An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP

Kaifeng Jiang [*], Defeng Sun [†], and Kim-Chuan Toh [‡]

Abstract

The accelerated proximal gradient (APG) method, first proposed by Nesterov, and later refined by Beck and Teboulle, and studied in a unifying manner by Tseng has proven to be highly efficient in solving some classes of large scale structured convex optimization (possibly nonsmooth) problems, including nuclear norm minimization problems in matrix completion and l_1 minimization problems in compressed sensing. The method has superior worst-case iteration complexity over the classical projected gradient method, and usually has good practical performance on problems with appropriate structures. In this paper, we extend the APG method to the inexact setting where the subproblem in each iteration is only solved approximately, and show that it enjoys the same worst-case iteration complexity as the exact counterpart if the subproblems are progressively solved to sufficient accuracy. We apply our inexact APG method to solve large scale convex quadratic semidefinite programming (QSDP) problems of the form: $\min\{\frac{1}{2}\langle x,\, \mathcal{Q}(x)\rangle + \langle c,\, x\rangle \mid \mathcal{A}(x) = b, x \succeq 0\}$, where $\mathcal{Q}, \mathcal{A}$ are given linear maps and b, c are given data. The subproblem in each iteration is solved by a semismooth Newton-CG (SSNCG) method with warm-start using the iterate from the previous iteration. Our APG-SSNCG method is demonstrated to be efficient for QSDP problems whose positive semidefinite linear maps $\mathcal{Q}$ are highly ill-conditioned or rank deficient.

Keywords: Inexact accelerated proximal gradient method, Semismooth Newton CG method, Convex quadratic SDP, Nearest correlation matrix estimation

---

[*]Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. (`kaifengjiang@nus.edu.sg`)

[†]Department of Mathematics and NUS Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (`matsundf@nus.edu.sg`).

[‡]Department of Mathematics National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (`mattohkc@nus.edu.sg`); and Singapore-MIT Alliance, 4 Engineering Drive 3, Singapore 117576.

1   Introduction

Let $\mathcal{S}^n$ be the space of $n \times n$ real symmetric matrices endowed with the standard trace inner product $\langle \cdot, \cdot \rangle$ and Frobenius norm $\| \cdot \|$, and $\mathcal{S}^n_+$ ($\mathcal{S}^n_{++}$) the set of positive semidefinite (definite) matrices in $\mathcal{S}^n$. We consider the following linearly constrained convex semidefinite programming problem:

$$(P) \quad \min\{f(x) \,:\, \mathcal{A}(x) = b, \ x \succeq 0, \ x \in \mathcal{S}^n\}$$

where f is a smooth convex function on $\mathcal{S}^n_+$, $\mathcal{A} : \mathcal{S}^n \to \mathcal{R}^m$ is a linear map, $b \in \mathcal{R}^m$, and $x \succeq 0$ means that $x \in \mathcal{S}^n_+$. Let $\mathcal{A}^*$ be the adjoint of $\mathcal{A}$. The dual problem associated with (P) is given by

$$(D) \quad \max\{f(x) - \langle \nabla f(x), \, x \rangle + \langle b, \, p \rangle \,:\, \nabla f(x) - \mathcal{A}^* p - z = 0, p \in \mathcal{R}^m, z \succeq 0, x \succeq 0\}.$$

We assume that the linear map $\mathcal{A}$ is surjective, and that strong duality holds for (P) and (D). Let x_* be an optimal solution of (P) and (x_*, p_*, z_*) be an optimal solution of (D). Then, as a consequence of strong duality, they must satisfy the following KKT conditions:

$$\mathcal{A}(x) = b, \quad \nabla f(x) - \mathcal{A}^* p - z = 0, \quad \langle x, \, z \rangle = 0, \quad x, z \succeq 0.$$

The problem (P) contains the following important special case of convex quadratic semidefinite programming (QSDP):

$$\min \left\{ \frac{1}{2} \langle x, \, \mathcal{Q}(x) \rangle + \langle c, \, x \rangle \,:\, \mathcal{A}(x) = b, x \succeq 0 \right\}, \tag{1}$$

where $\mathcal{Q} : \mathcal{S}^n \to \mathcal{S}^n$ is a given self-adjoint positive semidefinite linear operator and $c \in \mathcal{S}^n$. Note that the Lagrangian dual problem of (1) is given by

$$\max \left\{ -\frac{1}{2} \langle x, \, \mathcal{Q}(x) \rangle + \langle b, p \rangle \,:\, \mathcal{A}^*(p) - \mathcal{Q}(x) + z = c, z \succeq 0 \right\}. \tag{2}$$

A typical example of QSDP is the nearest correlation matrix problem, where given a symmetric matrix $u \in \mathcal{S}^n$ and a linear map $\mathcal{L} : \mathcal{S}^n \to \mathcal{R}^{n \times n}$, one intends to solve

$$\min \left\{ \frac{1}{2} \|\mathcal{L}(x - u)\|^2 \,:\, \operatorname{diag}(x) = e, \ x \succeq 0 \right\}, \tag{3}$$

where $e \in \mathcal{R}^n$ is the vector of all ones and $u \in \mathcal{S}^n$ is given. If we let $\mathcal{Q} = \mathcal{L}^* \mathcal{L}$ and $c = -\mathcal{L}^* \mathcal{L}(u)$ in (3), then we get the QSDP problem (1). A well studied special case of (3) is the W-weighted nearest correlation matrix problem, where $\mathcal{L} = W^{1/2} \circledast W^{1/2}$ for a given $W \in \mathcal{S}^n_{++}$ and $\mathcal{Q} = W \circledast W$. Note that for $U \in \mathcal{R}^{n \times r}$, $V \in \mathcal{R}^{n \times s}$, $U \circledast V : \mathcal{R}^{r \times s} \to \mathcal{S}^n$ is the symmetrized Kronecker product linear map defined by $U \circledast V(M) = (UMV^T + VM^TU^T)/2$.

There are several methods available for solving this special case of (3), which include the alternating projection method [5], the quasi-Newton method [7], the inexact semismooth Newton-CG method [10] and the inexact interior-point method [13]. All these methods, excluding the inexact interior-point method, rely critically on the fact that the projection of a given matrix $x \in \mathcal{S}^n$ onto $\mathcal{S}^n_+$ has an analytical formula with respect to the norm $\|W^{1/2}(\cdot)W^{1/2}\|$. However, all above mentioned techniques cannot be extended to efficiently solve the H-weighted case [5] of (3), where $\mathcal{L}(x) = H \circ x$ for some $H \in \mathcal{S}^n$ with nonnegative entries and $\mathcal{Q}(x) = (H \circ H) \circ x$, with "$\circ$" denoting the Hardamard product of two matrices defined by $(A \circ B)_{ij} = A_{ij}B_{ij}$. The aforementioned methods are not well suited for the H-weighted case of (3) because there is no explicitly computable formula for the following problem

$$\min \left\{ \frac{1}{2}\|H \circ (x - u)\|^2 \ : \ x \succeq 0 \right\}, \tag{4}$$

where $u \in \mathcal{S}^n$ is a given matrix. To tackle the H-weighted case of (3), Toh [12] proposed an inexact interior-point method for a general convex QSDP including the H-weighted nearest correlation matrix problem. Recently, Qi and Sun [11] introduced an augmented Lagrangian dual method for solving the H-weighted version of (3), where the inner subproblem was solved by a semismooth Newton-CG (SSNCG) method. The augmented Lagrangian dual method avoids solving (4) directly and it can be much faster than the inexact interior-point method [12]. However, if the weight matrix H is very sparse or ill-conditioned, the conjugate gradient (CG) method would have great difficulty in solving the linear system of equations in the semismooth Newton method, and the augmented Lagrangian method would not be efficient or even fail.

Another example of QSDP comes from the civil engineering problem of estimating a positive semidefinite stiffness matrix for a stable elastic structure from r measurements of its displacements $\{u_1, \ldots, u_r\} \subset \mathcal{R}^n$ in response to a set of static loads $\{f_1, \ldots, f_r\} \subset \mathcal{R}^n$ [17]. In this application, one is interested in the QSDP problem: $\min\{\|f - \mathcal{L}(x)\|^2 \mid x \in \mathcal{S}^n_+\}$, where $\mathcal{L} : \mathcal{S}^n \to \mathcal{R}^{n \times r}$ is defined by $\mathcal{L}(x) = xu$, and $f = [f_1, \ldots, f_r]$, $u = [u_1, \ldots, u_r]$. In this case, the corresponding map $\mathcal{Q} = \mathcal{L}^*\mathcal{L}$ is given by $\mathcal{Q}(x) = (xB + Bx)/2$ with $B = uu^T$.

