Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

# 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 MPEC (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:

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

where $\lambda$ and $\kappa$ are strictly positive scalars, $\mathbf{X} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^m$, the linear map $\boldsymbol{\mathcal{A}}(\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is defined as $\boldsymbol{\mathcal{A}}(\mathbf{X}) = [\langle \mathbf{A}^{(1)}, \mathbf{X}\rangle, ..., \langle \mathbf{A}^{(m)}, \mathbf{X}\rangle]^T$, 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 $\text{prox}_g(\mathbf{c}) \triangleq \min_{\mathbf{z}} \; g(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{c}\|_2^2$ can be efficiently computed.

Note that we constrain $\mathbf{X}$ with a ball of radius $\kappa$ 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 $\kappa$ 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 important 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 sub-optimality of the semi-definite optimization for sensor network localization. Existing approximation solutions, such as Schatten's $\ell_p$ norm method (Ji et al. 2013), only give sub-optimal 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 KKT 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 $\boldsymbol{\chi}_{kj}$ between $\mathbf{a_k}$ and $\mathbf{s_j}$ for some $k, j$, and $\boldsymbol{\chi}_{ij}$ between $\mathbf{s_i}$ and $\mathbf{s_j}$ for some $i, j$. Specifically, we model the noisy distance measurements as:

$$\|\mathbf{a_k} - \mathbf{s_j}\|_2^2 = \boldsymbol{\chi}_{kj}^2 + \varepsilon_{kj},\ \|\mathbf{s_i} - \mathbf{s_j}\|_2^2 = \boldsymbol{\chi}_{ij}^2 + \varepsilon_{ij},$$

where each $(k, j) \in \Pi_{as}$ and each $(i, j) \in \Pi_{ss}$ are some selected pairs of the known (noisy) distances $\boldsymbol{\chi}$. We denote the noise variable as $\boldsymbol{\varepsilon} \in \mathbb{R}^{|\Pi|}$, 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:

$$\begin{aligned} \|\mathbf{s_i} - \mathbf{s_j}\|_2^2 &= \mathbf{e_{ij}}^T \mathbf{S}\mathbf{S}^T \mathbf{e_{ij}}, \\ \|\mathbf{a_k} - \mathbf{s_j}\|_2^2 &= \left(\begin{smallmatrix} \mathbf{a_k} \\ \mathbf{e_j} \end{smallmatrix}\right)^T \begin{pmatrix} \mathbf{I}_d & \mathbf{S} \\ \mathbf{S}^T & \mathbf{S}^T\mathbf{S} \end{pmatrix} \left(\begin{smallmatrix} \mathbf{a_k} \\ \mathbf{e_j} \end{smallmatrix}\right), \end{aligned}$$

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:

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

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

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

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 non-convex 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,

Table 1: Semi-definite rank minimization algorithms.

| 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) | $rank(\mathbf{X}) \leq k$, with $\mathbf{X} \approx \mathbf{L}\mathbf{L}^{\mathbf{T}}$, $\mathbf{L} \in \mathbb{R}^{n \times k}$ |
| **(c)** Schatten $\ell_p$ 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) | $rank(\mathbf{X}) \leq k \Leftrightarrow tr(\mathbf{X}) = \|\sigma(\mathbf{X})\|_{\text{top-k}}$ |
| **(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) | $rank(\mathbf{X}) = \min_{0 \preceq \mathbf{V} \preceq \mathbf{I}} tr(\mathbf{I} - \mathbf{V}),\ s.t.\ \langle \mathbf{V}, \mathbf{X} \rangle = 0$ |

SDP relaxation always produces this maximal rank dimensional solution. The classical way to obtain a low dimensional solution is to project 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 $\ell_p$-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 $\mathbf{L}\mathbf{L}^{\mathbf{T}}$. One important feature of this approach is avoiding the need to perform eigenvalue decomposition. **(c)** Schatten $\ell_p$ 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

