A Proximal Alternating Direction Method for Semi-Definite Rank Minimization

Ganzhao Yuan and Bernard Ghanem

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

Abstract

Semi-definite rank minimization problems model a wide range of applications in both signal processing and machine learning fields. This class of problem is NP-hard in general. In this paper, we propose a proximal Alternating Direction Method (ADM) for the well-known semi-definite rank regularized minimization problem. Specifically, we first reformulate this NP-hard problem as an equivalent biconvex M-PEC (Mathematical Program with Equilibrium Constraints), and then solve it using proximal ADM, which involves solving a sequence of structured convex semi-definite subproblems to find a desirable solution to the original rank regularized optimization problem. Moreover, based on the Kurdyka-Łojasiewicz inequality, we prove that the proposed method always converges to a KKT stationary point under mild conditions. We apply the proposed method to the widely studied and popular sensor network localization problem. Our extensive experiments demonstrate that the proposed algorithm outperforms state-of-the-art low-rank semi-definite minimization algorithms in terms of solution quality.

Keywords: Semidefinite Rank Minimization, MPEC, Sensor Network Localization, Kurdyka-Łojasiewicz Inequality, Proximal ADM, Convergence Analysis

1 Introduction

In this paper, we mainly focus on the following composite rank regularized semi-definite optimization problem:

$$\min_{0 \leq \mathbf{X} \leq \kappa \mathbf{I}} g\left(\mathbf{\mathcal{A}}(\mathbf{X}) - \mathbf{b} \right) + \lambda \operatorname{rank}(\mathbf{X}), \tag{1}$$

where λ and κ are strictly positive scalars, $\mathbf{X} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^m$, the linear map $\mathcal{A}(\cdot) : \mathbb{R}^{n \times n} \to \mathbb{R}^m$ is defined as $\mathcal{A}(\mathbf{X}) = [\langle \mathbf{A}^{(1)}, \mathbf{X} \rangle, ..., \langle \mathbf{A}^{(m)}, \mathbf{X} \rangle]^T$, and the matrices $\mathbf{A}^{(i)} \in \mathbb{R}^{n \times n}, i = 1, ..., m$ are given. Moreover, $g(\cdot)$ is a simple proper lower semi-continuous convex function such that its Moreau proximal operator $\operatorname{prox}_g(\mathbf{c}) \triangleq \min_{\mathbf{z}} g(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{c}\|_2^2$ can be efficiently computed.

Note that we constrain X with a ball of radius κ to ensure the boundedness of the solution. This is to guarantee convergence; however, it interestingly does not increase the computational complexity of our proposed solver much. If

no prior information on κ is known, one can set it to a sufficiently large value in practice. We remark that another equally popular optimization model is to formulate Eq (1) into a rank-constrianted/fixed-rank optimization problem. However, in real applications, the true rank is usually unknown or, for the constrained problem the low-rank solution may not even exist. In this sense, Eq (1) is more appealing.

The optimization problem in Eq(1) describes many applications of interest to both the signal processing and machine learning communities, including sensor network localization (Biswas et al. 2006b), near-isometric embedding (Chinmay Hegde 2015), low-dimensional Euclidean embedding (Dattorro 2011; Recht, Fazel, and Parrilo 2010), non-metric multidimensional scaling (Agarwal et al. 2007), low-rank metric learning (Law, Thome, and Cord 2014; Liu et al. 2015; Cong et al. 2013), low-rank kernel learning (Meka et al. 2008), optimal beamforming (Huang and Palomar 2010), ellipsoid fitting (Saunderson et al. 2012), optimal power flow (Louca, Seiler, and Bitar 2013), and cognitive radio networks (Yu and Lau 2011), to name a few.

We mainly focus on positive semi-definite (PSD) optimization. However, there are many applications (Candès and Recht 2009; Zhang et al. 2013; 2012) such as matrix completion and image classification, where the solutions are not necessarily PSD. Fortunately, one can resolve this issue by embedding any general matrix with a larger PSD hull (refer to Semi-definite Embedding Lemma in the supplementary material). Moreover, many SDP optimization problems are inherently low-rank. For example, for standard semi-definite programming it has been proven that the rank of the solution is upper-bounded by $\lfloor \frac{1}{2}(\sqrt{8m+1}-1) \rfloor$, where m is the number of equality constraints (Moscato, Norman, and Pataki 1998). For metric learning (Roweis and Saul 2000) and sensor network localization problems (Biswas and Ye 2004), the data distance metric often lives in a much lower dimensional space.

In this paper, we give specific attention to solving the popular sensor network localization problem (Biswas et al. 2006b; Zhang et al. 2010; Ji et al. 2013; Wang et al. 2008; Krislock and Wolkowicz 2010), which falls into the low-rank semi-definite optimization framework of Eq (1). The problem of finding the positions of all the nodes given a few anchor nodes and the relative distance information between the nodes is called sensor network localization. It is an im-

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

portant task in wireless network applications such as target tracking, emergency response, logistics support and mobile advertising (Ji et al. 2013).

Challenges and Contributions: There are mainly three challenges of existing work. (a) The general rank minimization problem in Eq (1) is NP-hard due to the non-convexity and discontinuous nature of the rank function. There is little hope of finding the global minimum efficiently in all instances. In order to deal with this issue, we reformulate the rank minimization problem as an equivalent augmented optimization problem with a bilinear equality constraint using a variational characterization of the rank function. Then, we propose a proximal Alternating Direction Method (ADM) to solve it. The resulting algorithm seeks a desirable solution to the original optimization problem without requiring any approximation. (b) The second aspect is the suboptimality of the semi-definite optimization for sensor network localization. Existing approximation solutions, such as Schatten's ℓ_p norm method (Ji et al. 2013), only give suboptimal solutions. We resolve this issue by considering an exact method for solving general rank regularized optimization. Experimental results show that our method is more effective than the state-of-the-art. (c) The third aspect is the convergence of the optimization algorithm. Many existing convergence results for non-convex rank minimization problems tend to be either limited to unconstrained problems or unapplicable to constrained optimization. We resolve this issue by combining the complementarity reformulation of the problem and a recent non-convex analysis tool called the Kurdyka-Łojasiewicz inequality(Attouch and Bolte 2009; Bolte, Sabach, and Teboulle 2014). In fact, we prove that the proposed ADM algorithm converges to a first-order KK-T point under mild conditions. To the best of our knowledge, this is the first multiplier method for solving rank minimization problem with guaranteed convergence.

Organization and Notations: This paper is organized as follows. Section 2 provides a brief description of the sensor location network problem and the related work. Section 3 presents our proposed proximal ADM optimization framework and Section 4 summarizes the experimental results. Finally, Section 5 concludes this paper. Throughout this paper, we use lowercase and uppercase boldfaced letters to denote real vectors and matrices respectively. $\langle \mathbf{X}, \mathbf{Y} \rangle$ is the Euclidean inner product of \mathbf{X} with $\mathbf{Y}, \boldsymbol{\sigma}(\mathbf{X})$ is the eigenvalues of \mathbf{X} , and $diag(\mathbf{x})$ is a diagonal matrix with \mathbf{x} in the main diagonal entries. Finally, $\|\mathbf{X}\|_{\mathbf{H}} \triangleq (vec(\mathbf{X})^T \mathbf{H} vec(\mathbf{X}))^{1/2}$ denotes the generalized vector norm.

2 Preliminaries and Related Work

2.1 Preliminaries

The sensor network localization problem is defined as follows. We are given c anchor points $\mathbf{A} = [\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_c}] \in \mathbb{R}^{c \times d}$, whose locations are known, and u sensor points $\mathbf{S} = [\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_u}] \in \mathbb{R}^{u \times d}$ whose locations we wish to determine. Furthermore, we are given the Euclidean distance values χ_{kj} between $\mathbf{a_k}$ and $\mathbf{s_j}$ for some k, j, and χ_{ij} between $\mathbf{s_i}$ and $\mathbf{s_j}$ for some i, j. Specifically, we model the noisy distance measurements as:

 $\|\mathbf{a}_{\mathbf{k}} - \mathbf{s}_{\mathbf{j}}\|_2^2 = \chi_{kj}^2 + \varepsilon_{kj}, \ \|\mathbf{s}_{\mathbf{i}} - \mathbf{s}_{\mathbf{j}}\|_2^2 = \chi_{ij}^2 + \varepsilon_{ij},$

where each $(k, j) \in \Pi_{as}$ and each $(i, j) \in \Pi_{ss}$ are some selected pairs of the known (noisy) distances χ . We denote the noise variable as $\varepsilon \in \mathbb{R}^{|\Pi|}$, where $|\Pi|$ is the total number of elements in $\Pi \triangleq \Pi_{as} \cup \Pi_{ss}$. Then, we formulate the distances in the following matrix representation:

$$\begin{aligned} \|\mathbf{s}_{\mathbf{i}} - \mathbf{s}_{\mathbf{j}}\|_{2}^{2} &= \mathbf{e}_{\mathbf{i}\mathbf{j}}^{T}\mathbf{S}\mathbf{S}^{T}\mathbf{e}_{\mathbf{i}\mathbf{j}}, \\ \|\mathbf{a}_{\mathbf{k}} - \mathbf{s}_{\mathbf{j}}\|_{2}^{2} &= \left(\begin{smallmatrix}\mathbf{a}_{\mathbf{k}}\\\mathbf{e}_{\mathbf{j}}\end{smallmatrix}\right)^{T}\left(\begin{smallmatrix}\mathbf{I}_{d} & \mathbf{S}\\\mathbf{S}^{T} & \mathbf{S}^{T}\mathbf{S}\end{smallmatrix}\right)\left(\begin{smallmatrix}\mathbf{a}_{\mathbf{k}}\\\mathbf{e}_{\mathbf{j}}\end{smallmatrix}\right), \end{aligned}$$

where $\mathbf{e_{ij}} \in \mathbb{R}^u$ has 1 at the i^{th} position, -1 at the j^{th} position and zero everywhere else. Hence, we formulate sensor network localization as the following optimization:

Find
$$\mathbf{S} \in \mathbb{R}^{d \times u}$$
,
s.t. $\mathbf{e_{ij}}^T \mathbf{S}^T \mathbf{S} \mathbf{e_{ij}} = \mathbf{d}_{ij}^2 + \boldsymbol{\varepsilon}_{ij}, \|\boldsymbol{\varepsilon}_{ij}^T \, \boldsymbol{\varepsilon}_{kj}^T \|_q \leq \delta$
 $\begin{pmatrix} \mathbf{a_k} \\ \mathbf{e_j} \end{pmatrix}^T \begin{pmatrix} \mathbf{I}_d & \mathbf{S} \\ \mathbf{S}^T & \mathbf{S}^T \mathbf{S} \end{pmatrix} \begin{pmatrix} \mathbf{a_k} \\ \mathbf{e_j} \end{pmatrix} = \mathbf{d}_{kj}^2 + \boldsymbol{\varepsilon}_{kj}.$ (2)

Here q can be 1 (for laplace noise), 2 (for Gaussian noise) or ∞ (for uniform noise), see e.g. (Yuan and Ghanem 2015). The parameter δ which depends on the noise level needs to be specified by the user. By introducing the PSD hull $\mathbf{X} = \begin{pmatrix} \mathbf{I}_d & \mathbf{S} \\ \mathbf{S}^T & \mathbf{S}^T \mathbf{S} \end{pmatrix} \in \mathbb{R}^{(u+d) \times (u+d)}$, we have the following rank minimization problem (Ji et al. 2013):

$$\begin{array}{ll} \min_{\mathbf{X}} & rank(\mathbf{X}) \\ s.t. & \langle (0; \mathbf{e_{ij}})(0; \mathbf{e_{ij}})^T, \mathbf{X} \rangle = \mathbf{d}_{ij}^2 + \varepsilon_{ij} \\ & \langle (\mathbf{a_k}; \mathbf{e_j})(\mathbf{a_k}; \mathbf{e_j})^T, \mathbf{X} \rangle = \mathbf{d}_{kj}^2 + \varepsilon_{kj} \\ & \mathbf{X}_{1:d,1:d} = \mathbf{I}_d, \| \boldsymbol{\varepsilon}_{ij}^T \, \boldsymbol{\varepsilon}_{kj}^T \|_q \leq \delta, \ \mathbf{X} \succeq 0, \quad (3) \end{array}$$

It is not hard to validate that Eq (3) is a special case of the general optimization framework in Eq (1).

2.2 Related Work

This subsection presents a brief review of existing related work, from the viewpoint of sensor network localization and semi-definite rank minimization algorithms.

Sensor network localization is a well studied problem in distance geometry (Biswas et al. 2006a; Rallapalli et al. 2010; Krislock and Wolkowicz 2010; Dattorro 2011). Several convex/non-convex approaches have been proposed in the literature. Semi-definite programming relaxation for this problem was initially proposed by (Biswas and Ye 2004). The basic idea of this approach is to convert the nonconvex quadratic constraints into linear matrix inequality constraints by using an SDP lifting technique to remove the quadratic term in the optimization problem. It was subsequently shown that if there is no noise in the measured distances, the sensor network localization problem can be solved in polynomial time under a unique solution assumption (So and Ye 2007). However, if the sensor network problem does not have a unique solution, there must exist a higher rank localization solution that minimizes the objective function. In this case, SDP relaxation always produces this maximal rank dimensional solution. The classical way to obtain a low dimensional solution is to project

Optimization Algorithms and References	Description
(a) convex trace norm (Fazel 2002; Candès and Tao 2010)	$rank(\mathbf{X}) \approx tr(\mathbf{X})$
(b) nonlinear factorization (Burer and Monteiro 2003)	$rank(\mathbf{X}) \leq k, ext{ with } \mathbf{X} pprox \mathbf{LL^T}, \ \mathbf{L} \in \mathbb{R}^{n imes k}$
(c) Schatten ℓ_p norm (Lu 2014; Nie, Huang, and Ding 2012)	$rank(\mathbf{X}) \approx \ \sigma(\mathbf{X})\ _p$
(d) log-det heuristic (Fazel, Hindi, and Boyd 2003; Deng et al. 2013)	$rank(\mathbf{X}) \approx \log \det(\mathbf{X} + \epsilon \mathbf{I})$
(e) truncated nuclear norm (Hu et al. 2013; Miao, Pan, and Sun 2015)	$rank(\mathbf{X}) \leq k \ \Leftrightarrow \ tr(\mathbf{X}) = \ \sigma(\mathbf{X})\ _{ ext{top-k}}$
(f) pseudo-inverse reformulation(Zhao 2012)	$rank(\mathbf{A}) = rank(\mathbf{A}^{\dagger}\mathbf{A}) = tr(\mathbf{A}^{\dagger}\mathbf{A})$
(g) iterative hard thresholding (Zhang and Lu 2011; Lu and Zhang 2013)	$\min_{\mathbf{X}} \frac{1}{2} \ \mathbf{X} - \mathbf{X}'\ _F^2 + rank(\mathbf{X})$
(h) MPEC reformulation [this paper],(Yuan and Ghanem 2015)	$rank(\mathbf{X}) = \min_{0 \leq \mathbf{V} \leq \mathbf{I}} tr(\mathbf{I} - \mathbf{V}), \ s.t. \ \langle \mathbf{V}, \mathbf{X} \rangle = 0$

Table 1: Semi-definite rank minimization algorithms.

the high dimensional solution to the desirable space using eigenvalue decomposition, but this generally only produces sub-optimal results. Second-order cone programming relaxation was proposed in (Tseng 2007), which has superior polynomial complexity. However, this technique obtains good results only when the anchor nodes are placed on the outer boundary, since the positions of the estimated remaining nodes lie within the convex hull of the anchor nodes. Due to the high computational complexity of the standard SDP algorithm, the work of (Wang et al. 2008; Pong and Tseng 2011) considers further relaxations of the semi-definite programming approach to address the sensor network localization problem. Very recently, the work of (Ji et al. 2013) explores the use of a nonconvex surrogate of the rank function, namely the Schatten ℓ_p -norm, in network localization. Although the resulting optimization is nonconvex, they show that a first-order critical point can be approximated in polynomial time by an interior-point algorithm.

Several semi-definite rank minimization algorithms have been studied in the literature (See Table 1). (a) Convex trace norm (Fazel 2002) is a lower bound of the rank function in the sense of operator (or spectral) norm. It is proven to lead to a near optimal low-rank solution (Candès and Tao 2010; Recht, Fazel, and Parrilo 2010) under certain incoherence assumptions. However, such assumptions may be violated in real applications. (b) Nonlinear factorization (Burer and Monteiro 2003; 2005) replaces the solution matrix \mathbf{X} by a nonlinear matrix multiplication LL^{T} . One important feature of this approach is avoiding the need to perform eigenvalue decomposition. (c) Schatten ℓ_p norm with $p \in (0, 1)$ was considered by (Lu 2014; Nie, Huang, and Ding 2012; Lu et al. 2014) to approximate the discrete rank function. It results in a local gradient Lipschitz continuous function, to which some smooth optimization algorithms can be applied. (d) Log-det heuristic (Fazel, Hindi, and Boyd 2003; Deng et al. 2013) minimizes the first-order Taylor series expansion of the objective function iteratively to find a local minimum. Since its first iteration is equivalent to solving the trace convex relaxation problem, it can be viewed as a refinement of the trace norm. (e) Truncated trace norm (Hu et al. 2013; Miao, Pan, and Sun 2015; Law, Thome, and Cord 2014) minimizes the summation of the smallest (n - k)eigenvalues, where k is the matrix rank. This is due to the fact that these eigenvalues have little effect on the approximation of the matrix rank. (f) Pseudo-inverse reformulations (Zhao 2012) consider an equivalent formulation to the rank function: $rank(\mathbf{A}) = tr(\mathbf{A}^{\dagger}\mathbf{A})$. However, similar to matrix rank, the pseudo-inverse function is not continuous. Fortunately, one can use a Tikhonov regularization technique ¹ to approximate the pseudo-inverse. Inspired by this fact, the work of (Zhao 2012) proves that rank minimization can be approximated to any level of accuracy via continuous optimization. (g) Iterative hard thresholding (Zhang and Lu 2011) considers directly and iteratively setting the largest (in magnitude) elements to zero in a gradient descent format. It has been incorporated into the Penalty Decomposition Algorithm (PDA) framework (Lu and Zhang 2013). Although PDA is guaranteed to converge to a local minimum, it lacks stability. The value of the penalty function can be very large, and the solution can be degenerate when the minimization subproblem is not exactly solved.

From above, we observe that existing methods either produce approximate solutions (method (a), (c), (d) and (g)), or limited to solving feasibility problems (method (b) and (e)). The only existing exact method (method (g)) is the penalty method. However, it often gives much worse results even as compared with the simple convex methods, as shown in our experiments. This unappealing feature motivates us to design a new exact multiplier method in this paper. Recently, the work of (Li and Qi 2011) considers a continuous variational reformulation of the low-rank problem to solve symmetric semi-definite optimization problems subject to a rank constraint. They design an ADM algorithm that finds a stationary point of the rank-constrained optimization problem. Inspired by this work, we consider a augmented Lagrangian method to solve the general semi-definite rank minimization problem by handling its equivalent MPEC reformulation. Note that the formulation in (Li and Qi 2011) can be viewed as a special case of ours, since it assumes that the solution has unit diagonal entries, i.e. $diag(\mathbf{X}) = \mathbf{1}$.