The main purpose of this paper is to design an efficient algorithm to solve the problem (P). The algorithm we propose here is based on the accelerated proximal gradient (APG) method of Beck and Teboulle [1] (the method is called FISTA in [1]), where in the kth iteration with iterate $\bar{x}_k$, a subproblem of the following form must be solved:

$$\min \left\{ \langle \nabla f(\bar{x}_k), \ x - \bar{x}_k \rangle + \frac{1}{2}\langle x - \bar{x}_k, \ \mathcal{H}_k(x - \bar{x}_k) \rangle : \mathcal{A}(x) = b, \ x \succeq 0 \right\} \tag{5}$$

where $\mathcal{H}_k : \mathcal{S}^n \to \mathcal{S}^n$ is a given self-adjoint positive definite linear operator. In FISTA [1], $\mathcal{H}_k$ is restricted to $L\mathcal{I}$, where $\mathcal{I}$ denotes the identity map and L is a Lipschitz constant for ∇f. More significantly, for FISTA in [1], the subproblem (5) must be solved exactly to generate the next iterate x_{k+1}. In this paper, we design an inexact APG method which

overcomes the two limitations just mentioned. Specifically, in our inexact algorithm, the subproblem (5) is only solved approximately and $\mathcal{H}_k$ is not restricted to be a scalar multiple of $\mathcal{I}$. In addition, we are able to show that if the subproblem (5) is progressively solved with sufficient accuracy, then the number of iterations needed to achieve ε-optimality (in terms of the function value) is also proportional to $1/\sqrt{\varepsilon}$, just as in the exact algorithm.

Another strong motivation for designing an inexact APG algorithm comes from the recent paper [2], which considered the following regularized inverse problem:

$$\min_x \left\{ \frac{1}{2}\|\Phi x - y\|^2 + \lambda\|x\|_{\mathcal{B}} \right\} \tag{6}$$

where $\Phi : \mathcal{R}^p \to \mathcal{R}^n$ is a given linear map and $\|x\|_{\mathcal{B}}$ is the atomic norm induced by a given compact set of atoms $\mathcal{B}$ in $\mathcal{R}^p$. It appears that the APG algorithm is highly suited for solving (6). But note that in each iteration of the APG algorithm, a subproblem of the form, $\min_z\{\frac{1}{2}\|z - x\|^2 + \mu\|z\|_{\mathcal{B}}\} \equiv \min\{\frac{1}{2}\|y - x\|^2 \mid \|y\|_{\mathcal{B}}^* \leq \mu\}$, must be solved. However, for most choices of $\mathcal{B}$, the subproblem does not admit an analytical solution and has to be solved numerically. As a result, the subproblem is never solved exactly. In fact, it could be computational very expensive to solve the subproblem to high accuracy. Our inexact APG algorithm thus has the attractive computationally advantage that the subproblems need only be solved with progressively better accuracy while still maintaining the global iteration complexity.

We should mention that the fast gradient method of Nesterov [9] has also been extended in [4] to the problem, $\min\{f(x) \mid x \in Q\}$, where the function f is convex (not necessarily smooth) on the closed convex set Q, and is equipped with the so-called first-order (δ, L)-oracle where for any $y \in Q$, we can compute a pair $(f_{\delta,L}(y), g_{\delta,L}(y))$ such that $0 \leq f(x) - f_{\delta,L}(y) - \langle g_{\delta,L}(y),\, x - y \rangle \leq \frac{L}{2}\|x - y\|^2 + \delta \quad \forall\, x \in Q$. In the inexact-oracle fast gradient method in [4], the subproblem of the form $\min\{\langle g,\, x - y \rangle + \frac{\gamma}{2}\|x - y\|^2 \mid x \in Q\}$ in each iteration must be solved exactly. Thus the kind of the inexactness considered in [4] is very different from what we consider in this paper.

From a practical perspective, the extension of the APG algorithm in [1] to the inexact setting would not be as interesting if we cannot demonstrate it to be computationally viable or offer any computational advantage over alternative algorithms for solving a problem such as (1). Hence, even though the focus of this paper is on designing some theoretical inexact APG algorithms for solving the problem (P), and on establishing their iteration complexities, we do present some preliminary numerical results to demonstrate the practical viability of our proposed inexact algorithms. In particular, as we will demonstrate later in the paper, if the linear operator $\mathcal{H}_k$ is chosen appropriately so that the subproblem (5) is amenable to computation via the SSNCG method, then our inexact APG algorithm can be much more efficient than the state-of-the-art algorithm (the augmented Lagrangian method in [11]) for solving some convex QSDP problems arising from the H-weighted case of the nearest correlation matrix problem (3). In fact, from our preliminary numerical results, we observe that when the weight matrix H in (4) is

highly ill-conditioned, our inexact APG algorithm can be 15-70 times faster than the state-of-the-art augmented Lagrangian method designed in [11].

The paper is organized as follows. In section 2, we propose an inexact APG algorithm for solving a minimization problem of the form $\min\{f(x) + g(x) : x \in \mathcal{X}\}$ where f is a smooth convex function with Lipschitz continuous gradient and g is a proper lower semi-continuous convex function. We also prove that the proposed inexact APG algorithm enjoys the same iteration complexity as the FISTA algorithm in [1]. In section 3, we propose and analyse an inexact APG algorithm for the problem (P) for which the semidefinite least squares subproblem in each iteration is not required to satisfy a stringent primal feasibility condition. In section 4, we conduct some preliminary numerical experiments to evaluate the practical performance of our proposed inexact APG algorithms for solving QSDP problems (1) arising from H-weighted nearest correlation matrix problems. We also evaluate the performance of the proposed algorithms on randomly generated QSDP problems for which the map $\mathcal{Q}$ takes the form as in the stiffness matrix estimation problem in [17].

2   An inexact accelerated proximal gradient method

For more generality, we consider the following minimization problem

$$\min\{F(x) := f(x) + g(x) : x \in \mathcal{X}\} \tag{7}$$

where $\mathcal{X}$ is a finite-dimensional Hilbert space. The functions $f : \mathcal{X} \to \mathcal{R}$, $g : \mathcal{X} \to \mathcal{R} \cup \{+\infty\}$ are proper, lower semi-continuous convex functions (possibly nonsmooth). We assume that $\text{dom}(g) := \{x \in \mathcal{X} : g(x) < \infty\}$ is closed, f is continuously differentiable on $\mathcal{X}$ and its gradient ∇f is Lipschitz continuous with modulus L on $\mathcal{X}$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \quad \forall \, x, y \in \mathcal{X}.$$

We also assume that the problem (7) is solvable with an optimal solution $x_* \in \text{dom}(g)$. The inexact APG algorithm we propose for solving (7) is described as follows.

Algorithm 1. Given a tolerance $\varepsilon > 0$. Input $y_1 = x_0 \in \text{dom}(g)$, $t_1 = 1$. Set $k = 1$. Iterate the following steps.

Step 1. Find an approximate minimizer

$$x_k \approx \arg\min_{y \in \mathcal{X}} \left\{ f(y_k) + \langle \nabla f(y_k),\, y - y_k \rangle + \frac{1}{2}\langle y - y_k,\, \mathcal{H}_k(y - y_k)\rangle + g(y)\right\}, \qquad (8)$$

where $\mathcal{H}_k$ is a self-adjoint positive definite linear operator that is chosen by the user.

Step 2. Compute $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$.

Step 3. Compute $y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})$.

Notice that Algorithm 1 is an inexact version of the algorithm FISTA in [1], where x_k need not be the exact minimizer of the subproblem (8). In addition, the quadratic term is not restricted to the form $\frac{L_k}{2}\|y - y_k\|^2$ where L_k is a positive scalar.

Given any positive definite linear operator $\mathcal{H}_j : \mathcal{X} \to \mathcal{X}$, and $y_j \in \mathcal{X}$, we define $q_j(\cdot) : \mathcal{X} \to \mathcal{R}$ by

$$q_j(x) = f(y_j) + \langle \nabla f(y_j),\, x - y_j \rangle + \frac{1}{2}\langle x - y_j,\, \mathcal{H}_j(x - y_j)\rangle. \qquad (9)$$

Note that if we choose $\mathcal{H}_j = L\mathcal{I}$, then we have $f(x) \leq q_j(x)$ for all $x \in \text{dom}(g)$.

Let $\{\xi_k\}, \{\epsilon_k\}$ be given convergent sequences of nonnegative numbers such that

$$\sum_{k=1}^{\infty} \xi_k < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \epsilon_k < \infty.$$

Suppose for each j, we have an approximate minimizer:

$$x_j \approx \arg\min\{q_j(x) + g(x) : x \in \mathcal{X}\} \qquad (10)$$

that satisfies the conditions

$$F(x_j) \leq q_j(x_j) + g(x_j) + \frac{\xi_j}{2t_j^2}, \qquad (11)$$

$$\nabla f(y_j) + \mathcal{H}_j(x_j - y_j) + \gamma_j = \delta_j \quad \text{with } \|\mathcal{H}_j^{-1/2}\delta_j\| \leq \epsilon_j/(\sqrt{2}t_j) \qquad (12)$$

where $\gamma_j \in \partial g(x_j; \frac{\xi_j}{2t_j^2})$ (the set of $\frac{\xi_j}{2t_j^2}$-subgradients of g at x_j). Note that for x_j to be an approximate minimizer, we must have $x_j \in \text{dom}(g)$. We should mention that the condition (11) is usually easy to satisfy. For example, if $\mathcal{H}_j$ is chosen such that $f(x) \leq q_j(x)$ for all $x \in \text{dom}(g)$, then (11) is automatically satisfied.

