The Minimum-Rank Gram Matrix Completion via Modified Fixed Point Continuation Method*

Yue Ma and Lihong Zhi
Key Laboratory of Mathematics Mechanization
AMSS, Beijing 100190, China
{yma, lzhi}@mmrc.iss.ac.cn

ABSTRACT

The problem of computing a representation for a real polynomial as a sum of minimum number of squares of polynomials can be casted as finding a symmetric positive semidefinite real matrix of minimum rank subject to linear equality constraints. In this paper, we propose algorithms for solving the minimum-rank Gram matrix completion problem, and show the convergence of these algorithms. Our methods are based on the fixed point continuation method. We also use the Barzilai-Borwein technique and a specific linear combination of two previous iterates to accelerate the convergence of modified fixed point continuation algorithms. We demonstrate the effectiveness of our algorithms for computing approximate and exact rational sum of squares decompositions of polynomials with rational coefficients.

Categories and Subject Descriptors: I.1.2 [Symbolic and Algebraic Manipulation]: Algorithms; G.1.6 [Numerical Analysis]: Global optimization

General Terms: algorithms, experimentation

Keywords: Gram matrix completion, nuclear norm minimization, Schur decomposition, sum of squares, fixed point iterative method

1. INTRODUCTION

Let $x = [x_1, \ldots, x_s]$ and $f(x) \in \mathbb{R}[x]$, then f is a sum of squares (SOS) in $\mathbb{R}[x]$ if and only if it can be written in the form

$$f(x) = m_d(x)^T \cdot W \cdot m_d(x), \tag{1}$$

in which $m_d(x)$ is a column vector of monomials of degree less than or equal to d and W is a real positive semidefinite matrix [42, Theorem 1] (see also [10]). W is also called a *Gram matrix* for f. If W has rational entries, then f is a sum of squares in $\mathbb{Q}[x]$.

*This material is based on work supported by a NKBRPC 2011CB302400, the Chinese National Natural Science Foundation under Grants 60821002/F02, 60911130369 and 10871194.

ISSAC'11, June 8–11, 2011, San Jose, California, USA.

Problem 1 *Let $f \in \mathbb{Q}[x_1, \ldots, x_s]$ be a polynomial of degree $2d$, compute a representation for it as a sum of minimum number of squares of polynomials in $\mathbb{Q}[x_1, \ldots, x_s]$.*

The set of all matrices W for which (1) holds is an affine subspace of the set of symmetric matrices. If the intersection of this affine subspace with the cone of positive semidefinite (PSD) matrices is nonempty, then f can be written as a sum of squares. Since the components of $m_d(x)$ are not algebraically independent, W is in general not unique. Problem 1 can be restated as finding a Gram matrix with minimum rank satisfying a given set of constraints:

$$\left.\begin{array}{ll} \min & \operatorname{rank}(W) \\ s.t. & f(x) = m_d(x)^T \cdot W \cdot m_d(x) \\ & W \succeq 0,\ W^T = W \end{array}\right\} \tag{2}$$

For $s = 1$, Pourchet's main theorem [41] implies that every positive definite univariate polynomial in $\mathbb{Q}[x]$ is a sum of five squares in $\mathbb{Q}[x]$. Therefore, the minimum rank of the Gram matrix satisfying (2) is bounded by 5 for $s = 1$. For $s > 1$, Pfister's general theorem [40] shows that every positive definite polynomial in $\mathbb{R}[x_1, \ldots, x_s]$ is a sum of 2^s squares of rational functions in $\mathbb{R}(x_1, \ldots, x_s)$. It is well known that there exist positive semidefinite polynomials which cannot be written as sums of polynomial squares. However, as shown in [22], various exceptional SOS problems in the literature can be written as sums of less than 10 squares of polynomials after multiplying by suitable polynomials.

In general, the rank minimization is an intractable problem and is in fact provably NP-hard due to the combinational nature of the non-convex rank function [9]. In [12, 13, 43], they showed that $\operatorname{rank}(W)$ can be replaced by the nuclear norm of W, which is the best convex approximation of the rank function over the unit ball of matrices.

Expanding the right-hand side of the equality condition of (2), matching coefficients of the monomials, we obtain a set of linear equations which can be written as

$$\mathcal{A}(W) = b, \tag{3}$$

where $b = (b_1, \ldots, b_p) \in \mathbb{R}^p$ and b_i is the coefficient of the monomial $x^{\alpha_i} = x_1^{\alpha_{i,1}} \cdots x_s^{\alpha_{i,s}}$ in $f(x)$. The action of the linear operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^p$ on W is described by the inner product $\langle A_i, W \rangle := \operatorname{Tr}(A_i^T W)$ for $A_1, \ldots, A_p \in \mathbb{S}^n$. We use exponent tuples for indexing the matrices, then the entry in A_i with row index β_i and column index γ_i is equal to one if $\beta_i + \gamma_i = \alpha_i$ and zero otherwise. Therefore, there are at most n nonzero entries (being 1) in $A_i, i = 1, \ldots, p$. We use $\mathcal{A}^* : \mathbb{R}^p \to \mathbb{S}^n$ to denote the adjoint operator of $\mathcal{A}$.

The rank minimization problem (2) can be relaxed to a nuclear norm minimization problem

$$\left.\begin{array}{ll}\min & \|W\|_* \\ s.t. & \mathcal{A}(W) = b \\ & W \succeq 0,\, W^T = W\end{array}\right\} \tag{4}$$

where the nuclear norm $\|W\|_*$ is defined as the sum of its singular values. The constraint $\mathcal{A}(W) = b$ can also be relaxed, resulting in either the problem

$$\left.\begin{array}{ll}\min & \|W\|_* \\ s.t. & \|\mathcal{A}(W) - b\|_2 \le \epsilon \\ & W \succeq 0,\, W^T = W\end{array}\right\} \tag{5}$$

or its Lagrangian version

$$\min_{W \in \mathbb{S}^n_+} \ \mu\|W\|_* + \frac{1}{2}\|\mathcal{A}(W) - b\|_2^2, \tag{6}$$

where $\mathbb{S}^n_+$ is the set of symmetric positive semidefinite matrices and $\mu > 0$ is a parameter.

