \in \mathbb{R}^{m \times n}} \quad \xi \|S\|_1 + \frac{1}{2} \|S - (D - q(\tilde{X}))\|_F^2. \quad (3.17)$$

Then (3.6) trivially follows from first-order optimality conditions for the above problem and the fact that $X^* = D - S^*$. $\square$

The following lemma will be used later in Section 5. However, we give its proof here, since it uses some equations from the proof of Lemma 3.1. Let $\mathbf{1}_\chi(\cdot, \cdot)$ denote the indicator function of the closed convex set $\chi \subset \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, i.e. if $(Z, S) \in \chi$, then $\mathbf{1}_\chi(Z, S) = 0$; otherwise, $\mathbf{1}_\chi(Z, S) = \infty$.

LEMMA 3.2. *Suppose that $\delta > 0$. Let (X^*, S^*) be an optimal solution to problem (P_{ns}) and θ^* be an optimal Lagrangian multiplier such that (X^*, S^*) and θ^* together satisfy the KKT conditions, i-v in the proof of Lemma 3.1. Then $(W^*, W^*) \in \partial \mathbf{1}_\chi(X^*, S^*)$, where $W^* := -Q + \rho(\tilde{X} - X^*) = \theta^*(X^* + S^* - D)$.*

Proof. Let $W^* := -Q + \rho(\tilde{X} - X^*)$, then from i and v of the KKT optimality conditions in the proof of Lemma 3.1, we have $W^* = \theta^*(X^* + S^* - D)$ and

$$\|W^*\|_F = \theta^* \|X^* + S^* - D\| = \theta^* (\|X^* + S^* - D\| - \delta) + \theta^* \delta = \theta^* \delta. \quad (3.18)$$

Moreover, for all $(X, S) \in \chi$, it follows from the definition of χ that $\langle W^*, \theta^*(X + S - D) \rangle \leq \theta^* \|W^*\|_F \|X + S - D\|_F \leq \theta^* \delta \|W^*\|_F$. Thus, for all $(X, S) \in \chi$, we have $\langle W^*, W^* \rangle = \|W^*\|_F^2 = \theta^* \delta \|W^*\|_F \geq \langle W^*, \theta^*(X + S - D) \rangle$. Hence,

$$0 \geq \langle W^*, \theta^*(X + S - D) - W^* \rangle = \langle W^*, \theta^*(X - X^* + S - S^*) \rangle \quad \forall (X, S) \in \chi. \quad (3.19)$$

It follows from the proof of Lemma 3.1 that if $\|D - q(\tilde{X})\|_F > \delta$, then $\theta^* > 0$, where $q(\tilde{X}) = \tilde{X} - \frac{1}{\rho}Q$. Therefore, (3.19) implies that

$$0 \geq \langle W^*, X - X^* + S - S^* \rangle \quad \forall (X, S) \in \chi. \quad (3.20)$$

On the other hand, if $\|D - q(\tilde{X})\|_F \leq \delta$, then $\theta^* = 0$. Hence $W^* = \theta^*(X^* + S^* - D) = 0$, and (3.20) follows trivially. Therefore, (3.20) always holds and this shows that $(W^*, W^*) \in \partial \mathbf{1}_\chi(X^*, S^*)$. $\square$

4. Alternating Linearization and Augmented Lagrangian Algorithms. In this and the next section we present algorithms for solving problems (3.1) and (1.4) that are based on partial variable splitting combined with alternating minimization of a suitably linearized augmented Lagrangian function. We can write problems (1.4) and (3.1) generically as

$$\min_{X, S \in \mathbb{R}^{m \times n}} \{ \phi(X) + \xi g(S) : (X, S) \in \chi \}. \quad (4.1)$$

For problem (1.4), $\phi(X) = f(X) = \|X\|_*$, while for problem (3.1), $\phi(X) = f_\mu(X)$ given in (2.1).

In this section, we first assume that $\phi : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ and $g : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ are any closed convex functions such that $\nabla \phi$ is Lipschitz continuous, and χ is a general closed convex set. Here we use partial variable splitting, i.e. we only split the X variables in (4.1), to arrive at the following equivalent problem

$$\min_{X, S, Z \in \mathbb{R}^{m \times n}} \{ \phi(X) + \xi g(S) : X = Z, (Z, S) \in \chi \}. \quad (4.2)$$

Let $\psi(Z, S) := \xi g(S) + \mathbf{1}_\chi(Z, S)$ and define the augmented Lagrangian function

$$\mathcal{L}_\rho(X, Z, S; Y) = \phi(X) + \psi(Z, S) + \langle Y, X - Z \rangle + \frac{\rho}{2} \|X - Z\|_F^2. \quad (4.3)$$

Then minimizing (4.3) by alternating between X and then (Z, S) leads to several possible methods that can compute a solution to (4.2). These include the alternating linearization method (ALM) with skipping step

Algorithm 3 ALM-S(Y_0)

```
1: input:  $X_0 \in \mathbb{R}^{m \times n}$ ,  $S_0 \in \mathbb{R}^{m \times n}$ ,  $Y_0 \in \mathbb{R}^{m \times n}$ 
2:  $Z_0 \leftarrow X_0$ ,  $k \leftarrow 0$ 
3: while  $k \geq 0$  do
4:    $X_{k+1} \leftarrow \operatorname{argmin}_X \mathcal{L}_\rho(X, Z_k, S_k; Y_k)$ 
5:   if  $\phi(X_{k+1}) + \psi(X_{k+1}, S_k) > \mathcal{L}_\rho(X_{k+1}, Z_k, S_k; Y_k)$  then
6:      $X_{k+1} \leftarrow Z_k$ 
7:   end if
8:    $(Z_{k+1}, S_{k+1}) \leftarrow \operatorname{argmin}_{Z, S} \psi(Z, S) + \phi(X_{k+1}) + \langle \nabla \phi(X_{k+1}), Z - X_{k+1} \rangle + \frac{\rho}{2} \|Z - X_{k+1}\|_F^2$ 
9:    $Y_{k+1} \leftarrow -\nabla \phi(X_{k+1}) + \rho(X_{k+1} - Z_{k+1})$ 
10:   $k \leftarrow k + 1$ 
11: end while
```

that has an $\mathcal{O}(\frac{\rho}{k})$ convergence rate, and the fast version of this method with an $\mathcal{O}(\frac{\rho}{k^2})$ rate (see [3] for full splitting versions of these methods). In this paper, we only provide a proof of the complexity result for the alternating linearization method with skipping steps (ALM-S) in Theorem 4.1 below. One can easily extend the proof of Theorem 4.1 to an ALM method based on (4.3) with the function $g(S)$ replaced by a suitably smoothed version (see [3] for the details of ALM algorithm).

THEOREM 4.1. *Let $\phi : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ and $\psi : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ be closed convex functions such that $\nabla \phi$ is Lipschitz continuous with Lipschitz constant L , and χ be a closed convex set. Let $\Phi(X, S) := \phi(X) + \psi(X, S)$. For $\rho \geq L$, the sequence $\{Z_k, S_k\}_{k \in \mathbb{Z}_+}$ in Algorithm ALM-S satisfies*

$$\Phi(Z_k, S_k) - \Phi(X^*, S^*) \leq \rho \frac{\|X_0 - X^*\|_F^2}{2(k + n_k)}, \quad (4.4)$$

where $(X^*, S^*) = \operatorname{argmin}_{X, S \in \mathbb{R}^{m \times n}} \Phi(X, S)$, $n_k := \sum_{i=0}^{k-1} \mathbf{1}_{\{\Phi(X_{i+1}, S_i) > \mathcal{L}_\rho(X_{i+1}, Z_i, S_i; Y_i)\}}$ and $\mathbf{1}_{\{\cdot\}}$ is 1 if its argument is true; otherwise, 0.

Proof. See Appendix A for the proof. $\square$

We obtain **Algorithm 4** by applying **Algorithm 3** to solve problem (3.1), where the smooth function $\phi(X) = f_\mu(X)$, defined in (2.1), the non-smooth closed convex function is $\xi \|S\|_1 + \mathbf{1}_\chi(X, S)$ and $\chi = \{(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} : \|X + S - D\|_F \leq \delta\}$. Theorem 4.1 shows that **Algorithm 4** has an iteration complexity of $\mathcal{O}(\frac{1}{\epsilon^2})$ to obtain ϵ -optimal and feasible solution of (P).

Algorithm 4 PARTIALLY SMOOTH ALM(Y_0)

```
1: input:  $Y_0 \in \mathbb{R}^{m \times n}$ 
2:  $Z_0 \leftarrow 0$ ,  $S_0 \leftarrow D$ ,  $k \leftarrow 0$ 
3: while  $k \geq 0$  do
4:    $X_{k+1} \leftarrow \operatorname{argmin}_X f_\mu(X) + \langle Y_k, X - Z_k \rangle + \frac{\rho}{2} \|X - Z_k\|_F^2$ 
5:    $B_k \leftarrow f_\mu(X_{k+1}) + \xi \|S_k\|_1 + \langle Y_k, X_{k+1} - Z_k \rangle + \frac{\rho}{2} \|X_{k+1} - Z_k\|_F^2$ 
6:   if  $f_\mu(X_{k+1}) + \xi \|S_k\|_1 + \mathbf{1}_\chi(X_{k+1}, S_k) > B_k$  then
7:      $X_{k+1} \leftarrow Z_k$ 
8:   end if
9:    $(Z_{k+1}, S_{k+1}) \leftarrow \operatorname{argmin}_{Z, S} \{\xi \|S\|_1 + \langle \nabla f_\mu(X_{k+1}), Z - X_{k+1} \rangle + \frac{\rho}{2} \|Z - X_{k+1}\|_F^2 : (Z, S) \in \chi\}$ 
10:   $Y_{k+1} \leftarrow -\nabla f_\mu(X_{k+1}) + \rho(X_{k+1} - Z_{k+1})$ 
11:   $k \leftarrow k + 1$ 
12: end while
```

Using the fast version of **Algorithm 3**, a fast version of **Algorithm 4** with $\mathcal{O}(\rho/k^2)$ convergence rate, employing partial splitting and alternating linearization, can be constructed. This fast version can compute an ϵ -optimal and feasible solution to problem (P) in $\mathcal{O}(1/\epsilon)$ iterations. Moreover, like the proximal gradient methods described earlier, each iteration for these methods can be computed efficiently. The subproblems

to be solved at each iteration of **Algorithm 4** and its fast version have the following generic form:

$$\min_{X \in \mathbb{R}^{m \times n}} f_\mu(X) + \langle Q, X - \tilde{X} \rangle + \frac{\rho}{2} \|X - \tilde{X}\|_F^2, \quad (4.5)$$

$$\min_{Z, S \in \mathbb{R}^{m \times n}} \{\xi \|S\|_1 + \langle Q, Z - \tilde{Z} \rangle + \frac{\rho}{2} \|Z - \tilde{Z}\|_F^2 : (Z, S) \in \chi\}. \quad (4.6)$$

Let $U \text{ diag}(\sigma) V^T$ denote the singular value decomposition of the matrix $\tilde{X} - Q/\rho$, then X^* , the minimizer of the subproblem in (4.5), can be easily computed as $U \text{ diag}\left(\sigma - \frac{\sigma}{\max\{\rho\sigma, 1+\rho\mu\}}\right) V^T$. And Lemma 3.1 shows how to solve the subproblem in (4.6).

5. Non-smooth Augmented Lagrangian Algorithm. **Algorithm 5** is a Non-Smooth Augmented Lagrangian Algorithm (NSA) that solves the non-smooth problem (P). The subproblem in Step 4 of **Algorithm 5** is a matrix shrinkage problem and can be solved efficiently by computing a singular value decomposition (SVD) of an $m \times n$ matrix; and Lemma 3.1 shows that the subproblem in Step 6 can also be solved efficiently.

Algorithm 5 NSA(Z_0, Y_0)

```

1: input:  $Z_0 \in \mathbb{R}^{m \times n}, Y_0 \in \mathbb{R}^{m \times n}$ 
2:  $k \leftarrow 0$ 
3: while  $k \leq 0$  do
4:    $X_{k+1} \leftarrow \text{argmin}_X \{\|X\|_* + \langle Y_k, X - Z_k \rangle + \frac{\rho_k}{2} \|X - Z_k\|_F^2\}$ 
5:    $\hat{Y}_{k+1} \leftarrow Y_k + \rho_k (X_{k+1} - Z_k)$ 
6:    $(Z_{k+1}, S_{k+1}) \leftarrow \text{argmin}_{\{(Z, S) : \|Z + S - D\|_F^2 \leq \delta^2\}} \{\xi \|S\|_1 + \langle -Y_k, Z - X_{k+1} \rangle + \frac{\rho_k}{2} \|Z - X_{k+1}\|_F^2\}$ 
7:   Let  $\theta_k$  be an optimal Lagrangian dual variable for the  $\frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}$  constraint
8:    $Y_{k+1} \leftarrow Y_k + \rho_k (X_{k+1} - Z_{k+1})$ 
9:   Choose  $\rho_{k+1}$  such that  $\rho_{k+1} \geq \rho_k$ 
10:   $k \leftarrow k + 1$ 
11: end while

```

We now prove that Algorithm NSA converges under fairly mild conditions on the sequence $\{\rho_k\}_{k \in \mathbb{Z}_+}$ of penalty parameters. We first need the following lemma, which extends the similar result given in [6] to partial splitting of variables.