To establish the iteration complexity result analogous to the one in [1] for Algorithm 1, we need to establish a series of lemmas whose proofs are extensions of those in [1] to account for the inexactness in x_k. We should note that though the ideas in the proofs are similar, but as the reader will notice later, the technical details become much more involved due to the error terms induced by the inexact solutions of the subproblems.

Lemma 2.1. *Given $y_j \in \mathcal{X}$ and a positive definite linear operator $\mathcal{H}_j$ on $\mathcal{X}$ such that the conditions (11) and (12) hold. Then for any $x \in \mathcal{X}$, we have*

$$F(x) - F(x_j) \geq \frac{1}{2}\langle x_j - y_j, \, \mathcal{H}_j(x_j - y_j)\rangle + \langle y_j - x, \, \mathcal{H}_j(x_j - y_j)\rangle + \langle \delta_j, \, x - x_j\rangle - \frac{\xi_j}{t_j^2}. \tag{13}$$

Proof. By condition (11), we have

$$F(x) - F(x_j) \geq F(x) - q_j(x_j) - g(x_j) - \xi_j/(2t_j^2)$$

$$= g(x) - g(x_j) + f(x) - f(y_j) - \langle \nabla f(y_j), \, x_j - y_j\rangle - \frac{1}{2}\langle x_j - y_j, \, \mathcal{H}_j(x_j - y_j)\rangle - \xi_j/(2t_j^2).$$

By the convexity of f and the definition of γ_j, we have

$$f(x) \geq f(y_j) + \langle \nabla f(y_j), \, x - y_j\rangle, \quad g(x) \geq g(x_j) + \langle \gamma_j, \, x - x_j\rangle - \xi_j/(2t_j^2).$$

Hence

$$F(x) - F(x_j)$$

$$\geq \langle \gamma_j, \, x - x_j\rangle + \langle \nabla f(y_j), \, x - y_j\rangle - \langle \nabla f(y_j), \, x_j - y_j\rangle - \frac{1}{2}\langle x_j - y_j, \, \mathcal{H}_j(x_j - y_j)\rangle - \xi_j/t_j^2$$

$$= \langle \gamma_j + \nabla f(y_j), \, x - x_j\rangle - \frac{1}{2}\langle x_j - y_j, \, \mathcal{H}_j(x_j - y_j)\rangle - \xi_j/t_j^2$$

$$= \langle \delta_j - \mathcal{H}_j(x_j - y_j), \, x - x_j\rangle - \frac{1}{2}\langle x_j - y_j, \, \mathcal{H}_j(x_j - y_j)\rangle - \xi_j/t_j^2.$$

From here, the required inequality (13) follows readily. $\qquad\qquad\square$

For later purpose, we define the following quantities:

$$v_k = F(x_k) - F(x_*) \geq 0, \quad u_k = t_k x_k - (t_k - 1)x_{k-1} - x_*, \tag{14}$$

$$a_k = t_k^2 v_k \geq 0, \quad b_k = \tfrac{1}{2}\langle u_k, \, \mathcal{H}_k(u_k)\rangle \geq 0, \quad e_k = t_k\langle \delta_k, \, u_k\rangle \tag{15}$$

$$\tau = \tfrac{1}{2}\langle x_0 - x_*, \, \mathcal{H}_1(x_0 - x_*)\rangle, \quad \bar{\epsilon}_k = \sum_{j=1}^k \epsilon_j, \quad \bar{\xi}_k = \sum_{j=1}^k (\xi_j + \epsilon_j^2). \tag{16}$$

Note that for the choice where $\epsilon_j = 1/j^\alpha = \xi_j$ for all $j \geq 1$, where $\alpha > 1$ is fixed, we have

$$\bar{\epsilon}_k \leq \frac{1}{\alpha - 1}, \quad \bar{\xi}_k \leq \frac{3}{2}\frac{1}{\alpha - 1} \;\forall\, k \geq 1.$$

7

Lemma 2.2. *Suppose that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ for all k. Then*

$$a_{k-1} + b_{k-1} \geq a_k + b_k - e_k - \xi_k. \tag{17}$$

Proof. By applying the inequality (13) to $x = x_{k-1} \in \operatorname{dom}(g)$ with $j = k$, we get

$$v_{k-1} - v_k = F(x_{k-1}) - F(x_k) \tag{18}$$

$$\geq \frac{1}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + \langle y_k - x_{k-1}, \mathcal{H}_k(x_k - y_k)\rangle + \langle \delta_k, x_{k-1} - x_k \rangle - \xi_k/t_k^2.$$

Similarly, by applying the inequality (13) to $x = x_* \in \operatorname{dom}(g)$ with $j = k$, we get

$$-v_k = F(x_*) - F(x_k)$$

$$\geq \frac{1}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + \langle y_k - x_*, \mathcal{H}_k(x_k - y_k)\rangle + \langle \delta_k, x_* - x_k \rangle - \xi_k/t_k^2. \tag{19}$$

By multiplying (18) throughout by $t_k - 1$ (note that $t_k \geq 1$ for all $k \geq 1$) and adding that to (19), we get

$$(t_k - 1)v_{k-1} - t_k v_k$$

$$\geq \frac{t_k}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + \langle t_k y_k - (t_k - 1)x_{k-1} - x_*, \mathcal{H}_k(x_k - y_k)\rangle$$

$$- \langle \delta_k, t_k x_k - (t_k - 1)x_{k-1} - x_* \rangle - \xi_k/t_k. \tag{20}$$

Now, by multiplying (20) throughout by t_k and using the fact that $t_{k-1}^2 = t_k(t_k - 1)$, we have

$$a_{k-1} - a_k = t_{k-1}^2 v_{k-1} - t_k^2 v_k$$

$$\geq \frac{t_k^2}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + t_k\langle t_k y_k - (t_k - 1)x_{k-1} - x_*, \mathcal{H}_k(x_k - y_k)\rangle$$

$$- \langle \delta_k, t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_* \rangle - \xi_k.$$

Let $\mathbf{a} = t_k y_k$, $\mathbf{b} = t_k x_k$, and $\mathbf{c} = (t_k - 1)x_{k-1} + x_*$. By using the fact that $\langle \mathbf{b} - \mathbf{a}, \mathcal{H}_k(\mathbf{b} - \mathbf{a})\rangle + 2\langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{a})\rangle = \langle \mathbf{b} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{c})\rangle - \langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{a} - \mathbf{c})\rangle$, we get

$$a_{k-1} - a_k \geq \frac{1}{2}\langle \mathbf{b} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{c})\rangle - \frac{1}{2}\langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{a} - \mathbf{c})\rangle$$

$$- \langle \delta_k, t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_* \rangle - \xi_k. \tag{21}$$

Now $\mathbf{a} - \mathbf{c} = t_k y_k - \mathbf{c} = t_k x_{k-1} + (t_{k-1} - 1)(x_{k-1} - x_{k-2}) - \mathbf{c} = u_{k-1}$, $\mathbf{b} - \mathbf{c} = u_k$, and $t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_* = t_k u_k$. Thus (21) implies that

$$a_{k-1} - a_k \geq \frac{1}{2}\langle u_k, \mathcal{H}_k(u_k)\rangle - \frac{1}{2}\langle u_{k-1}, \mathcal{H}_k(u_{k-1})\rangle - \langle \delta_k, t_k u_k \rangle - \xi_k$$

$$\geq b_k - b_{k-1} - \langle \delta_k, t_k u_k \rangle - \xi_k. \tag{22}$$

8

Note that in deriving (22), we have used the fact that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k$. $\qquad \square$

Lemma 2.3. *Suppose that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ for all k. Then*

$$a_k \leq (\sqrt{\tau} + \bar{\epsilon}_k)^2 + 2\bar{\xi}_k. \tag{23}$$

Proof. Note that we have $|e_k| \leq \|\mathcal{H}_k^{-1/2}\delta_k\|\|\mathcal{H}_k^{1/2}u_k\|t_k \leq \epsilon_k \|\mathcal{H}_k^{1/2}u_k\|/\sqrt{2} = \epsilon_k\sqrt{b_k}$.