[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 $\mathbf{X}$-subproblem with $\mathbf{D} \triangleq \mu\mathbf{I}$:

$$\mathbf{X}^{k+1} = \arg\min_{0\preceq\mathbf{X}\preceq\kappa\mathbf{I}} \mathcal{L}(\mathbf{X}, \mathbf{V}^k, \pi^k) + \frac{1}{2}\|\mathbf{X} - \mathbf{X}^k\|_{\mathbf{D}}^2 \quad (4)$$

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

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

(S.3) Update the Lagrange multiplier:

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

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

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.

### 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:*

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

*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. □

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

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

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 $\mathbf{X}$ and $\mathbf{V}$ 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 $\boldsymbol{\sigma}_j(\mathbf{V})$ and $\boldsymbol{\sigma}_j(\mathbf{X})$ 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 Eq (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:

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

where $\pi$ 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 $\mathbf{X}$ in Eq (4) and $\mathbf{V}$ 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 $\pi^k$, we conclude that $\pi^k$ is monotone non-increasing. Moreover, if we initialize $\pi^0 = 0$ in the first iteration, $\pi$ is always non-negative.

**(b) Initialization Strategy**. We initialize both $\mathbf{N}^0$ to $\mathbf{I}$ and $\pi^0$ 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 $\mathbf{X}$-subproblem.

**(c) V-Subproblem**. Variable $\mathbf{V}$ in Eq (5) is updated by solving the following problem:

$$\begin{aligned}\mathbf{V}^{k+1} = \arg\min_{0\preceq\mathbf{V}\preceq\mathbf{I}}\ & -\lambda tr(\mathbf{V}) + \pi\langle\mathbf{V}, \mathbf{X}^{k+1}\rangle \\ & + \tfrac{\alpha}{2}(\langle\mathbf{V}, \mathbf{X}^{k+1}\rangle)^2 + \tfrac{1}{2}\|\mathbf{V} - \mathbf{V}^k\|_{\mathbf{E}}^2\end{aligned} \quad (9)$$

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

$$\mathbf{V}^{k+1} = \arg\min_{0\preceq\mathbf{V}\preceq\mathbf{I}} \frac{L}{2}\|\mathbf{V} - \mathbf{W}\|_F^2 \quad (10)$$

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 $\mathbf{X}$ in Eq (4) is updated by solving the following structured convex optimization problem:

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

where $\boldsymbol{\mathcal{B}}(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is another linear map defined as $\boldsymbol{\mathcal{B}}(\mathbf{X}) \triangleq \langle \mathbf{V}, \mathbf{X} \rangle$, $\mathbf{C} = \pi^k \mathbf{V}^k$. The $\mathbf{X}$-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 non-convex, 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 $\pi^k$ is bounded for all $k$. As $k \rightarrow +\infty$, 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:

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

where $\boldsymbol{\sigma}(\mathbf{X})$ 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 MPEC 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). **(ii)** 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):

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

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 $\ell_p$ 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 $\tilde{\boldsymbol{\chi}} \in \mathbb{R}^o$ from $\Pi$ and inject them with noise by $\tilde{\boldsymbol{\chi}} \leftarrow \tilde{\boldsymbol{\chi}} + s \times \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \in \mathbb{R}^o$ 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]:

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

where $\|\mathbf{x}\|_{0\text{-}\epsilon}$ is the soft $\ell_0$ norm which counts the number of elements whose magnitude is greater than a threshold $\epsilon = 0.01 \cdot \|\mathbf{x}\|$, $\forall \mathbf{x} \in \mathbb{R}^n$. $\bar{\mathbf{S}}$ 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**$, 0.4, ..., 1.0$ |

[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).

[3]Note that we need to retrieve $\mathbf{S}$ from $\mathbf{X}$ (See Eq (2)).

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.

