A Proximal Alternating Direction Method for Semi-Definite Rank Minimization (Supplementary Material)

Ganzhao Yuan and Bernard Ghanem
King Abdullah University of Science and Technology (KAUST), Saudi Arabia
yuanganzhao@gmail.com, bernard.ghanem@kaust.edu.sa

The supplementary material is organized as follows. Section 1 presents the details of our proofs. Section 2 presents the convergence analysis of the proposed MPEC-based proximal ADM algorithm. Section 3 presents the classical ADM for solving the subproblem. Finally, Section 4 presents additional experimental results.

1 Proofs

Lemma 1. [Semidefinite Embedding Lemma (Fazel, Hindi, and Boyd 2003)] Let \( R \in \mathbb{R}^{m \times n} \) be a given matrix. Then \( \text{rank}(R) \leq r \) if and only if there exist matrices \( S = S^T \in \mathbb{R}^{m \times m} \) and \( T = T^T \in \mathbb{R}^{n \times n} \) such that

\[
\begin{bmatrix}
S & R \\
R^T & T
\end{bmatrix} \succeq 0, \quad \text{rank}(S) + \text{rank}(T) \leq 2r.
\]

Remark. We remark that the semidefinite optimization problem discussed in this paper is very general. To illustrate this point, we consider the following optimization problem:

\[
\min_{R} \quad g(R) + \lambda \text{rank}(R)
\]

By the Semidefinite Embedding Lemma, we have the following equivalent semi-definite optimization problem :

\[
\min_{R} \quad g(R) + \frac{1}{2} \text{rank} \left( \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix} \right), \quad \text{s.t.} \quad \begin{bmatrix} S & R \\ R^T & T \end{bmatrix} \succeq 0.
\]

Since \( \begin{bmatrix} S & R \\ R^T & T \end{bmatrix} \) is positive semidefinite, by the Schur complement condition, it holds that \( S \succeq 0 \) and \( T \succeq 0 \). Then the variable matrix \( \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix} \) is also positive semidefinite. Using the MPEC reformulation, we have the following equivalent optimization problem:

\[
\min_{R,s.t.} \quad g(R) + \frac{1}{2} \text{tr}(V),
\]

\[
s.t. \quad \langle I - V, \begin{bmatrix} S & 0 \\ 0 & T \end{bmatrix} \rangle = 0, \quad \text{rank} \left( \begin{bmatrix} S & R \\ R^T & T \end{bmatrix} \right) \geq 0, \quad I \succeq V \succeq 0.
\]

Clearly, the equivalent optimization problem in Eq (2) can be solved by our proposed PADM algorithm. \( \square \)

Here, we prove the variational formulation of the NP-hard rank function.

Lemma 2. For any given positive semidefinite matrix \( X \in \mathbb{R}^{n \times n} \), it holds that

\[
\text{rank}(X) = \min_{V} \quad \text{tr}(I - V), \quad \text{s.t.} \quad \langle V, X \rangle = 0, \quad 0 \preceq V \preceq I
\]

and the unique optimal solution of the minimization problem in Eq (3) is given by \( V^* = U \text{diag}(\text{sign}(\sigma))U^T \), where \( X = U \text{diag}(\sigma)U^T \) is the eigenvalue decomposition of \( X \), \text{sign} is the standard sign function.

Proof. First of all, we let \( X \) and \( V \) be arbitrary feasible solutions with their eigenvalue decomposition given by \( X = U \text{diag}(\sigma)U^T \) and \( V = S \text{diag}(\delta)S^T \), respectively. It always holds that:

\[
\langle X, V \rangle = ||U \text{diag}(\sqrt{\sigma})U \text{diag}(\sqrt{\delta})S^T||_F^2 \geq 0
\]