LEMMA 5.1. *Suppose that $\delta > 0$. Let $\{X_k, Z_k, S_k, Y_k, \theta_k\}_{k \in \mathbb{Z}_+}$ be the sequence produced by Algorithm NSA. $(X^*, X^*, S^*) = \text{argmin}_{X, Z, S} \{\|X\|_* + \xi \|S\|_1 : \frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}, X = Z\}$ be any optimal solution, $Y^* \in \mathbb{R}^{m \times n}$ and $\theta^* \geq 0$ be any optimal Lagrangian duals corresponding to the constraints $X = Z$ and $\frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}$, respectively. Then $\{\|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2\}_{k \in \mathbb{Z}_+}$ is a non-increasing sequence and*

$$\begin{aligned} \sum_{k \in \mathbb{Z}_+} \|Z_{k+1} - Z_k\|_F^2 &< \infty & \sum_{k \in \mathbb{Z}_+} \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2 &< \infty, \\ \sum_{k \in \mathbb{Z}_+} \rho_k^{-1} \langle -Y_{k+1} + Y^*, S_{k+1} - S^* \rangle &< \infty & \sum_{k \in \mathbb{Z}_+} \rho_k^{-1} \langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle &< \infty, \\ \sum_{k \in \mathbb{Z}_+} \rho_k^{-1} \langle Y^* - Y_{k+1}, X^* + S^* - Z_{k+1} - S_{k+1} \rangle &< \infty. \end{aligned}$$

Proof. See Appendix B for the proof. $\square$

Given partially split SPCP problem, $\min_{X, Z, S} \{\|X\|_* + \xi \|S\|_1 : X = Z, (Z, S) \in \chi\}$, let $\mathcal{L}$ be its Lagrangian function

$$\mathcal{L}(X, Z, S; Y, \theta) = \|X\|_* + \xi \|S\|_1 + \langle Y, X - Z \rangle + \frac{\theta}{2} (\|Z + S - D\|_F^2 - \delta^2). \quad (5.1)$$

THEOREM 5.2. *Suppose that $\delta > 0$. Let $\{X_k, Z_k, S_k, Y_k, \theta_k\}_{k \in \mathbb{Z}_+}$ be the sequence produced by Algorithm NSA. Choose $\{\rho_k\}_{k \in \mathbb{Z}_+}$ such that*

- (i) $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k} = \infty$: Then $\lim_{k \in \mathbb{Z}_+} Z_k = \lim_{k \in \mathbb{Z}_+} X_k = X^*$, $\lim_{k \in \mathbb{Z}_+} S_k = S^*$ such that $(X^*, S^*) = \operatorname{argmin}\{\|X\|_* + \xi \|S\|_1 : \|X + S - D\|_F \leq \delta\}$.
- (ii) $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k^2} = \infty$: If $\|D - X^*\|_F \neq \delta$, then $\lim_{k \in \mathbb{Z}_+} \theta_k = \theta^* \geq 0$ and $\lim_{k \in \mathbb{Z}_+} Y_k = Y^*$ such that $(X^*, X^*, S^*, Y^*, \theta^*)$ is a saddle point of the Lagrangian function $\mathcal{L}$ in (5.1). Otherwise, if $\|D - X^*\|_F = \delta$, then there exists a limit point, (Y^*, θ^*) , of the sequence $\{Y_k, \theta_k\}_{k \in \mathbb{Z}_+}$ such that $(Y^*, \theta^*) = \operatorname{argmax}_{Y, \theta} \{\mathcal{L}(X^*, X^*, S^*; Y, \theta) : \theta \geq 0\}$.

Remark 5.1. Requiring $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k} = \infty$ is similar to the condition in Theorem 2 in [6], which is needed to show that Algorithm I-ALM converges to an optimal solution of the robust PCA problem.

Remark 5.2. Let $D = X^0 + S^0 + \zeta^0$ such that $\|\zeta^0\|_F \leq \delta$ and (X^0, S^0) satisfies the assumptions of Theorem 1.2. If $\|S^0\|_F > \sqrt{Cmn}\delta$, then with very high probability, $\|D - X^*\|_F > \delta$, where C is the numerical constant defined in Theorem 1.2. Therefore, most of the time in applications, one does not encounter the case where $\|D - X^*\|_F = \delta$.

Proof. From Lemma 5.1 and the fact that $X_{k+1} - Z_{k+1} = \frac{1}{\rho_k} (Y_{k+1} - Y_k)$ for all $k \geq 1$, we have

$$\infty > \sum_{k \in \mathbb{Z}_+} \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2 = \sum_{k \in \mathbb{Z}_+} \|X_{k+1} - Z_{k+1}\|_F^2.$$

Hence, $\lim_{k \in \mathbb{Z}_+} (X_k - Z_k) = 0$.

Let $(X^\#, X^\#, S^\#) = \operatorname{argmin}_{X, Z, S} \{\|X\|_* + \xi \|S\|_1 : \frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}, X = Z\}$ be any optimal solution, $Y^\# \in \mathbb{R}^{m \times n}$ and $\theta^\# \geq 0$ be any optimal Lagrangian duals corresponding to $X = Z$ and $\frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}$ constraints, respectively and $f^* := \|X^\#\|_* + \xi \|S^\#\|_1$.

Moreover, let $\chi = \{(Z, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} : \|Z + S - D\|_F \leq \delta\}$ and $\mathbf{1}_\chi(Z, S)$ denote the indicator function of the closed convex set χ , i.e. $\mathbf{1}_\chi(Z, S) = 0$ if $(Z, S) \in \chi$; otherwise, $\mathbf{1}_\chi(Z, S) = \infty$. Since the sequence $\{(Z_k, S_k)\}_{k \in \mathbb{Z}_+}$ produced by NSA is a feasible sequence for the set χ , we have $\mathbf{1}_\chi(Z_k, S_k) = 0$ for all $k \geq 1$. Hence, the following inequality is true for all $k \geq 0$

$$\begin{aligned} & \|X_k\|_* + \xi \|S_k\|_1 \\ &= \|X_k\|_* + \xi \|S_k\|_1 + \mathbf{1}_\chi(Z_k, S_k), \\ &\leq \|X^\#\|_* + \xi \|S^\#\|_1 + \mathbf{1}_\chi(X^\#, S^\#) - \langle -\hat{Y}_k, X^\# - X_k \rangle - \langle -Y_k, S^\# - S_k \rangle - \langle Y_k, X^\# + S^\# - Z_k - S_k \rangle, \\ &= f^* + \langle -\hat{Y}_k + Y^\#, X_k - X^\# \rangle + \langle -Y_k + Y^\#, S_k - S^\# \rangle + \langle Y^\# - Y_k, X^\# + S^\# - Z_k - S_k \rangle \\ &\quad + \langle Y^\#, Z_k - X_k \rangle, \end{aligned} \tag{5.2}$$

where the inequality follows from the convexity of norms and the fact that $-Y_k \in \xi \partial \|S_k\|_1$, $-\hat{Y}_k \in \partial \|X_k\|_*$ and $(Y_k, Y_k) \in \partial \mathbf{1}_\chi(Z_k, S_k)$; the final equality follows from rearranging the terms and the fact that $(X^\#, S^\#) \in \chi$.

From Lemma 5.1, we have

$$\sum_{k \in \mathbb{Z}_+} \rho_{k-1}^{-1} \left(\langle -\hat{Y}_k + Y^\#, X_k - X^\# \rangle + \langle -Y_k + Y^\#, S_k - S^\# \rangle + \langle Y^\# - Y_k, X^\# + S^\# - Z_k - S_k \rangle \right) < \infty.$$

Since $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k} = \infty$, there exists $\mathcal{K} \subset \mathbb{Z}_+$ such that

$$\lim_{k \in \mathcal{K}} \left(\langle -\hat{Y}_k + Y^\#, X_k - X^\# \rangle + \langle -Y_k + Y^\#, S_k - S^\# \rangle + \langle Y^\# - Y_k, X^\# + S^\# - Z_k - S_k \rangle \right) = 0. \tag{5.3}$$

(5.3) and the fact that $\lim_{k \in \mathbb{Z}_+} Z_k - X_k = 0$ imply that along $\mathcal{K}$ (5.2) converges to $f^* = \|X^\#\|_* + \xi \|S^\#\|_1 = \min\{\|X\|_* + \xi \|S\|_1 : (X, S) \in \chi\}$; hence along $\mathcal{K}$ subsequence, $\{\|X_k\|_* + \xi \|S_k\|_1\}_{k \in \mathcal{K}}$ is a bounded sequence. Therefore, there exists $\mathcal{K}^* \subset \mathcal{K} \subset \mathbb{Z}_+$ such that $\lim_{k \in \mathcal{K}^*} (X_k, S_k) = (X^*, S^*)$. Also, since $\lim_{k \in \mathbb{Z}_+} Z_k - X_k = 0$ and $(Z_k, S_k) \in \chi$ for all $k \geq 1$, we also have $(X^*, S^*) = \lim_{k \in \mathcal{K}^*} (Z_k, S_k) \in \chi$. Since the limit of both sides of (5.2) along $\mathcal{K}^*$ gives $\|X^*\|_* + \xi \|S^*\|_1 = \lim_{k \in \mathcal{K}^*} \|X_k\|_* + \xi \|S_k\|_1 \leq f^*$ and $(X^*, S^*) \in \chi$, we conclude that $(X^*, S^*) = \operatorname{argmin}\{\|X\|_* + \xi \|S\|_1 : (X, S) \in \chi\}$.

It is also true that (X^*, X^*, S^*) is an optimal solution to an equivalent problem: $\operatorname{argmin}_{X, Z, S} \{\|X\|_* + \xi \|S\|_1 : \frac{1}{2}\|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}, X = Z\}$. Now, let $\bar{Y} \in \mathbb{R}^{m \times n}$ and $\bar{\theta} \geq 0$ be optimal Lagrangian duals corresponding to $X = Z$ and $\frac{1}{2}\|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}$ constraints, respectively. From Lemma 5.1, it follows that $\{\|Z_k - X^*\|_F^2 + \rho_k^{-2}\|Y_k - \bar{Y}\|_F^2\}_{k \in \mathbb{Z}_+}$ is a bounded non-increasing sequence. Hence, it has a unique limit point, i.e.

$$\lim_{k \in \mathbb{Z}_+} \|Z_k - X^*\|_F^2 = \lim_{k \in \mathbb{Z}_+} \|Z_k - X^*\|_F^2 + \rho_k^{-2}\|Y_k - \bar{Y}\|_F^2 = \lim_{k \in \mathcal{K}^*} \|Z_k - X^*\|_F^2 + \rho_k^{-2}\|Y_k - \bar{Y}\|_F^2 = 0,$$

where the equalities follow from the facts that $\lim_{k \in \mathcal{K}^*} Z_k = X^*$, $\mu_k \nearrow \infty$ as $k \rightarrow \infty$ and $\{\hat{Y}_k\}_{k \in \mathbb{Z}_+}$, $\{Y_k\}_{k \in \mathbb{Z}_+}$ are bounded sequences. $\lim_{k \in \mathbb{Z}_+} \|Z_k - X^*\|_F = 0$ and $\lim_{k \in \mathbb{Z}_+} Z_k - X_k = 0$ imply that $\lim_{k \in \mathbb{Z}_+} X_k = X^*$.

Using Lemma 3.1 for the k -th subproblem given in Step 6 in **Algorithm 5**, we have

$$S_{k+1} = \operatorname{sign} \left(D - \left(X_{k+1} + \frac{1}{\rho_k} Y_k \right) \right) \odot \max \left\{ \left| D - \left(X_{k+1} + \frac{1}{\rho_k} Y_k \right) \right| - \xi \frac{(\rho_k + \theta_k)}{\rho_k \theta_k} E, \mathbf{0} \right\}, \quad (5.4)$$

$$Z_{k+1} = \frac{\theta_k}{\rho_k + \theta_k} (D - S_{k+1}) + \frac{\rho_k}{\rho_k + \theta_k} \left(X_{k+1} + \frac{1}{\rho_k} Y_k \right). \quad (5.5)$$

If $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F \leq \delta$, then $\theta_k = 0$; otherwise, $\theta_k > 0$ is the unique solution such that $\phi_k(\theta_k) = \delta$, where

$$\phi_k(\theta) := \left\| \min \left\{ \frac{\xi}{\theta} E, \frac{\rho_k}{\rho_k + \theta} \left| D - \left(X_{k+1} + \frac{1}{\rho_k} Y_k \right) \right| \right\} \right\|_F. \quad (5.6)$$

In the following, it is shown that the sequence $\{S_k\}_{k \in \mathbb{Z}_+}$ has a unique limit point S^* . Since $\lim_{k \in \mathbb{Z}_+} X_k = X^*$, $\{Y_k\}_{k \in \mathbb{Z}_+}$ is a bounded sequence and $\rho_k \nearrow \infty$ as $k \rightarrow \infty$, we have $\lim_{k \in \mathbb{Z}_+} X_{k+1} + \frac{1}{\rho_k} Y_k = X^*$.