### 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 $\pi$ 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.

## 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.

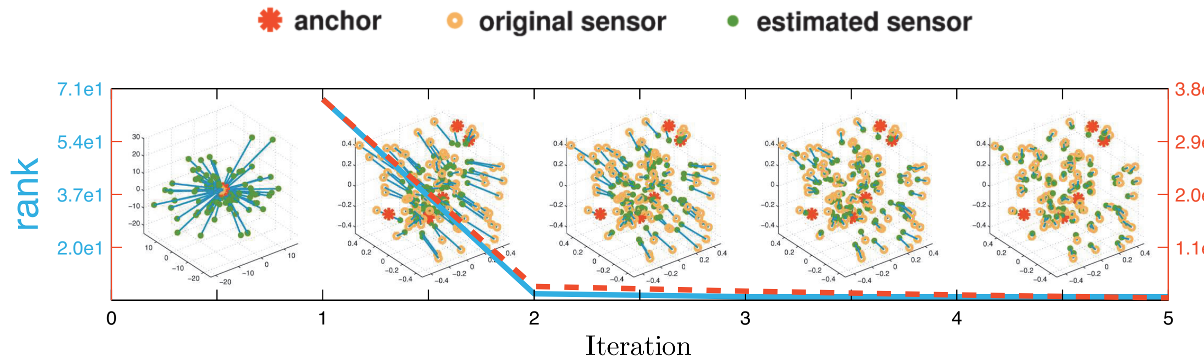
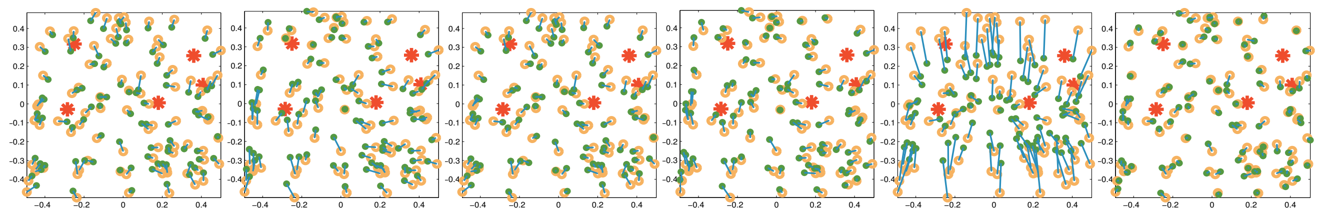
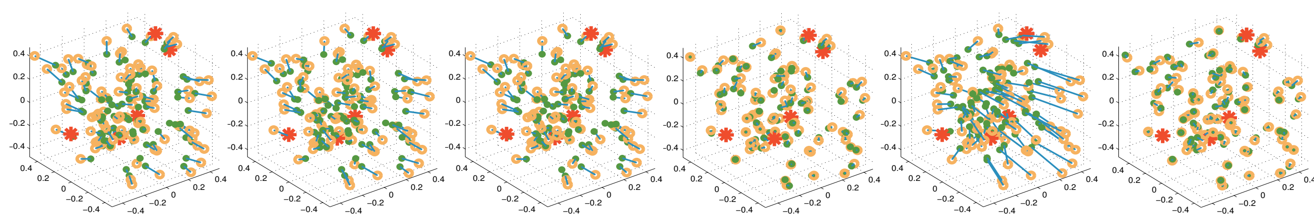



Figure 1: Asymptotic behavior of optimizing Eq (3) for the minimum-rank sensor network localization problem in the presence of Gaussian noise. We plot the values of $rank$ (blue) and $dist$ (red) against the number of iterations, as well as, the sensor locations at different stages of the optimization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



(a) FM, $dist = 0.19$, $rank = 7$. (b) TAM, $dist = 0.21$, $rank = 4$. (c) LPAM, $dist = 0.19$, $rank = 8$. (d) LDHM, $dist = 0.19$, $rank = 2$. (e) PDA, $dist = 0.33$, $rank = 8$. (f) PADM, $dist = 0.14$, $rank = 2$.