The second equality is achieved only when \( V \) and \( X \) are simultaneously unitarily diagonalizable, i.e. both \( V \) and \( X \) share the same spectral decomposition. Therefore, the feasible set for \( V \) in Eq (3) must be contained in the set \( \{ V : V = U \text{diag}(\nu)U^T, \quad 0 \preceq \nu \preceq 1, \quad \nu \in \mathbb{R}^r \} \). Using the fact that \( \text{tr}(V) = \langle v, 1 \rangle, \quad \text{tr}(X) = \langle \sigma, 1 \rangle \) and \( \langle V, X \rangle = \langle U \text{diag}(\nu)U^T, U \text{diag}(\sigma)U^T \rangle = \langle \nu, \sigma \rangle \), Eq (3) reduces to the vector case:

\[
\text{rank}(X) = \min_{\nu} \quad \langle 1, 1 - \nu \rangle, \quad \text{s.t.} \quad \langle \nu, \sigma \rangle = 0, \quad 0 \preceq \nu \preceq 1
\]

Since \( \nu \preceq 0 \) and \( \sigma \preceq 0 \), Eq (4) is equivalent to Eq (5):

\[
\text{rank}(X) = \min_{\nu} \quad \langle 1, 1 - \nu \rangle, \quad \text{s.t.} \quad \nu \odot \sigma = 0, \quad 0 \preceq \nu \preceq 1
\]

where \( \odot \) denotes the Hadamard product (also known as the entrywise product). Note that for all \( i \in [r] \) when \( \sigma_i = 0, v_i = 1 \) will be achieved by minimization, when \( \sigma_i \neq 0, v_i \) will be enforced by the constraint. Since the objective function in Eq (4) is linear, minimization is always achieved at the boundaries of the feasible solution space. Thus, \( V^* = U \text{diag}(1 - \text{sign}(\sigma))U^T \). We thus complete the proof of this lemma. \( \square \)

\( ^2 \text{sign}(x) = \begin{cases} -1, & x < 0; \\ 0, & x = 0; \\ 1, & x > 0. \end{cases} \)
The following lemma shows how to compute the generalized Singular Value Thresholding (SVT) operator which is involved in $V$-subproblem in our PADM Algorithm.

**Lemma 3.** Assume that $W$ has the SVD decomposition that $W = U \text{diag}(\sigma) U^T$. The optimal solution of the following optimization problem

$$\arg \min_{V} \frac{1}{2} \|V - W\|_F^2 + I_{\Delta}(V)$$

(6)

can be computed as $U \text{diag}(\min(1, \max(0, \sigma))) U^T$. Here $I_{\Delta}$ is an indicator function of the convex set $\Delta \triangleq \{ V \mid 0 \preceq V \leq I \}$ with $I_{\Delta}(V) \triangleq \{0, V \in \Delta, \text{otherwise} \}$.

**Proof.** The proof of this lemma is very natural. For completeness, we present our proof here. For notation convenience, we use $\sigma$ and $z$ to denote the singular values of $W$ and $V$, respectively. We naturally derive the following inequalities:

$$\frac{1}{2} \| V - W \|_F^2 + I_{\Delta}(V) = \frac{1}{2} (\| z \|^2 + \| \sigma \|^2 - 2(V, W)) + I_{\Theta}(z) \leq \frac{1}{2} (\| z \|^2 + \| \sigma \|^2 - 2(z, \sigma)) + I_{\Theta}(z) = \frac{1}{2} \| z - \sigma \|^2 + I_{\Theta}(z)$$

From the von Neumann’s trace inequality, the solution set of (12) must be contained in the set $\{ V \mid V = U \text{diag}(\sigma^+) U^T \}$, where $\sigma^+$ is given by

$$\sigma^+ = \arg \min_z \frac{1}{2} \| z - \sigma \|^2 + I_{\Theta}(z)$$

(7)

Since the optimization problem in Eq (7) is decomposable, a simple computation yields that the solution can be computed in closed form as: $\sigma^+ = \min(1, \max(0, \sigma))$. Therefore, $V^* = U \text{diag}(\min(1, \max(0, \sigma))) U^T$. We thus complete the proof of this lemma.

\[ \square \]

2 Convergence Analysis

The global convergence of ADM for convex problems was given by He and Yuan in [He and Yuan 2012] under an elegant variational inequality framework. However, since our MPEC optimization problem is non-convex, the convergence analysis for ADM needs additional conditions. In non-convex optimization, convergence to a stationary point (local minimum) is the best convergence property that we can hope for. Under boundedness condition, we show that the sequence generated by the proximal ADM converges to a KKT point.