Case 1: $\|D - X^*\|_F \leq \delta$. Previously, we have shown that there exists a subsequence $\mathcal{K}^* \subset \mathbb{Z}_+$ such that $\lim_{k \in \mathcal{K}^*} (X_k, S_k) = (X^*, S^*) = \operatorname{argmin}_{X, S} \{\|X\|_* + \xi \|S\|_1 : \|X + S - D\|_F \leq \delta\}$. On the other hand, since $\|D - X^*\|_F \leq \delta$, $(X^*, \mathbf{0})$ is a feasible solution. Hence, $\|X^*\|_* + \xi \|S^*\| \leq \|X^*\|_*$, which implies that $S^* = \mathbf{0}$.

$$\begin{aligned} & \|X_k\|_* + \xi \|S_k\|_1 \\ &= \|X_k\|_* + \xi \|S_k\|_1 + \mathbf{1}_X(Z_k, S_k), \\ &\leq \|X^*\|_* + \xi \|\mathbf{0}\|_1 + \mathbf{1}_X(X^*, \mathbf{0}) - \langle -\hat{Y}_k, X^* - X_k \rangle - \langle -Y_k, \mathbf{0} - S_k \rangle - \langle Y_k, X^* + \mathbf{0} - Z_k - S_k \rangle, \\ &= \|X^*\|_* + \langle \hat{Y}_k, X^* - X_k \rangle + \langle Y_k, Z_k - X^* \rangle. \end{aligned} \quad (5.7)$$

Since the sequences $\{Y_k\}_{k \in \mathbb{Z}_+}$ and $\{\hat{Y}_k\}_{k \in \mathbb{Z}_+}$ are bounded and $\lim_{k \in \mathbb{Z}_+} X_k = \lim_{k \in \mathbb{Z}_+} Z_k = X^*$, taking the limit on both sides of (5.7), we have

$$\begin{aligned} & \|X^*\|_* + \xi \lim_{k \in \mathbb{Z}_+} \|S_k\|_1 = \lim_{k \in \mathbb{Z}_+} \|X_k\|_* + \xi \|S_k\|_1 \\ &= \lim_{k \in \mathbb{Z}_+} \|X^*\|_* + \langle \hat{Y}_k, X^* - X_k \rangle + \langle Y_k, Z_k - X^* \rangle = \|X^*\|_*. \end{aligned}$$

Therefore, $\lim_{k \in \mathbb{Z}_+} \|S_k\|_1 = 0$, which implies that $\lim_{k \in \mathbb{Z}_+} S_k = S^* = \mathbf{0}$.

Case 2: $\|D - X^*\|_F > \delta$. Since $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F \rightarrow \|D - X^*\|_F > \delta$, there exists $K \in \mathbb{Z}_+$ such that for all $k \geq K$, $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F > \delta$. For all $k \geq K$, $\phi_k(\cdot)$ is a continuous and strictly decreasing function of θ for $\theta \geq 0$. Hence, inverse function $\phi_k^{-1}(\cdot)$ exists around δ for all $k \geq K$. Thus, $\phi_k(0) = \|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F > \delta$ and $\lim_{\theta \rightarrow \infty} \phi_k(\theta) = 0$ imply that $\theta_k = \phi_k^{-1}(\delta) > 0$ for all $k \geq K$. Moreover, $\phi_k(\theta) \leq \phi(\theta) := \left\| \frac{\xi}{\theta} E \right\|_F$ implies that $\theta_k \leq \frac{\xi \sqrt{mn}}{\delta}$ for all $k \geq K$. Therefore, $\{\theta_k\}_{k \in \mathbb{Z}_+}$ is a bounded sequence, which has a convergent subsequence $\mathcal{K}_\theta \subset \mathbb{Z}_+$ such that $\lim_{k \in \mathcal{K}_\theta} \theta_k = \theta^*$. We also have $\phi_k(\theta) \rightarrow \phi_\infty(\theta)$ pointwise for all $0 \leq \theta \leq \frac{\xi \sqrt{mn}}{\delta}$, where

$$\phi_\infty(\theta) := \left\| \min \left\{ \frac{\xi}{\theta} E, \|D - X^*\| \right\} \right\|_F. \quad (5.8)$$

Since $\phi_k(\theta_k) = \delta$ for all $k \geq K$, we have

$$\delta = \lim_{k \in \mathcal{K}} \phi_k(\theta_k) = \left\| \min \left\{ \frac{\xi}{\theta_k} E, \frac{\rho_k}{\rho_k + \theta_k} \left| D - \left(X_{k+1} + \frac{1}{\rho_k} Y_k \right) \right| \right\} \right\|_F = \phi_\infty(\theta^*). \quad (5.9)$$

Note that since $\|D - X^*\|_F > \delta$, ϕ_∞ is invertible around δ , i.e. ϕ_∞^{-1} exists around δ . Thus, $\theta^* = \phi_\infty^{-1}(\delta)$. Since $\mathcal{K}_\theta$ is an arbitrary subsequence, we can conclude that $\theta^* := \lim_{k \in \mathbb{Z}_+} \theta_k = \phi_\infty^{-1}(\delta)$. Since there exists $\theta^* > 0$ such that $\theta^* = \lim_{k \in \mathbb{Z}_+} \theta_k$, taking the limit on both sides of (5.4), we have

$$S^* := \lim_{k \in \mathbb{Z}_+} S_{k+1} = \text{sign}(D - X^*) \odot \max \left\{ |D - X^*| - \frac{\xi}{\theta^*} E, \mathbf{0} \right\}, \quad (5.10)$$

and this completes the first part of the theorem.

Now, we will show that if $\|D - X^*\|_F \neq \delta$, then the sequences $\{\theta_k\}_{k \in \mathbb{Z}_+}$ and $\{Y_k\}_{k \in \mathbb{Z}_+}$ have unique limits. Note that from (B.3), it follows that $Y_k = \theta_{k-1}(Z_k + S_k - D)$ for all $k \geq 1$. First suppose that $\|D - X^*\|_F < \delta$. Since $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F \rightarrow \|D - X^*\|_F < \delta$, there exists $K \in \mathbb{Z}_+$ such that for all $k \geq K$, $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F < \delta$. Thus, from Lemma 3.1 for all $k \geq K$, $\theta_k = 0$, $S_{k+1} = 0$, $Z_{k+1} = X_{k+1} + \frac{1}{\rho_k} Y_k$, which implies that $\theta^* := \lim_{k \in \mathbb{Z}_+} \theta_k = 0$ and $Y^* = \lim_{k \in \mathbb{Z}_+} Y_k = \lim_{k \in \mathbb{Z}_+} \theta_{k-1}(Z_k + S_k - D) = \mathbf{0}$ since $S^* = \lim_{k \in \mathbb{Z}_+} S_k = \lim_{k \in \mathbb{Z}_+} S_k = 0$, $\lim_{k \in \mathbb{Z}_+} Z_k = X^*$ and $\|D - X^*\|_F < \delta$. Now suppose that $\|D - X^*\|_F > \delta$. In Case 2 above we have shown that $\theta^* = \lim_{k \in \mathbb{Z}_+} \theta_k$. Hence, there exists $Y^* \in \mathbb{R}^{m \times n}$ such that $Y^* = \lim_{k \in \mathbb{Z}_+} \theta_{k-1}(Z_k + S_k - D) = \theta^*(X^* + S^* - D)$.

Suppose that $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k^2} = \infty$. From Lemma 5.1, we have $\sum_{k \in \mathbb{Z}_+} \|Z_{k+1} - Z_k\|_F^2 < \infty$. Equivalently, the series can be written as

$$\infty > \sum_{k \in \mathbb{Z}_+} \|Z_{k+1} - Z_k\|_F^2 = \sum_{k \in \mathbb{Z}_+} \rho_k^{-2} \|\hat{Y}_{k+1} - Y_{k+1}\|_F^2. \quad (5.11)$$

Since $\sum_{k \in \mathbb{Z}_+} \frac{1}{\rho_k^2} = \infty$, there exists a subsequence $\mathcal{K} \subset \mathbb{Z}_+$ such that $\lim_{k \in \mathcal{K}} \|\hat{Y}_{k+1} - Y_{k+1}\|_F^2 = 0$. Hence, $\lim_{k \in \mathcal{K}} \rho_k^2 \|Z_{k+1} - Z_k\|_F^2 = 0$, i.e. $\lim_{k \in \mathcal{K}} \rho_k (Z_{k+1} - Z_k) = 0$.

Using (B.1), (B.2) and (B.3), we have

$$0 \in \partial \|X_{k+1}\|_* + \theta_k (Z_{k+1} + S_{k+1} - D) + \rho_k (Z_{k+1} - Z_k), \quad (5.12)$$

$$0 \in \xi \partial \|S_{k+1}\|_1 + \theta_k (Z_{k+1} + S_{k+1} - D). \quad (5.13)$$

If $\|D - X^*\| \neq \delta$, then there exists $Y^* \in \mathbb{R}^{m \times n}$ such that $Y^* = \lim_{k \in \mathbb{Z}_+} \theta_{k-1}(Z_k + S_k - D) = \theta^*(X^* + S^* - D)$. Taking the limit of (5.12), (5.13) along $\mathcal{K} \subset \mathbb{Z}_+$ and using the fact that $\lim_{k \in \mathcal{K}} \rho_k (Z_{k+1} - Z_k) = 0$, we have

$$0 \in \partial \|X^*\|_* + \theta^*(X^* + S^* - D), \quad (5.14)$$

$$0 \in \xi \partial \|S^*\|_1 + \theta^*(X^* + S^* - D). \quad (5.15)$$

(5.14) and (5.15) together imply that (X^*, S^*) , $Y^* = \theta^*(X^* + S^* - D)$ and θ^* satisfy KKT optimality conditions for the problem $\min_{X, Z, S} \{\|X\|_* + \xi \|S\|_1 : \frac{1}{2} \|Z + S - D\|_F^2 \leq \frac{\delta^2}{2}, X = Z\}$. Hence, $(X^*, X^*, S^*, Y^*, \theta^*)$ is a saddle point of the Lagrangian function

$$\mathcal{L}(X, Z, S; Y, \theta) = \|X\|_* + \xi \|S\|_1 + \langle Y, X - Z \rangle + \frac{\theta}{2} (\|Z + S - D\|_F^2 - \delta^2).$$

Suppose that $\|D - X^*\|_F = \delta$. Fix $k > 0$. If $\|D - (X_{k+1} + \frac{1}{\rho_k} Y_k)\|_F \leq \delta$, then $\theta_k = 0$. Otherwise, $\theta_k > 0$ and as shown in case 2 in the first part of the proof $\theta_k \leq \frac{\xi \sqrt{mn}}{\delta}$. Thus, for any $k > 0$, $0 \leq \theta_k \leq \frac{\xi \sqrt{mn}}{\delta}$. Since $\{\theta_k\}_{k \in \mathbb{Z}_+}$ is a bounded sequence, there exists a further subsequence $\mathcal{K}_\theta \subset \mathcal{K}$ such that $\theta^* := \lim_{k \in \mathcal{K}_\theta} \theta_{k-1}$ and $Y^* := \lim_{k \in \mathcal{K}_\theta} \theta_{k-1}(Z_k + S_k - D) = \theta^*(X^* + S^* - D)$ exist. Thus, taking the limit of (5.12), (5.13) along $\mathcal{K}_\theta \subset \mathbb{Z}_+$ and using the facts that $\lim_{k \in \mathcal{K}} \rho_k (Z_{k+1} - Z_k) = 0$ and $X^* = \lim_{k \in \mathbb{Z}_+} X_k = \lim_{k \in \mathbb{Z}_+} Z_k$, $S^* = \lim_{k \in \mathbb{Z}_+} S_k$ exist, we conclude that $(X^*, X^*, S^*, Y^*, \theta^*)$ is a saddle point of the Lagrangian function $\mathcal{L}(X, Z, S; Y, \theta)$. $\square$

6. Numerical experiments. Our preliminary numerical experiments showed that among the four algorithms discussed in this paper, NSA is the fastest. It also has very few parameters that need to be tuned. Therefore, we only report the results for NSA. We conducted two sets of numerical experiments with NSA to solve (1.4), where $\xi = \frac{1}{\sqrt{\max\{m,n\}}}$. In the first set we solved randomly generated instances of the stable principle component pursuit problem. In this setting, first we tested only NSA to see how the run times scale with respect to problem parameters and size; then we compared NSA with another alternating direction augmented Lagrangian algorithm ASALM [10]. In the second set of experiments, we ran NSA and ASALM to extract moving objects from an airport security noisy video [5].