Prior work. In [1, 16, 25, 26, 27], they studied how to determine whether partially specified positive semidefinite matrices can be completed to fully specified matrices satisfying certain prescribed properties. A number of recent work has also shown that the low-rank solution can be recovered exactly via minimizing the nuclear norm under certain conditions [7, 8, 43, 44]. Notice that the nuclear norm of a symmetric positive semidefinite matrix is actually the trace, the nuclear norm minimization problem (4) is just a standard linear SDP problem which can be directly solved by interior-point methods in [4, 5, 13, 29, 45, 47] or projection methods in [18, 19, 23, 33, 37, 53] for large-scale SDP problems. Since most of these methods use second-order information, the memory requirement for computing descent directions quickly becomes too large as the problem size increases. Recently, several fast algorithms using only first-order information have been developed in [6, 14, 31, 32]. Moreover, some accelerated gradient algorithms were also proposed in [3, 20, 34, 35, 36, 49, 50] which have an attractive convergence rate of $O(1/k^2)$, where k is the iteration counter. These first-order methods, based on function values and gradient evaluation, cannot yield as high accuracy as interior point methods, but much larger problems can be solved since no second-order information needs to be computed and stored.

Main results. In this paper, we present two algorithms for solving the minimum-rank Gram matrix completion problem (4). Our algorithms are based on the fixed point continuation method (FPC). By modifying the shrinkage operator in FPC and using the Barzilai-Borwein (BB) technique to compute explicit dynamically updated step sizes, we get an algorithm, called modified fixed point continuation method with the Barzilai-Borwein technique (MFPC-BB). We prove the convergence of our algorithm under certain conditions.

We incorporate an accelerating technique in the MFPC-BB algorithm by computing the next iterate based not only on the previous one, but also on two previously computed iterates to get the AFPC-BB algorithm, which keeps its simplicity but shares the improved rate $O(1/k^2)$ of the optimal gradient method.

Numerical experiments demonstrate that modified FPC algorithms outperform SDP solvers SeDuMi [48] (in YALMIP [30]) and SDPNAL [53] for computing approximate and exact rational SOS of polynomials with rational coefficients, especially for large-scale sparse examples.

Structure of the paper. In Section 2, we derive the modified fixed point iterative algorithm for the minimum-rank Gram matrix completion problem. In Section 3, we establish the convergence result for our algorithm and prove that it converges to the optimal solution of problem (6). In Section 4, we introduce two techniques to accelerate the convergence of our algorithm and present MFPC-BB and AFPC-BB algorithms for solving problem (6). We demonstrate the performance and effectiveness of our algorithms through numerical examples for computing approximate and exact rational sum of squares decompositions of polynomials with rational coefficients in Section 5.

2. MODIFIED FIXED POINT ITERATIVE ALGORITHM

Let $f : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}$ be a convex function, the subdifferential of f at $X^* \in \mathbb{R}^{n_1 \times n_2}$ denoted by $\partial f(X^*)$ is the compact convex set defined by

$$\partial f(X^*) := \{Z \in \mathbb{R}^{n_1 \times n_2} : f(Y) \ge f(X^*) + \langle Z, Y - X^* \rangle, \\ \forall\ Y \in \mathbb{R}^{n_1 \times n_2}\}.$$

Following discussions in [28, Theorem 3.1] and [51], we derive the expression of the subdifferential of the nuclear norm at a symmetric matrix.

Theorem 1 *Let $W \in \mathbb{S}^n$, then*

$$\partial\|W\|_* = \{Q^{(1)}Q^{(1)T} - Q^{(2)}Q^{(2)T} + Z : Q^{(i)T}Z = 0, \\ i = 1, 2,\ \textit{and}\ \|Z\|_2 \le 1\},$$

where $Q^{(1)}$ and $Q^{(2)}$ are orthogonal eigenvectors associated with the positive and negative eigenvalues of W respectively.

Proof. Suppose that the eigenvalues of a symmetric matrix W can be ordered as $\lambda_1 \ge \cdots \ge \lambda_t > 0 > \lambda_{t+1} \ge \cdots \ge \lambda_s$, $\lambda_{s+1} = \cdots = \lambda_n = 0$. Let $W = Q\Lambda Q^T$ be a Schur decomposition of W, where $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. These matrices can be partitioned as

$$Q = \left(Q^{(1)}, Q^{(2)}, Q^{(3)}\right), \quad \Lambda = \begin{pmatrix} \Lambda^{(1)} & 0 & 0 \\ 0 & \Lambda^{(2)} & 0 \\ 0 & 0 & \Lambda^{(3)} \end{pmatrix},$$

with $Q^{(1)}, Q^{(2)}, Q^{(3)}$ having $t, s-t, n-s$ columns and being associated with $\Lambda^{(1)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_t)$, $\Lambda^{(2)} = \mathrm{diag}(\lambda_{t+1}, \ldots, \lambda_s)$, and $\Lambda^{(3)} = \mathrm{diag}(\lambda_{s+1}, \ldots, \lambda_n)$, respectively.

Let $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ and recall that

$$\partial\|\lambda\|_1 = \{y \in \mathbb{R}^n : y_i = 1, i = 1, \ldots, t;\ y_j = -1, \\ j = t+1, \ldots, s;\ |y_k| < 1, k = s+1, \ldots, n\}.$$

Let $Y \in \partial\|W\|_*$, by [28, Theorem 3.1], we have

$$Y = Q\,\mathrm{diag}(d)\,Q^T,$$

where $d \in \partial\|\lambda\|_1$. Therefore

$$Y = Q^{(1)}Q^{(1)T} - Q^{(2)}Q^{(2)T} + Q^{(3)}DQ^{(3)T},$$

where D is an $(n-s)\times(n-s)$ diagonal matrix with diagonal elements less than 1 in modulus.

Let $Z = Q^{(3)}DQ^{(3)T}$, we have $Q^{(i)T}Z = 0, i = 1, 2$. Let $\sigma_1(\cdot)$ denote the largest singular value of a given matrix, then we have

$$\|Z\|_2 = Q^{(3)}DQ^{(3)T} \leq \sigma_1(D) < 1,$$

which completes the proof. □

The optimality condition in [31, Theorem 2] can be generalized to the optimality condition for the constrained convex optimization problem (6).

Theorem 2 *Let $f : \mathbb{S}^n \to \mathbb{R}$ be a proper convex function, i.e. $f < +\infty$ for at least one point and $f > -\infty$ for every point in its domain. Then W^* is an optimal solution to the problem*

$$\min_{W\in\mathbb{S}^n_+} \ f(W) \tag{7}$$

if and only if $W^* \in \mathbb{S}^n_+$, and there exists a matrix $U \in \partial f(W^*)$ such that

$$\langle U, V - W^*\rangle \geq 0, \ \textit{for all } V \in \mathbb{S}^n_+. \tag{8}$$

Proof. Suppose $U \in \partial f(W^*)$ and satisfies the inequality condition (8), then

$$f(V) \geq f(W^*) + \langle U, V - W^*\rangle, \quad \forall\ V \in \mathbb{S}^n_+,$$

we have $f(V) \geq f(W^*)$, for all $V \in \mathbb{S}^n_+$. This shows that W^* is an optimal solution of the problem (7).