3 Proposed Optimization Algorithm

This section presents our proposed optimization algorithm. Specifically, we first reformulate the optimization problem in Eq (1) as an equivalent MPEC (Mathematical Program with Equilibrium Constraints) in Section 3.1, and then solve the equality constrained optimization problem by a proximal Alternating Direction Method (ADM) in Section 3.2. In Subsection 3.3, we discuss the merits of the MPEC reformulation and the proximal ADM algorithm.

$${}^{1}\mathbf{A}^{\dagger} = \lim_{\epsilon \to 0} (\mathbf{A}^{T}\mathbf{A} + \epsilon \mathbf{I})^{-1}\mathbf{A}^{T} = \lim_{\epsilon \to 0} \mathbf{A}^{T} (\mathbf{A}\mathbf{A}^{T} + \epsilon \mathbf{I})^{-1}$$

Algorithm 1 A Proximal Alternating Direction Method for Solving the Non-Convex MPEC Problem (8)

(S.0) Initialize $\mathbf{X}^0 = \mathbf{0} \in \mathbb{R}^{n \times n}$, $\mathbf{V}^0 = \mathbf{I} \in \mathbb{R}^{n \times n}$, $\pi^0 = 0 \in \mathbb{R}$. Set k = 0 and $\mu > 0$.

(S.1) Solve the following X-subproblem with $\mathbf{D} \triangleq \mu \mathbf{I}$:

$$\mathbf{X}^{k+1} = \underset{0 \leq \mathbf{X} \leq \kappa \mathbf{I}}{\operatorname{arg\,min}} \mathcal{L}(\mathbf{X}, \mathbf{V}^k, \pi^k) + \frac{1}{2} \|\mathbf{X} - \mathbf{X}^k\|_{\mathbf{D}}^2 \qquad (4)$$

(S.2) Solve the following V-subproblem with $\mathbf{E} \triangleq \mu \mathbf{I} + \alpha \|\mathbf{X}^{k+1}\|_F^2 \mathbf{I} - \alpha vec(\mathbf{X}^{k+1})vec(\mathbf{X}^{k+1})^T$:

$$\mathbf{V}^{k+1} = \underset{0 \leq \mathbf{V} \leq \mathbf{I}}{\operatorname{arg\,min}} \ \mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}, \pi^k) + \frac{1}{2} \|\mathbf{V} - \mathbf{V}^k\|_{\mathbf{E}}^2$$
(5)

(S.3) Update the Lagrange multiplier:

$$\pi^{k+1} = \pi^k + \alpha(\langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle) \tag{6}$$

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

3.1 Equivalent MPEC Reformulation

We reformulate the semi-definite rank minimization problem in Eq (1) as an equivalent MPEC from the primal-dual viewpoint. We provide the variational characterization of the rank function in the following lemma.

Lemma 1. For any PSD matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, it holds that:

$$rank(\mathbf{X}) = \min_{0 \le \mathbf{V} \le \mathbf{I}} tr(\mathbf{I} - \mathbf{V}), \ s.t. \ \langle \mathbf{V}, \mathbf{X} \rangle = 0,$$
(7)

and the unique optimal solution of the minimization problem in Eq (7) is given by $\mathbf{V}^* = \mathbf{U} diag(\mathbf{1} - sign(\boldsymbol{\sigma}))\mathbf{U}^T$, where $\mathbf{X} = \mathbf{U} diag(\boldsymbol{\sigma})\mathbf{U}^T$ is the SVD of \mathbf{X} .

Proof. Refer to the supplementary material. \Box

The result of Lemma 1 implies that the rank regularized problem in Eq(1) is equivalent to

$$\min_{\substack{0 \leq \mathbf{X} \leq \kappa \mathbf{I} \\ 0 \leq \mathbf{V} \leq \mathbf{I}}} g\left(\mathbf{A}(\mathbf{X}) - \mathbf{b}\right) + \lambda tr(\mathbf{I} - \mathbf{V}), s.t. \langle \mathbf{V}, \mathbf{X} \rangle = 0.$$
(8)

in a sense that if \mathbf{X}^* is a global optimal solution of Eq (1), then $(\mathbf{X}^*, \mathbf{U}diag(\mathbf{1} - sign(\boldsymbol{\sigma}))\mathbf{U}^T)$ is globally optimal for Eq (8). The converse is also true.

Eq (8) is a bi-convex problem since it is convex with respect to each of the two variables X and V when the other is fixed. The equality $\langle \mathbf{V}, \mathbf{X} \rangle = 0$ is an equilibrium/complementarity constraint. This is because for all $j \in$ [n] (i) both $\sigma_i(\mathbf{V})$ and $\sigma_i(\mathbf{X})$ are non-negative and (ii) the equality only holds when either component is zero. Compared with Eq (1), Eq (8) is a non-smooth non-convex minimization problem and its non-convexity is only caused by the complementarity constraint. Although the MPEC problem in Eq (8) is obtained by increasing the dimension of the original rank-regularized problem in Eq (1), this does not lead to additional local optimal solutions. In the following section, we will develop an algorithm to solve Eq (8) using proximal ADM and show that such a "lifting" technique can achieve a desirable solution of the original rank regularized optimization problem.

3.2 Proximal ADM Optimization Framework

Here, we give a detailed description of the solution algorithm to the optimization in Eq (8). This problem is rather difficult to solve because it is neither convex nor smooth. To curtail these issues, we propose a solution that is based on proximal ADM (PADM), which updates the primal and dual variables of the augmented Lagrangian function in E-q (8) in an alternating way. The augmented Lagrangian $\mathcal{L} : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R} \to \mathbb{R}$ is defined as:

$$\mathcal{L}(\mathbf{X}, \mathbf{V}, \pi) \triangleq g(\mathcal{A}(\mathbf{X}) - \mathbf{b}) + \lambda tr(\mathbf{I} - \mathbf{V}) + \pi \langle \mathbf{V}, \mathbf{X} \rangle + \frac{\alpha}{2} (\langle \mathbf{V}, \mathbf{X} \rangle)^2, \ s.t. \ 0 \leq \mathbf{X} \leq \kappa \mathbf{I}, \ 0 \leq \mathbf{V} \leq \mathbf{I},$$

where π is the Lagrange multiplier associated with the constraint $\langle \mathbf{V}, \mathbf{X} \rangle = 0$, and $\alpha > 0$ is the penalty parameter. We detail the PADM iteration steps for Eq (8) in Algorithm 1. In simple terms, PADM updates are performed by optimizing for a set of primal variables at a time, while keeping all other primal and dual variables fixed. The dual variables are updated by gradient ascent on the resulting dual problem.

At first glance, Algorithm 1 might seem to be merely an application of PADM on the MPEC reformulation in Eq(8). However, it has some interesting properties that are worth commenting on.

(a) Monotone property. For any feasible solution of variables X in Eq (4) and V in Eq (5), it can be used to show that $\langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \geq 0$. Using the fact that $\alpha^k > 0$ and due to the update rule of π^k , we conclude that π^k is monotone non-increasing. Moreover, if we initialize $\pi^0 = 0$ in the first iteration, π is always non-negative.

(b) Initialization Strategy. We initialize both N^0 to I and π^0 to 0. This is for the sake of finding a reasonable good local minimum in the first iteration as it reduces to a convex trace norm minimization problem for the X-subproblem.

(c) V-Subproblem. Variable V in Eq (5) is updated by solving the following problem:

$$\mathbf{V}^{k+1} = \underset{\substack{0 \leq \mathbf{V} \leq \mathbf{I}}}{\arg\min} - \lambda tr(\mathbf{V}) + \pi \langle \mathbf{V}, \mathbf{X}^{k+1} \rangle + \frac{\alpha}{2} (\langle \mathbf{V}, \mathbf{X}^{k+1} \rangle)^2 + \frac{1}{2} \|\mathbf{V} - \mathbf{V}^k\|_{\mathbf{E}}^2$$
(9)

Introducing the proximal term in the V-subproblem enables finding a closed-form solution. After an elementary calculation, subproblem (9) can be simplified as

$$\mathbf{V}^{k+1} = \underset{0 \le \mathbf{V} \le \mathbf{I}}{\operatorname{arg\,min}} \ \frac{L}{2} \|\mathbf{V} - \mathbf{W}\|_F^2 \tag{10}$$

where $\mathbf{W} = \mathbf{V}^k - \mathbf{G}/L$, with $\mathbf{G} = -\lambda \mathbf{I} + \pi^k \mathbf{X}^{k+1} + \alpha \cdot \mathbf{X}^{k+1} \cdot \langle \mathbf{X}^{k+1}, \mathbf{V}^k \rangle$ and $L = \mu + \alpha ||\mathbf{X}^{k+1}||_F^2$. Assume that $\mathbf{W} = \mathbf{U} diag(\boldsymbol{\chi}) \mathbf{U}^T$. Clearly, Eq (10) has a closed-form solution: $\mathbf{V}^{k+1} = \mathbf{U} diag(\min(\mathbf{1}, \max(\mathbf{0}, \boldsymbol{\chi}))) \mathbf{U}^T$.

(d) X-Subproblem. Variable X in Eq (4) is updated by solving the following structured convex optimization problem:

$$\min_{0 \leq \mathbf{X} \leq \kappa \mathbf{I}} g(\boldsymbol{\mathcal{A}}(\mathbf{X}) - \mathbf{b}) + \frac{\alpha}{2} \|\boldsymbol{\mathcal{B}}(\mathbf{X})\|_F^2 + \frac{\mu}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{X}, \mathbf{C} \rangle,$$

where $\mathcal{B}(\cdot) : \mathbb{R}^{n \times n} \to \mathbb{R}$ is another linear map defined as $\mathcal{B}(\mathbf{X}) \triangleq \langle \mathbf{V}, \mathbf{X} \rangle$, $\mathbf{C} = \pi^k \mathbf{V}^k$. The **X**-subproblem has no closed-form solution, but it can be solved by classical/linearized ADM (He and Yuan 2012; Lin, Liu, and Su 2011). Refer to the supplementary material for more details.