6.1. Random Stable Principle Component Pursuit Problems. We tested NSA on randomly generated stable principle component pursuit problems. The data matrices for these problems, $D = X^0 + S^0 + \zeta^0$, were generated as follows

- i. $X^0 = UV^T$, such that $U \in \mathbb{R}^{n \times r}$, $V \in \mathbb{R}^{n \times r}$ for $r = c_r n$ and $U_{ij} \sim \mathcal{N}(0, 1)$, $V_{ij} \sim \mathcal{N}(0, 1)$ for all i, j are independent standard Gaussian variables and $c_r \in \{0.05, 0.1\}$,
- ii. $\Lambda \subset \{(i, j) : 1 \leq i, j \leq n\}$ such that cardinality of Λ , $|\Lambda| = p$ for $p = c_p n^2$ and $c_p \in \{0.05, 0.1\}$,
- iii. $S_{ij}^0 \sim \mathcal{U}[-100, 100]$ for all $(i, j) \in \Lambda$ are independent uniform random variables between -100 and 100 ,
- iv. $\zeta_{ij}^0 \sim \mathcal{N}(0, 1)$ for all i, j are independent Gaussian variables.

We created 10 random problems of size $n \in \{500, 1000, 1500\}$, i.e. $D \in \mathbb{R}^{n \times n}$, for each of the two choices of c_r and c_p using the procedure described above, where ϱ was set such that signal-to-noise ratio of D is either 80dB or 45dB. Signal-to-noise ratio of D is given by

$$\text{SNR}(D) = 10 \log_{10} \left(\frac{\mathbb{E} [\|X^0 + S^0\|_F^2]}{\mathbb{E} [\|\zeta^0\|_F^2]} \right) = 10 \log_{10} \left(\frac{c_r n + c_s 100^2/3}{\varrho^2} \right). \quad (6.1)$$

Hence, for a given SNR value, we selected ϱ according to (6.1). Table 6.1 displays the ϱ value we have used in our experiments. As in [10], we set $\delta = \sqrt{(n + \sqrt{8n})} \varrho$ in (1.4) in the first set of experiments for both

TABLE 6.1
 ϱ values depending on the experimental setting

SNR	n	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
80dB	500	0.0014	0.0019	0.0015	0.0020
	1000	0.0015	0.0020	0.0016	0.0021
	1500	0.0016	0.0020	0.0018	0.0022
45dB	500	0.0779	0.1064	0.0828	0.1101
	1000	0.0828	0.1101	0.0918	0.1171
	1500	0.0874	0.1136	0.1001	0.1236

NSA and ASALM.

Our code for NSA was written in MATLAB 7.2 and can be found at <http://www.columbia.edu/~nsa2106>. We terminated the algorithm when

$$\frac{\|(X_{k+1}, S_{k+1}) - (X_k, S_k)\|_F}{\|(X_k, S_k)\|_F + 1} \leq \varrho. \quad (6.2)$$

The results of our experiments are displayed in Tables 6.2 and 6.3. In Table 6.2, the row labeled **CPU** lists the running time of NSA in *seconds* and the row labeled **SVD#** lists the number of partial singular value decomposition (SVD) computed by NSA. The minimum, average and maximum CPU times and number of partial SVD taken over the 10 random instances are given for each choice of n , c_r and c_p values. Table C.3 and Table C.4 in the appendix list additional error statistics.

With the stopping condition given in (6.2), the solutions produced by NSA have $\frac{\|X^{sol} + S^{sol} - D\|_F}{\|D\|_F}$ approximately 1.5×10^{-4} when $\text{SNR}(D) = 80\text{dB}$ and 5×10^{-3} when $\text{SNR}(D) = 45\text{dB}$, regardless of the problem dimension n and the problem parameters related to the rank and sparsity of D , i.e. c_r and c_p . After thresholding the singular values of X^{sol} that were less than 1×10^{-12} , NSA found the true rank in all 120

random problems solved when $\text{SNR}(D) = 80\text{dB}$, and it found the true rank for 113 out of 120 problems when $\text{SNR}(D) = 45\text{dB}$, while for 6 of the remaining problems $\text{rank}(X^{\text{sol}})$ is off from $\text{rank}(X^0)$ only by 1. Table 6.2 shows that the number of partial SVD was a very slightly increasing function of n , c_r and c_p . Moreover, Table 6.3 shows that the relative error of the solution $(X^{\text{sol}}, S^{\text{sol}})$ was almost constant for different n , c_r and c_p values.

TABLE 6.2
NSA: Solution time for decomposing $D \in \mathbb{R}^{n \times n}$, $n \in \{500, 1000, 1500\}$

SNR	n	Field	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
			min/avg/max	min/avg/max	min/avg/max	min/avg/max
80dB	500	SVD#	9/9.0/9	9/9.5/10	10/10.0/10	11/11/11
		CPU	3.2/4.4/5.1	3.6/5.1/6.6	4.3/5.2/6.4	5.0/6.2/8.1
	1000	SVD#	9/9.9/10	10/10.0/10	11/11/11	12/12.0/12
		CPU	16.5/19.6/22.4	14.6/20.7/24.3	25.2/26.9/29.1	27.9/31.2/36.3
	1500	SVD#	10/10.0/10	10/10.9/11	12/12.0/12	12/12.2/13
		CPU	38.6/44.1/46.6	43.7/48.6/51.9	78.6/84.1/90.8	80.7/97.7/155.2
45dB	500	SVD#	6/6/6	6/6.9/7	7/7.1/8	8/8/8
		CPU	2.3/2.9/4.2	2.9/3.6/4.5	2.9/3.9/6.2	3.5/4.2/6.0
	1000	SVD#	7/7.0/7	7/7.0/7	8/8.1/9	9/9.0/9
		CPU	11.5/13.4/17.4	10.6/13.3/17.9	17.1/18.7/20.7	19.7/23.8/28.9
	1500	SVD#	7/7.9/8	8/8.0/8	9/9.0/9	9/9.0/9
		CPU	34.1/37.7/44.0	30.7/37.1/45.6	55.6/59.0/63.7	55.9/59.7/64.8

TABLE 6.3
NSA: Solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n \in \{500, 1000, 1500\}$

SNR	n	Relative Error	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
			avg / max	avg / max	avg / max	avg / max
80dB	500	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	4.0E-4 / 4.2E-4	5.8E-4 / 8.5E-4	3.6E-4 / 3.9E-4	4.4E-4 / 4.5E-4
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	1.7E-4 / 1.8E-4	1.6E-4 / 2.5E-4	1.6E-4 / 1.8E-4	1.3E-4 / 1.3E-4
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	2.0E-4 / 2.4E-4	3.8E-4 / 4.1E-4	2.2E-4 / 2.2E-4	2.8E-4 / 2.9E-4
	1000	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	1.2E-4 / 1.4E-4	1.5E-4 / 1.6E-4	1.2E-4 / 1.3E-4	1.1E-4 / 1.1E-4
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	1.8E-4 / 2.2E-4	2.1E-4 / 2.6E-4	1.3E-4 / 1.3E-4	2.8E-4 / 2.9E-4
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	1.3E-4 / 1.6E-4	9.6E-5 / 1.1E-4	8.1E-5 / 8.5E-5	1.3E-4 / 1.4E-4
	1500	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	6.0E-3 / 6.2E-3	8.0E-3 / 9.2E-3	6.1E-3 / 6.3E-3	8.1E-3 / 8.2E-3
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	2.1E-3 / 2.2E-3	2.3E-3 / 2.7E-3	2.2E-3 / 2.3E-3	2.7E-3 / 2.9E-3
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	4.1E-3 / 4.2E-3	6.1E-3 / 6.2E-3	4.6E-3 / 4.7E-3	6.0E-3 / 6.5E-3
45dB	500	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	1.9E-3 / 1.9E-3	2.4E-3 / 2.5E-3	2.3E-3 / 3.5E-3	3.1E-3 / 3.7E-3
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	3.4E-3 / 3.6E-3	4.7E-3 / 4.7E-3	3.9E-3 / 4.0E-3	5.3E-3 / 5.3E-3
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	1.8E-3 / 1.8E-3	2.3E-3 / 2.3E-3	2.6E-3 / 3.5E-3	3.1E-3 / 3.1E-3
	1000	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	4.1E-3 / 4.2E-3	6.1E-3 / 6.2E-3	4.6E-3 / 4.7E-3	6.0E-3 / 6.5E-3
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	1.9E-3 / 1.9E-3	2.4E-3 / 2.5E-3	2.3E-3 / 3.5E-3	3.1E-3 / 3.7E-3
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	3.4E-3 / 3.6E-3	4.7E-3 / 4.7E-3	3.9E-3 / 4.0E-3	5.3E-3 / 5.3E-3
	1500	$\frac{\ X^{\text{sol}} - X^0\ _F}{\ X^0\ _F}$	1.8E-3 / 1.8E-3	2.3E-3 / 2.3E-3	2.6E-3 / 3.5E-3	3.1E-3 / 3.1E-3
		$\frac{\ X^{\text{sol}} - X^0\ _F}{\ S^{\text{sol}} - S^0\ _F}$	3.4E-3 / 3.6E-3	4.7E-3 / 4.7E-3	3.9E-3 / 4.0E-3	5.3E-3 / 5.3E-3
		$\frac{\ S^{\text{sol}} - S^0\ _F}{\ S^0\ _F}$	1.8E-3 / 1.8E-3	2.3E-3 / 2.3E-3	2.6E-3 / 3.5E-3	3.1E-3 / 3.1E-3

Next, we compared NSA with ASALM [10] for a fixed problem size, i.e. $n = 1500$ where $D \in \mathbb{R}^{n \times n}$. In all the numerical experiments, we terminated NSA according to (6.2). For random problems with $\text{SNR}(D) = 80\text{dB}$, we terminated ASALM according to (6.2). However, for random problems with $\text{SNR}(D) = 45\text{dB}$, ASALM produced solutions with 99% relative errors when (6.2) was used. Therefore, for random problems with $\text{SNR}(D) = 45\text{dB}$, we terminated ASALM either when it computed a solution with better relative errors comparing to NSA solution for the same problem or when an iterate satisfied (6.2) with the righthand side replaced by 0.1ρ . The code for ASALM was obtained from the authors of [10].

The comparison results are displayed in Table 6.5 and Table 6.6. In Table 6.5, the row labeled **CPU** lists the running time of each algorithm in *seconds* and the row labeled **SVD#** lists the number of partial SVD computation of each algorithm. In Table 6.5, the minimum, average and maximum of CPU times and the number of partial SVD computation of each algorithm taken over the 10 random instances are given for each two choices of c_r and c_p . Moreover, Table C.1 and Table C.2 given in the appendix list different error statistics.

We used PROPACK [4] for computing partial singular value decompositions. In order to estimate the rank of X^0 , we followed the scheme proposed in Equation (17) in [6].

Both NSA and ASALM found the true rank in all 40 random problems solved when $\text{SNR}(D) = 80\text{dB}$. NSA found the true rank for 39 out of 40 problems with $n = 1500$ when $\text{SNR}(D) = 45\text{dB}$, while for the remaining 1 problem $\text{rank}(X^{\text{sol}})$ is off from $\text{rank}(X^0)$ only by 1. On the other hand, when $\text{SNR}(D) = 45\text{dB}$, ASALM could not find the true rank in any of the test problems. For each of the four problem settings corresponding to different c_r and c_p values, in Table 6.4 we report the average and maximum of $\text{rank}(X^{\text{sol}})$ over 10 random instances, after thresholding the singular values of X^{sol} that were less than 1×10^{-12} . Table 6.5 shows that for all of the problem classes, the number of partial SVD required by ASALM was

TABLE 6.4
NSA vs ASALM: $\text{rank}(X^{\text{sol}})$ values for problems with $n = 1500$, $\text{SNR}(D) = 45\text{dB}$

	$\text{rank}(X^0) = 75$		$\text{rank}(X^0) = 150$	
	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
Alg.	avg / max	avg / max	avg / max	avg / max
NSA	75 / 75	75 / 75	150.1 / 151	150 / 150
ASALM	175.8 / 177	179 / 207	222.4 / 224	201.9 / 204

more than twice the number that NSA required. On the other hand, there was a big difference in CPU times; this difference can be explained by the fact that ASALM required more leading singular values than NSA did per partial SVD computation. Table 6.6 shows that although the relative errors of the low-rank components produced by NSA were slightly better, the relative errors of the sparse components produced by NSA were significantly better than those produced by ASALM. Finally, in Figure 6.1, we plot the decomposition of $D = X^0 + S^0 + \zeta^0 \in \mathbb{R}^{n \times n}$ generated by NSA, where $\text{rank}(X^0) = 75$, $\|S^0\|_0 = 112,500$ and $\text{SNR}(D) = 45$. In the first row, we plot randomly selected 1500 components of S^0 and 100 leading singular values of X^0 in the first row. In the second row, we plot the same components of S^{sol} and 100 singular of X^{sol} produced by NSA. In the third row, we plot the absolute errors of S^{sol} and X^{sol} . Note that the scales of the graphs showing absolute errors of S^{sol} and X^{sol} are larger than those of S^0 and X^0 . And in the fourth row, we plot the same 1500 random components of ζ^0 . When we compare the absolute error graphs of S^{sol} and X^{sol} with the graph showing ζ^0 , we can confirm that the solution produced by NSA is inline with Theorem 1.2.