For the ease of discussions, we define:

$$u \triangleq \{ X, V \}, \quad s \triangleq \{ X, V, \pi \}$$

(8)

and

$$\Omega \triangleq \{ X \mid 0 \preceq X \preceq \kappa I \}, \quad \Delta \triangleq \{ V \mid 0 \preceq V \leq I \}$$

(9)

First of all, we present the first-order KKT conditions of the MPEC reformulation optimization problem. Based on the augmented Lagrangian function of the MPEC reformulation, we naturally derive the following KKT conditions of the optimization problem for $\{ x^*, V^*, \pi^* \}$:

$$\begin{align*}
0 & \in \partial I_{\Theta}(X^*) + A^T \partial g(A(X^*) - b) + \pi V^* \\
0 & \in \partial I_{\Delta}(V^*) - \lambda I + \pi X^* \\
0 & = (V^*, X^*)
\end{align*}$$

whose existence can be guaranteed by Robinson’s constraint qualification [Rockafellar, Wets, and Wets 1998].

First of all, we prove the subgradient lower bound for the iterates gap by the following lemma.

**Lemma 4.** Assume that $\pi^k$ is bounded for all $k$, then there exists a constant $\varpi > 0$ such that the following inequality holds:

$$\| \partial L(s^{k+1}) \| \leq \varpi \| s^{k+1} - s^k \|$$

(10)

**Proof.** By the optimal condition of the $X$-subproblem and $V$-subproblem, we have:

$$\begin{align*}
0 & \in D(X^{k+1} - X^k) + A^T \partial g(A(X^{k+1}) - b) + \pi^k V^k + \alpha(V^k, X^{k+1}) V^k + \partial I_{\Theta}(X^{k+1}) \\
0 & \in E(V^{k+1} - V^k) - \lambda I + \pi^k X^{k+1} + \alpha(V^{k+1}, X^k) X^{k+1} + \partial I_{\Delta}(X^{k+1})
\end{align*}$$

(11)

By the definition of $L$ we have that

$$\begin{align*}
\partial L_X(s^{k+1}) &= A^T \partial g(A(X^{k+1} - b) + \pi^k V^{k+1} + \alpha(X^{k+1}, X^k) V^{k+1} + \partial I_{\Theta}(X^{k+1}) \\
&= - (\pi + \alpha(V^k, X^{k+1})) V^k + (\pi^{k+1} + \alpha(V^{k+1}, X^{k+1})) V^{k+1} \\
&\quad - D(X^{k+1} - X^k) \\
&= -(\pi^k + \alpha(V^k, X^{k+1})) V^k + (\pi^{k+1} + \alpha(V^{k+1}, X^{k+1})) V^{k+1} \\
&\quad + (\pi^{k+1} + \alpha(V^{k+1}, X^{k+1})) V^{k+1} - D(X^{k+1} - X^k) \\
&= -(\pi^{k+1} + \alpha(V^k - V^{k+1}, X^{k+1})) V^k + (\pi^{k+1} - \pi^{k}) V^{k+1} \\
&\quad + \pi^{k+1} (V^{k+1} - V^k) - D(X^{k+1} - X^k)
\end{align*}$$

(12)