Proximal ADM has excellent convergence properties in practice, but the optimization problem in Eq (8) is nonconvex, so additional conditions are needed to guarantee convergence to a KKT point. In what follows and based on the Kurdyka-Łojasiewicz inequality, we prove that under broad assumptions, our proximal ADM algorithm always converges to a KKT point. Specifically, we have the following convergence result.

Theorem 1. Convergence of Algorithm 1. Assume that π^k is bounded for all k. As $k \to +\infty$, Algorithm 1 converges to a first order KKT point of Eq (8).

Proof. Refer to the supplementary material.

3.3 Discussions

In this paper, we consider a variational characterization of the rank function in Lemma 1. However, other alternative MPEC reformulation exists. Using the result in our previous work (Yuan and Ghanem 2015), we have:

$$rank(X) = \min_{\mathbf{0} \le \mathbf{v} \le \mathbf{1}} \langle \mathbf{1}, \mathbf{1} - \mathbf{v} \rangle, \ s.t. \ \langle \mathbf{v}, \boldsymbol{\sigma}(\mathbf{X}) \rangle = 0$$

where $\sigma(\mathbf{X})$ denotes the eigenvalues of \mathbf{X} . However, such a reformulation is non-convex with respect to \mathbf{X} for general \mathbf{v} . The proposed reformulation in Eq (8) is convex with respect to \mathbf{X} , which is very helpful for convergence. The key strategy of the *biconvex* formulation is enforcing \mathbf{X} and \mathbf{V} to share the same spectral decomposition.

There are two merits behind the MPEC reformulation. (i) Eq (8) is a continuous optimization reformulation. This facilitates analyzing its KKT condition and utilizing existing continuous optimization algorithms to solve the resulting convex sub-problems. (ii) MPEC is an effective way to model certain classes of discrete optimization (Yuan and Ghanem 2015; Bi, Liu, and Pan 2014; Luo, Pang, and Ralph 1996). We argue that, from a practical point of view, improved solutions to Eq (1) can be obtained by reformulating the problem in terms of complementarity constraints.

We propose a proximal ADM algorithm to solve the M-PEC problem. There are three reasons that explain the good performance of our proposed optimization algorithm. (i) It targets a solution to the *original* problem in Eq (1). (i-i) It finds a good initialization. It reduces to the classical convex relaxation method in the first iteration. (iii) It has a monotone/greedy property owing to the complimentarity constraints brought on by the MPEC. The complimentary system characterizes the optimality of the KKT solution. We let $u \triangleq \{\mathbf{X}, \mathbf{V}\}$. Our solution directly handles the complimentary system of Eq (1) which takes the following form (on eigenvalues for the matrix case):

$$\langle f(u), g(u) \rangle = 0, \ f(u) \ge 0, \ g(u) \ge 0$$

The complimentary constraint is the source of all the special properties of MPEC that distinguishes it from general nonlinear optimization. We penalize the complimentary error of $\langle f(u), g(u) \rangle$ (which is always non-negative) and ensure that the error is decreasing in every iteration.

4 Experimental Results

In this section, we provide empirical validation for our proposed method by conducting extensive sensor network localization experiments and performing a thorough comparative analysis with the state-of-the-art. We compare our method (denoted as PADM) with five state-of-the-art and popular algorithms: Feasibility Method (FM) (Biswas et al. 2006a), Trace Approximation Method (TAM) (Biswas et al. 2006a), Schatten ℓ_p Approximation Method (LPAM)²(Ji et al. 2013; Lu et al. 2015), Log-Det Heuristic Method (LDHM) (Fazel, Hindi, and Boyd 2003), and Penalty Decomposition Algorithm (PDA) (Zhang and Lu 2011). We provide our supplementary material and MATLAB implementation online at: http://yuanganzhao.weebly.com/.

4.1 Experimental Setup

Following the experimental setting in (Biswas et al. 2006b), we uniformly generate c anchors (c = 5 in all our experiments) and u sensors in the range [-0.5, 0.5] to generate d-dimensional data points. To generate random and noisy distance measure, we uniformly select $o \triangleq (r \times |\Pi|)$ subset measurements $\tilde{\chi} \in \mathbb{R}^o$ from Π and inject them with noise by $\tilde{\chi} \leftarrow \tilde{\chi} + s \times \varepsilon$, where $\varepsilon \in \mathbb{R}^o$ is noise of unit scale. Here s and r can be viewed as the noise level and sampling ratio, respectively. We consider two ways to measure the quality of the recovered solution \mathbf{X}^3 :

$$\begin{aligned} \operatorname{rank}(\mathbf{X}) &\triangleq \|\boldsymbol{\sigma}(\mathbf{X})\|_{0\epsilon},\\ \operatorname{dist}(\mathbf{S}) &\triangleq (1/n \cdot \sum_{i=1}^{n} \|\mathbf{S}(i,:) - \bar{\mathbf{S}}(i,:)\|_2)^{1/2} \end{aligned}$$

where $\|\mathbf{x}\|_{0-\epsilon}$ is the soft ℓ_0 norm which counts the number of elements whose magnitude is greater than a threshold $\epsilon = 0.01 \cdot \|\mathbf{x}\|, \forall \mathbf{x} \in \mathbb{R}^n$. $\mathbf{\bar{S}}$ is the true position of the sensors.

Table 2: Varying parameters used in the experiments

dimension (d)	2, 3, 7
noise type (q)	$2, 1, \infty$
# sensors (u)	30, 50, 70, 90 , 110, 130, 150, 170, 190
noise level (s)	0.01, 0.03, 0.05 , 0.07,, 0.19
sampling ratio (r)	0.1, 0.2, 0.3 , 0.4,, 1.0

In our experiments, we test the impact of five parameters: d, q, u, s, and r. Although we are mostly interested in d-dimensional (d = 2 or 3) localization problems, Problem (3) is also strongly related to Euclidean distance matrix completion, a larger dimension (e.g. d = 7) is also interesting. The range of all these five parameters is summarized in Table 2. Unless otherwise specified, the default parameters in bold are used. Due to space limitation, we only present our experimental localization results in the presence of Gaussian noise (p = 2). For more experimental results on laplace noise (i.e. p = 1) and uniform noise ($p = \infty$), please refer to supplementary material.

 $^{^{2}}$ Since the interior-point method (Ji et al. 2013) is not convenient to solve the general composite rank minimization problem, we consider an alternative ADM algorithm which is based on generalized singular value thresholding (Lu et al. 2015).

³Note that we need to retrieve **S** from **X** (See Eq (2)).



Figure 1: Asymptotic behavior of optimizing Eq (3) for the minimum-rank sensor network localization problem in the presence of Gaussian noise. We plot the values of rank (blue) and dist (red) against the number of iterations, as well as, the sensor locations at different stages of the optimization (1, 2, 3, 4, 5).



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

4.2 Convergence Behavior and Examples

First of all, we verify the convergence property of our proposed PADM algorithm by considering the d = 3 sensor network localization problem. We record rank and dist values for PADM at every iteration k and plot these results in Figure 1. We observe that both the rank and dist values decrease monotonically, and we attribute this to the monotone property of the dual variable π in Algorithm 1. Moreover, the rank and dist values stabilize after the 5^{th} iteration, which means that our algorithm has converged. The decrease of the values is negligible after this iteration. This implies that a looser stopping criterion can be used without sacrificing much localization quality. Second, we show two localization examples on d = 2 and d = 3 data to demonstrate the effectiveness of PADM. As can be seen in Figure 2 and Figure 3, LPAM improves upon the convex/non-convex methods, while our PADM achieves the lowest rank and dist values in the experiments.

4.3 Varying the Parameter u, s and r

We now evaluate the performance of all the methods with varying number of sensor u, noise levels s and sampling ratio r. We report the recovered results in Figure 4, Figure 5 and Figure 6, respectively. We make the following observations. (i) For the convex methods TAM and FM, TAM often achieves a lower rank solution and gives better performance. (ii) LDHM generally outperforms the convex methods TAM and FM because it can often refine the solution of the trace relaxation method when using appropriate initialization. However, this method is still unstable in the varying sampling ratio test cases. (iii) For all our experiments, PDA fails to localize the sensors and generates much worse results than the other methods. (iv) For all the methods, the dist value tends to increase (decrease) as the noise level (sampling ratio) increases. Our proposed PADM generally and relatively gives better performance than all the remaining methods, i.e. it often achieves lower rank and dist values.



Figure 4: Performance comparison with varying number of sensors u in the presence of Gaussian noise.





Figure 6: Performance comparison with varying sampling ratio r in the presence of Gaussian noise.

5 Conclusions

In this paper, we propose an MPEC approach for solving the semi-definite rank minimization problem. Although the optimization problem is non-convex, we design an effective proximal ADM algorithm to solve the equivalent MPEC problem. We also prove that our method is convergent to a first-order KKT point. We apply our method to the problem of sensor network localization, where extensive experimental results demonstrate that our method generally achieves better solution quality than existing methods. This is due to the fact that the original rank problem is not approximated.

Acknowledgments

Research reported in this publication was supported by competitive research funding from King Abdullah University of Science and Technology (KAUST). Yuan is also supported by Natural Science Foundation of China (61402182), Postdoctoral Science Foundation of China (2015M572317), and Fundamental Research Funds for Central Universities. A special thanks is also extended to Prof. Shaohua Pan for her helpful discussion on the MPEC techniques.

References

Agarwal, S.; Wills, J.; Cayton, L.; Lanckriet, G. R. G.; Kriegman, D. J.; and Belongie, S. 2007. Generalized non-metric multidimensional scaling. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 11–18. Attouch, H., and Bolte, J. 2009. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming* 116(1-2):5–16.

Bi, S.; Liu, X.; and Pan, S. 2014. Exact penalty decomposition method for zero-norm minimization based on MPEC formulation. *SIAM Journal on Scientific Computing (SISC)* 36(4).

Biswas, P., and Ye, Y. 2004. Semidefinite programming for ad hoc wireless sensor network localization. In *Information Processing in Sensor Networks (IPSN)*, 46–54.