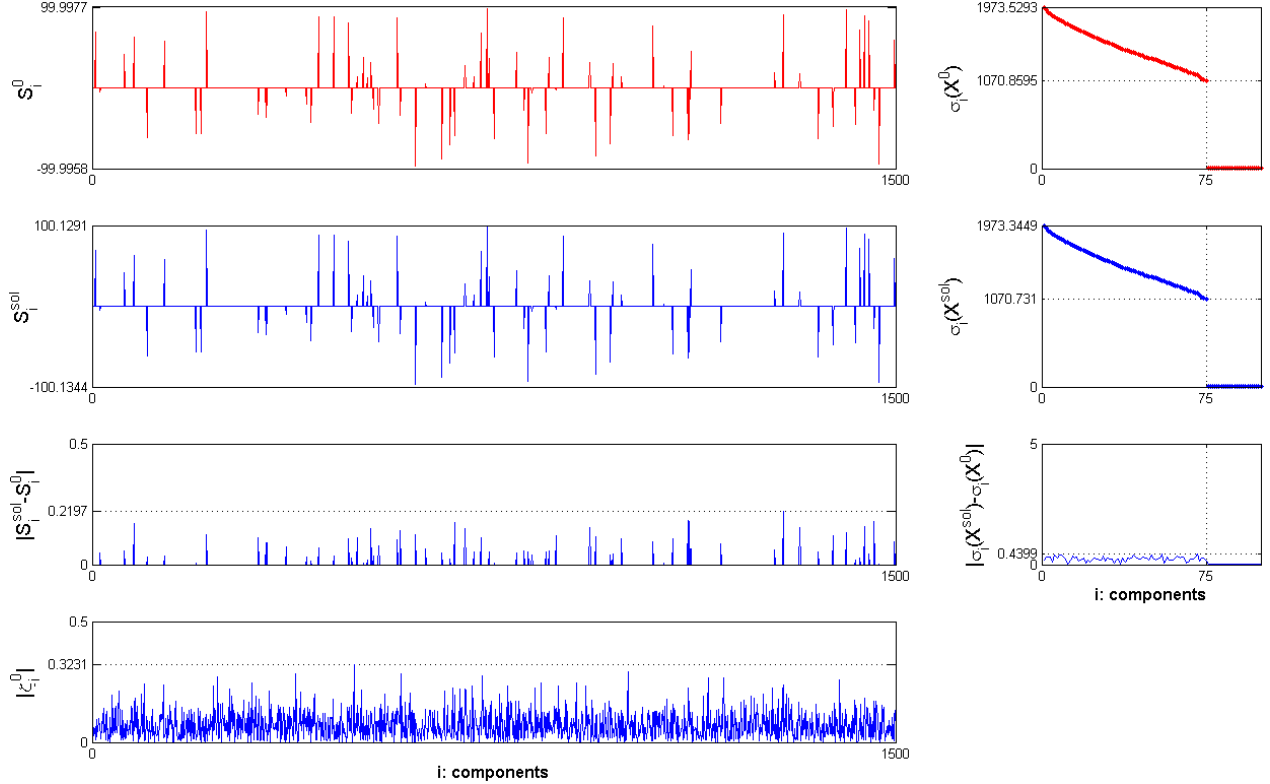
TABLE 6.5
NSA vs ASALM: Solution time for decomposing $D \in \mathbb{R}^{n \times n}$, $n = 1500$

SNR	Alg.	Field	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
			min/avg/max	min/avg/max	min/avg/max	min/avg/max
80dB	NSA	SVD#	10/10.0/10	10/10.9/11	12/12.0/12	12/12.2/13
		CPU	38.6/44.1/46.6	43.7/48.6/51.9	78.6/84.1/90.8	80.7/97.7/155.2
	ASALM	SVD#	22/22.0/22	20/20.0/20	29/29.0/29	29/29.4/30
		CPU	657.3/677.8/736.2	809.7/850.0/874.7	1277.3/1316.1/1368.6	1833.2/1905.2/2004.7
45dB	NSA	SVD#	7/7.9/8	8/8.0/8	9/9.0/9	9/9.0/9
		CPU	34.1/37.7/44.0	30.7/37.1/45.6	55.6/59.0/63.7	55.9/59.7/64.8
	ASALM	SVD#	21/21/21	18/18.5/19	28/28.0/28	27/27.3/28
		CPU	666.6/686.9/708.9	835.7/857.1/887.2	1201.9/1223.2/1277.5	1677.1/1739.1/1846.5

TABLE 6.6
NSA vs ASALM: Solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n = 1500$

SNR	Alg.	Relative Error	$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
			avg / max	avg / max	avg / max	avg / max
80dB	NSA	$\ X^{sol} - X^0\ _F$	1.8E-4 / 2.2E-4	2.1E-4 / 2.6E-4	1.3E-4 / 1.3E-4	2.8E-4 / 2.9E-4
		$\ S^{sol} - S^0\ _F$	1.3E-4 / 1.6E-4	9.6E-5 / 1.1E-4	8.1E-5 / 8.5E-5	1.3E-4 / 1.4E-4
	ASALM	$\ X^{sol} - X^0\ _F$	3.9E-4 / 4.2E-4	8.4E-4 / 8.8E-4	6.6E-4 / 6.8E-4	1.4E-3 / 1.4E-3
		$\ S^{sol} - S^0\ _F$	5.7E-4 / 6.2E-4	7.6E-4 / 8.0E-4	1.1E-3 / 1.1E-3	1.4E-3 / 1.4E-3
45dB	NSA	$\ X^{sol} - X^0\ _F$	3.4E-3 / 3.6E-3	4.7E-3 / 4.7E-3	3.9E-3 / 4.0E-3	5.3E-3 / 5.3E-3
		$\ S^{sol} - S^0\ _F$	1.8E-3 / 1.8E-3	2.3E-3 / 2.3E-3	2.6E-3 / 3.5E-3	3.1E-3 / 3.1E-3
	ASALM	$\ X^{sol} - X^0\ _F$	4.6E-3 / 4.8E-3	7.3E-3 / 8.4E-3	4.7E-3 / 4.7E-3	7.8E-3 / 7.9E-3
		$\ S^{sol} - S^0\ _F$	4.8E-3 / 4.9E-3	5.8E-3 / 7.0E-3	5.5E-3 / 5.5E-3	7.3E-3 / 7.5E-3

FIG. 6.1. NSA: Comparison of randomly selected 1500 components of ζ^0 with absolute errors of those components in S^{sol} and $\sigma(X^{sol})$. $D \in \mathbb{R}^{n \times n}$, $n = 1500$, $\text{SNR}(D) = 45\text{dB}$



6.2. Foreground Detection on a Noisy Video. We used NSA and ASALM to extract moving objects in an airport security video [5], which is a sequence of 201 grayscale frames of size 144×176 . We assume that the airport security video [5] was not corrupted by Gaussian noise. We formed the i -th column of the data matrix D by stacking the columns of the i^{th} frame into a long vector, i.e. D is in $\mathbb{R}^{25344 \times 201}$. In order to have a noisy video with $\text{SNR} = 20\text{dB}$ signal-to-noise ratio (SNR), given D , we chose $\varrho = \|D\|_F / (\sqrt{144 \times 176 \times 201} 10^{\text{SNR}/20})$ and then obtained a noisy D by $D = D + \varrho \text{randn}(144 * 176, 201)$, where $\text{randn}(m, n)$ produces a random matrix with independent standard Gaussian entries. Solving for $(X^*, S^*) = \arg\min_{X, S \in \mathbb{R}^{25344 \times 201}} \{\|X\|_* + \xi \|S\|_1 : \|X + S - D\|_F \leq \delta\}$, we decompose D into a low rank

matrix X^* and a sparse matrix S^* . We estimate the i -th frame background image with the i -th column of X^* and estimate the i -th frame moving object with the i -th column of S^* . Both algorithms are terminated when $\frac{\|(X_{k+1}, S_{k+1}) - (X_k, S_k)\|_F}{\|(X_k, S_k)\|_F + 1} \leq \varrho \times 10^{-4}$.

The recovery statistics of each algorithm are displayed in Table 6.7. (X^{sol}, S^{sol}) denote the variables corresponding to the low-rank and sparse components of D , respectively, when the algorithm of interest terminates. Figure 7 and Figure 7 show the 35-th, 100-th and 125-th frames of the noise added airport security video [5] in their first row of images. The second and third rows in these tables have the recovered background and foreground images of the selected frames, respectively. Even though the visual quality of recovered background and foreground are very similar, Table 6.7 shows that both the number of partial SVDs and the CPU time of NSA are significantly less than those for ASALM.

TABLE 6.7
NSA vs ASALM: Recovery statistics for foreground detection on a noisy video

Alg.	CPU	SVD#	$\ X^{sol}\ _*$	$\ S^{sol}\ _1$	$\text{rank}(X^{sol})$	$\frac{\ X^{sol} + S^{sol} - D\ _F}{\ D\ _F}$
NSA	160.8	19	398662.9	76221854.1	81	0.00068
ASALM	910.0	94	401863.6	75751977.1	89	0.00080

7. Acknowledgements. We would like to thank to Min Tao for providing the code ASALM.



FIG. 7.1. Background extraction from a video with 20dB SNR using NSA



FIG. 7.2. Background extraction from a video with 20dB SNR using ASALM

REFERENCES

- [1] A. BECK AND M. TEBoulLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.
- [2] E. J. CANDÈS, X. LI, Y. MA, AND WRIGHT J., *Robust principle component analysis?*, Journal of ACM, 58 (2011), pp. 1–37.
- [3] D. GOLDFARB, S. MA, AND K. SCHEINBERG, *Fast alternating linearization methods for minimizing the sum of two convex functions*. arXiv:0912.4571v2, October 2010.
- [4] R.M. LARSEN, *Lanczos bidiagonalization with partial reorthogonalization*, Technical report DAIMI PB-357, Department of Computer Science, Aarhus University, 1998.
- [5] L. LI, W. HUANG, I. GU, AND Q. TIAN, *Statistical modeling of complex backgrounds for foreground object detection*, IEEE Trans. on Image Processing, 13 (2004), pp. 1459–1472.
- [6] Z. LIN, M. CHEN, L. WU, AND Y. MA, *The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices*, arXiv:1009.5055v2, (2011).
- [7] Z. LIN, A. GANESH, J. WRIGHT, L. WU, M. CHEN, AND Y. MA, *Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix*, tech. report, UIUC Technical Report UILU-ENG-09-2214, 2009.
- [8] Y. NESTEROV, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004.
- [9] ———, *Smooth minimization of nonsmooth functions*, Mathematical Programming, 103 (2005), pp. 127–152.
- [10] M. TAO AND X. YUAN, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM Journal on Optimization, 21 (2011), pp. 57–81.
- [11] P. TSENG, *On accelerated proximal gradient methods for convex-concave optimization*, submitted to SIAM Journal on Optimization, (2008).
- [12] J. WRIGHT, A. GANESH, S. RAO, Y. PENG, AND Y. MA, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*, in Proceedings of Neural Information Processing Systems (NIPS), December 2009.
- [13] Z. ZHOU, X. LI, J. WRIGHT, E. CANDÈS, AND Y. MA, *Stable principle component pursuit*, Proceedings of International Symposium on Information Theory, (2010).

Appendix A. Proof of Theorem 4.1.

DEFINITION A.1. Let $\phi : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ and $\psi : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ be closed convex functions and define

$$Q^\phi(Z, S|X) := \psi(Z, S) + \phi(X) + \langle \gamma^\phi(X), Z - X \rangle + \frac{\rho}{2} \|Z - X\|_F^2, \quad (\text{A.1})$$

$$Q^\psi(Z|X, S) := \phi(Z) + \psi(X, S) + \langle \gamma_x^\psi(X, S), Z - X \rangle + \frac{\rho}{2} \|Z - X\|_F^2, \quad (\text{A.2})$$

and

$$(p_x^\phi(X), p_s^\phi(X)) := \underset{Z, S \in \mathbb{R}^{m \times n}}{\operatorname{argmin}} Q^\phi(Z, S|X), \quad (\text{A.3})$$

$$p^\psi(X, S) := \underset{Z \in \mathbb{R}^{m \times n}}{\operatorname{argmin}} Q^\psi(Z|X, S), \quad (\text{A.4})$$

where $\gamma^\phi(X)$ is any subgradient in the subdifferential $\partial\phi$ at the point X and $(\gamma_x^\psi(X, S), \gamma_s^\psi(X, S))$ is any subgradient in the subdifferential $\partial\psi$ at the point (X, S) .