Conversely, suppose W^* is the optimal solution of the problem (7), and (8) does not hold, i.e., for each $U \in \partial f(W^*)$,

$$\exists\ V \in \mathbb{S}^n_+, \ \text{s.t. } \langle U, V - W^*\rangle < 0. \tag{9}$$

Consider $Z(t) = tW^* + (1-t)V$, where $t \in [0, 1]$ is a parameter. Since $Z(t)$ is on the line segment between W^* and V, and $\mathbb{S}^n_+$ is a convex set, $Z(t) \in \mathbb{S}^n_+, \forall\ t \in [0, 1]$. By [46, Theorem 23.4], the one-sided directional derivative of f at $Z(1)$ with respect to the vector $W^* - V$ satisfies

$$\begin{aligned} f'(Z(t); W^* - V)|_{t=1} &= f'(W^*; W^* - V) \\ &= \sup\{\langle W, W^* - V\rangle : \forall\ W \in \partial f(W^*)\} \\ &\geq \langle U, W^* - V\rangle > 0, \ \text{by (9)}. \end{aligned}$$

Therefore, for a small value $\epsilon > 0$, we have $f(Z(1-\epsilon)) < f(W^*)$, which is contradict to the fact that W^* is optimal to the problem (7). □

Based on above theorems, we can introduce a thresholding operator and extend the fixed point iterative scheme for solving (6).

Definition 1 *Suppose $W = Q\Lambda Q^T$ is a Schur decomposition of a matrix $W \in \mathbb{S}^n$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and Q is a real orthogonal matrix. For any $\nu \geq 0$, the matrix thresholding operator $\mathcal{T}_\nu(\cdot)$ is defined as*

$$\mathcal{T}_\nu(W) := Q\mathcal{T}_\nu(\Lambda)Q^T, \quad \mathcal{T}_\nu(\Lambda) = \mathrm{diag}(\{\lambda_i - \nu\}_+),$$

where $t_+ = \max(0, t)$.

We should point out that the idea of using the eigenvalue decomposition of Y^k has also appeared in [49, Remark 3]. However, to our best knowledge, there exists no convergence analysis about the eigenvalue thresholding operator in the literature.

Let μ and τ be positive real numbers and X^0 be an initial starting matrix. For $k = 0, 1, 2, \cdots$, we compute

$$\left\{ \begin{array}{rcl} Y^k & = & X^k - \tau\mathcal{A}^*(\mathcal{A}(X^k) - b), \\ X^{k+1} & = & \mathcal{T}_{\tau\mu}(Y^k), \end{array} \right. \tag{10}$$

until a stopping criterion is reached.

Theorem 3 *For the operator $\mathcal{A}$ defined by (3), suppose a matrix $W^* \in \mathbb{S}^n_+$ satisfies*

1. *$\|\mathcal{A}(W^*) - b\|_2 < \mu/n$ for a small positive number μ.*
2. *$W^* = \mathcal{T}_{\tau\mu}(h(W^*))$, where $h(\cdot) = I(\cdot) - \tau\mathcal{A}^*(\mathcal{A}(\cdot) - b)$ and $I(\cdot)$ is an identity operator.*

Then W^* is the unique optimal solution of the problem (6).

Proof. Let $\nu = \tau\mu$ and $Y^* = h(W^*) = W^* + E \in \mathbb{S}^n$, where $E = -\tau\mathcal{A}^*(\mathcal{A}(W^*) - b)$. We claim that $\mathcal{T}_\nu(Y^*)$ is the unique optimal solution to the following problem

$$\min_{W\in\mathbb{S}^n_+} \ \nu\|W\|_* + \frac{1}{2}\|W - Y^*\|_F^2, \tag{11}$$

In fact, since the objective function $\nu\|W\|_* + \frac{1}{2}\|W - Y^*\|_F^2$ is strictly convex, there exists a unique minimizer, and we only need to prove that it is equal to $\mathcal{T}_\nu(Y^*)$. Without loss of generality, we assume that the eigenvalues of Y^* can be ordered as

$$\begin{aligned} &\lambda_1(Y^*) \geq \cdots \geq \lambda_t(Y^*) \geq \nu > \lambda_{t+1}(Y^*) \geq \cdots > 0 > \\ &\cdots \geq \lambda_s(Y^*), \lambda_{s+1}(Y^*) = \cdots = \lambda_n(Y^*) = 0. \end{aligned}$$

We compute a Schur decomposition of Y^* as

$$Y^* = Q^{(1)}\Lambda^{(1)}Q^{(1)T} + Q^{(2)}\Lambda^{(2)}Q^{(2)T},$$

where $\Lambda^{(1)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_t)$, $\Lambda^{(2)} = \mathrm{diag}(\lambda_{t+1}, \ldots, \lambda_s)$, $Q^{(1)}$ and $Q^{(2)}$ are block matrices corresponding to $\Lambda^{(1)}$ and $\Lambda^{(2)}$ respectively. Let $\widehat{X} = \mathcal{T}_\nu(Y^*)$, we have

$$\widehat{X} = Q^{(1)}(\Lambda^{(1)} - \nu I)Q^{(1)T},$$

therefore,

$$Y^* - \widehat{X} = \nu(Q^{(1)}Q^{(1)T} + Z), \quad Z = \nu^{-1}Q^{(2)}\Lambda^{(2)}Q^{(2)T}.$$

By definition, $Q^{(1)T}Z = 0$.

- If $\lambda_{t+1}(Y^*) \geq |\lambda_s(Y^*)|$, then $\|Z\|_2 = \lambda_{t+1}(Y^*)/\nu < 1$.
- Otherwise, let $y = (y_1, \ldots, y_p)^T = \mathcal{A}(W^*) - b \in \mathbb{R}^p$. Since $\mathcal{A}^*y = A_1y_1 + \cdots + A_py_p$ and there are at most n nonzero entries (being 1) in A_i, we have

  $$\|E\|_F^2 = \tau^2\|\mathcal{A}^*y\|_F^2 \leq \tau^2n^2(y_1^2 + \cdots + y_p^2) < \tau^2\mu^2.$$

Notice that $E \in \mathbb{S}^n$ and $W^* \in \mathbb{S}^n_+$, by [15, Theorem 8.1.5], we have

$$\|Z\|_2 = \frac{|\lambda_s(Y^*)|}{\nu} = \frac{\max\{|\lambda_1(E)|, |\lambda_n(E)|\}}{\nu} \leq \frac{\|E\|_F}{\nu} < 1.$$

Hence, according to Theorem 1, we have $Y^* - \widehat{X} \in \nu\partial\|\widehat{X}\|_*$, which means that $0 \in \nu\partial\|\widehat{X}\|_* + \widehat{X} - Y^*$. By Theorem 2, we immediately conclude that $\mathcal{T}_\nu(Y^*)$ is an optimal solution of the problem (11).