The first step uses the definition of $L_X(s^{k+1})$, the second step uses Eq (11), the third step uses $V^k + V^{k+1} - V^{k+1} = V^k$, the fourth step uses the multiplier update rule for $\pi^k$. Assume that $\pi^{k+1}$ is bounded by a constant $\rho$ that $\pi^{k+1} \leq \rho$. We have:

$$\begin{align*}
\| \partial L_X(s^{k+1}) \|_F &\leq \| \alpha(V^k - V^{k+1}, X^{k+1}) V^k \| + \| (\pi^{k+1} - \pi^{k}) V^{k+1} \| \\
&\quad + \| \pi^{k+1} (V^{k+1} - V^k) \| + \| D(X^{k+1} - X^k) \| \\
&\leq 2 n \alpha \| V^k - V^{k+1} \|_F + 2 \| \pi^{k+1} - \pi^{k} \| \\
&\quad + \rho \| V^k - V^{k+1} \|_F + \| D \| \cdot \| X^{k+1} - X^k \|_F
\end{align*}$$

(13)
Similarly, we have
\[ \partial \mathcal{L}(s^{k+1}) = \partial I_\Delta(V^{k+1}) - \lambda I + \pi^{k+1} X^{k+1} + \alpha (V^{k+1}, X^{k+1}) X^{k+1} \]
\[ = (\pi^{k+1} - \pi^k) X^{k+1} - E(V^{k+1} - V^k) \]
\[ \partial \mathcal{L}_\pi(s^{k+1}) = \langle I - V^{k+1}, X^{k+1} \rangle = \frac{1}{\alpha} (\pi^{k+1} - \pi^k) \]

Then we derive the following inequalities:
\[ ||\partial \mathcal{L}(s^{k+1})||_F \leq \kappa ||\pi^k - \pi^{k+1}|| + ||E|| \cdot ||V^{k+1} - V^k||_F \]  \hspace{1cm} (14)
\[ ||\partial \mathcal{L}_\pi(s^{k+1})|| \leq \frac{1}{\alpha} ||\pi^{k+1} - \pi^k|| \]  \hspace{1cm} (15)

Combining Eqs. (13) and (15), we conclude that there exists \( \omega > 0 \) such that the following inequality holds:
\[ ||\partial \mathcal{L}(s)||_F \leq \omega ||s - s^k||. \]

Thus, we complete the proof of this lemma.

The following lemma is useful in our convergence analysis.

Lemma 5. Assume that \( \pi^k \) is bounded for all \( k \), then we have the following inequality:
\[ \sum_{k=0}^{\infty} ||s^k - s^{k+1}||^2 < +\infty \]  \hspace{1cm} (16)

In particular the sequence \( ||s^k - s^{k+1}|| \) is asymptotically regular, namely \( ||s^k - s^{k+1}|| \to 0 \) as \( k \to \infty \). Moreover any cluster point of \( s^k \) is a stationary point of \( \mathcal{L} \).

Proof. Due to the initialization and the update rule of \( \pi \), we conclude that \( \pi^k \) is nonnegative and monotone non-decreasing. Moreover, as \( k \to \infty \), we have:
\[ \langle X^{k+1}, V^{k+1} \rangle = 0. \]  \hspace{1cm} (17)

Suppose that \( \langle X^{k+1}, V^{k+1} \rangle \neq 0 \), then \( \pi^k = +\infty \) as \( k \to \infty \). This contradicts our assumption that \( \pi^k \) is bounded. Therefore, we conclude that as \( k \to +\infty \) it holds that:
\[ \sum_{i=1}^{k} ||\pi^{i+1} - \pi^i|| < +\infty \]  \hspace{1cm} (18)

On the other hand, we naturally derive the following inequalities:
\[ \mathcal{L}(X^{k+1}, V^{k+1}, \pi^{k+1}) = \mathcal{L}(X^{k+1}, V^{k+1}, \pi^k) + (\pi^{k+1} - \pi^k, \langle V^{k+1}, X^{k+1} \rangle) \]
\[ \leq \mathcal{L}(X^k, V^{k+1}, \pi^k) + \frac{1}{\alpha} ||\pi^{k+1} - \pi^k||^2 \]
\[ + \frac{1}{\alpha} ||\pi^{k+1} - \pi^k||^2 \]
\[ \leq \mathcal{L}(X^k, V^k, \pi^k) - \frac{\mu}{2} ||V^{k+1} - X^k||^2 \]
\[ - \frac{\mu}{2} ||V^{k+1} - V^k||^2 + \frac{1}{\alpha} ||\pi^{k+1} - \pi^k||^2 \]  \hspace{1cm} (19)