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



(a) FM, $dist = 0.28$, $rank = 8$. (b) TAM, $dist = 0.29$, $rank = 6$. (c) LPAM, $dist = 0.27$, $rank = 10$. (d) LDHM, $dist = 0.21$, $rank = 3$. (e) PDA, $dist = 0.38$, $rank = 9$. (f) PADM, $dist = 0.17$, $rank = 3$.

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


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.

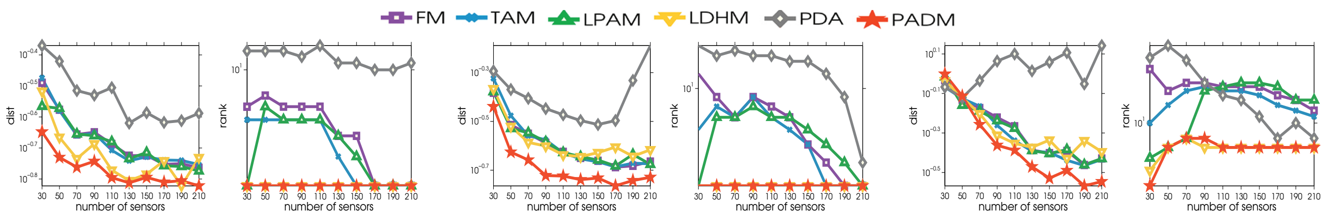
Krislock, N., and Wolkowicz, H. 2010. Explicit sensor network localization using semidefinite representations and facial reductions. *SIOPT* 20(5):2679–2708.

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. *SIOPT* 21(4):1641–1666.

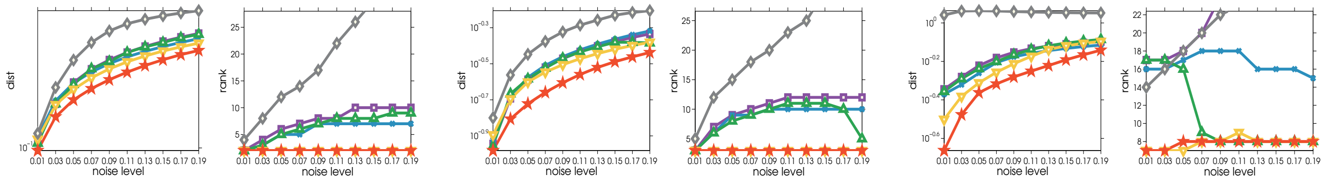
Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, 612–620.

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.



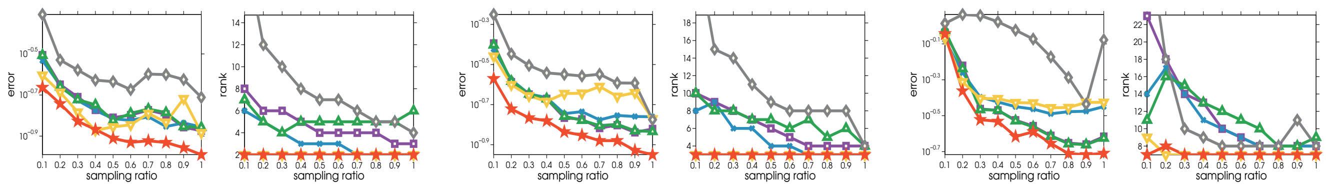

(b) $dist$ and $rank$ comparisons on 2d data (c) $dist$ and $rank$ comparisons on 3d data (d) $dist$ and $rank$ comparisons on 7d data

Figure 4: Performance comparison with varying number of sensors $u$ in the presence of Gaussian 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 noise level $s$ in the presence of Gaussian 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 sampling ratio $r$ in the presence of Gaussian noise.


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 $\ell_p$ 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 as-rigid-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.