Since the objective function of the problem (6) is strictly convex, its optimal solution is also unique. If $W^* = \mathcal{T}_{\tau\mu}(Y^*)$, by Theorem 2, there exists a matrix $U \in \nu\partial\|W^*\|_* + W^* - Y^*$ such that

$$\langle U, V - W^* \rangle \geq 0, \quad \forall\, V \in \mathbb{S}^n_+.$$

Let $\widetilde{U} = U/\tau$, by substituting $\nu = \tau\mu$ and $Y^* = W^* - \tau\mathcal{A}^*(\mathcal{A}(W^*) - b)$ into the above subdifferential function, we have $\widetilde{U} \in \mu\partial\|W^*\|_* + \mathcal{A}^*(\mathcal{A}(W^*) - b)$ satisfying

$$\langle \widetilde{U}, V - W^* \rangle \geq 0, \quad \forall\, V \in \mathbb{S}^n_+.$$

By applying Theorem 2 once again, it is true that W^* is the optimal solution of the problem (6). □

3. CONVERGENCE ANALYSIS

In this section, we analyze the convergence properties of the modified fixed point iterative scheme (10). We begin by recording two lemmas which establish the non-expansivity of the thresholding operator $\mathcal{T}_\nu(h(\cdot))$.

Lemma 1 *The thresholding operator $\mathcal{T}_\nu$ is non-expansive, i.e., for any $X_1, X_2 \in \mathbb{S}^n$,*

$$\|\mathcal{T}_\nu(X_1) - \mathcal{T}_\nu(X_2)\|_F \leq \|X_1 - X_2\|_F. \tag{12}$$

Moreover,

$$\begin{aligned}\|X_1 - X_2\|_F &= \|\mathcal{T}_\nu(X_1) - \mathcal{T}_\nu(X_2)\|_F \\ \Longleftrightarrow X_1 - X_2 &= \mathcal{T}_\nu(X_1) - \mathcal{T}_\nu(X_2).\end{aligned}$$

Lemma 2 *Suppose that the step size τ satisfies $\tau \in (0, 2/\|\mathcal{A}\|_2^2)$. Then the operator $h(\cdot) = I(\cdot) - \tau\mathcal{A}^*(\mathcal{A}(\cdot) - b)$ is non-expansive, i.e., for any $X_1, X_2 \in \mathbb{S}^n$,*

$$\|h(X_1) - h(X_2)\|_F \leq \|X_1 - X_2\|_F.$$

Moreover, we have

$$\begin{aligned}\|h(X_1) - h(X_2)\|_F &= \|X_1 - X_2\|_F \\ \Longleftrightarrow h(X_1) - h(X_2) &= X_1 - X_2,\end{aligned}$$

where $I(\cdot)$ is an identity operator.

The proof of these two lemmas follows the similar strategy presented in [31]. Instead of the inequality (3.3) in [31], we use the fact that for $X, Y \in \mathbb{S}^n$,

$$\mathrm{Tr}(X\,Y) \leq \lambda(X)^T\lambda(Y),$$

where $\lambda(X), \lambda(Y)$ are the vectors of eigenvalues of X and Y respectively [28, Theorem 2.2].

We now claim that the modified fixed point iterations (10) converge to the optimal solution of the problem (6).

Theorem 4 *Let $\tau \in (0, 2/\|\mathcal{A}\|_2^2)$ and $W^* \in \mathbb{S}^n_+$ satisfy*

1. *$\|\mathcal{A}(W^*) - b\|_2 < \mu/n$ for a small positive number μ.*
2. *$W^* = \mathcal{T}_{\tau\mu}(h(W^*))$, where $h(\cdot) = I(\cdot) - \tau\mathcal{A}^*(\mathcal{A}(\cdot) - b)$.*

Then the sequence $\{X^k\}$ obtained via modified fixed point iterations (10) converges to W^*.

PROOF. Let $\nu = \tau\mu$. Since both $\mathcal{T}_\nu(\cdot)$ and $h(\cdot)$ are non-expansive, $\mathcal{T}_\nu(h(\cdot))$ is also non-expansive. Therefore, $\{X^k\}$ lies in a compact set and must have a limit point. Suppose $\widetilde{X} = \lim_{j\longrightarrow\infty} X^{k_j}$ satisfying $\|\mathcal{A}(\widetilde{X}) - b\|_2 < \mu/n$. By $W^* = \mathcal{T}_\nu(h(W^*))$, we have

$$\begin{aligned}\|X^{k+1} - W^*\|_F &= \|\mathcal{T}_\nu(h(X^k)) - \mathcal{T}_\nu(h(W^*))\|_F \\ &\leq \|h(X^k) - h(W^*)\|_F \leq \|X^k - W^*\|_F,\end{aligned}$$

which means that the sequence $\{\|X^k - W^*\|_F\}$ is monotonically non-increasing. Therefore

$$\lim_{k\longrightarrow\infty} \|X^k - W^*\|_F = \|\widetilde{X} - W^*\|_F,$$

where $\widetilde{X}$ can be any limit point of $\{X^k\}$. By the continuity of $\mathcal{T}_\nu(h(\cdot))$, we have

$$\mathcal{T}_\nu(h(\widetilde{X})) = \lim_{j\longrightarrow\infty} \mathcal{T}_\nu(h(X^{k_j})) = \lim_{j\longrightarrow\infty} X^{k_j+1},$$

i.e., $\mathcal{T}_\nu(h(\widetilde{X}))$ is also a limit point of $\{X^k\}$. Therefore, we have

$$\begin{aligned}\|\mathcal{T}_\nu(h(\widetilde{X})) - \mathcal{T}_\nu(h(W^*))\|_F &= \|\mathcal{T}_\nu(h(\widetilde{X})) - W^*\|_F \\ &= \|\widetilde{X} - W^*\|_F.\end{aligned}$$

Using Lemma 1 and Lemma 2 we obtain

$$\mathcal{T}_\nu(h(\widetilde{X})) - \mathcal{T}_\nu(h(W^*)) = h(\widetilde{X}) - h(W^*) = \widetilde{X} - W^*,$$

which implies $\mathcal{T}_\nu(h(\widetilde{X})) = \widetilde{X}$. By Theorem 3, $\widetilde{X}$ is the optimal solution to the problem (6), i.e., $\widetilde{X} = W^*$. Hence, we have

$$\lim_{k\longrightarrow\infty} \|X^k - W^*\|_F = 0,$$

i.e., $\{X^k\}$ converges to its unique limit point W^*. □

4. IMPLEMENTATION

This section provides implementation details of the modified FPC algorithm for solving the minimum-rank Gram matrix completion problem.