The first step uses the definition of \( \mathcal{L} \); the second step uses update rule of the Lagrangian multiplier \( \pi \); the third and fourth step use the \( \mu \)-strongly convexity of \( \mathcal{L} \) with respect to \( X \) and \( V \), respectively. We define \( C = \mathcal{L}(X^0, V^0, \pi^0) - \mathcal{L}(X^{k+1}, V^{k+1}, \pi^{k+1}) + \frac{1}{\alpha} \sum_{i=1}^{k} ||\pi^{i+1} - \pi^i||^2 \). Clearly, by the boundedness of \( X^k, V^k \) and \( \pi^k \), both \( \mathcal{L}(X^{k+1}, V^{k+1}, \pi^{k+1}) \) and \( C \) are bounded. Summing Eq. (19) over \( i = 1, 2, \ldots, k \), we have:
\[ \frac{\mu}{2} \sum_{i=1}^{k} ||u^{i+1} - u^i||^2 \leq C \]  \hspace{1cm} (20)

Thus, combining Eq. (17) and Eq. (19), we have:
\[ \sum_{k=1}^{\infty} ||s^k - s^{k+1}||^2 < +\infty \]  \hspace{1cm} (21)

In particular, \[ ||s^k - s^k|| \to 0 \] as \( k \to \infty \). Thus, we complete the proof of this lemma.

Remarks: Lemma 5 states that any cluster point is the KKT point. Strictly speaking, this result does not imply the convergence of the algorithm. This is because the boundedness of \( \sum_{k=0}^{\infty} ||s^k - s^{k+1}||^2 \) does not imply that the sequence \( s^k \) is convergent\(^3\). In what follows, we aim to prove stronger result in Theorem 1.

Our analysis is mainly based on a recent non-convex analysis tool called Kurdyka-Łojasiewicz inequality (Attouch et al. 2010; Bolte, Sabach, and Teboulle 2014). One key condition of our proof requires that the Lagrangian function \( \mathcal{L}(s) \) satisfies the so-call (KL) property in its effective domain. It is so-called the semi-algebraic function satisfy the Kurdyka-Łojasiewicz property. It is not hard to validate that the Lagrangian function \( \mathcal{L}(s) \) is a semi-algebraic function\(^4\). This is not surprising since semi-algebraic function is ubiquitous in applications. Interested readers can refer to (Xu and Yin 2013) for more details. We now present the following proposition established in (Attouch et al. 2010).

Proposition 1. For a given semi-algebraic function \( \mathcal{L}(s) \), for all \( s \in \text{dom}\mathcal{L} \), there exists \( \theta \in (0, 1) \), \( \eta \in (0, +\infty) \) a neighborhood \( S \) of \( s \) and a concave and continuous function \( \varphi(t) = c \cdot t^{1-\eta}, t \in [0, \eta] \) such that for all \( s \in S \) and satisfies \( \mathcal{L}(s) \in (\mathcal{L}(s), \mathcal{L}(s) + \eta) \), the following inequality holds:
\[ \text{dist}(0, \partial \mathcal{L}(s)) \varphi(\mathcal{L}(s) - \mathcal{L}(s)) \geq 1, \forall s \]
where \( \text{dist}(0, \partial \mathcal{L}(s)) = \min \{||u^*|| : u^* \in \partial \mathcal{L}(s)\} \).