Biswas, P.; Lian, T.-C.; Wang, T.-C.; and Ye, Y. 2006a. Semidefinite programming based algorithms for sensor network localization. *ACM Transactions on Sensor Networks (TOSN)* 2(2):188– 220.

Biswas, P.; Liang, T.; Toh, K.; Ye, Y.; and Wang, T. 2006b. Semidefinite programming approaches for sensor network localization with noisy distance measurements. *IEEE Transactions on Automation Science and Engineering* 3(4):360–371.

Bolte, J.; Sabach, S.; and Teboulle, M. 2014. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming* 146(1-2):459–494.

Burer, S., and Monteiro, R. D. C. 2003. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming* 95(2):329–357.

Burer, S., and Monteiro, R. D. C. 2005. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming* 103(3):427–444. Candès, E. J., and Recht, B. 2009. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* (*FoCM*) 9(6):717–772.

Candès, E. J., and Tao, T. 2010. The power of convex relaxation: near-optimal matrix completion. *IEEE Transactions on Information Theory* 56(5):2053–2080.

Chinmay Hegde, Aswin C. Sankaranarayanan, W. Y. R. G. B. 2015. Numax: A convex approach for learning near-isometric linear embeddings. *IEEE Transactions on Signal Processing (to appear)*.

Cong, Y.; Liu, J.; Yuan, J.; and Luo, J. 2013. Self-supervised online metric learning with low rank constraint for scene categorization. *IEEE Transactions on Image Processing (TIP)* 22(8):3179–3191.

Dattorro, J. 2011. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing USA.

Deng, Y.; Dai, Q.; Liu, R.; Zhang, Z.; and Hu, S. 2013. Low-rank structure learning via nonconvex heuristic recovery. *TNNLS* 24(3):383–396.

Fazel, M.; Hindi, H.; and Boyd, S. P. 2003. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the American Control Conference*, volume 3, 2156–2162.

Fazel, M. 2002. Matrix rank minimization with applications. *PhD thesis, Stanford University.*

He, B., and Yuan, X. 2012. On the $\mathcal{O}(1/n)$ convergence rate of the douglas-rachford alternating direction method. *SIAM Journal on Numerical Analysis (SINUM)* 50(2):700–709.

Hu, Y.; Zhang, D.; Ye, J.; Li, X.; and He, X. 2013. Fast and accurate matrix completion via truncated nuclear norm regularization. *TPAMI* 35(9):2117–2130.

Huang, Y., and Palomar, D. P. 2010. Rank-constrained separable semidefinite programming with applications to optimal beamforming. *IEEE Transactions on Signal Processing* 58(2):664–678.

Ji, S.; Sze, K.; Zhou, Z.; So, A. M.; and Ye, Y. 2013. Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In *IEEE International Conference* on Computer Communications (INFOCOM), 2499–2507.

Krislock, N., and Wolkowicz, H. 2010. Explicit sensor network localization using semidefinite representations and facial reductions. *SIOPT* 20(5):2679–2708.

Law, M. T.; Thome, N.; and Cord, M. 2014. Fantope regularization in metric learning. In *CVPR*, 1051–1058.

Li, Q., and Qi, H. 2011. A sequential semismooth newton method for the nearest low-rank correlation matrix problem. *SIOP*-T 21(4):1641–1666.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 612–620.

Liu, W.; Mu, C.; Ji, R.; Ma, S.; Smith, J. R.; and Chang, S. 2015. Low-rank similarity metric learning in high dimensions. In *AAAI*, 2792–2799.

Louca, R.; Seiler, P.; and Bitar, E. 2013. A rank minimization algorithm to enhance semidefinite relaxations of optimal power flow. In *Annual Allerton Conference on Communication, Control, and Computing*, 1010–1020.

Lu, Z., and Zhang, Y. 2013. Sparse approximation via penalty decomposition methods. *SIOPT* 23(4):2448–2478.

Lu, C.; Tang, J.; Yan, S.; and Lin, Z. 2014. Generalized nonconvex nonsmooth low-rank minimization. In *CVPR*, 4130–4137.

Lu, C.; Zhu, C.; Xu, C.; Yan, S.; and Lin, Z. 2015. Generalized singular value thresholding. In *AAAI*, 1805–1811.

Lu, Z. 2014. Iterative reweighted minimization methods for ℓ_p regularized unconstrained nonlinear programming. *Mathematical Programming* 147(1-2):277–307.

Luo, Z.-Q.; Pang, J.-S.; and Ralph, D. 1996. *Mathematical programs with equilibrium constraints*. Cambridge University Press.

Meka, R.; Jain, P.; Caramanis, C.; and Dhillon, I. S. 2008. Rank minimization via online learning. In *ICML*, 656–663.

Miao, W.; Pan, S.; and Sun, D. 2015. A rank-corrected procedure for matrix completion with fixed basis coefficients. *Mathematical Programming, to appear.*

Moscato, P.; Norman, M. G.; and Pataki, G. 1998. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research (MOR)* 23(2):339–358.

Nie, F.; Huang, H.; and Ding, C. H. Q. 2012. Low-rank matrix recovery via efficient schatten p-norm minimization. In *AAAI*.

Pong, T. K., and Tseng, P. 2011. (robust) edge-based semidefinite programming relaxation of sensor network localization. *Mathematical Programming* 130(2):321–358.

Rallapalli, S.; Qiu, L.; Zhang, Y.; and Chen, Y. 2010. Exploiting temporal stability and low-rank structure for localization in mobile networks. In *International Conference on Mobile Computing and Networking (MOBICOM)*, 161–172.

Recht, B.; Fazel, M.; and Parrilo, P. A. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* 52(3):471–501.

Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326.

Saunderson, J.; Chandrasekaran, V.; Parrilo, P. A.; and Willsky, A. S. 2012. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal on Matrix Analysis and Applications (SIMAX)* 33(4):1395–1416.

So, A. M.-C., and Ye, Y. 2007. Theory of semidefinite programming for sensor network localization. *Mathematical Programming* 109(2-3):367–384.

Tseng, P. 2007. Second-order cone programming relaxation of sensor network localization. *SIOPT* 18(1):156–185.

Wang, Z.; Zheng, S.; Ye, Y.; and Boyd, S. 2008. Further relaxations of the semidefinite programming approach to sensor network localization. *SIOPT* 19(2):655–673.

Yu, H., and Lau, V. K. N. 2011. Rank-constrained schur-convex optimization with multiple trace/log-det constraints. *IEEE Transactions on Signal Processing* 59(1):304–314.

Yuan, G., and Ghanem, B. 2015. $\ell_0 tv$: A new method for image restoration in the presence of impulse noise. In *CVPR*, 5369–5377.

Zhang, Y., and Lu, Z. 2011. Penalty decomposition methods for rank minimization. In *NIPS*, 46–54.

Zhang, L.; Liu, L.; Gotsman, C.; and Gortler, S. J. 2010. An asrigid-as-possible approach to sensor network localization. *ACM Transactions on Sensor Networks (TOSN)* 6(4):35:1–35:21.

Zhang, T.; Ghanem, B.; Liu, S.; and Ahuja, N. 2012. Robust visual tracking via multi-task sparse learning. In *CVPR*, 2042–2049.

Zhang, T.; Ghanem, B.; Liu, S.; Xu, C.; and Ahuja, N. 2013. Low-rank sparse coding for image classification. In *ICCV*, 281–288.

Zhao, Y.-B. 2012. An approximation theory of matrix rank minimization and its application to quadratic equations. *Linear Algebra and its Applications* 437(1):77 – 93.

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 $\mathbf{R} \in \mathbb{R}^{m \times n}$ be a given matrix. Then $rank(\mathbf{R}) \leq r$ if and only if there exist matrices $\mathbf{S} = \mathbf{S}^T \in \mathbb{R}^{m \times m}$ and $\mathbf{T} = \mathbf{T}^T \in \mathbb{R}^{n \times n}$ such that

$$\begin{bmatrix} \mathbf{S} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{T} \end{bmatrix} \succeq 0, \ rank(\mathbf{S}) + rank(\mathbf{T}) \le 2r.$$
(1)

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:

$$\min_{\mathbf{R}} g(\mathbf{R}) + \lambda \, rank(\mathbf{R}) \tag{2}$$

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

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

Since $\begin{bmatrix} \mathbf{S} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{T} \end{bmatrix}$ is positive semidefinite, by the Schur complement condition ¹, it holds that $\mathbf{S} \succeq 0$ and $\mathbf{T} \succeq 0$. Then the variable matrix $\begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{T} \end{bmatrix}$ is also positive semidefinite. Using the MPEC reformulation, we have the following equivalent optimization problem:

$$\min_{\mathbf{R},\mathbf{S},\mathbf{T},\mathbf{V}} g(\mathbf{R}) + \frac{1}{2} tr(\mathbf{V}),$$

s.t. $\langle \mathbf{I} - \mathbf{V}, \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{bmatrix} \rangle = 0, \begin{bmatrix} \mathbf{S} & \mathbf{R} \\ \mathbf{R}^T & \mathbf{T} \end{bmatrix} \succeq 0, \mathbf{I} \succeq \mathbf{V} \succeq 0.$

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

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹Schur complement condition implies that: $\mathbf{X} \triangleq \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \succeq 0 \Leftrightarrow \mathbf{A} \succeq 0, \mathbf{S} = \mathbf{C} - \mathbf{B}^T \mathbf{A}^{\dagger} \mathbf{B} \succeq 0$

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

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

 $rank(\mathbf{X}) = \min_{\mathbf{V}} tr(\mathbf{I} - \mathbf{V}), \ s.t. \langle \mathbf{V}, \mathbf{X} \rangle = 0, \ 0 \leq \mathbf{V} \leq I$ (3)

and the unique optimal solution of the minimization problem in Eq (3) is given by $\mathbf{V}^* = \mathbf{U} diag(sign(\boldsymbol{\sigma}))\mathbf{U}^T$, where $\mathbf{X} = \mathbf{U} diag(\boldsymbol{\sigma})\mathbf{U}^T$ is the eigenvalue decomposition of \mathbf{X} , sign is the standard signum function².