LEMMA A.2. Let $\phi, \psi, Q^\phi, Q^\psi, p_x^\phi, p_s^\phi, p^\psi, \gamma_x^\phi, \gamma_s^\phi, \gamma_x^\psi, \gamma_s^\psi$ be as given in Definition A.1. and $\Phi(X, S) := \phi(X) + \psi(X, S)$. Let $X^0 \in \mathbb{R}^{m \times n}$ and define $\hat{X} := p_x^\phi(X^0)$ and $\hat{S} := p_s^\phi(X^0)$. If

$$\Phi(\hat{X}, \hat{S}) \leq Q^\phi(\hat{X}, \hat{S}|X^0), \quad (\text{A.5})$$

then for any $(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$,

$$\frac{2}{\rho} \left(\Phi(X, S) - \Phi(\hat{X}, \hat{S}) \right) \geq \|X - \hat{X}\|_F^2 - \|X - X^0\|_F^2. \quad (\text{A.6})$$

Moreover, if

$$\Phi(p^\psi(\hat{X}, \hat{S}), \hat{S}) \leq Q^\psi(p^\psi(\hat{X}, \hat{S})|X, \hat{S}), \quad (\text{A.7})$$

then for any $(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$,

$$\frac{2}{\rho} \left(\Phi(X, S) - \Phi(p^\psi(\hat{X}, \hat{S}), \hat{S}) \right) \geq \|X - p^\psi(\hat{X}, \hat{S})\|_F^2 - \|X - \hat{X}\|_F^2. \quad (\text{A.8})$$

Proof. Let $X^0 \in \mathbb{R}^{m \times n}$ satisfy (A.5). Then for any $(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, we have

$$\Phi(X, S) - \Phi\left(p_x^\phi(X^0), \hat{S}\right) \geq \Phi(X, S) - Q^\phi\left(\hat{X}, \hat{S} | X^0\right). \quad (\text{A.9})$$

First order optimality conditions for (A.3) and ψ being a closed convex function guarantee that there exists $\left(\gamma_x^\psi(\hat{X}, \hat{S}), \gamma_s^\psi(\hat{X}, \hat{S})\right) \in \partial\psi(\hat{X}, \hat{S})$ such that

$$\gamma_x^\psi(\hat{X}, \hat{S}) + \gamma^\phi(X^0) + \rho(\hat{X} - X^0) = 0, \quad (\text{A.10})$$

$$\gamma_s^\psi(\hat{X}, \hat{S}) = 0, \quad (\text{A.11})$$

where $\partial\psi(\hat{X}, \hat{S})$ denotes the subdifferential of $\psi(\cdot, \cdot)$ at the point $(\hat{X}, \hat{S})$.

Moreover, using the convexity of $\psi(\cdot, \cdot)$ and $\phi(\cdot)$, we have

$$\begin{aligned} \psi(X, S) &\geq \psi(\hat{X}, \hat{S}) + \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle + \langle \gamma_s^\psi(\hat{X}, \hat{S}), S - \hat{S} \rangle, \\ \phi(X) &\geq \phi(X^0) + \langle \gamma^\phi(X^0), X - X^0 \rangle. \end{aligned}$$

These two inequalities and (A.11) together imply

$$\Phi(X, S) \geq \psi(\hat{X}, \hat{S}) + \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle + \phi(X^0) + \langle \gamma^\phi(X^0), X - X^0 \rangle. \quad (\text{A.12})$$

This inequality together with (A.5) and (A.10) gives

$$\begin{aligned} &\Phi(X, S) - \Phi(\hat{X}, \hat{S}) \\ &\geq \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle + \langle \gamma^\phi(X^0), X - X^0 \rangle - \langle \gamma^\phi(X^0), \hat{X} - X^0 \rangle - \frac{\rho}{2} \|X - X^0\|_F^2, \\ &= \langle \gamma^\phi(X^0) + \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle - \frac{\rho}{2} \|X - X^0\|_F^2, \\ &= \rho \langle X^0 - \hat{X}, X - \hat{X} \rangle - \frac{\rho}{2} \|X - X^0\|_F^2, \\ &= \frac{\rho}{2} (\|X - \hat{X}\|_F^2 - \|X - X^0\|_F^2). \end{aligned}$$

Hence, we have (A.6). Suppose that X^0 satisfies (A.7). Then for any $(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, we have

$$\Phi(X, S) - \Phi\left(p^\psi(\hat{X}, \hat{S}), \hat{S}\right) \geq \Phi(X, S) - Q^\psi\left(p^\psi(\hat{X}, \hat{S}) | \hat{X}, \hat{S}\right). \quad (\text{A.13})$$

First order optimality conditions for (A.4) and ϕ being a closed convex function guarantee that there exists $\gamma^\phi(p^\psi(\hat{X}, \hat{S})) \in \partial\phi(p^\psi(\hat{X}, \hat{S}))$ such that

$$\gamma^\phi(p^\psi(\hat{X}, \hat{S})) + \gamma_x^\psi(\hat{X}, \hat{S}) + \rho(p^\psi(\hat{X}, \hat{S}) - \hat{X}) = 0. \quad (\text{A.14})$$

Moreover, using the convexity of $\phi(\cdot)$ and $\psi(\cdot, \cdot)$, we have

$$\phi(X) \geq \phi(p^\psi(\hat{X}, \hat{S})) + \langle \gamma^\phi(p^\psi(\hat{X}, \hat{S})), X - p^\psi(\hat{X}, \hat{S}) \rangle, \quad (\text{A.15})$$

$$\psi(X, S) \geq \psi(\hat{X}, \hat{S}) + \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle, \quad (\text{A.16})$$

where (A.16) follows from the fact that $(\hat{X}, \hat{S}) = \operatorname{argmin}_{X, S} Q^\phi(X, S | X^0)$ implies $(\gamma_x^\psi(\hat{X}, \hat{S}), 0) \in \partial\psi(\hat{X}, \hat{S})$, i.e. we can set $\gamma_s^\psi(\hat{X}, \hat{S}) = 0$. Summing the two inequalities (A.15) and (A.16) give

$$\Phi(X, S) \geq \psi(\hat{X}, \hat{S}) + \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle + \phi(p^\psi(\hat{X}, \hat{S})) + \langle \gamma^\phi(p^\psi(\hat{X}, \hat{S})), X - p^\psi(\hat{X}, \hat{S}) \rangle. \quad (\text{A.17})$$

This inequality together with (A.7) and (A.14) gives

$$\begin{aligned} & \Phi(X, S) - \Phi(p^\psi(\hat{X}, \hat{S}), \hat{S}) \\ & \geq \langle \gamma_x^\psi(\hat{X}, \hat{S}), X - \hat{X} \rangle + \langle \gamma^\phi(p^\psi(\hat{X}, \hat{S})), X - p^\psi(\hat{X}, \hat{S}) \rangle \\ & \quad - \langle \gamma_x^\psi(\hat{X}, \hat{S}), p^\psi(\hat{X}, \hat{S}) - \hat{X} \rangle - \frac{\rho}{2} \|p^\psi(\hat{X}, \hat{S}) - \hat{X}\|_F^2, \\ & = \langle \gamma^\phi(p^\psi(\hat{X}, \hat{S})) + \gamma_x^\psi(\hat{X}, \hat{S}), X - p^\psi(\hat{X}, \hat{S}) \rangle - \frac{\rho}{2} \|p^\psi(\hat{X}, \hat{S}) - \hat{X}\|_F^2, \\ & = \rho \langle \hat{X} - p^\psi(\hat{X}, \hat{S}), X - p^\psi(\hat{X}, \hat{S}) \rangle - \frac{\rho}{2} \|p^\psi(\hat{X}, \hat{S}) - \hat{X}\|_F^2, \\ & = \frac{\rho}{2} (\|X - p^\psi(\hat{X}, \hat{S})\|_F^2 - \|X - \hat{X}\|_F^2). \end{aligned}$$

Hence, we have (A.8). $\square$

We are now ready to give the proof of Theorem 4.1.

Proof. Let $I := \{0 \leq i \leq k-1 : \Phi(X_{i+1}, S_i) \leq \mathcal{L}_\rho(X_{i+1}, Z_i, S_i; Y_i)\}$ and $I^c := \{0, 1, \dots, k-1\} \setminus I$. Since $\nabla\phi(\cdot)$ is Lipschitz continuous with Lipschitz constant L and $\rho \geq L$, $\Phi(p_x^\phi(X), p_s^\phi(X)) \leq Q^\phi(p_x^\phi(X), p_s^\phi(X) | X)$ is true for all $X \in \mathbb{R}^{m \times n}$. Since (A.5) in Lemma A.2 is true for all $X^0 \in \mathbb{R}^{m \times n}$, (A.6) is true for all $(X, S) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$. Particularly, since for all $i \in I \cup I^c$

$$(Z_{i+1}, S_{i+1}) = \operatorname{argmin}_{Z, S} Q^\phi(Z, S | X_{i+1}), \quad (\text{A.18})$$

setting $(X, S) := (X^*, S^*)$ and $X^0 := X_{i+1}$ in Lemma A.2 imply that $p_x^\phi(X_{i+1}) = Z_{i+1}$, $p_s^\phi(X_{i+1}) = S_{i+1}$ and we have

$$\frac{2}{\rho} (\Phi(X^*, S^*) - \Phi(Z_{i+1}, S_{i+1})) \geq \|Z_{i+1} - X^*\|_F^2 - \|X_{i+1} - X^*\|_F^2. \quad (\text{A.19})$$

Moreover, (A.18) implies that for all $i \in I \cup I^c$, there exists $(\gamma_x^\psi(Z_i, S_i), \gamma_s^\psi(Z_i, S_i)) \in \partial\psi(Z_i, S_i)$ such that

$$\gamma_x^\psi(Z_i, S_i) + \nabla\phi(X_i) + \rho(Z_i - X_i) = 0, \quad (\text{A.20})$$

$$\gamma_s^\psi(Z_i, S_i) = 0. \quad (\text{A.21})$$

(A.20) and the definition of Y_{i+1} of Algorithm ALM-S shown in **Algorithm 3** imply that

$$\gamma_x^\psi(Z_i, S_i) = -\nabla\phi(X_i) + \rho(X_i - Z_i) = Y_i.$$

Hence, by defining $Q^\psi(\cdot | Z_i, S_i)$ according to (A.2) using $\gamma_x^\psi(Z_i, S_i) = Y_i$, for all $X \in \mathbb{R}^{m \times n}$ we have

$$\mathcal{L}_\rho(X, Z_i, S_i; Y_i) = \phi(X) + \psi(Z_i, S_i) + \langle Y_i, X - Z_i \rangle + \frac{\rho}{2} \|X - Z_i\|_F^2 = Q^\psi(X | Z_i, S_i). \quad (\text{A.22})$$

for all $i \in I \cup I^c$. Hence, for all $i \in I$ $X_{i+1} = \operatorname{argmin}_X \mathcal{L}_\rho(X, Z_i, S_i; Y_i) = \operatorname{argmin}_X Q^\psi(X | Z_i, S_i)$. Thus, for all $i \in I$, setting $X^0 := X_i$ in Lemma A.2 imply $p_x^\phi(X_i) = Z_i$, $p_s^\phi(X_i) = S_i$ and $p^\psi(p_x^\phi(X_i), p_s^\phi(X_i)) = p^\psi(Z_i, S_i) = X_{i+1}$. For all $i \in I$ we have $\Phi(X_{i+1}, S_i) \leq \mathcal{L}_\rho(X_{i+1}, Z_i, S_i; Y_i) = Q^\psi(X_{i+1} | Z_i, S_i)$. Hence, for all $i \in I$ setting $X^0 := X_i$ in Lemma A.2 satisfies (A.7). Therefore, setting $(X, S) := (X^*, S^*)$ and $X^0 := X_i$ in Lemma A.2 implies that

$$\frac{2}{\rho} (\Phi(X^*, S^*) - \Phi(X_{i+1}, S_i)) \geq \|X_{i+1} - X^*\|_F^2 - \|Z_i - X^*\|_F^2. \quad (\text{A.23})$$

For any $i \in I$, summing (A.19) and (A.23) gives

$$\frac{2}{\rho} (2\Phi(X^*, S^*) - \Phi(X_{i+1}, S_i) - \Phi(Z_{i+1}, S_{i+1})) \geq \|Z_{i+1} - X^*\|_F^2 - \|Z_i - X^*\|_F^2. \quad (\text{A.24})$$