First, we show that $a_1 + b_1 \leq \tau + \epsilon_1\sqrt{b_1} + \xi_1$. Note that $a_1 = F(x_1) - F(x_*)$ and $b_1 = \frac{1}{2}\langle x_1 - x_*, \mathcal{H}_1(x_1 - x_*)\rangle$. By applying the inequality (13) to $x = x_*$ with $j = 1$, and noting that $y_1 = x_0$, we have that

$$
\begin{aligned}
-a_1 \;\geq\; & \frac{1}{2}\langle x_1 - y_1, \mathcal{H}_1(x_1 - y_1)\rangle + \langle y_1 - x_*, \mathcal{H}_1(x_1 - y_1)\rangle + \langle \delta_1, x_* - x_1\rangle - \xi_1 \\
=\; & \frac{1}{2}\langle x_1 - x_*, \mathcal{H}_1(x_1 - x_*)\rangle - \frac{1}{2}\langle y_1 - x_*, \mathcal{H}_1(y_1 - x_*)\rangle + \langle \delta_1, x_* - x_1\rangle - \xi_1 \\
=\; & b_1 - \frac{1}{2}\langle x_0 - x_*, \mathcal{H}_1(x_0 - x_*)\rangle + \langle \delta_1, x_* - x_1\rangle - \xi_1.
\end{aligned}
$$

Hence, by using the fact that $\|\mathcal{H}_1^{-1/2}\delta_1\| \leq \epsilon_1/\sqrt{2}$, we get

$$a_1 + b_1 \;\leq\; \frac{1}{2}\langle x_0 - x_*, \mathcal{H}_1(x_0 - x_*)\rangle - \langle \delta_1, x_* - x_1\rangle + \xi_1 \;\leq\; \tau + \epsilon_1\sqrt{b_1} + \xi_1. \tag{24}$$

Let

$$s_k = \epsilon_1\sqrt{b_1} + \cdots + \epsilon_k\sqrt{b_k} + \xi_1 + \cdots + \xi_k.$$

By Lemma 2.2, we have

$$
\begin{aligned}
\tau \;\geq\; & a_1 + b_1 - \epsilon_1\sqrt{b_1} - \xi_1 \;\geq\; a_2 + b_2 - \epsilon_1\sqrt{b_1} - \epsilon_2\sqrt{b_2} - \xi_1 - \xi_2 \\
\geq\; & \cdots \;\geq\; a_k + b_k - s_k. \tag{25}
\end{aligned}
$$

Thus we have $a_k + b_k \leq \tau + s_k$, and hence

$$s_k = s_{k-1} + \epsilon_k\sqrt{b_k} + \xi_k \;\leq\; s_{k-1} + \epsilon_k\sqrt{\tau + s_k} + \xi_k. \tag{26}$$

Note that since $\tau \geq b_1 - \epsilon_1\sqrt{b_1} - \xi_1$, we have $\sqrt{b_1} \leq \frac{1}{2}(\epsilon_1 + \sqrt{\epsilon_1^2 + 4(\tau + \xi_1)}) \leq \epsilon_1 + \sqrt{\tau + \xi_1}$.
Hence $s_1 = \epsilon_1\sqrt{b_1} + \xi_1 \leq \epsilon_1(\epsilon_1 + \sqrt{\tau + \xi_1}) + \xi_1 \leq \epsilon_1^2 + \xi_1 + \epsilon_1(\sqrt{\tau} + \sqrt{\xi_1})$.

The inequality (26) implies that

$$(\tau + s_k) - \epsilon_k\sqrt{\tau + s_k} - (\tau + s_{k-1} + \xi_k) \leq 0.$$

Hence we must have

$$\sqrt{\tau + s_k} \;\leq\; \frac{1}{2}\left(\epsilon_k + \sqrt{\epsilon_k^2 + 4(\tau + s_{k-1} + \xi_k)}\right).$$

9

Consequently

$$
\begin{aligned}
s_k &\leq s_{k-1} + \frac{1}{2}\epsilon_k^2 + \xi_k + \frac{1}{2}\epsilon_k\sqrt{\epsilon_k^2 + 4(\tau + s_{k-1} + \xi_k)} \\
&\leq s_{k-1} + \epsilon_k^2 + \xi_k + \epsilon_k\left(\sqrt{\tau} + \sqrt{s_{k-1} + \xi_k}\right).
\end{aligned}
$$

This implies that

$$
\begin{aligned}
s_k &\leq s_1 + \sum_{j=2}^{k}\epsilon_j^2 + \sum_{j=2}^{k}\xi_j + \sqrt{\tau}\sum_{j=2}^{k}\epsilon_j + \sum_{j=2}^{k}\epsilon_j\sqrt{s_{j-1} + \xi_j} \\
&\leq \bar{\xi}_k + \sqrt{\tau}\,\bar{\epsilon}_k + \sum_{j=1}^{k}\epsilon_j\sqrt{s_j} \\
&\leq \bar{\xi}_k + \sqrt{\tau}\,\bar{\epsilon}_k + \sqrt{s_k}\,\bar{\epsilon}_k. \tag{27}
\end{aligned}
$$

In the last inequality, we used the fact that $s_{j-1} + \xi_j \leq s_j$ and $0 \leq s_1 \leq \cdots \leq s_k$. The inequality (27) implies that

$$
\sqrt{s_k} \leq \tfrac{1}{2}\left(\bar{\epsilon}_k + (\bar{\epsilon}_k^2 + 4\bar{\xi}_k + 4\bar{\epsilon}_k\sqrt{\tau})^{1/2}\right).
$$

From here, we get $s_k \leq \bar{\epsilon}_k^2 + 2\bar{\xi}_k + 2\bar{\epsilon}_k\sqrt{\tau}$, and the required result follows from the fact that $a_k \leq \tau + s_k$ in (25). $\qquad\square$

Now we are ready to state the iteration complexity result for the inexact APG algorithm described in Algorithm 1.

Theorem 2.1. *Suppose the conditions (11) and (12) hold, and $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ for all k. Then*

$$
0 \leq F(x_k) - F(x_*) \leq \frac{4}{(k+1)^2}\left((\sqrt{\tau} + \bar{\epsilon}_k)^2 + 2\bar{\xi}_k\right). \tag{28}
$$

Proof. By Lemma 2.3 and the fact that $t_k \geq (k+1)/2$, we have

$$
F(x_k) - F(x_*) = a_k/t_k^2 \leq \frac{4}{(k+1)^2}((\sqrt{\tau} + \bar{\epsilon}_k)^2 + 2\bar{\xi}_k).
$$

From the assumption on the sequences $\{\xi_k\}$ and $\{\epsilon_k\}$, we know that both $\{\bar{\epsilon}_k\}$ and $\{\bar{\xi}_k\}$ are bounded. Then the required convergent complexity result follows. $\qquad\square$

Observe that in Theorem 2.1, we will recover the complexity result established in [1] if $\epsilon_j = 0 = \xi_j$ for all j.

10

2.1 Specialization to the case where $g = \delta_\Omega$

For the problem (P), it can be expressed in the form (7) with $g = \delta_\Omega$, where δ_Ω denotes the indicator function on the set

$$\Omega = \{x \in \mathcal{S}^n : \mathcal{A}(x) = b, \, x \succeq 0\}. \tag{29}$$

The sub-problem (8), for a fixed y_k, then becomes the following constrained minimization problem:

$$\min \left\{ \langle \nabla f(y_k), \, x - y_k \rangle + \frac{1}{2} \langle x - y_k, \, \mathcal{H}_k(x - y_k) \rangle \, : \, \mathcal{A}(x) = b, x \succeq 0 \right\}. \tag{30}$$

Suppose we have an approximate solution (x_k, p_k, z_k) to the KKT optimality conditions for (30). More specifically,

$$\nabla f(y_k) + \mathcal{H}_k(x_k - y_k) - \mathcal{A}^* p_k - z_k =: \delta_k \;\; \approx \;\; 0$$
$$\mathcal{A}(x_k) - b \;\; = \;\; 0 \tag{31}$$
$$\langle x_k, \, z_k \rangle =: \varepsilon_k \;\; \approx \;\; 0, \quad x_k, z_k \succeq 0.$$

To apply the complexity result established in Theorem 2.1, we need δ_k and ε_k to be sufficiently small so that the conditions (11) and (12) are satisfied. Observe that we need x_k to be contained in Ω in (31). Note that the first equation in (31) is the feasibility condition for the dual problem of (30), and it corresponds to the condition in (12) with $\gamma_k = -\mathcal{A}^* p_k - z_k$. Indeed, as we shall show next, γ_k is an ε_k-subgradient of g at $x_k \in \Omega$ if $z_k \succeq 0$. Now, given any $v \in \Omega$, we need to show that $g(v) \geq g(x_k) + \langle \gamma_k, \, v - x_k \rangle - \varepsilon_k$. We have $g(v) = 0$, $g(x_k) = 0$ since $v, x_k \in \Omega$, and

$$\langle \gamma_k, \, v - x_k \rangle \;\; = \;\; \langle \mathcal{A}^* p_k + z_k, \, x_k - v \rangle = \langle p_k, \, \mathcal{A}(x_k) - \mathcal{A}(v) \rangle + \langle z_k, \, x_k \rangle - \langle z_k, \, v \rangle$$
$$= \;\; \langle z_k, \, x_k \rangle - \langle z_k, \, v \rangle \;\; \leq \;\; \langle z_k, \, x_k \rangle \;\; = \;\; \varepsilon_k. \tag{32}$$

Note that in deriving (32), we used the fact that $\langle z_k, \, v \rangle \geq 0$ since $v \succeq 0$ and $z_k \succeq 0$. Thus the condition (12) is satisfied if $\|\mathcal{H}_k^{-1/2} \delta_k\| \leq \epsilon_k/(\sqrt{2} t_k)$ and $\varepsilon_k \leq \xi_k/(2t_k^2)$.