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

$$\langle \mathbf{X}, \mathbf{V} \rangle = \| \mathbf{U} diag(\sqrt{\sigma}) diag(\sqrt{\delta}) \mathbf{S}^T \|_F^2 \ge 0$$

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 $\{\mathbf{V} \mid \mathbf{V} = \mathbf{U} diag(\mathbf{v})\mathbf{U}^T, \mathbf{0} \leq \mathbf{v} \leq \mathbf{1}, \mathbf{v} \in \mathbb{R}^r\}$. Using the fact that $tr(\mathbf{V}) = \langle \mathbf{v}, \mathbf{1} \rangle$, $tr(\mathbf{X}) = \langle \boldsymbol{\sigma}, \mathbf{1} \rangle$ and $\langle \mathbf{V}, \mathbf{X} \rangle = \langle \mathbf{U} diag(\mathbf{v})\mathbf{U}^T, \mathbf{U} diag(\boldsymbol{\sigma})\mathbf{U}^T \rangle = \langle \mathbf{v}, \boldsymbol{\sigma} \rangle$, Eq (3) reduces to the vector case:

$$rank(\mathbf{X}) = \min_{\mathbf{v}} \langle \mathbf{1}, \mathbf{1} - \mathbf{v} \rangle, \ s.t. \ \langle \mathbf{v}, \boldsymbol{\sigma} \rangle = 0, \ \mathbf{0} \le \mathbf{v} \le \mathbf{1} \quad (4)$$

Since $\mathbf{v} \ge \mathbf{0}$ and $\boldsymbol{\sigma} \ge 0$, Eq (4) is equivalent to Eq (5):

$$rank(\mathbf{X}) = \min_{\mathbf{v}} \langle \mathbf{1}, \mathbf{1} - \mathbf{v} \rangle, \ s.t. \ \mathbf{v} \odot \boldsymbol{\sigma} = \mathbf{0}, \ \mathbf{0} \le \mathbf{v} \le \mathbf{1}$$
(5)

where \odot denotes the Hadamard product (also known as the entrywise product). Note that for all $i \in [r]$ when $\sigma_i = 0$, $\mathbf{v}_i = 1$ will be achieved by minimization, when $\sigma_i \neq 0$, $\mathbf{v}_i =$ 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, $\mathbf{v}^* = \mathbf{1} - sign(\sigma)$. Finally, we have $\mathbf{V}^* =$ $\mathbf{U}diag(\mathbf{1} - sign(\sigma))\mathbf{U}^T$. We thus complete the proof of this lemma.

$${}^{2} sign(x) = \begin{cases} -1, & x < 0; \\ 0, & x = 0; \\ 1, & x > 0. \end{cases}$$

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 $W = Udiag(\sigma)U^T$. The optimal solution of the following optimization problem

$$\underset{\mathbf{V}}{\operatorname{arg\,min}} \ \frac{1}{2} \|\mathbf{V} - \mathbf{W}\|_{F}^{2} + I_{\Delta}(\mathbf{V})$$
(6)

can be computed as Udiag(min(1, max(0, σ)))T. Here I_{Δ} is an indicator function of the convex set $\Delta \triangleq \{\mathbf{V} \mid 0 \preceq \mathbf{V} \leq I\}$ with $I_{\Delta}(\mathbf{V}) \triangleq \begin{cases} 0, & \mathbf{v} \in \Delta \\ \infty, & otherwise. \end{cases}$

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

$$\begin{aligned} &\frac{1}{2} \|\mathbf{V} - \mathbf{W}\|_F^2 + I_{\Delta}(\mathbf{V}) \\ &= &\frac{1}{2} (\|\mathbf{z}\|^2 + \|\boldsymbol{\sigma}\|^2 - 2\langle \mathbf{V}, \mathbf{W} \rangle) + I_{\Theta}(\mathbf{z}) \\ &\leq &\frac{1}{2} (\|\mathbf{z}\|^2 + \|\boldsymbol{\sigma}\|^2 - 2\langle \mathbf{z}, \boldsymbol{\sigma} \rangle) + I_{\Theta}(\mathbf{z}) \\ &= &\frac{1}{2} \|\mathbf{z} - \boldsymbol{\sigma}\|^2 + I_{\Theta}(\mathbf{z}) \end{aligned}$$

From the von Veumann's trace inequality, the solution set of (12) must be contained in the set $\{\mathbf{V} \mid \mathbf{V} = \mathbf{U}diag(\mathbf{v}^*)\mathbf{U}^T\}$, where \mathbf{v}^* is given by

$$\mathbf{v}^* = \arg\min_{\mathbf{z}} \ \frac{1}{2} \|\mathbf{z} - \boldsymbol{\sigma}\|^2 + I_{\Theta}(\mathbf{z})$$
(7)

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

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:

$$\mathbf{u} \triangleq \{\mathbf{X}, \mathbf{V}\}, \ \mathbf{s} \triangleq \{\mathbf{X}, \mathbf{V}, \pi\}$$
(8)

$$\Omega \triangleq \{ \mathbf{X} \mid 0 \preceq \mathbf{X} \preceq \kappa I \}, \ \Delta \triangleq \{ \mathbf{V} \mid 0 \preceq \mathbf{V} \preceq I \}$$
(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 { $\mathbf{X}^*, \mathbf{V}^*, \pi^*$ }:

$$\begin{array}{rcl} \mathbf{0} & \in & \partial I_{\Omega}(\mathbf{X}^{*}) + \mathcal{A}^{T} \partial g(\mathcal{A}(\mathbf{X}^{*}) - \mathbf{b}) + \pi \mathbf{V}^{*} \\ \mathbf{0} & \in & \partial I_{\Delta}(\mathbf{V}^{*}) - \lambda \mathbf{I} + \pi \mathbf{X}^{*} \\ \mathbf{0} & = & \langle \mathbf{V}^{*}, \mathbf{X}^{*} \rangle \end{array}$$

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 π^k is bounded for all k, then there exists a constant $\varpi > 0$ such that the following inequality holds:

$$\|\partial \mathcal{L}(\mathbf{s}^{k+1})\| \le \varpi \|\mathbf{s}^{k+1} - \mathbf{s}^k\|$$
(10)

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

$$\mathbf{0} \in \mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k}) + \mathcal{A}^{T} \partial g(\mathcal{A}(\mathbf{X}^{k+1}) - \mathbf{b}) + \pi^{k} \mathbf{V}^{k} + \alpha \langle \mathbf{V}^{k}, \mathbf{X}^{k+1} \rangle \mathbf{V}^{k} + \partial I_{\Omega}(\mathbf{X}^{k+1})$$
(11)
$$\mathbf{0} \in \mathbf{E}(\mathbf{V}^{k+1} - \mathbf{V}^{k}) - \lambda \mathbf{I} + \pi^{k} \mathbf{X}^{k+1} + \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \mathbf{X}^{k+1} + \partial I_{\Delta}(\mathbf{V}^{k+1})$$
(12)

By the definition of \mathcal{L} we have that

$$\begin{aligned} \partial \mathcal{L}_{\mathbf{X}}(\mathbf{s}^{k+1}) &= \mathbf{A}^{T} \partial g(\mathbf{A}\mathbf{X}^{k+1} - \mathbf{b}) + \pi^{k+1}\mathbf{V}^{k+1} + \\ & \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \mathbf{V}^{k+1} + \partial I_{\Omega}(\mathbf{X}^{k+1}) \\ &= -(\pi^{k} + \alpha \langle \mathbf{V}^{k}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k} + (\pi^{k+1} + \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k+1} \\ & -\mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k}) \\ &= -(\pi^{k} + \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle + \alpha \langle \mathbf{V}^{k} - \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k} \\ & + (\pi^{k+1} + \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k+1} - \mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k}) \\ &= -(\pi^{k+1} + \alpha \langle \mathbf{V}^{k} - \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k} + (2\pi^{k+1} - \pi^{k})\mathbf{V}^{k+1} \\ & -\mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k}) \\ &= -(\alpha \langle \mathbf{V}^{k} - \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle)\mathbf{V}^{k} + (\pi^{k+1} - \pi^{k})\mathbf{V}^{k+1} \\ & + \pi^{k+1}(\mathbf{V}^{k+1} - \mathbf{V}^{k}) - \mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k}) \end{aligned}$$

The first step uses the definition of $\mathcal{L}_{\mathbf{X}}(\mathbf{s}^{k+1})$, the second step uses Eq (11), the third step uses $\mathbf{V}^k + \mathbf{V}^{k+1} - \mathbf{V}^{k+1} = \mathbf{V}^k$, the fourth step uses the multiplier update rule for π .