\(^3\)One typical counter-example is \( s^k = \sum_{i=1}^{k} \frac{1}{k} \). Clearly, \( \sum_{k=1}^{\infty} ||s^k - s^{k+1}||^2 = \sum_{k=1}^{\infty} \left( \frac{1}{k} \right)^2 \) is bounded by \( \frac{\pi^2}{6} \); however, \( s^k \) is divergent since \( s^k = \ln(k) + C_n \), where \( C_n \) is the well-known Euler’s constant.

\(^4\)Note that semi-algebraic functions include (i) real polynomial functions, (ii) finite sums and products of semi-algebraic functions, and (iii) indicator functions of semi-algebraic sets. Using these definitions repeatedly, the graph of \( L(s) : \{(s, t) \mid t = L(s)\} \) can be proved to be a semi-algebraic set. Therefore, \( L(s) \) is a semi-algebraic function.
The following theorem establishes the convergence properties of the proposed algorithm under a boundedness condition.

**Theorem 1.** Assume that $\pi^k$ is bounded for all $k$. Then we have the following inequality:

$$\sum_{k=0}^{+\infty} \|s^k - s^{k+1}\| < \infty$$  (21)

Moreover, as $k \to \infty$, Algorithm 1 converges to the first order KKT point of the MPEC reformulation optimization problem.

**Proof.** For simplicity, we define $R^k = \varphi(\mathcal{L}(s^k) - \mathcal{L}(s^*)) - \varphi(\mathcal{L}(s^{k+1}) - \mathcal{L}(s^*))$. We naturally derive the following inequalities:

$$\frac{\mu}{2} \|u^{k+1} - u^k\|^2 - \frac{1}{\alpha} \|\pi^{k+1} - \pi^k\|^2$$

$$\leq \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$= \varphi(\mathcal{L}(s^k) - \mathcal{L}(s^*)) - \varphi(\mathcal{L}(s^{k+1}) - \mathcal{L}(s^*))$$

$$\leq R^k \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$\leq R^k \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$\leq R^k \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$\leq R^k \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$\leq R^k \mathcal{L}(s^k) - \mathcal{L}(s^{k+1})$$

$$= \frac{\mu}{2} \|u^{k+1} - u^k\|^2 + \frac{1}{\alpha} \|\pi^{k+1} - \pi^k\|^2$$

Summing the inequality above over $i = 1, 2, \ldots, k$, we have:

$$\sqrt{\frac{\mu}{8}} \sum_{i=1}^{k} \|u^{i+1} - u^i\| \leq Z + (\|u^1 - u^0\| + \|u^{k+1} - u^k\|)$$

where $Z = \frac{\alpha}{\mu} \sum_{i=1}^{k} (\varphi(\mathcal{L}(s^0) - \mathcal{L}(s^i)) - \varphi(\mathcal{L}(s^{k+1}) - \mathcal{L}(s^i)))$ is bounded real number. Therefore, we conclude that as $k \to \infty$, we obtain:

$$\sum_{i=1}^{k} \|u^{i+1} - u^i\| < +\infty$$  (22)

Moreover, by Eq (10) in Lemma 1, we have $\partial \mathcal{L}(s^{k+1}) = 0$. In other words, we have the following results:

$$0 \in \partial I_\Delta (X^{k+1} + A^T y_\Delta (A(X^{k+1}) - b) + \pi^{k+1} V^{k+1})$$

$$0 \in \partial I_\Delta (V^{k+1}) + A + \pi^{k+1} X^{k+1}$$

which imply that $\{X^{k+1}, V^{k+1}, \pi^{k+1}\}$ is a first-order KKT point.  

\[ \Box \]

**Algorithm 2** Classical Alternating Direction Method for Solving the Convex X-Subproblem in Eq (25).

(S.0) Initialize $X^0, Y^0 = 0 \in \mathbb{R}^{n \times n}, y^0 = 0 \in \mathbb{R}^m, z^0 = 0 \in \mathbb{R}^{m \times n}$. Set $t = 0$.