4.1 Evaluation of the eigenvalue thresholding operator

The main computational cost of the modified FPC algorithm is computing the Schur decompositions. Following the strategies in [6, 49], we use PROPACK [24] in Matlab to compute a partial Schur decomposition of a symmetric matrix.

PROPACK can not automatically compute only eigenvalues greater than a given threshold ν. To use this package, we must predetermine the number s_k of eigenvalues of Y^k to compute at the k-th iteration. Suppose $X^k = Q^{k-1}\Lambda^{k-1}(Q^{k-1})^T$, we set s_k equal to the number of diagonal entries of Λ^{k-1} that are no less than $\varepsilon_k\|\Lambda^{k-1}\|_2$, where ε_k is a small positive number. Notice that s_k is non-increasing. If s_k is too small, the non-expansive property (12) of the thresholding operator $\mathcal{T}_\nu$ may be violated. We increase s_k by 1 if the non-expansive property is violated 10 times [31].

4.2 Barzilai-Borwein technique

In [31], the authors always set the parameter $\tau = 1$ since their operator $\mathcal{A}$ is generated by randomly sampling a subset of p entries from matrices with i.i.d. standard Gaussian entries. For this linear map, the Lipschitz constant for

the objective function of (6) is 1. According to Theorem 4, convergence for the Gram matrix completion problem is guaranteed provided that $\tau \in (0,\ 2/\|\mathcal{A}\|_2^2)$. This choice is, however, too conservative and the convergence is typically slow.

There are many ways to select a step size. For simplicity, we describe a strategy, which is based on the Barzilai-Borwein method [2], for choosing the step size τ_k. Let $g(\cdot) = \mathcal{A}^*(\mathcal{A}(\cdot) - b)$ and $g^k = \mathcal{A}^*(\mathcal{A}(X^k) - b)$. We perform the shrinkage iteration (10) along the negative gradient direction g^k of the smooth function $\frac{1}{2}\|\mathcal{A}(X^k)-b\|_2^2$, then apply the thresholding operator $\mathcal{T}_\nu(\cdot)$ to accommodate the non-smooth term $\|X\|_*$. Hence, it is natural to choose τ_k based on the function $\frac{1}{2}\|\mathcal{A}(X^k) - b\|_2$ alone. Let

$$\Delta X = X^k - X^{k-1}, \quad \Delta g = g^k - g^{k-1}.$$

The BB step provides a two-point approximation to the secant equation underlying quasi-Newton method, specifically,

$$\tau_k = \frac{\langle \Delta X, \Delta g\rangle}{\langle \Delta g, \Delta g\rangle}, \quad \text{or} \quad \tau_k = \frac{\langle \Delta X, \Delta X\rangle}{\langle \Delta X, \Delta g\rangle}.$$

In order to avoiding the parameter τ_k being either too small or too large, we take

$$\tau_k = \max\{\tau_{min}, \min\{\tau_k, \tau_{max}\}\},$$

where $0 < \tau_{min} < \tau_{max} < \infty$ are fixed parameters.

The idea of using the BB step to accelerate the convergence of gradient algorithms has also appeared in [52].

4.3 Algorithms

As suggested in [17, 31, 49], we adopt a continuation strategy to solve the problem (6). For the problem (6) with a target parameter $\bar{\mu}$ being a moderately small number, we propose solving a sequence of problems (6) defined by an decreasing sequence μ_k. When a new problem, associated with μ_{k+1}, is to be solved, the approximate solution for the current problem with μ_k is used as the starting point. We use the parameter η to determine the rate of reduction of the consecutive μ_k, i.e.,

$$\mu_{k+1} = \max(\eta\mu_k, \bar{\mu}), \quad k = 1, \ldots, L-1.$$

Our modified fixed point continuation iterative scheme with the Barzilai-Borwein technique for solving (6) is outlined below.

Algorithm *MFPC-BB*

Input: ‣ Parameters $0 < \tau_{min} < \tau_0 < \tau_{max} < \infty$, $\mu_1 > \bar{\mu} > 0$, $\eta > 0$ and a tolerance $\epsilon > 0$

Output: ‣ A numeric Gram matrix.

- Set $X^0 = 0$.
- **For** $\mu = \mu_1, \ldots, \mu_L$, **do**
  1. Choose a step size τ_k via the BB technique such that $\tau_{min} \le \tau_k \le \tau_{max}$.
  2. Compute $Y^k = X^k - \tau_k\mathcal{A}^*(\mathcal{A}(X^k) - b)$ and a Schur decomposition of $Y^k = Q^k\ \Lambda^k\ (Q^k)^T$.
  3. Compute $X^{k+1} = Q^k\ \mathcal{T}_{\tau_k\mu_k}(\Lambda^k)\ (Q^k)^T$.
- **If** the stop criterion is true, **then return** $X_{\rm opt}$.
- **end for**.

However, as shown in [3, 20, 49], the above algorithm may converge as $O(1/k)$. Very recently, alternative algorithms that could speed up the performance of the gradient method FPC have been proposed in [20, 49]. These algorithms rely on computing the next iterate based not only on the previous one, but also on two or more previously computed iterates. We incorporate this new accelerating technique in our MFPC-BB algorithm to solve the affine constrained low-rank Gram matrix completion problem (6). The accelerated algorithm, called AFPC-BB, keeps the simplicity of MFPC-BB but shares the improved rate $O(1/k^2)$ of the optimal gradient method.

Algorithm *AFPC-BB*

Input: ‣ Parameters $0 < \tau_{min} < \tau_0 < \tau_{max} < \infty$, $\mu_1 > \bar{\mu} > 0$, $\eta > 0$ and tolerance $\epsilon > 0$

Output: ‣ A numeric Gram matrix.

- Set $X^0 = 0, t_0 = 1$.
- **For** $\mu = \mu_1, \ldots, \mu_L$, **do**
  1. Choose a step size τ_k via the BB technique such that $\tau_{min} \le \tau_k \le \tau_{max}$.
  2. Compute $Z^k = X^k + \frac{t_{k-1}-1}{t_k}(X^k - X^{k-1})$.
  3. Compute $Y^k = Z^k - \tau_k\mathcal{A}^*(\mathcal{A}(Z^k)-b)$ and a Schur decomposition of $Y^k = Q^k\ \Lambda^k\ (Q^k)^T$.
  4. Compute $X^{k+1} = Q^k\ \mathcal{T}_{\tau_k\mu_k}(\Lambda^k)\ (Q^k)^T$.
  5. Compute $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$.
- **If** the stop criterion is true, **then return** $X_{\rm opt}$.
- **end for**.