As we have already noted in the last paragraph, the approximate solution x_k obtained by solving the sub-problem (30) should be feasible, i.e. $x_k \in \Omega$. In practice we can maintain the positive semidefiniteness of x_k by performing projection onto $\mathcal{S}_+^n$. But the residual vector $r_k := \mathcal{A}(x_k) - b$ is usually not exactly equal to 0. In the following paragraph, we will propose a strategy to find a feasible solution $\tilde{x}_k \in \Omega$ given an approximate solution x_k of (31) for which r_k is not necessarily 0, but (x_k, p_k, z_k) satisfies that conditions that $x_k \succeq 0, z_k \succeq 0$, and $\|\mathcal{H}_k^{-1/2} \delta_k\| \leq \frac{1}{2} \epsilon_k/(\sqrt{2} t_k)$ and $\varepsilon_k \leq \frac{1}{2} \xi_k/(2t_k^2)$.

Suppose that there exists $\bar{x} \succ 0$ such that $\mathcal{A}(\bar{x}) = b$. Since $\mathcal{A}$ is surjective, $\mathcal{A}\mathcal{A}^*$ is nonsingular. Let $\omega_k = -\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}(r_k)$. We note that $\|\omega_k\|_2 \leq \|r_k\|/\sigma_{\min}(\mathcal{A})$, and

$\mathcal{A}(x_k + \omega_k) = b$, where $\|\cdot\|_2$ denotes the spectral norm. However, $x_k + \omega_k$ may not be positive semidefinite. Thus we consider the following iterate:

$$\tilde{x}_k = \lambda(x_k + \omega_k) + (1 - \lambda)\bar{x} = \lambda x_k + (\lambda\omega_k + (1 - \lambda)\bar{x}),$$

where $\lambda \in [0, 1]$. It is clear that $\mathcal{A}\tilde{x}_k = b$. By choosing $\lambda = 1 - \|\omega_k\|_2/(\|\omega_k\|_2 + \lambda_{\min}(\bar{x}))$, we can guarantee that $\tilde{x}_k$ is positive semidefinite. For $\tilde{x}_k$, we have

$$
\begin{aligned}
0 \leq \langle \tilde{x}_k, z_k \rangle &\leq \lambda\varepsilon_k + \lambda\sqrt{n}\|\omega_k\|_2\|z_k\| + \frac{\|\omega_k\|_2}{\|\omega_k\|_2 + \lambda_{\min}(\bar{x})}\sqrt{n}\lambda_{\max}(\bar{x})\|z_k\| \\
&\leq \varepsilon_k + \sqrt{n}\|\omega_k\|_2\|z_k\| + \sqrt{n}\frac{\|\omega_k\|_2}{\lambda_{\min}(\bar{x})}\lambda_{\max}(\bar{x})\|z_k\| \\
&\leq 2\varepsilon_k, \quad \text{if } \|\omega_k\|_2 \leq \frac{\varepsilon_k}{\sqrt{n}\|z_k\|}\left(1 + \frac{\lambda_{\max}(\bar{x})}{\lambda_{\min}(\bar{x})}\right)^{-1}.
\end{aligned}
$$

Moreover

$$\nabla f(y_k) + \mathcal{H}_k(\tilde{x}_k - y_k) - (\mathcal{A}^*p_k + z_k) = \delta_k + \mathcal{H}_k(\tilde{x}_k - x_k) =: \tilde{\delta}_k$$

Thus $\gamma_k = -\mathcal{A}^*p_k - z_k$ is an $2\varepsilon_k$-subgradient of g at $\tilde{x}_k \in \Omega$. Now $\|\mathcal{H}_k^{-1/2}\tilde{\delta}_k\| \leq \|\mathcal{H}_k^{-1/2}\delta_k\| + \|\mathcal{H}_k^{1/2}(\tilde{x}_k - x_k)\|$, and

$$\|\mathcal{H}_k^{1/2}(\tilde{x}_k - x_k)\|^2 = \langle \tilde{x}_k - x_k, \mathcal{H}_k(\tilde{x}_k - x_k)\rangle \leq n\|\omega_k\|_2^2\lambda_{\max}(H_1)\left(1 + \frac{\|\bar{x} - x_k\|_2}{\lambda_{\min}(\bar{x})}\right)^2.$$

Thus we have

$$\|\mathcal{H}_k^{-1/2}\tilde{\delta}_k\| \leq \epsilon_k/(\sqrt{2}t_k) \quad \text{if } \|w_k\|_2 \leq \frac{\epsilon_k}{2\sqrt{2n}\,t_k}(\lambda_{\max}(\mathcal{H}_1))^{-1/2}\left(1 + \frac{\|\bar{x} - x_k\|_2}{\lambda_{\min}(\bar{x})}\right)^{-1}.$$

To conclude $(\tilde{x}_k, p_k, z_k)$ would satisfy the condition (12) if

$$\|\omega_k\|_2 \leq \min\left\{\frac{\xi_k}{4t_k^2\sqrt{n}\|z_k\|}\left(1 + \frac{\lambda_{\max}(\bar{x})}{\lambda_{\min}(\bar{x})}\right)^{-1}, \frac{\epsilon_k}{2\sqrt{2n\lambda_{\max}(\mathcal{H}_1)}\,t_k}\left(1 + \frac{\|\bar{x} - x_k\|_2}{\lambda_{\min}(\bar{x})}\right)^{-1}\right\}. \quad (33)$$

We should note that even though we have succeeded in constructing a feasible $\tilde{x}_k$ in Ω. The accuracy requirement in (33) could be too stringent for computational efficiency. For example, when $\sigma_{\min}(\mathcal{A})$ is small, or $\|z_k\|$ is large, or $\bar{x}$ has a large condition number, or $\lambda_{\max}(\mathcal{H}_1)$ is large, we would expect that x_k must be computed to rather high accuracy so that $\|r_k\|$ is small enough for (33) to be satisfied.

3 Analysis of an inexact APG method for (P)

To apply Algorithm 1 to solve the problem (P), the requirement that x_k must be primal feasible, i.e., $x_k \in \Omega$, can be restrictive as it limits our flexibility of choosing a non-primal feasible algorithm for solving (30). Even though the modification outlined in last paragraph of section 2.1 is able to produce a primal feasible $\tilde{x}_k$, the main drawback is that the residual norm $\|\omega_k\|$ must satisfy the stringent accuracy condition in (33). To overcome the drawbacks just mentioned, here we propose an inexact APG algorithm for solving (P) for which the iterate x_k need not be strictly contained in Ω. As the reader will observe later, the analysis of the iteration complexity of the proposed inexact APG becomes even more challenging than the analysis done in the previous section.

We let (x_*, p_*, z_*) be an optimal solution of (P) and (D). In this section, we let

$$q_k(x) = f(y_k) + \langle \nabla f(y_k),\ x - y_k \rangle + \frac{1}{2}\langle x - y_k,\ \mathcal{H}_k(x - y_k) \rangle, \quad x \in \mathcal{X}. \tag{34}$$

Note that $\mathcal{X} = \mathcal{S}^n$. The inexact APG algorithm we propose for solving (P) is given as follows.

Algorithm 2. Given a tolerance $\varepsilon > 0$. Input $y_1 = x_0 \in \mathcal{X}$, $t_1 = 1$. Set $k = 1$. Iterate the following steps.

Step 1. Find an approximate minimizer

$$x_k \approx \arg\min_{x \in \mathcal{X}} \left\{ q_k(x)\ :\ x \in \Omega \right\}, \tag{35}$$

where $\mathcal{H}_k$ is a self-adjoint positive definite operator that is chosen by the user, and x_k is allowed to be contained in a suitable enlargement Ω_k of Ω.

Step 2. Compute $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$.

Step 3. Compute $y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}} \right)(x_k - x_{k-1})$.