(S.1) Solve the following $(y, X)$-subproblem:

$$X^{t+1} = \arg\min_{X} J(X, y^t, Y^t, Z^t, z^t)$$  (23)

(S.2) Solve the following Y-subproblem:

$$y^{t+1} = \arg\min_{y, 0 \leq Y \leq \kappa I} J(X^{t+1}, y, Y, Z^t, z^t)$$  (24)

(S.3) Update the Lagrange multiplier via the following formula:

$$Z^{t+1} = Z^t + \beta(X^{t+1} - Y^{t+1})$$

$$z^{t+1} = z^t + \gamma(A(X^{t+1}) - b - y^{t+1})$$

(S.4) Set $t := t + 1$ and then go to Step (S.1).

### 3 Solving the Convex Subproblem

The efficiency of Algorithm proximal ADM in Algorithm 1 relies whether the convex subproblem can be efficiently solved. In this section, we aim to solve the following semidefinite optimization subproblem involved in the proposed PADM algorithm:

$$\min_{0 \preceq X_{s+1}} g(A(X) - b) + \frac{\alpha_1}{2} \|B(X)\|_F^2 + \frac{\mu}{2} \|X\|_F^2 + \langle X, C \rangle,$$  (25)

Our solution is naturally based on the classical ADM [He and Yuan 2012, Lin, Liu, and Su 2011]. For completeness,
we include our algorithm details here. First, we introduce two auxiliary vectors \( y \in \mathbb{R}^m \) and \( Y \in \mathbb{R}^{n \times n} \) to reformulate Eq (25) as:

\[
\min_{y, X, Y} \; g(y) + \frac{\alpha}{2} \|B(X)\|_F^2 + \frac{\mu}{2} \|X\|_F^2 + \langle X, C \rangle,
\]

s.t. \( A(X) - b = y, \; X = Y, \; 0 \leq Y \leq \kappa I. \) (26)

Let \( J_{\beta, \gamma} : \mathbb{R}^m \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^m \times \mathbb{R}^{n \times n} \to \mathbb{R} \) be the augmented Lagrangian function in Eq (26)

\[ J(y, X, Y; z, Z) = g(y) + \frac{\alpha}{2} \|B(X)\|_F^2 + \frac{\mu}{2} \|X\|_F^2 + \langle X, C \rangle + \langle z, A(X) - b - y \rangle + \frac{\gamma}{2} \|A(X) - b - y\|_F^2 + \langle Z, X - Y \rangle + \beta \|X - Y\|_F^2, \; s.t. \; 0 \leq Y \leq \kappa I \]

\( z \) and \( Z \) are the Lagrange multipliers associated to the constraints \( A(Y) - b - y = 0 \) and \( X - Y = 0 \), respectively, and \( \gamma, \beta > 0 \) are the penalty parameters. The detailed iteration steps of the classical ADM for Eq (26) are described in Algorithm 2.

Next, we focus our attention on the solutions of subproblems (23) and (24) arising in Algorithm 2.

(i) \( X \)-subproblem. The first-order optimality condition for the variable \( X \) is:

\[ \alpha B^* B (X^{t+1}) + \gamma A^* A (X^{t+1}) + (\mu + \beta) X^{t+1} = E \]

where \( E = \beta Y + \gamma A^* (b + y) - C - A^* z - Z \). Solving this linear system gives:

\[ X^{t+1} = (\alpha B^* B + \gamma A^* A + (\mu + \beta) I)^{-1} E \] (27)

When the dimension of the solution is high, solving this linear system may dominate the computation time. However, one can use iterative conjugate gradient to alleviate this computational burden. We remark that it is also possible to utilize linearized ADM to address this issue (He and Yuan 2012; Lin, Liu, and Su 2011).

(ii) \( (y, Y) \)-subproblem. Variable \( y \) in Eq (24) is updated by solving the following problem:

\[ y^{t+1} = \arg \min_{y \in \mathbb{R}^m} \; g(y) + \frac{\gamma}{2} \|q - y\|^2, \]

with \( q = A(X) - b + z / \gamma \). It reduces to the Moreau proximal operator \( \text{prox}_{\gamma g} \) that can be evaluated efficiently by our assumption in the introduction section.