The following theorem shows that by performing the gradient step at the matrix Z^k instead of at the approximate solution X^k, the convergence rate of the MFPC-BB method can be accelerated to $O(1/k^2)$.

Theorem 5 *[20, 49] Let $\{X^k\}$ be the sequence generated by the AFPC-BB algorithm. Then for any $k > 1$, we have*

$$F(X^k) - F(X^*) \le \frac{C\|X^* - X^0\|_F^2}{(k+1)^2},$$

where C is a constant, $F(X)$ is the objective function and X^* is the optimal solution of the problem (6).

5. NUMERICAL EXPERIMENTS

In this section, we report the performance of our modified FPC algorithms for writing a real positive semidefinite polynomial as a sum of minimum number of squares of polynomials. In our tests, we generate positive semidefinite matrices $W \in \mathbb{Q}^{n\times n}$ with rank r by sampling an $n \times r$ factor L with random integers ranging from -10 to 10, and setting $W = LL^{\rm T}$. We construct the column vector $m_d(x)$ by choosing monomials in $x_1, \ldots, x_s$ of degree less than or equal to d for $2 \le s \le 4$ and $5 \le d \le 20$. Therefore, a positive semidefinite polynomial is obtained

$$f(x) = m_d(x)^T \cdot W \cdot m_d(x) \in \mathbb{Q}[x].$$

Replacing entries in W by parameters, expanding the right-hand side of the equality and matching coefficients of the

Problems				MFPC		MFPC-BB		AFPC-BB	
n	r	p	FR	# iter	error	# iter	error	# iter	error
100	10	579	1.6494	527	9.98e-4	434	9.97e-4	50	9.46e-4
200	10	1221	1.6011	797	9.99e-4	512	9.99e-4	59	9.84e-4
500	10	5124	0.9670	632	4.99e-3	499	4.99e-3	66	4.90e-3

Table 1: Comparison of MFPC, MFPC-BB and AFPC-BB, without using continuation technique.

monomials, we obtain a set of linear equations (3), which defines the linear map $\mathcal{A}$ from $\mathbb{S}^n$ to $\mathbb{R}^p$.

We know that an $n \times n$ symmetric matrix of rank r depends on $d_r = r(2n - r + 1)/2$ degrees of freedom. Let FR (degrees of freedom ratio) be d_r/p, where p is the number of linear constrains. If FR is large (close to 1), recovering W becomes harder as the number of measurements is close to the degree of freedom. Conversely, if FR is close to zero, recovering W becomes easier. Note that if $FR > 1$, there might have an infinite number of matrices with rank r satisfying given affine constraints.

The stopping criterion for the MFPC, MFPC-BB, AFPC-BB algorithms in our numerical experiments is given as follows:

$$\text{error} := \frac{\|\mathcal{A}(X_{\text{opt}}) - b\|_2}{\|b\|_2} < \epsilon, \tag{13}$$

where ϵ is a moderately small number. Throughout the experiments, we choose the initial matrix $X^0 = 0$. For each test, we make an initial estimate of the value $L = \|\mathcal{A}\|_2^2$ which is the smallest Lipschitz constant of the gradient of $\frac{1}{2}\|\mathcal{A}X - b\|_2^2$. We set the Barzilai-Borwein parameters $\tau_{max} = 10/L$ and $\tau_{min} = 10^{-3}/L$. The thresholds 10 and 10^{-3} are found after some experiments.

We have implemented the MFPC-BB and AFPC-BB algorithms in MATLAB. All runs are conducted on a HP xw8600 workstation with an Inter Xeon(R) 2.67GHz CPU and 3.00 GB of RAM. The codes can be downloaded from http://www.mmrc.iss.ac.cn/~lzhi/Research/hybrid/FPCs/

5.1 Numerical experiments on random Gram matrix completion problems

In the first series of test, we set $\epsilon = 10^{-3}$ and compare the performance of the MFPC, MFPC-BB and AFPC-BB algorithms without continuation technique to solve problem (6) for randomly generated Gram matrix completion problems with moderate dimensions. In order to see the convergence behaviors of these algorithms clearly, we compute the full Schur decompositions at each iteration.

Table 1 reports the degree of freedom ratio FR, the number of iterations, and the error (13) of the three algorithms. Computational efficiency is measured by the number of iterations. As can be seen from this table, on the condition that these three algorithms achieve similar errors, MFPC-BB provides better performance with less number of iterations than MFPC, which shows that the Barzilai-Borwein technique is quite effective in accelerating the convergence of the MFPC algorithm. Moreover, AFPC-BB outperforms the other two algorithms greatly in terms of the number of iterations.

In Table 2, we report the performance of the AFPC-BB algorithm with continuation technique on randomly generated Gram matrix completion problems. We use PROPACK to compute partial eigenvalues and eigenvectors. For the continuation technique, we set the target parameter $\bar{\mu}$ to be $10^{-4}\|\mathcal{A}^*b\|$ and $\mu_1 = 1/4\|\mathcal{A}^*b\|$. The update strategy for μ_k is $\max(1/4\mu_{k-1}, \bar{\mu})$ until the stopping criterion is satisfied with $\epsilon = 10^{-3}$. The running time here and hereafter is shown in seconds. The rank of the Gram matrix is computed for the given tolerance 10^{-5} for all the following numerical experiments.

Problems				Results		
n	r	p	FR	# iter	rank	time
100	10	579	1.6494	76	10	1.48e+0
500	10	3309	1.4974	139	27	6.13e+1
1000	50	10621	4.5923	127	59	1.53e+2
1500	50	25573	2.8849	196	77	5.41e+2

Table 2: Numerical results for AFPC-BB, with continuation technique.

As indicated in this table, it takes the AFPC-BB algorithm fewer than 200 iterations and less than 10 minutes to reach convergence. For all problems in this set, FR is larger than 1. It is rather surprising that the low-rank Gram matrix can be recovered given only such a small number of affine constraints. To our best knowledge, nobody has considered solving matrix completion problems in this situation yet.

5.2 Exact rational sum of squares certificates

The numerical Gram matrix W returned by the AFPC-BB algorithm satisfies

$$f(x) \approx m_d(x)^T \cdot W \cdot m_d(x), \quad W \succeq 0.$$

In order to derive an exact SOS decomposition of f, we might need to start with an approximate Gram matrix with high accuracy [21, 22, 38, 39] and convert it into a rational matrix.