Note that when $\Omega_k = \Omega$, the dual problem of (35) is given by

$$\max \left\{ q_k(x) - \langle \nabla q_k(x),\ x \rangle + \langle b,\ p \rangle \mid \nabla q_k(x) - \mathcal{A}^* p - z = 0,\ z \succeq 0, x \succeq 0 \right\}. \tag{36}$$

Let $\{\xi_k\}, \{\epsilon_k\}, \{\mu_k\}$ be given convergent sequences of nonnegative numbers such that

$$\sum_{k=1}^{\infty} \xi_k < \infty, \quad \sum_{k=1}^{\infty} \epsilon_k < \infty, \quad \text{and } \sum_{k=1}^{\infty} \mu_k < \infty,$$

and Δ be a given positive number. We assume that the approximate minimizer x_k in (35) has the property that x_k and its corresponding dual variables (p_k, z_k) satisfy the following conditions:

$$f(x_k) \;\leq\; q_k(x_k) + \xi_k/(2t_k^2)$$

$$|\langle \nabla q_k(x_k),\, x_k \rangle - \langle b,\, p_k \rangle| \leq \Delta$$

$$\nabla q_k(x_k) - \mathcal{A}^* p_k - z_k \;=\; \delta_k, \text{ with } \|H_k^{-1/2}\delta_k\| \leq \epsilon_k/(\sqrt{2}t_k) \tag{37}$$

$$\|r_k\| \;\leq\; \mu_k/t_k^2$$

$$\langle x_k,\, z_k \rangle \;\leq\; \xi_k/(2t_k^2)$$

$$x_k \succeq 0, \; z_k \succeq 0,$$

where $r_k := \mathcal{A}(x_k) - b$. We assume that $\mu_k/t_k^2 \geq \mu_{k+1}/t_{k+1}^2$ and $\epsilon_k/t_k \geq \epsilon_{k+1}/t_{k+1}$ for all k. Observe that the last five conditions in (37) stipulate that (x_k, p_k, z_k) is an approximate optimal solution of (35) and (36).

Just as in the previous section, we need to establish a series of lemmas to analyse the iteration complexity of Algorithm 2. However, we should mention that the lack of feasibility in x_k (i.e., x_k may not be contained in Ω) introduces nontrivial technical difficulties in the proof of the complexity result for Algorithm 2. For example, $F(x_k) \geq F(x_*)$ no longer hold as in the feasible case when $x_k \in \Omega$.

Lemma 3.1. *Given $y_j \in \mathcal{X}$ and a positive definite linear operator $\mathcal{H}_j$ on $\mathcal{X}$ such that the conditions in (37) hold. Then for any $x \in \mathcal{S}_+^n$, we have*

$$f(x) - f(x_j) \;\geq\; \frac{1}{2}\langle x_j - y_j,\, \mathcal{H}_j(x_j - y_j) \rangle + \langle y_j - x,\, \mathcal{H}_j(x_j - y_j) \rangle$$
$$+ \langle \delta_j + \mathcal{A}^* p_j,\, x - x_j \rangle - \xi_j/t_j^2. \tag{38}$$

Proof. Since $f(x_j) \leq q_j(x_j) + \xi_j/(2t_j^2)$, we have

$$f(x) - f(x_j) \geq f(x) - q_j(x_j) - \xi_j/(2t_j^2)$$

$$= f(x) - f(y_j) - \langle \nabla f(y_j),\, x_j - y_j \rangle - \frac{1}{2}\langle x_j - y_j,\, \mathcal{H}_j(x_j - y_j) \rangle - \xi_j/(2t_j^2)$$

$$\geq \langle \nabla f(y_j),\, x - x_j \rangle - \frac{1}{2}\langle x_j - y_j,\, \mathcal{H}_j(x_j - y_j) \rangle - \xi_j/(2t_j^2).$$

Note that in the last inequality, we have used the fact that $f(x) - f(y_j) \geq \langle \nabla f(y_j),\, x - y_j \rangle$

14

for all $x \in \mathcal{X}$. Now, by using (37), we get

$$f(x) - f(x_j)$$

$$\geq \langle \nabla f(y_j), x - x_j \rangle - \frac{1}{2}\langle x_j - y_j, \mathcal{H}_j(x_j - y_j) \rangle - \xi_j/(2t_j^2)$$

$$= \langle \delta_j + \mathcal{A}^* p_j - \mathcal{H}_j(x_j - y_j), x - x_j \rangle + \langle z_j, x - x_j \rangle - \frac{1}{2}\langle x_j - y_j, \mathcal{H}_j(x_j - y_j) \rangle - \xi_j/(2t_j^2)$$

$$\geq \langle \delta_j + \mathcal{A}^* p_j - \mathcal{H}_j(x_j - y_j), x - x_j \rangle - \frac{1}{2}\langle x_j - y_j, \mathcal{H}_j(x_j - y_j) \rangle - \xi_j/t_j^2.$$

From here, the required inequality (38) follows readily. Note that in deriving the last inequality, we have used the fact that $\langle z_j, x_j \rangle \leq \xi_j/(2t_j^2)$ and $\langle z_j, x \rangle \geq 0$. $\qquad\square$

For later purpose, we define the following quantities for $k \geq 1$:

$$v_k = f(x_k) - f(x_*), \quad u_k = t_k x_k - (t_k - 1)x_{k-1} - x_*,$$

$$a_k = t_k^2 v_k, \quad b_k = \tfrac{1}{2}\langle u_k, \mathcal{H}_k(u_k) \rangle \geq 0, \quad e_k = t_k\langle \delta_k, u_k \rangle, \tag{39}$$

$$\eta_k = \langle p_k, t_k^2 r_k - t_{k-1}^2 r_{k-1} \rangle, \text{ with } \eta_1 = \langle p_1, r_1 \rangle,$$

$$\chi_k = \|p_{k-1} - p_k\|\mu_{k-1}, \text{ with } \chi_1 = 0,$$

$$\bar{\epsilon}_k = \textstyle\sum_{j=1}^k \epsilon_j, \quad \bar{\xi}_k = \sum_{j=1}^k (\xi_j + \epsilon_j^2), \quad \bar{\chi}_k = \sum_{j=1}^k \chi_j,$$

$$\tau = \tfrac{1}{2}\langle x_0 - x_*, \mathcal{H}_1(x_0 - x_*) \rangle.$$

Note that unlike the analysis in the previous section, a_k may be negative.

Lemma 3.2. *Suppose that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ for all k. Then*

$$a_{k-1} + b_{k-1} \geq a_k + b_k - e_k - \xi_k - \eta_k. \tag{40}$$

Proof. By applying the inequality (38) to $x = x_{k-1} \succeq 0$ with $j = k$, we get

$$v_{k-1} - v_k = f(x_{k-1}) - f(x_k) \tag{41}$$

$$\geq \frac{1}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k) \rangle + \langle y_k - x_{k-1}, \mathcal{H}_k(x_k - y_k) \rangle + \langle \delta_k + \mathcal{A}^* p_k, x_{k-1} - x_k \rangle - \xi_k/t_k^2.$$

Similarly, by applying the inequality (38) to $x = x_* \succeq 0$ with $j = k$, we get

$$-v_k = f(x_*) - f(x_k) \tag{42}$$

$$\geq \frac{1}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k) \rangle + \langle y_k - x_*, \mathcal{H}_k(x_k - y_k) \rangle + \langle \delta_k + \mathcal{A}^* p_k, x_* - x_k \rangle - \xi_k/t_k^2.$$

By multiplying (41) throughout by $t_k - 1$ (note that $t_k \geq 1$ for all $k \geq 1$) and adding that to (42), we get

$$(t_k - 1)v_{k-1} - t_k v_k$$

$$\geq \frac{t_k}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + \langle t_k y_k - (t_k - 1)x_{k-1} - x_*, \mathcal{H}_k(x_k - y_k)\rangle$$

$$- \langle \delta_k + \mathcal{A}^* p_k, t_k x_k - (t_k - 1)x_{k-1} - x_*\rangle - \xi_k/t_k. \tag{43}$$

Now, by multiplying (20) throughout by t_k and using the fact that $t_{k-1}^2 = t_k(t_k - 1)$, we get

$$a_{k-1} - a_k = t_{k-1}^2 v_{k-1} - t_k^2 v_k$$

$$\geq \frac{t_k^2}{2}\langle x_k - y_k, \mathcal{H}_k(x_k - y_k)\rangle + t_k\langle t_k y_k - (t_k - 1)x_{k-1} - x_*, \mathcal{H}_k(x_k - y_k)\rangle$$

$$- \langle \delta_k + \mathcal{A}^* p_k, t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_*\rangle - \xi_k.$$

Let $\mathbf{a} = t_k y_k$, $\mathbf{b} = t_k x_k$, and $\mathbf{c} = (t_k - 1)x_{k-1} + x_*$. By using the fact that $\langle \mathbf{b} - \mathbf{a}, \mathcal{H}_k(\mathbf{b} - \mathbf{a})\rangle + 2\langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{a})\rangle = \langle \mathbf{b} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{c})\rangle - \langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{a} - \mathbf{c})\rangle$, we get

$$a_{k-1} - a_k \geq \frac{1}{2}\langle \mathbf{b} - \mathbf{c}, \mathcal{H}_k(\mathbf{b} - \mathbf{c})\rangle - \frac{1}{2}\langle \mathbf{a} - \mathbf{c}, \mathcal{H}_k(\mathbf{a} - \mathbf{c})\rangle$$

$$- \langle \delta_k + \mathcal{A}^* p_k, t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_*\rangle - \xi_k. \tag{44}$$