Variable \( Y \) in Eq (24) is updated by solving the following problem:

\[ Y^{t+1} = \arg \min_{0 \leq Y \leq \kappa I} \; \frac{\beta}{2} \|Y - S\|_F^2 \]

with \( S = X^{t+1} + Z / \beta \). Assume that \( S \) has the spectral decomposition \( S = V \text{diag}(s)V^T \). A simple computation yields that the solution \( Y^{t+1} \) can be computed in closed form as:

\[ Y^{t+1} = V \text{diag}(\max(0, \min(\kappa I, s))) V^T \]

The exposition above shows that the computation required in each iteration of Algorithm 2 is insignificant.

Classical ADM has excellent convergence both in theory and in practice for convex problems. The convergence of Algorithm 2 can be obtained since the Féjer monotonicity of iterative sequences \( \{y^t, X^t, Y^t, Z^t\} \) holds due to convexity. For the proof of convergence of Algorithm 2 interested readers can refer to (He and Yuan 2012) for more details.

### 4 Additional Experimental Results

In this section, we present some additional experimental results to demonstrate the superiority of our proposed proximal ADM algorithm. Due to page limitations, we were not able to add these results in the submission. We extend our method for sensor network localization problem in the presence of laplace noise and uniform noise. We show our results of laplace noise in Figure 7 and uniform noise in Figure 8. These experimental results strengthen our conclusions drawn in our submission.

### References

Figure 1: Asymptotic behavior on minimum-rank sensor network localization problem in the presence of laplace noise. We plot the values of rank (blue) and dist (red) against the number of iterations, as well as how the sensors have been located at different stages of the process (1, 2, 3, 4, 5).

(a) FM, dist = 0.20, rank = 8.
(b) TAM, dist = 0.21, rank = 5.
(c) LPAM, dist = 0.19, rank = 6.
(d) LDHM, dist = 0.18, rank = 2.
(e) PDA, dist = 0.27, rank = 11.
(f) PADM, dist = 0.15, rank = 2.

Figure 2: Performance comparison on 2d data in the presence of laplace noise.

(a) FM, dist = 0.32, rank = 10.
(b) TAM, dist = 0.29, rank = 8.
(c) LPAM, dist = 0.21, rank = 3.
(d) LDHM, dist = 0.32, rank = 3.
(e) PDA, dist = 0.65, rank = 9.
(f) PADM, dist = 0.21, rank = 3.

Figure 3: Performance comparison on 3d data in the presence of laplace noise.

(b) dist and rank comparisons on 2d data
(c) dist and rank comparisons on 3d data
(d) dist and rank comparisons on 7d data

Figure 4: Performance comparison with varying the number of sensor $n$ in the presence of laplace noise.

(a) dist and rank comparisons on 2d data
(b) dist and rank comparisons on 3d data
(c) dist and rank comparisons on 7d data

Figure 5: Performance comparison with varying the noise level $s$ in the presence of laplace noise.

(a) dist and rank comparisons on 2d data
(b) dist and rank comparisons on 3d data
(c) dist and rank comparisons on 7d data

Figure 6: Performance comparison with varying the sampling ratio $r$ in the presence of laplace noise.
Figure 7: Asymptotic behavior on minimum-rank sensor network localization problem in the presence of uniform noise. We plot the values of rank (blue) and dist (red) against the number of iterations, as well as how the sensors have been located at different stages of the process (1, 2, 3, 4, 5).

Figure 8: Performance comparison on 2d data in the presence of uniform noise.

Figure 9: Performance comparison on 3d data in the presence of uniform noise.

Figure 10: Performance comparison with varying the number of sensor $u$ in the presence of uniform noise.

Figure 11: Performance comparison with varying the noise level $s$ in the presence of uniform noise.

Figure 12: Performance comparison with varying the sampling ratio $r$ in the presence of uniform noise.