Although first-order methods are often the only practical option for solving large-scale problems, it is rather difficult for them to achieve high accuracy. Therefore, we apply the structure-preserving Gauss-Newton iterations (see [21, 22]) to refine the low-rank Gram matrix W returned by the AFPC-BB algorithm: we choose a rank r which is less than or equal to the rank of W and compute the truncated $\mathrm{L^TDL}$ decomposition of W to obtain an approximate SOS decomposition

$$f(x) \approx \sum_{i=1}^{r} (\sum_{\alpha} c_{i,\alpha} x^{\alpha})^2,$$

then apply the standard Gauss-Newton iteration to compute $\Delta c_{i,\alpha} x^{\alpha}$ such that

$$f(x) = \sum_{i=1}^{r} (\sum_{\alpha} c_{i,\alpha} x^{\alpha} + \Delta c_{i,\alpha} x^{\alpha})^2 + O(\sum_{i=1}^{r} (\sum_{\alpha} \Delta c_{i,\alpha} x^{\alpha})^2).$$

Examples				Results				Gauss-Newton iteration		
n	r	p	FR	solvers	rank	θ	time	rank	θ	time
100	5	579	0.8463	AFPC-BB	9	8.415e-1	1.75e+0	5	1.935e-9	2.98e+1
				SDPNAL	16	2.600e-1	1.50e+0	5	8.852e-10	2.63e+1
				SeDuMi	100	5.373e-2	4.03e+0	5	1.102e-10	3.22e+1
200	5	1221	0.8108	AFPC-BB	14	3.629e+0	1.07e+1	5	6.950e-10	4.02e+2
				SDPNAL	21	2.828e+0	1.06e+1	5	6.912e-10	5.57e+2
				SeDuMi	200	2.579e-1	5.56e+1	5	7.176e-10	1.10e+3
300	5	1932	0.7712	AFPC-BB	14	2.232e+1	2.32e+1	5	1.379e-9	5.61e+2
				SDPNAL	25	2.505e+0	2.69e+1	5	1.075e-9	7.05e+2
				SeDuMi	300	4.748e-1	2.62e+2	5	1.131e-9	6.89e+2
400	5	2610	0.7624	AFPC-BB	15	1.252e+1	6.23e+1	5	5.825e-7	1.22e+3
				SDPNAL	27	2.086e+0	8.69e+1	5	2.341e-8	5.03e+3
				SeDuMi	399	3.384e-1	4.88e+2	5	4.390e-8	5.03e+3
500	5	5124	0.4859	AFPC-BB	17	2.483e+1	5.33e+1	5	1.479e-5	7.92e+3
				SDPNAL	38	6.333e+0	2.53e+2	5	4.913e-8	1.84e+4
				SeDuMi	–	–	–	–	–	–

Table 3: Exact SOS certificates via AFPC-BB, SDPNAL, SeDuMi and Gauss-Newton iterations.

Problems				AFPC-BB			SDPNAL		
n	r	p	FR	rank	θ	time	rank	θ	time
400	10	10078	0.3924	10	1.712e+1	2.46e+1	66	1.093e+1	1.43e+2
500	20	24240	0.4047	20	1.497e+1	4.48e+1	113	4.232e+1	6.72e+2
1000	10	27101	0.3673	10	2.207e+1	3.70e+2	99	8.801e+1	2.70e+3
1000	50	95367	0.5114	50	1.009e+1	6.56e+2	218	9.200e+1	9.92e+3
1500	10	45599	0.3280	10	3.310e+1	1.00e+3	121	3.408e+1	3.72e+4
1500	50	122742	0.6011	50	1.508e+1	3.84e+3	226	3.790e+1	1.36e+4

Table 4: Exact SOS certificates via AFPC-BB and SDPNAL.

The matrix W is updated accordingly to $W + \Delta W$ and the iteration is stopped when the backward error

$$\theta = \|f(x) - m_d(x)^T \cdot W \cdot m_d(x)\|_2$$

is less than the given tolerance. If it doesn't reach convergence after several Gauss-Newton iterations, we may increase the precision or use different r and try Gauss-Newton iterations again. Since these iterations may be run with multi-precision, it will be very expensive if r is large. This motivates us to find a Gram matrix with the minimum rank.

We notice that the AFPC-BB algorithm provides a low-rank Gram matrix to seed Gauss-Newton iterations while the SDP solver SeDuMi [48] (in YALMIP [30]) usually returns a Gram matrix with the maximum rank (see [11, Theorem 2.1]). It is interesting to notice that the newly developed SDP solver SDPNAL [53] can return a Gram matrix with relatively low rank.

In Table 3, we construct random examples with dense monomial vector $m_d(x)$ and compare the performance of the AFPC-BB algorithm with the SDP solvers SeDuMi and SDPNAL for recovering the low-rank Gram matrices. We also show the effectiveness of Gauss-Newton iterations run in Maple 13 with $Digits = 14$ in refining the numerical Gram matrix. The tolerance ϵ for the three solvers is set to 5×10^{-4} for all the numerical experiments, which is small enough to guarantee very good recoverability.

As indicated in this table, for the first four examples, we can use Gauss-Newton iterations (5.2) to refine the Gram matrices returned by all three algorithms to relatively high accuracy. After rounding every entry of the refined matrix to the nearest integer, we can easily recover a rational Gram matrix with rank 5 which gives the exact SOS representation of the nonnegative polynomial. However, when $n = 500$, SeDuMi have troubles to recover the exact SOS decomposition for the given precision.

In Table 4, we construct random examples with sparse monomial vector $m_d(x)$, therefore, the degree of freedom ratio FR is reduced. It is surprising to notice that, without running Gauss-Newton iterations to achieve high accuracy, it is possible to recover the exact SOS representation of nonnegative polynomials directly from the numerical low-rank Gram matrix returned by the AFPC-BB algorithm. However, we are not yet able to recover exact SOS decompositions of polynomials directly from the matrices returned by SDPNAL, which have relatively large rank for the given tolerance.

Acknowledgments: We thank Ju Sun (National University of Singapore) for pointing out the reference [49]. We also thank all the referees for their useful comments.

6. REFERENCES

[1] Barrett, W., Johnson, C. R., and Tarazaga, P. The real positive definite completion problem for a simple cycle. *Linear Algebra Appl. 192* (1993), 3–31.

[2] Barzilai, J., and Borwein, J. Two-point step size gradient methods. *IMA J. Numer. Anal. 8* (1988), 141–148.

[3] Beck, A., and Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci. 2*, 1 (2009), 183–202.

[4] Burer, S., and Monteiro, R. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program. 95*, 2 (2003), 329–357.

[5] Burer, S., and Monteiro, R. Local minima and convergence in low-rank semidefinite programming. *Math. Program. 103*, 3 (2005), 427–444.

[6] Cai, J.-F., Candès, E. J., and Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim. 20*, 4 (2010), 1956–1982.

[7] Candès, E. J., and Recht, B. Exact matrix completion via convex optimization. *Found. Comput. Math. 9*, 6 (2009), 717–772.

[8] Candès, E. J., and Tao, T. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory 56*, 5 (2010), 2053–2080.