Moreover, since $X_{i+1} = Z_i$ for $i \in I^c$ and (A.19) holds for all $i \in I \cup I^c$, we trivially have

$$\frac{2}{\rho} (\Phi(X^*, S^*) - \Phi(Z_{i+1}, S_{i+1})) \geq \|Z_{i+1} - X^*\|_F^2 - \|Z_i - X^*\|_F^2. \quad (\text{A.25})$$

Summing (A.24) and (A.25) over $i = 0, 1, \dots, k-1$ gives

$$\frac{2}{\rho} \left((2|I| + |I^c|) \Phi(X^*, S^*) - \sum_{i \in I} \Phi(X_{i+1}, S_i) - \sum_{i=0}^{k-1} \Phi(Z_{i+1}, S_{i+1}) \right) \geq \|Z_k - X^*\|_F^2 - \|Z_0 - X^*\|_F^2. \quad (\text{A.26})$$

For any $i \in I \cup I^c$, setting $(X, S) := (X_{i+1}, S_i)$ and $X^0 := X_{i+1}$ in Lemma A.2 gives

$$\frac{2}{\rho} (\Phi(X_{i+1}, S_i) - \Phi(Z_{i+1}, S_{i+1})) \geq \|Z_{i+1} - X_{i+1}\|_F^2 \geq 0. \quad (\text{A.27})$$

Trivially, for $i = 1, \dots, k$ we also have

$$\frac{2}{\rho} (\Phi(X_i, S_{i-1}) - \Phi(Z_i, S_i)) \geq \|Z_i - X_i\|_F^2 \geq 0. \quad (\text{A.28})$$

Moreover, since for all $i \in I$ setting $X^0 := X_i$ in Lemma A.2 satisfies (A.7), setting $(X, S) := (Z_i, S_i)$ and $X^0 := X_i$ in Lemma A.2 implies that

$$\frac{2}{\rho} (\Phi(Z_i, S_i) - \Phi(X_{i+1}, S_i)) \geq \|X_{i+1} - Z_i\|_F^2 \geq 0. \quad (\text{A.29})$$

And since $X_{i+1} = Z_i$ for all $i \in I^c$, (A.29) trivially holds for all $i \in I^c$. Thus, for all $i \in I \cup I^c$ we have

$$\frac{2}{\rho} (\Phi(Z_i, S_i) - \Phi(X_{i+1}, S_i)) \geq 0. \quad (\text{A.30})$$

Adding (A.27) and (A.30) yields $\Phi(Z_i, S_i) \geq \Phi(Z_{i+1}, S_{i+1})$ for all $i \in I \cup I^c$ and adding (A.28) and (A.30) yields $\Phi(X_i, S_{i-1}) \geq \Phi(X_{i+1}, S_i)$ for all $i = 1, \dots, k-1$. Hence,

$$\sum_{i=0}^{k-1} \Phi(Z_{i+1}, S_{i+1}) \geq k\Phi(Z_k, S_k), \text{ and } \sum_{i \in I} \Phi(X_{i+1}, S_i) \geq n_k \Phi(X_k, S_{k-1}). \quad (\text{A.31})$$

These two inequalities, (A.26) and the fact that $X_0 = Z_0$ imply

$$\frac{2}{\rho} ((2|I| + |I^c|) \Phi(X^*, S^*) - n_k \Phi(X_k, S_{k-1}) - k\Phi(Z_k, S_k)) \geq -\|X_0 - X^*\|_F^2. \quad (\text{A.32})$$

Hence, (4.4) follows from the facts: $2|I| + |I^c| = k + n_k$ and $n_k \Phi(X_k, S_{k-1}) + k\Phi(Z_k, S_k) \geq (k + n_k) \Phi(Z_k, S_k)$ due to (A.27). $\square$

Appendix B. Proof of Lemma 5.1.

Proof. Since Y^* and θ^* are optimal Lagrangian dual variables, we have

$$(X^*, X^*, S^*) = \operatorname{argmin}_{X, Z, S} \|X\|_* + \xi \|S\|_1 + \langle Y^*, X - Z \rangle + \frac{\theta^*}{2} (\|Z + S - D\|_F^2 - \delta^2).$$

Then from first-order optimality conditions, we have

$$\begin{aligned} 0 &\in \partial \|X^*\|_* + Y^*, \\ 0 &\in \xi \partial \|S^*\|_1 + \theta^*(X^* + S^* - D), \\ -Y^* + \theta^*(X^* + S^* - D) &= 0. \end{aligned}$$

Hence, $-Y^* \in \partial \|X^*\|_*$ and $-Y^* \in \xi \partial \|S^*\|_1$.

For $k \geq 0$, since X_{k+1} is the optimal solution for the k -th subproblem given in Step 4 in **Algorithm 5**, from the first-order optimality conditions it follows that

$$0 \in \partial \|X_{k+1}\|_* + Y_k + \rho_k(X_{k+1} - Z_k). \quad (\text{B.1})$$

For $k \geq 0$, let $\theta_k \geq 0$ be the optimal Lagrange multiplier for the quadratic constraint in the k -th subproblem given in Step 6 in **Algorithm 5**. Since (S_{k+1}, Z_{k+1}) is the optimal solution, from the first-order optimality conditions it follows that

$$0 \in \xi \partial \|S_{k+1}\|_1 + \theta_k(Z_{k+1} + S_{k+1} - D), \quad (\text{B.2})$$

$$-Y_k + \rho_k(Z_{k+1} - X_{k+1}) + \theta_k(Z_{k+1} + S_{k+1} - D) = 0. \quad (\text{B.3})$$

From (B.1), it follows that $-\hat{Y}_{k+1} \in \partial \|X_{k+1}\|_*$. Hence, $\{\hat{Y}_k\}_{k \in \mathbb{Z}_+}$ is a bounded sequence. From (B.2) and (B.3), it follows that $-Y_{k+1} \in \xi \partial \|S_{k+1}\|_1$. Hence, $\{Y_k\}_{k \in \mathbb{Z}_+}$ is also a bounded sequence.

Furthermore, since $Y_{k+1} - Y_k = \rho_k(X_{k+1} - Z_{k+1})$ and $Y_{k+1} - \hat{Y}_{k+1} = \rho_k(Z_k - Z_{k+1})$, we have

$$\begin{aligned} &\rho_k^{-1} \langle Y_{k+1} - Y_k, Y_{k+1} - Y^* \rangle \\ &= \langle X_{k+1} - Z_{k+1}, Y_{k+1} - Y^* \rangle, \\ &= \langle X_{k+1} - X^*, Y_{k+1} - Y^* \rangle + \langle X^* - Z_{k+1}, Y_{k+1} - Y^* \rangle, \\ &= \langle X_{k+1} - X^*, Y_{k+1} - \hat{Y}_{k+1} \rangle + \langle X_{k+1} - X^*, \hat{Y}_{k+1} - Y^* \rangle + \langle X^* - Z_{k+1}, Y_{k+1} - Y^* \rangle, \\ &= \rho_k \langle X_{k+1} - X^*, Z_k - Z_{k+1} \rangle + \langle X_{k+1} - X^*, \hat{Y}_{k+1} - Y^* \rangle + \langle X^* - Z_{k+1}, Y_{k+1} - Y^* \rangle. \end{aligned}$$

Using the above equality, for all $k \geq 1$, we trivially have

$$\begin{aligned} &\|Z_{k+1} - X^*\|_F^2 + \rho_k^{-2} \|Y_{k+1} - Y^*\|_F^2 \\ &= \|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - Z_k\|_F^2 - \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2, \\ &\quad + 2 \langle Z_{k+1} - X^*, Z_{k+1} - Z_k \rangle + 2 \rho_k^{-2} \langle Y_{k+1} - Y_k, Y_{k+1} - Y^* \rangle, \\ &= \|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - Z_k\|_F^2 - \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2, \\ &\quad + 2 \langle Z_{k+1} - X^*, Z_{k+1} - Z_k \rangle + 2 \langle X_{k+1} - X^*, Z_k - Z_{k+1} \rangle \\ &\quad - 2 \rho_k^{-1} \left(\langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle + \langle -Y_{k+1} + Y^*, X^* - Z_{k+1} \rangle \right) \\ &= \|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - Z_k\|_F^2 - \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2, \\ &\quad + 2 \langle Z_{k+1} - X_{k+1}, Z_{k+1} - Z_k \rangle - 2 \rho_k^{-1} \left(\langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle + \langle -Y_{k+1} + Y^*, X^* - Z_{k+1} \rangle \right) \\ &= \|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - Z_k\|_F^2 - \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2, \\ &\quad - 2 \rho_k^{-1} \left(\langle Y_{k+1} - Y_k, Z_{k+1} - Z_k \rangle + \langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle + \langle -Y_{k+1} + Y^*, X^* - Z_{k+1} \rangle \right) \quad (\text{B.4}) \end{aligned}$$

Since $-Y_k \in \xi \partial \|S_k\|_1$ for all $k \geq 1$ and $-Y^* \in \xi \partial \|S^*\|_1$, we have for all $k \geq 1$

$$\langle -Y_{k+1} + Y_k, S_{k+1} - S_k \rangle \geq 0, \quad (\text{B.5})$$

$$\langle -Y_{k+1} + Y^*, S_{k+1} - S^* \rangle \geq 0. \quad (\text{B.6})$$

Since $\rho_{k+1} \geq \rho_k$ for all $k \geq 1$, adding (B.5), (B.6) to (B.4) and subtracting (B.6) from (B.4), we have

$$\begin{aligned}
& \|Z_{k+1} - X^*\|_F^2 + \rho_{k+1}^{-2} \|Y_{k+1} - Y^*\|_F^2 \\
& \leq \|Z_{k+1} - X^*\|_F^2 + \rho_k^{-2} \|Y_{k+1} - Y^*\|_F^2 \\
& \leq \|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - Z_k\|_F^2 - \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2 \\
& \quad - 2\rho_k^{-1} \left(\langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle + \langle -Y_{k+1} + Y^*, S_{k+1} - S^* \rangle \right) \\
& \quad - 2\rho_k^{-1} (\langle Y_{k+1} - Y_k, Z_{k+1} + S_{k+1} - Z_k - S_k \rangle + \langle -Y_{k+1} + Y^*, X^* + S^* - Z_{k+1} - S_{k+1} \rangle) \tag{B.7}
\end{aligned}$$

Applying Lemma 3.2 on the k -th subproblem given in Step 6 in Algorithm 5, it follows that

$$(Y_{k+1}, Y_{k+1}) \in \partial \mathbf{1}_\chi(Z_{k+1}, S_{k+1}).$$

Using arguments similar to those used in the proof of Lemma 3.2, one can also show that

$$(Y^*, Y^*) \in \partial \mathbf{1}_\chi(X^*, S^*).$$

Moreover, since $-Y_k \in \xi \partial \|S_k\|_1$, $-\hat{Y}_k \in \xi \partial \|X_k\|_*$ for all $k \geq 1$, $-Y^* \in \xi \partial \|S^*\|_1$ and $-Y^* \in \partial \|X^*\|_*$, we have

$$\begin{aligned}
\langle Y_{k+1} - Y_k, Z_{k+1} + S_{k+1} - Z_k - S_k \rangle & \geq 0, \\
\langle -Y_{k+1} + Y^*, X^* + S^* - Z_{k+1} - S_{k+1} \rangle & \geq 0, \\
\langle -Y_{k+1} + Y^*, S_{k+1} - S^* \rangle & \geq 0, \\
\langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle & \geq 0,
\end{aligned}$$

for all $k \geq 1$. Therefore, the above inequalities and (B.7) together imply that $\{\|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2\}_{k \in \mathbb{Z}_+}$ is a non-increasing sequence. Moreover, we also have

$$\begin{aligned}
& \sum_{k \in \mathbb{Z}_+} \|Z_{k+1} - Z_k\|_F^2 + \rho_k^{-2} \|Y_{k+1} - Y_k\|_F^2 \\
& + 2 \sum_{k \in \mathbb{Z}_+} \rho_k^{-1} \left(\langle -\hat{Y}_{k+1} + Y^*, X_{k+1} - X^* \rangle + \langle -Y_{k+1} + Y^*, S_{k+1} - S^* \rangle \right) \\
& + 2 \sum_{k \in \mathbb{Z}_+} \rho_k^{-1} (\langle Y_{k+1} - Y_k, Z_{k+1} + S_{k+1} - Z_k - S_k \rangle + \langle -Y_{k+1} + Y^*, X^* + S^* - Z_{k+1} - S_{k+1} \rangle) \\
& = \sum_{k \in \mathbb{Z}_+} (\|Z_k - X^*\|_F^2 + \rho_k^{-2} \|Y_k - Y^*\|_F^2 - \|Z_{k+1} - X^*\|_F^2 - \rho_{k+1}^{-2} \|Y_{k+1} - Y^*\|_F^2) < \infty
\end{aligned}$$

□

Appendix C. Additional Statistics for Numerical Experiments.