Now $\mathbf{a} - \mathbf{c} = t_k y_k - \mathbf{c} = t_k x_{k-1} + (t_{k-1} - 1)(x_{k-1} - x_{k-2}) - \mathbf{c} = u_{k-1}$, $\mathbf{b} - \mathbf{c} = u_k$, and $t_k^2 x_k - t_{k-1}^2 x_{k-1} - t_k x_* = t_k u_k$. Thus (44) implies that

$$a_{k-1} - a_k \geq \frac{1}{2}\langle u_k, \mathcal{H}_k(u_k)\rangle - \frac{1}{2}\langle u_{k-1}, \mathcal{H}_k(u_{k-1})\rangle - \langle \delta_k + \mathcal{A}^* p_k, t_k u_k\rangle - \xi_k$$

$$\geq b_k - b_{k-1} - \langle \delta_k, t_k u_k\rangle - \langle p_k, \mathcal{A}(t_k u_k)\rangle - \xi_k. \tag{45}$$

Note that in deriving (45), we have used the fact that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k$. Now

$$\langle p_k, \mathcal{A}(t_k u_k)\rangle = \langle p_k, t_k^2(\mathcal{A}x_k - b) - t_{k-1}^2(\mathcal{A}x_{k-1} - b)\rangle = \langle p_k, t_k^2 r_k - t_{k-1}^2 r_{k-1}\rangle.$$

From here, the required result is proved. $\qquad\square$

Lemma 3.3. *Suppose that $\mathcal{H}_{k-1} \succeq \mathcal{H}_k \succ 0$ and the conditions in (37) are satisfied for all k. Then*

$$a_k + b_k \leq (\sqrt{\tau} + \bar{\epsilon}_k)^2 + \|p_k\|\mu_k + 2(\bar{\xi}_k + \bar{\chi}_k + \omega_k) \tag{46}$$

where $\omega_k = \sum_{j=1}^k \epsilon_j \sqrt{A_j}$, and

$$A_j = \|p_j\|\mu_j + a_j^-, \quad \text{with } a_j^- = \max\{0, -a_j\}.$$

Proof. Note that we have $|e_k| \leq \|\mathcal{H}_k^{-1/2}\delta_k\|\|\mathcal{H}_k^{1/2}u_k\|t_k \leq \epsilon_k\|\mathcal{H}_k^{1/2}u_k\|/\sqrt{2} = \epsilon_k\sqrt{b_k}$.

First, we show that $a_1 + b_1 \leq \tau + |\langle p_1, r_1\rangle| + \epsilon_1\sqrt{b_1} + \xi_1$. Note that $a_1 = f(x_1) - f(x_*)$ and $b_1 = \frac{1}{2}\langle x_1 - x_*, \mathcal{H}_1(x_1 - x_*)\rangle$. By applying the inequality (38) to $x = x_*$ with $j = 1$, and noting that $y_1 = x_0$, we have that

$$
\begin{aligned}
-a_1 &\geq \frac{1}{2}\langle x_1 - y_1, \mathcal{H}_1(x_1 - y_1)\rangle + \langle y_1 - x_*, \mathcal{H}_1(x_1 - y_1)\rangle + \langle \delta_1 + \mathcal{A}^*p_1, x_* - x_1\rangle - \xi_1 \\
&= \frac{1}{2}\langle x_1 - x_*, \mathcal{H}_1(x_1 - x_*)\rangle - \frac{1}{2}\langle y_1 - x_*, \mathcal{H}_1(y_1 - x_*)\rangle + \langle \delta_1 + \mathcal{A}^*p_1, x_* - x_1\rangle - \xi_1 \\
&= b_1 - \frac{1}{2}\langle x_0 - x_*, \mathcal{H}_1(x_0 - x_*)\rangle + \langle \delta_1 + \mathcal{A}^*p_1, x_* - x_1\rangle - \xi_1.
\end{aligned}
$$

Hence, by using the fact that $\|\mathcal{H}_1^{-1/2}\delta_1\| \leq \epsilon_1/\sqrt{2}$, we get

$$
\begin{aligned}
a_1 + b_1 &\leq \frac{1}{2}\langle x_0 - x_*, \mathcal{H}_1(x_0 - x_*)\rangle - \langle \delta_1 + \mathcal{A}^*p_1, x_* - x_1\rangle + \xi_1 \\
&\leq \tau + \epsilon_1\sqrt{b_1} + \langle p_1, r_1\rangle + \xi_1 \leq \tau + |\langle p_1, r_1\rangle| + \epsilon_1\sqrt{b_1} + \xi_1.
\end{aligned}
$$

Let $s_1 = \epsilon_1\sqrt{b_1} + \xi_1$ and for $k \geq 2$,

$$
s_k = \sum_{j=1}^{k}\epsilon_j\sqrt{b_j} + \sum_{j=1}^{k}\xi_j + \sum_{j=1}^{k}\chi_j.
$$

By Lemma 3.2, we have

$$
\begin{aligned}
\tau &\geq a_1 + b_1 - \epsilon_1\sqrt{b_1} - \xi_1 - \eta_1 \\
&\geq a_2 + b_2 - \epsilon_2\sqrt{b_2} - \epsilon_1\sqrt{b_1} - \xi_1 - \xi_2 - \eta_1 - \eta_2 \\
&\geq \cdots \\
&\geq a_k + b_k - \sum_{j=1}^{k}\epsilon_j\sqrt{b_j} - \sum_{j=1}^{k+1}\xi_j - \sum_{j=1}^{k}\eta_j \\
&\geq a_k + b_k - \sum_{j=1}^{k}\epsilon_j\sqrt{b_j} - \sum_{j=1}^{k}\xi_j - |\langle p_k, t_k^2 r_k\rangle| - \sum_{j=1}^{k}\chi_j.
\end{aligned}
$$

Note that in the last inequality, we used the fact that

$$
\sum_{j=1}^{k}\eta_j = \langle p_k, t_k^2 r_k\rangle + \sum_{j=1}^{k-1}\langle p_j - p_{j+1}, t_j^2 r_j\rangle \leq |\langle p_k, t_k^2 r_k\rangle| + \sum_{j=1}^{k}\chi_j.
$$

Thus we have $a_k + b_k \leq \tau + |\langle p_k, t_k^2 r_k\rangle| + s_k$, and this implies that

$$
b_k \leq \tau_k + s_k \quad \text{where } \tau_k := \tau + |\langle p_k, t_k^2 r_k\rangle| - a_k \leq \tau + A_k. \tag{47}
$$

17

Hence

$$s_k = s_{k-1} + \epsilon_k \sqrt{b_k} + \xi_k + \chi_k \leq s_{k-1} + \epsilon_k \sqrt{\tau_k + s_k} + \xi_k + \chi_k. \tag{48}$$

Note that since $\tau_1 \geq b_1 - \epsilon_1 \sqrt{b_1} - \xi_1$, we have $\sqrt{b_1} \leq \frac{1}{2}(\epsilon_1 + \sqrt{\epsilon_1^2 + 4(\tau_1 + \xi_1)}) \leq \epsilon_1 + \sqrt{\tau_1 + \xi_1}$. Hence $s_1 = \epsilon_1 \sqrt{b_1} + \xi_1 \leq \epsilon_1(\epsilon_1 + \sqrt{\tau_1 + \xi_1}) + \xi_1 \leq \epsilon_1^2 + \xi_1 + \epsilon_1(\sqrt{\tau_1} + \sqrt{\xi_1})$.