[9] Chistov, A., and Grigoriev, D. Complexity of quantifier elimination in the theory of algebraically closed fields. In *Proceedings of the Mathematical Foundations of Computer Science 1984* (London, UK, 1984), Springer-Verlag, pp. 17–31.

[10] Choi, M., Lam, T., and Reznick, B. Sums of squares of real polynomials. *Symp. in Pure Math. 58*, 2 (1995), 103–126.

[11] de Klerk, E., Roos, C., and Terlaky, T. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Oper. Res. Lett. 20*, 5 (1997), 213–221.

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

[13] Fazel, M., Hindi, H., and Boyd, S. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference* (2001), pp. 4734–4739.

[14] Goldfarb, D., and Ma, S. Convergence of fixed point continuation algorithms for matrix rank minimization. *Foundations of Computational Mathematics 11*, 2 (2011), 183–210.

[15] Golub, G. H., and Van Loan, C. F. *Matrix computations*, third ed. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1996.

[16] Grone, R., Johnson, C. R., de Sá, E. M., and Wolkowicz, H. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl. 58* (1984), 109–124.

[17] Hale, E. T., Yin, W., and Zhang, Y. Fixed-point continuation for l_1-minimization: methodology and convergence. *SIAM J. Optim. 19*, 3 (2008), 1107–1130.

[18] Henrion, D., and Malick, J. Projection methods in convex optimization. *LAAS-CNRS Research Report 10730* (2010).

[19] Henrion, D., and Malick, J. Projection methods for conic feasibility problems; application to sum-of-squares decompositions. *Optimization Methods and Software 26*, 1 (2011), 23–46.

[20] Ji, S., and Ye, J. An accelerated gradient method for trace norm minimization. In *Proceedings of the 26th Annual International Conference on Machine Learning* (New York, NY, USA, 2009), ICML '09, ACM, pp. 457–464.

[21] Kaltofen, E., Li, B., Yang, Z., and Zhi, L. Exact certification of global optimality of approximate factorizations via rationalizing sums-of-squares with floating point scalars. In *ISSAC '08: Proceedings of the twenty-first international symposium on Symbolic and algebraic computation* (New York, NY, USA, 2008), ACM, pp. 155–164.

[22] Kaltofen, E., Li, B., Yang, Z., and Zhi, L. Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients, 2009. Accepted for publication in J. Symbolic Comput.

[23] Kočvara, M., and Stingl, M. On the solution of large-scale SDP problems by the modified barrier method using iterative solvers. *Math. Program. 109*, 2-3, Ser. B (2007), 413–444.

[24] Larsen, R. PROPACK - software for large and sparse SVD calculations. Available from: http://soi.stanford.edu/~rmunk/PROPACK/.

[25] Laurent, M. The real positive semidefinite completion problem for series-parallel graphs. *Linear Algebra Appl. 252* (1997), 347–366.

[26] Laurent, M. Polynomial instances of the positive semidefinite and Euclidean distance matrix completion problems. *SIAM J. Matrix Anal. Appl. 22*, 3 (2000), 874–894.

[27] Laurent, M. Matrix completion problems. *Encyclopedia of Optimization III (Interior - M)* (2001), 221–229.

[28] Lewis, A. S. Convex analysis on the Hermitian matrices. *SIAM J. Optim. 6*, 1 (1996), 164–177.

[29] Liu, Z., and Vandenberghe, L. Interior-point method for nuclear norm approximation with application to system identification. *SIAM J. Matrix Anal. Appl. 31* (2009), 1235–1256.

[30] Löfberg, J. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proc. IEEE CCA/ISIC/CACSD Conf.* (Taipei, Taiwan, 2004). URL: http://control.ee.ethz.ch/~joloef/yalmip.php.

[31] Ma, S., Goldfarb, D., and Chen, L. Fixed point and bregman iterative methods for matrix rank minimization. *Math. Program.* (2009), 1–33.

[32] Ma, Y. The minimum-rank Gram matrix completion via fixed point continuation method (in Chinese). *Journal of Systems Science and Mathematical Sciences 30*, 11 (2010), 1501–1511.

[33] Malick, J., Povh, J., Rendl, F., and Wiegele, A. Regularization methods for semidefinite programming. *SIAM J. Optim. 20*, 1 (2009), 336–356.

[34] Nesterov, Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady 27* (1983), 372–376.

[35] Nesterov, Y. Smooth minimization of non-smooth functions. *Math. Program. 103*, 1 (2005), 127–152.

[36] Nesterov, Y. Gradient methods for minimizing composite objective function. Tech. rep., 2007.

[37] Nie, J. Regularization methods for sum of squares relaxations in large scale polynomial optimization. Tech. rep., 2009. Available: http://arxiv.org/abs/0909.3551.

[38] Peyrl, H., and Parrilo, P. A Macaulay 2 package for computing sum of squares decompositions of polynomials with rational coefficients. In *ProcSNC07* (2007), pp. 207–208.

[39] Peyrl, H., and Parrilo, P. Computing sum of squares decompositions with rational coefficients. *Theoretical Computer Science 409* (2008), 269–281.

[40] Pfister, A. Zur Darstellung definiter Funktionen als Summe von Quadraten. *Inventiones Math. 4*, 4 (1967), 229–236.

[41] Pourchet, Y. Sur la représentation en somme de carrés des polynômes à une indéterminée sur un corps de nombres algébriques. *Acta Arith. 19* (1971), 89–104.

[42] Powers, V., and Wörmann, T. An algorithm for sums of squares of real polynomials. *J. Pure Appl. Algebra 127*, 1 (1998), 99–104.

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

[44] Recht, B., Xu, W., and Hassibi, B. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *CDC* (2008), pp. 3065–3070.

[45] Rennie, J., and Srebro, N. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning* (2005), ICML '05, pp. 713–719.

[46] Rockafellar, R. *Convex Analysis*. Princeton University Press, 1972.

[47] Srebro, N., Rennie, J., and Jaakkola, T. Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems* (2005).

[48] Sturm, J. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software 11/12* (1999), 625–653.

[49] Toh, K.-C., and Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Tech. rep., 2009. Available: http://www.optimization-online.org/DBHTML/2009/03/2268.html.

[50] Tseng, P. On accelerated proximal gradient methods for convex-concave optimization. *Submitted to SIAM J. Optim* (2008).

[51] Watson, G. A. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl. 170* (1992), 33–45.

[52] Wen, Z., Yin, W., Goldfarb, D., and Zhang, Y. A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation. *SIAM J. Sci. Comput. 32*, 4 (2010), 1832–1857.

[53] Zhao, X.-Y., Sun, D., and Toh, K.-C. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim. 20*, 4 (2010), 1737–1765.