TABLE C.1

NSA vs ASALM: Additional statistics on solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n = 1500$, $\text{SNR}(D) = 80\text{dB}$

	$c_r=0.05$ $c_p=0.05$		$c_r=0.05$ $c_p=0.1$	
	NSA	ASALM	NSA	ASALM
Error Type	avg / max	avg / max	avg / max	avg / max
$ \ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _* /\ \mathbf{X}^0\ _*$	1.7E-6 / 5.2E-6	6.9E-6 / 1.0E-5	5.0E-6 / 3.2E-5	2.3E-5 / 3.8E-5
$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	4.1E-2 / 5.2E-2	3.9E-2 / 4.7E-2	5.2E-2 / 1.8E-1	1.1E-1 / 1.7E-1
$\max\{ \sigma_i : \sigma_i^0 = 0\}$	7.9E-13 / 2.2E-12	6.3E-13 / 1.6E-12	8.6E-13 / 2.0E-12	1.1E-12 / 2.0E-12
$ \ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1 /\ \mathbf{S}^0\ _1$	1.1E-5 / 1.4E-5	6.2E-6 / 9.7E-6	9.7E-6 / 1.5E-5	8.6E-5 / 9.7E-5
$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	2.9E-1 / 3.5E-1	5.9E-1 / 8.0E-1	2.2E-1 / 2.4E-1	5.9E-1 / 7.4E-1
$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	0 / 0	4.0E-1 / 7.2E-1	8.3E-3 / 1.1E-2	1.9E-1 / 5.5E-1

	$c_r=0.1$ $c_p=0.05$		$c_r=0.1$ $c_p=0.1$	
	NSA	ASALM	NSA	ASALM
Error Type	avg / max	avg / max	avg / max	avg / max
$ \ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _* /\ \mathbf{X}^0\ _*$	5.6E-6 / 6.4E-6	4.6E-5 / 4.9E-5	6.2E-6 / 7.1E-6	1.2E-4 / 1.4E-4
$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	5.7E-2 / 6.2E-2	1.2E-1 / 1.3E-1	8.8E-2 / 1.0E-1	3.0E-1 / 3.7E-1
$\max\{ \sigma_i : \sigma_i^0 = 0\}$	6.9E-13 / 1.5E-12	6.2E-13 / 9.9E-13	6.2E-13 / 1.3E-12	3.9E-13 / 1.0E-12
$ \ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1 /\ \mathbf{S}^0\ _1$	1.2E-5 / 1.6E-5	1.6E-4 / 1.7E-4	3.4E-5 / 3.7E-5	2.5E-4 / 2.7E-4
$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	1.6E-1 / 1.9E-1	6.7E-1 / 8.3E-1	1.7E-1 / 2.0E-1	7.9E-1 / 9.5E-1
$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	7.0E-3 / 1.1E-2	1.5E-1 / 2.5E-1	1.3E-2 / 1.9E-2	1.2E-1 / 2.5E-1

TABLE C.2

NSA vs ASALM: Additional statistics on solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n = 1500$, $\text{SNR}(D) = 45\text{dB}$

	$c_r=0.05$ $c_p=0.05$		$c_r=0.05$ $c_p=0.1$	
	NSA	ASALM	NSA	ASALM
Error Type	avg / max	avg / max	avg / max	avg / max
$ \ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _* /\ \mathbf{X}^0\ _*$	1.8E-4 / 3.6E-4	1.4E-3 / 1.5E-3	2.2E-4 / 2.4E-4	2.4E-3 / 2.6E-3
$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	5.9E-1 / 1.8E+0	1.1E+0 / 1.5E+0	9.8E-1 / 1.1E+0	2.3E+0 / 2.6E+0
$\max\{ \sigma_i : \sigma_i^0 = 0\}$	6.4E-13 / 1.3E-12	3.7E+0 / 3.8E+0	6.1E-13 / 1.0E-12	4.7E+0 / 5.5E+0
$ \ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1 /\ \mathbf{S}^0\ _1$	1.7E-4 / 1.9E-4	4.2E-3 / 4.3E-3	1.3E-4 / 1.3E-4	2.9E-3 / 3.6E-3
$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	1.0E+0 / 1.2E+0	3.0E+0 / 3.6E+0	1.3E+0 / 1.4E+0	3.2E+0 / 3.8E+0
$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	3.6E-1 / 4.0E-1	2.2E+0 / 2.6E+0	5.3E-1 / 6.1E-1	2.3E+0 / 3.1E+0

	$c_r=0.1$ $c_p=0.05$		$c_r=0.1$ $c_p=0.1$	
	NSA	ASALM	NSA	ASALM
Error Type	avg / max	avg / max	avg / max	avg / max
$ \ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _* /\ \mathbf{X}^0\ _*$	3.7E-4 / 6.5E-4	9.7E-5 / 1.3E-4	6.7E-4 / 6.8E-4	8.4E-4 / 9.0E-4
$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	1.3E+0 / 1.5E+0	1.2E+0 / 1.3E+0	2.5E+0 / 2.8E+0	1.3E+0 / 1.5E+0
$\max\{ \sigma_i : \sigma_i^0 = 0\}$	1.6E-1 / 1.6E+0	3.6E+0 / 3.7E+0	7.3E-13 / 1.7E-12	3.2E+0 / 3.3E+0
$ \ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1 /\ \mathbf{S}^0\ _1$	8.1E-4 / 3.2E-3	4.7E-3 / 4.8E-3	8.9E-4 / 9.0E-4	4.4E-3 / 4.5E-3
$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	9.3E-1 / 1.1E+0	2.7E+0 / 3.3E+0	1.1E+0 / 1.2E+0	3.2E+0 / 3.5E+0
$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	5.7E-1 / 6.6E-1	1.1E+0 / 1.4E+0	7.1E-1 / 7.9E-1	1.3E+0 / 1.6E+0

TABLE C.3

NSA: Additional statistics on solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n \in \{500, 1000, 1500\}$, $\text{SNR}(D) = 80\text{dB}$

		$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
n	Error Type	avg / max	avg / max	avg / max	avg / max
500	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	7.2E-6 / 1.1E-5	2.0E-5 / 2.7E-5	5.6E-6 / 8.2E-6	2.1E-5 / 3.1E-5
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	1.7E-2 / 2.4E-2	3.4E-2 / 5.6E-2	2.1E-2 / 2.7E-2	3.2E-2 / 3.8E-2
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	1.6E-13 / 2.9E-13	2.0E-13 / 5.6E-13	1.1E-13 / 2.5E-13	8.6E-14 / 1.7E-13
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	1.6E-5 / 1.7E-5	1.5E-5 / 1.8E-5	2.9E-5 / 3.2E-5	2.6E-5 / 3.0E-5
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	3.2E-1 / 4.0E-1	3.0E-1 / 4.3E-1	2.6E-1 / 3.2E-1	1.8E-1 / 2.3E-1
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	9.5E-3 / 2.2E-2	1.5E-2 / 2.5E-2	1.5E-2 / 2.5E-2	1.8E-2 / 3.4E-2
1000	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	5.6E-6 / 1.7E-5	6.2E-6 / 1.7E-5	6.9E-6 / 8.6E-6	1.5E-6 / 2.6E-6
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	1.8E-2 / 4.0E-2	3.1E-2 / 4.8E-2	5.1E-2 / 6.0E-2	5.9E-2 / 6.8E-2
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	3.3E-13 / 4.8E-13	3.3E-13 / 5.0E-13	2.9E-13 / 6.6E-13	2.8E-13 / 4.8E-13
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	1.1E-5 / 1.5E-5	1.7E-5 / 1.9E-5	2.8E-5 / 3.0E-5	2.9E-5 / 3.0E-5
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	2.7E-1 / 3.1E-1	3.1E-1 / 3.8E-1	2.2E-1 / 2.8E-1	1.6E-1 / 1.7E-1
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	1.7E-4 / 9.7E-4	1.2E-2 / 1.7E-2	7.8E-3 / 1.2E-2	1.2E-2 / 1.5E-2
1500	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	1.7E-6 / 5.2E-6	5.0E-6 / 3.2E-5	5.6E-6 / 6.4E-6	6.2E-6 / 7.1E-6
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	4.1E-2 / 5.2E-2	5.2E-2 / 1.8E-1	5.7E-2 / 6.2E-2	8.8E-2 / 1.0E-1
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	7.9E-13 / 2.2E-12	8.6E-13 / 2.0E-12	6.9E-13 / 1.5E-12	6.2E-13 / 1.3E-12
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	1.1E-5 / 1.4E-5	9.7E-6 / 1.5E-5	1.2E-5 / 1.6E-5	3.4E-5 / 3.7E-5
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	2.9E-1 / 3.5E-1	2.2E-1 / 2.4E-1	1.6E-1 / 1.9E-1	1.7E-1 / 2.0E-1
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	0 / 0	8.3E-3 / 1.1E-2	7.0E-3 / 1.1E-2	1.3E-2 / 1.9E-2

TABLE C.4

NSA: Additional statistics on solution accuracy for decomposing $D \in \mathbb{R}^{n \times n}$, $n \in \{500, 1000, 1500\}$, $\text{SNR}(D) = 45\text{dB}$

		$c_r=0.05$ $c_p=0.05$	$c_r=0.05$ $c_p=0.1$	$c_r=0.1$ $c_p=0.05$	$c_r=0.1$ $c_p=0.1$
n	Error Type	avg / max	avg / max	avg / max	avg / max
500	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	6.0E-4 / 9.3E-4	5.5E-4 / 6.2E-4	7.4E-4 / 8.8E-4	1.0E-3 / 1.3E-3
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	5.1E-1 / 7.8E-1	5.4E-1 / 7.7E-1	8.2E-1 / 8.9E-1	9.2E-1 / 1.2E+0
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	1.7E-13 / 2.7E-13	1.6E-13 / 3.0E-13	1.0E-13 / 2.1E-13	1.1E-1 / 6.0E-1
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	3.0E-4 / 3.4E-4	2.1E-4 / 2.9E-4	3.0E-4 / 1.2E-3	6.4E-4 / 1.1E-3
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	1.6E+0 / 1.9E+0	1.4E+0 / 1.8E+0	1.2E+0 / 1.6E+0	1.0E+0 / 1.3E+0
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	2.3E-1 / 2.9E-1	4.0E-1 / 4.7E-1	3.5E-1 / 4.6E-1	5.4E-1 / 6.1E-1
1000	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	2.8E-4 / 3.1E-4	4.4E-4 / 7.5E-4	5.6E-4 / 8.0E-4	7.4E-4 / 8.4E-4
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	5.2E-1 / 6.2E-1	8.6E-1 / 1.2E+0	1.7E+0 / 1.9E+0	1.8E+0 / 1.9E+0
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	2.5E-13 / 5.3E-13	4.3E-13 / 9.0E-13	2.0E-1 / 2.0E+0	6.3E-1 / 3.9E+0
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	2.2E-4 / 2.3E-4	1.4E-4 / 1.7E-4	5.5E-4 / 3.7E-3	1.1E-3 / 2.5E-3
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	1.3E+0 / 1.5E+0	1.5E+0 / 1.8E+0	1.1E+0 / 1.3E+0	9.6E-1 / 1.1E+0
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	2.7E-1 / 3.2E-1	4.6E-1 / 5.2E-1	4.6E-1 / 5.1E-1	6.4E-1 / 6.7E-1
1500	$\frac{\ \mathbf{X}^{\text{sol}}\ _* - \ \mathbf{X}^0\ _*}{\ \mathbf{X}^0\ _*}$	1.8E-4 / 3.6E-4	2.2E-4 / 2.4E-4	3.7E-4 / 6.5E-4	6.7E-4 / 6.8E-4
	$\max\{ \sigma_i - \sigma_i^0 : \sigma_i^0 > 0\}$	5.9E-1 / 1.8E+0	9.8E-1 / 1.1E+0	1.3E+0 / 1.5E+0	2.5E+0 / 2.8E+0
	$\max\{ \sigma_i : \sigma_i^0 = 0\}$	6.4E-13 / 1.3E-12	6.1E-13 / 1.0E-12	1.6E-1 / 1.6E+0	7.3E-13 / 1.7E-12
	$\frac{\ \mathbf{S}^{\text{sol}}\ _1 - \ \mathbf{S}^0\ _1}{\ \mathbf{S}^0\ _1}$	1.7E-4 / 1.9E-4	1.3E-4 / 1.3E-4	8.1E-4 / 3.2E-3	8.9E-4 / 9.0E-4
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} - \mathbf{S}_{ij}^0 : \mathbf{S}_{ij}^0 \neq 0\}$	1.0E+0 / 1.2E+0	1.3E+0 / 1.4E+0	9.3E-1 / 1.1E+0	1.1E+0 / 1.2E+0
	$\max\{ \mathbf{S}_{ij}^{\text{sol}} : \mathbf{S}_{ij}^0 = 0\}$	3.6E-1 / 4.0E-1	5.3E-1 / 6.1E-1	5.7E-1 / 6.6E-1	7.1E-1 / 7.9E-1