Assume that π^{k+1} is bounded by a constant ρ that $\pi^{k+1} \leq \rho$. We have:

$$\begin{aligned} \|\partial \mathcal{L}_{\mathbf{X}}(\mathbf{s}^{k+1})\|_{F} \\ &\leq \|\alpha \langle \mathbf{V}^{k} - \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \mathbf{V}^{k}\| + \|(\pi^{k+1} - \pi^{k})\mathbf{V}^{k+1}\| \\ &+ \|\pi^{k+1}(\mathbf{V}^{k+1} - \mathbf{V}^{k})\| + \|\mathbf{D}(\mathbf{X}^{k+1} - \mathbf{X}^{k})\| \\ &\leq 2n\kappa\alpha \|\mathbf{V}^{k} - \mathbf{V}^{k+1}\|_{F} + 2\|\pi^{k+1} - \pi^{k}\| \\ &\rho \|\mathbf{V}^{k} - \mathbf{V}^{k+1}\|_{F} + \|\mathbf{D}\| \cdot \|\mathbf{X}^{k+1} - \mathbf{X}^{k}\|_{F} \end{aligned}$$
(13)

Similarly, we have

$$\partial \mathcal{L}_{\mathbf{V}}(\mathbf{s}^{k+1}) = \partial I_{\Delta}(\mathbf{V}^{k+1}) - \lambda \mathbf{I} + \pi^{k+1} \mathbf{X}^{k+1} \\ + \alpha \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \mathbf{X}^{k+1} \\ = (\pi^{k+1} - \pi^k) \mathbf{X}^{k+1} - \mathbf{E}(\mathbf{V}^{k+1} - \mathbf{V}^k) \\ \partial \mathcal{L}_{\pi}(\mathbf{s}^{k+1}) = \langle \mathbf{I} - \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle = \frac{1}{\alpha} (\pi^{k+1} - \pi^k)$$

Then we derive the following inequalities:

$$\|\partial \mathcal{L}_{\mathbf{V}}(\mathbf{s}^{k+1})\|_{F} \le \kappa \|\pi^{k} - \pi^{k+1}\| + \|\mathbf{E}\| \cdot \|\mathbf{V}^{k+1} - \mathbf{V}^{k}\|_{F}$$
(14)

$$\|\partial \mathcal{L}_{\pi}(\mathbf{s}^{k+1})\| \le \frac{1}{\alpha} \|\pi^{k+1} - \pi^k\|$$
(15)

Combining Eqs (13-15), we conclude that there exists $\varpi > 0$ such that the following inequality holds

$$\|\partial \mathcal{L}(\mathbf{s})\|_F \le \varpi \|\mathbf{s} - \mathbf{s}^k\|.$$

Thus, we complete the proof of this lemma.

The following lemma is useful in our convergence analysis. **Lemma 5.** Assume that π^k is bounded for all k, then we have the following inequality:

$$\sum_{k=0}^{\infty} \|\mathbf{s}^k - \mathbf{s}^{k+1}\|^2 < +\infty$$
 (16)

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

Proof. Due to the initialization and the update rule of π , we conclude that π^k is nonnegative and monotone non-decreasing. Moreover, as $k \to \infty$, we have: $\langle \mathbf{X}^{k+1}, \mathbf{V}^{k+1} \rangle = 0$. This can be proved by contradiction. Suppose that $\langle \mathbf{X}^{k+1}, \mathbf{V}^{k+1} \rangle \neq 0$, then $\pi^k = +\infty$ as $k \to \infty$. This contradicts our assumption that π^k is bounded. Therefore, we conclude that as $k \to +\infty$ it holds that

$$\langle \mathbf{X}^{k+1}, \mathbf{V}^{k+1} \rangle = 0,$$

$$\sum_{i=1}^{k} \|\pi^{i+1} - \pi^{i}\| < +\infty$$

$$\sum_{i=1}^{k} \|\pi^{i+1} - \pi^{i}\|^{2} < +\infty$$
(17)

On the other hand, we naturally derive the following inequalities:

$$\mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}^{k+1}; \pi^{k+1}) = \mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}^{k+1}; \pi^{k}) + \langle \pi^{k+1} - \pi^{k}, \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \rangle
= \mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}^{k+1}; \pi^{k}) + \frac{1}{\alpha} \| \pi^{k+1} - \pi^{k} \|^{2}
\leq \mathcal{L}(\mathbf{X}^{k}, \mathbf{V}^{k+1}; \pi^{k}) - \frac{\mu}{2} \| \mathbf{X}^{k+1} - \mathbf{X}^{k} \|^{2}
+ \frac{1}{\alpha} \| \pi^{k+1} - \pi^{k} \|^{2}
\leq \mathcal{L}(\mathbf{X}^{k}, \mathbf{V}^{k}; \pi^{k}) - \frac{\mu}{2} \| \mathbf{X}^{k+1} - \mathbf{X}^{k} \|^{2}
- \frac{\mu}{2} \| \mathbf{V}^{k+1} - \mathbf{V}^{k} \|^{2} + \frac{1}{\alpha} \| \pi^{k+1} - \pi^{k} \|^{2}$$
(18)

The first step uses the definition of \mathcal{L} ; the second step uses update rule of the Lagrangian multiplier π ; the third and fourth step use the μ -strongly convexity of \mathcal{L} with respect to **X** and **v**, respectively. We define C = $\mathcal{L}(\mathbf{X}^0, \mathbf{V}^0; \pi^0) - \mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}^{k+1}; \pi^{k+1}) + \frac{1}{\alpha} \sum_{i=1}^k ||\pi^{i+1} - \pi^i||^2$. Clearly, by the boundedness of \mathbf{X}^k , \mathbf{V}^k and π^k , both $\mathcal{L}(\mathbf{X}^{k+1}, \mathbf{V}^{k+1}; \pi^{k+1})$ and C are bounded. Summing Eq (18) over i = 1, 2..., k, we have:

$$\frac{\mu}{2} \sum_{i=1}^{k} \|\mathbf{u}^{i+1} - \mathbf{u}^{i}\|^{2} \leq C$$
 (19)

Therefore, combining Eq(17) and Eq (19), we have $\sum_{k=1}^{+\infty} \|\mathbf{s}^{k+1} - \mathbf{s}^k\|^2 < +\infty$; in particular $\|\mathbf{s}^{k+1} - \mathbf{s}^k\| \to 0$. By Eq (10), we have that:

$$\|\partial \mathcal{L}(\mathbf{s}^{k+1})\| \le \varpi \|\mathbf{s}^{k+1} - \mathbf{s}^k\| \to 0$$
(20)

which implies that any cluster point of s^k is a stationary point of \mathcal{L} . 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 $\sum_{k=0}^{\infty} ||\mathbf{s}^k - \mathbf{s}^{k+1}||^2$ does not imply that the sequence \mathbf{s}^k is convergent³. 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 $s \in dom\mathcal{L}$, there exists $\theta \in [0, 1)$, $\eta \in (0, +\infty]$ a neighborhood S of s and a concave and continuous function $\varphi(t) = c \cdot t^{1-\theta}, t \in [0, \eta)$ such that for all $\bar{s} \in S$ and satisfies $\mathcal{L}(\bar{s}) \in (\mathcal{L}(s), \mathcal{L}(s) + \eta)$, the following inequality holds:

$$dist(0,\partial \mathcal{L}(\bar{\mathbf{s}}))\varphi^k(\mathcal{L}(\mathbf{s}) - \mathcal{L}(\bar{\mathbf{s}})) \ge 1, \ \forall \mathbf{s}$$

where $dist(0, \partial \mathcal{L}(\bar{\mathbf{s}})) = \min\{||u^*|| : u^* \in \partial \mathcal{L}(\bar{\mathbf{s}})\}.$

³One typical counter-example is $\mathbf{s}^{k} = \sum_{i=1}^{k} \frac{1}{i}$. Clearly, $\sum_{k=0}^{\infty} \|\mathbf{s}^{k} - \mathbf{s}^{k+1}\|^{2} = \sum_{k=1}^{\infty} (\frac{1}{k})^{2}$ is bounded by $\frac{\pi^{2}}{6}$; however, \mathbf{s}^{k} is divergent since $s^{k} = \ln(k) + C_{e}$, where C_{e} is the well-known Euler's constant.

⁴Note 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 π^k is bounded for all k. Then we have the following inequality:

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

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 $R^k = \varphi(\mathcal{L}(\mathbf{s}^k) - \mathcal{L}(\mathbf{s}^*)) - \varphi(\mathcal{L}(\mathbf{s}^{k+1}) - \mathcal{L}(\mathbf{s}^*))$. We naturally derive the following inequalities:

$$\begin{split} & \frac{\mu}{2} \| \mathbf{u}^{k+1} - \mathbf{u}^k \|^2 - \frac{1}{\alpha} \| \pi^{k+1} - \pi^k \|^2 \\ & \leq \quad \mathcal{L}(\mathbf{s}^k) - \mathcal{L}(\mathbf{s}^{k+1}) \\ & = \quad (\mathcal{L}(\mathbf{s}^k) - \mathcal{L}(\mathbf{s}^*)) - (\mathcal{L}(\mathbf{s}^{k+1}) - \mathcal{L}(\mathbf{s}^*)) \\ & \leq \quad \frac{R^k}{\varphi^k (\mathcal{L}(\mathbf{s}^k) - \mathcal{L}(\mathbf{s}^*))} \\ & \leq \quad R^k dist(0, \partial \mathcal{L}(\mathbf{s}^k)) \\ & \leq \quad R^k \varpi (\| \mathbf{X}^k - \mathbf{X}^{k-1} \| + \| \mathbf{V}^k - \mathbf{V}^{k-1} \| + \| \pi^k - \pi^{k-1} \|) \\ & \leq \quad R^k \varpi (\sqrt{2} \| \mathbf{u}^k - \mathbf{u}^{k-1} \| + \| \pi^k - \pi^{k-1} \|) \\ & = \quad R^k \varpi \sqrt{\frac{16}{\mu}} (\sqrt{\frac{\mu}{8}} \| \mathbf{u}^k - \mathbf{u}^{k-1} \| + \sqrt{\frac{\mu}{16}} \| \pi^k - \pi^{k-1} \|) \\ & \leq \quad \frac{4(R^k \varpi)^2}{\mu} + (\sqrt{\frac{\mu}{8}} \| \mathbf{u}^k - \mathbf{u}^{k-1} \| + \sqrt{\frac{\mu}{16}} \| \pi^k - \pi^{k-1} \|)^2 \end{split}$$

The first step uses Eq (18); the third step uses the concavity of φ such that $\varphi(a) - \varphi(b) \geq \varphi^k(a)(a-b)$ for all $a, b \in \mathbb{R}$; the fourth step uses the KL property such that $dist(0, \partial \mathcal{L}(\mathbf{s}^k))\varphi^k(\mathcal{L}(\mathbf{s}^k) - \mathcal{L}(\mathbf{s}^*)) \geq 1$ as in Proposition 1; the fifth step uses Eq (10); the sixth step uses the inequality that $||x; y|| \leq ||x|| + ||y|| \leq \sqrt{2}||x; y||$, where ';' in [·] denotes the row-wise partitioning indicator as in Matlab; the last step uses that fact that $2ab \leq a^2 + b^2$ for all $a, b \in \mathbb{R}$. As $k \to \infty$, we have:

$$\frac{\mu}{2} \|\mathbf{u}^{k+1} - \mathbf{u}^{k}\|^{2} \le \frac{4(R^{k}\kappa)^{2}}{\mu} + \left(\sqrt{\frac{\mu}{8}} \|\mathbf{u}^{k} - \mathbf{u}^{k-1}\|\right)^{2}$$

Take the squared root of both side and use the inequality that $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for all $a, b \in \mathbb{R}_+$, we have

$$\begin{split} & \sqrt{\frac{\mu}{8}} \| \mathbf{u}^{k+1} - \mathbf{u}^k \| \\ & \leq \quad \frac{R^k \kappa}{\sqrt{\mu}} + \sqrt{\frac{\mu}{32}} \| \mathbf{u}^k - \mathbf{u}^{k-1} \| \\ & \leq \quad \frac{R^k \kappa}{\sqrt{\mu}} + \sqrt{\frac{\mu}{8}} (\| \mathbf{u}^k - \mathbf{u}^{k-1} \| - \| \mathbf{u}^{k+1} - \mathbf{u}^k \|) \end{split}$$

Summing the inequality above over i = 1, 2..., k, we have:

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

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

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

Moreover, by Eq (10) in Lemma 4 we have $\partial \mathcal{L}(\mathbf{s}^{k+1}) = 0$. In other words, we have the following results:

$$\begin{aligned} \mathbf{0} &\in \partial I_{\Delta}(\mathbf{X}^{k+1}) + \mathcal{A}^{T} \partial g(\mathcal{A}(\mathbf{X}^{k+1}) - \mathbf{b}) + \pi^{k+1} \mathbf{V}^{k+1} \\ \mathbf{0} &\in \partial I_{\Delta}(\mathbf{V}^{k+1}) + \lambda \mathbf{I} + \pi^{k+1} \mathbf{X}^{k+1} \\ \mathbf{0} &= \langle \mathbf{V}^{k+1}, \mathbf{X}^{k+1} \rangle \end{aligned}$$

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

(S.0) Initialize $\mathbf{X}^0, \mathbf{Y}^0 = \mathbf{0} \in \mathbb{R}^{n \times n}, \mathbf{y}^0 = \mathbf{0} \in \mathbb{R}^m, \mathbf{z}^0 = \mathbf{0} \in \mathbb{R}^m, \mathbf{Z}^0 = \mathbf{0} \in \mathbb{R}^{n \times n}$. Set t = 0. (S.1) Solve the following (\mathbf{y}, \mathbf{X}) -subproblem:

$$(\mathbf{y}, \mathbf{x})$$
-subproblem.

$$\mathbf{X}^{t+1} = \underset{\mathbf{X}}{\arg\min} \mathcal{J}(\mathbf{X}, \mathbf{y}^{t}, \mathbf{Y}^{t}; \mathbf{Z}^{t}, \mathbf{z}^{t})$$
(23)

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

$$(\mathbf{y}^{t+1}, \mathbf{Y}^{t+1}) = \operatorname*{arg\,min}_{\mathbf{y}, 0 \leq \mathbf{Y} \leq \kappa I} \mathcal{J}(\mathbf{X}^{t+1}, \mathbf{y}, \mathbf{Y}; \mathbf{Z}^{t}, \mathbf{z}^{t}) \quad (24)$$

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

$$\mathbf{Z}^{t+1} = \mathbf{Z}^t + \beta(\mathbf{X}^{t+1} - \mathbf{Y}^{t+1})$$
$$z^{t+1} = z^t + \gamma(\mathcal{A}(\mathbf{X}^{k+1}) - \mathbf{b} - \mathbf{y}^{k+1})$$

(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:

$$\min_{0 \leq \mathbf{X} \leq \kappa \mathbf{I}} g\left(\mathcal{A}(\mathbf{X}) - \mathbf{b} \right) + \frac{\alpha}{2} \| \mathcal{B}(\mathbf{X}) \|_F^2 + \frac{\mu}{2} \| \mathbf{X} \|_F^2 + \langle \mathbf{X}, \mathbf{C} \rangle,$$
(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 $\mathbf{y} \in \mathbb{R}^m$ and $\mathbf{Y} \in \mathbb{R}^{n \times n}$ to reformulate Eq (25) as:

$$\min_{\mathbf{y},\mathbf{X},\mathbf{Y}} g(\mathbf{y}) + \frac{\alpha}{2} \|\mathcal{B}(\mathbf{X})\|_F^2 + \frac{\mu}{2} \|\mathbf{X}\|_F^2 + \langle \mathbf{X}, \mathbf{C} \rangle,$$

s.t. $\mathcal{A}(\mathbf{X}) - \mathbf{b} = \mathbf{y}, \mathbf{X} = \mathbf{Y}, 0 \leq \mathbf{Y} \leq \kappa I.$ (26) Let $\mathcal{J}_{\beta,\gamma} : \mathbb{R}^m \times \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \times \mathbb{R}^m \times \mathbb{R}^{n \times n} \to \mathbb{R}$ be the augmented Lagrangian function in Eq(26)

$$\begin{aligned} \mathcal{J}(\mathbf{y}, \mathbf{X}, \mathbf{Y}; \mathbf{z}, \mathbf{Z}) &= g\left(\mathbf{y}\right) + \frac{\alpha}{2} \|\mathcal{B}(\mathbf{X})\|_{F}^{2} + \frac{\mu}{2} \|\mathbf{X}\|_{F}^{2} \\ &+ \langle \mathbf{X}, \mathbf{C} \rangle + \langle \mathbf{z}, \mathcal{A}(\mathbf{X}) - \mathbf{b} - \mathbf{y} \rangle + \frac{\gamma}{2} \|\mathcal{A}(\mathbf{X}) - \mathbf{b} - \mathbf{y}\|_{F}^{2} \\ &+ \langle \mathbf{Z}, \mathbf{X} - \mathbf{Y} \rangle + \frac{\beta}{2} \|\mathbf{X} - \mathbf{Y}\|_{F}^{2}, \ s.t. \ 0 \leq \mathbf{Y} \leq \kappa I \end{aligned}$$

z and Z are the Lagrange multipliers associated to the constraints $\mathcal{A}(\mathbf{Y}) - \mathbf{b} - \mathbf{y} = 0$ and $\mathbf{X} - \mathbf{Y} = 0$, 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:

 $\alpha \mathcal{B}^* \mathcal{B}(\mathbf{X}^{t+1}) + \gamma \mathcal{A}^* \mathcal{A}(\mathbf{X}^{t+1}) + (\mu + \beta) \mathbf{X}^{t+1} = \mathbf{E}$ where $\mathbf{E} = \beta \mathbf{Y} + \gamma \mathcal{A}^*(\mathbf{b} + \mathbf{y}) - \mathbf{C} - \mathcal{A}^* z - \mathbf{Z}.$ Solving this linear system gives:

$$\mathbf{X}^{t+1} = \left(\alpha \mathcal{B}^* \mathcal{B} + \gamma \mathcal{A}^* \mathcal{A} + (\mu + \beta) \mathbf{I}\right)^{-1} \mathbf{E}$$
(27)

When the dimension of the solution is high, solving this linear system may dominant 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:

$$\mathbf{y}^{t+1} = \operatorname*{arg\,min}_{\mathbf{y} \in \mathbb{R}^m} g\left(\mathbf{y}\right) + \frac{\gamma}{2} \|\mathbf{q} - \mathbf{y}\|^2,$$

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

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

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

with $\mathbf{S} = \mathbf{X}^{t+1} + \mathbf{Z}^t / \beta$. Assume that \mathbf{S} has the spectral decomposition that $\mathbf{S} = \mathbf{V} diag(\mathbf{s}) \mathbf{V}^T$. A simple computation yields that the solution \mathbf{Y}^{t+1} can be computed in closed form as: $\mathbf{Y}^{t+1} = \mathbf{V} diag(\max(\mathbf{0}, \min(\kappa \mathbf{1}, \mathbf{s}))) \mathbf{V}^T$. 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 $\{\mathbf{y}^t, \mathbf{X}^t, \mathbf{Y}^t, \mathbf{z}^t, \mathbf{Z}^t\}$ 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 1-6 and uniform noise in Figure 7-12, these experimental results strengthen our conclusions drawn in our submission.

References

- [Attouch et al. 2010] Attouch, H.; Bolte, J.; Redont, P.; and Soubeyran, A. 2010. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality. *Mathematics of Operations Research (MOR)* 35(2):438–457. 3
- [Bolte, Sabach, and Teboulle 2014] Bolte, J.; Sabach, S.; and Teboulle, M. 2014. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming* 146(1-2):459–494. 3
- [Fazel, Hindi, and Boyd 2003] Fazel, M.; Hindi, H.; and Boyd, S. P. 2003. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the American Control Conference*, volume 3, 2156–2162. 1
- [He and Yuan 2012] He, B., and Yuan, X. 2012. On the $\mathcal{O}(1/n)$ convergence rate of the douglas-rachford alternating direction method. *SIAM Journal on Numerical Analysis (SINUM)* 50(2):700–709. 2, 4, 5
- [Lin, Liu, and Su 2011] Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 612–620. 4, 5
- [Rockafellar, Wets, and Wets 1998] Rockafellar, R. T.; Wets, R. J.-B.; and Wets, M. 1998. Variational analysis. springe, berlin, heidelberg, new york, 1998. 2
- [Xu and Yin 2013] Xu, Y., and Yin, W. 2013. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences (SIIMS)* 6(3):1758–1789. 3



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) *dist* and *rank* comparisons on 2d 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).