The inequality (48) implies that

$$(\tau_k + s_k) - \epsilon_k \sqrt{\tau_k + s_k} - \left(\tau_k + s_{k-1} + \xi_k + \chi_k\right) \leq 0.$$

Hence we must have

$$\sqrt{\tau_k + s_k} \leq \frac{1}{2}\left(\epsilon_k + \sqrt{\epsilon_k^2 + 4(\tau_k + s_{k-1} + \xi_k + \chi_k)}\right).$$

Consequently, by using the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for any $a, b \geq 0$, we have

$$\begin{aligned}
s_k &\leq s_{k-1} + \frac{1}{2}\epsilon_k^2 + \xi_k + \chi_k + \frac{1}{2}\epsilon_k \sqrt{\epsilon_k^2 + 4(\tau_k + s_{k-1} + \xi_k + \chi_k)} \\
&\leq s_{k-1} + \frac{1}{2}\epsilon_k^2 + \xi_k + \chi_k + \frac{1}{2}\epsilon_k \sqrt{\epsilon_k^2 + 4(\tau + A_k + s_{k-1} + \xi_k + \chi_k)} \\
&\leq s_{k-1} + \epsilon_k^2 + \xi_k + \chi_k + \epsilon_k\left(\sqrt{\tau + A_k} + \sqrt{s_{k-1} + \xi_k + \chi_k}\right). \tag{49}
\end{aligned}$$

This implies that

$$\begin{aligned}
s_k &\leq s_1 + \sum_{j=2}^{k}(\epsilon_j^2 + \xi_j + \chi_j) + \sum_{j=2}^{k} \epsilon_j \sqrt{\tau + A_j} + \sum_{j=2}^{k} \epsilon_j \sqrt{s_{j-1} + \xi_j + \chi_j} \\
&\leq \bar{\xi}_k + \bar{\chi}_k + \sum_{j=1}^{k} \epsilon_j \sqrt{\tau + A_j} + \sum_{j=1}^{k} \epsilon_j \sqrt{s_j} \\
&\leq \bar{\xi}_k + \bar{\chi}_k + \omega_k + \bar{\epsilon}_k \sqrt{\tau} + \bar{\epsilon}_k \sqrt{s_k}. \tag{50}
\end{aligned}$$

In the last inequality, we used the fact that $s_{j-1} + \xi_j + \chi_j \leq s_j$, and $0 \leq s_1 \leq \cdots \leq s_k$. The inequality (50) implies that

$$\sqrt{s_k} \leq \frac{1}{2}\left(\bar{\epsilon}_k + \sqrt{\bar{\epsilon}_k^2 + 4\theta_k}\right), \tag{51}$$

where $\theta_k = \bar{\xi}_k + \bar{\chi}_k + \omega_k + \bar{\epsilon}_k \sqrt{\tau}$. From here, we get

$$s_k \leq \bar{\epsilon}_k^2 + 2\theta_k. \tag{52}$$

The required result follows from (52) and the fact that $a_k + b_k \leq \tau + s_k + |\langle p_k, t_k^2 r_k \rangle| \leq \tau + s_k + \|p_k\|\mu_k$. $\qquad \square$

Let

$$\Omega_k := \{x \in \mathcal{S}^n : \|\mathcal{A}(x) - b\| \leq \mu_k/t_k^2, \ x \succeq 0\} \tag{53}$$

and

$$x_*^k := \operatorname{argmin}\{f(x) : x \in \Omega_k\}. \tag{54}$$

Since $x_*, x_k \in \Omega_k$, we have $f(x_*) \geq f(x_*^k)$ and $f(x_k) \geq f(x_*^k)$. Hence $v_k = f(x_k) - f(x_*) \leq f(x_k) - f(x_*^k)$. Also, since $\mu_k/t_k^2 \geq \mu_{k+1}/t_{k+1}^2$, we have $f(x_*^{k+1}) \geq f(x_*^k)$ and $\Omega_{k+1} \subseteq \Omega_k$.

Lemma 3.4. *For all $k \geq 1$, we have*

$$0 \leq f(x_*) - f(x_*^k) \leq \|p_*\|\mu_k/t_k^2. \tag{55}$$

Proof. By the convexity of f, we have

$$\begin{aligned}
f(x_*) - f(x_*^k) &\leq \langle \nabla f(x_*), x_* - x_*^k \rangle = \langle \mathcal{A}^* p_* + z_*, x_* - x_*^k \rangle \\
&= \langle p_*, \mathcal{A}(x_*) - \mathcal{A}(x_*^k) \rangle + \langle z_*, x_* \rangle - \langle z_*, x_*^k \rangle \\
&\leq \|p_*\|\|b - \mathcal{A}(x_*^k)\| \leq \|p_*\|\mu_k/t_k^2.
\end{aligned}$$

Note that in deriving the second last inequality, we have used the fact that $\langle z_*, x_* \rangle = 0$, $\langle z_*, x_*^k \rangle \geq 0$, and $\mathcal{A}(x_*) = b$. $\qquad \square$

Theorem 3.1. *Suppose $M_k = \max_{1 \leq j \leq k}\{\sqrt{(\|p_*\| + \|p_j\|)\mu_j}\}$. Then we have*

$$-\frac{4\|p_*\|\mu_k}{(k+1)^2} \leq f(x_k) - f(x_*) \leq \frac{4}{(k+1)^2}\left((\sqrt{\tau} + \bar{\epsilon}_k)^2 + \|p_k\|\mu_k + 2\bar{\epsilon}_k M_k + 2(\bar{\xi}_k + \bar{\chi}_k)\right). \tag{56}$$

Proof. The inequality on the left-hand side of (56) follows from Lemma 3.4 and the fact that $t_k \geq (k+1)/2$ and $f(x_*^k) - f(x_*) \leq f(x_k) - f(x_*)$. Next, we prove the inequality on the right-hand side of (56). By Lemma 3.3 and noting that $b_k \geq 0$, we have

$$t_k^2(f(x_k) - f(x_*)) = a_k \leq (\sqrt{\tau} + \bar{\epsilon}_k)^2 + \|p_k\|\mu_k + 2(\bar{\xi}_k + \bar{\chi}_k) + 2\omega_k.$$

Now $-a_j = t_j^2(f(x_*) - f(x_j)) \leq t_j^2(f(x_*) - f(x_*^j)) \leq \|p_*\|\mu_j$. Hence $a_j^- \leq \|p_*\|\mu_j$, and

$$\omega_k \leq \sum_{j=1}^k \epsilon_j \sqrt{\|p_j\|\mu_j + \|p_*\|\mu_j} \leq M_k\bar{\epsilon}_k. \tag{57}$$

19

From here, the required result follows. □

From the assumption on the sequences $\{\epsilon_k\}$, $\{\xi_k\}$, and $\{\mu_k\}$, we know that the sequences $\{\bar{\epsilon}_k\}$ and $\{\bar{\xi}_k\}$ are bounded. In order to show that the sequence of function values $f(x_k)$ converges to the optimal function value $f(x_*)$ with the convergent rate $O(1/k^2)$, it is enough to show that the sequence $\{\|p_k\|\}$ is bounded under certain conditions, from which we can also have the boundedness of $\{M_k\}$ and $\{\bar{\chi}_k\}$. Then the desired convergent rate of $O(1/k^2)$ for our inexact APG method follows.

3.1 Boundedness of $\{p_k\}$

In this subsection, we consider sufficient conditions to ensure the boundedness of the sequence $\{p_k\}$.

Lemma 3.5. *Suppose that there exists $(\bar{x}, \bar{p}, \bar{z})$ such that*

$$\mathcal{A}(\bar{x}) = b, \ \bar{x} \succeq 0, \quad \nabla f(\bar{x}) = \mathcal{A}^* \bar{p} + \bar{z}, \ \bar{z} \succ 0. \tag{58}$$

If the sequence $\{f(x_k)\}$ is bounded from above, then the sequence $\{x_k\}$ is bounded.

Proof. By using the convexity of f, we have

$$
\begin{aligned}
f(\bar{x}) - f(x_k) &\leq \langle \nabla f(\bar{x}), \ \bar{x} - x_k \rangle = \langle \mathcal{A}^* \bar{p} + \bar{z}, \ \bar{x} - x_k \rangle \\
&= \langle \bar{p}, \ \mathcal{A}(\bar{x}) - \mathcal{A}(x_k) \rangle + \langle \bar{z}, \ \bar{x} \rangle - \langle \bar{z}, \ x_k \rangle \\
&\leq \|\bar{p}\| \|b - \mathcal{A}(x_k)\| + \langle \bar{z}, \ \bar{x} \rangle - \langle \bar{z}, \ x_k \rangle \\
&\leq \|\bar{p}\| \mu_k / t_k^2 + \langle \bar{z}, \ \bar{x} \rangle - \langle \bar{z}, \ x_k \rangle \leq \|\bar{p}\| \mu_1 + \langle \bar{z}, \ \bar{x} \rangle - \langle \bar{z}, \ x_k \rangle.
\end{aligned}
$$

Thus

$$\lambda_{\min}(\bar{z}) \mathrm{Tr}(x_k) \leq \langle \bar{z}, \ x_k \rangle \leq \|\bar{p}\| \mu_1 + \langle \bar{z}, \ \bar{x} \rangle - f(\bar{x}) + f(x_k). \tag{59}$$

From here, the required result is proved. □

Remark 3.1. *The condition that $\{f(x_k)\}$ is bounded from above appears to be fairly weak. But unfortunately we are not able to prove that this condition holds true. In many cases of interest, such as the nearest correlation matrix problem (3), the condition that $\{f(x_k)\}$ is bounded above or that $\{x_k\}$ is bounded can be ensured since Ω_1 is bounded.*

Lemma 3.6. *Suppose that $\{x_k\}$ is bounded and there exists $\hat{x}$ such that*

$$\mathcal{A}(\hat{x}) = b, \ \hat{x} \succ 0.$$

Then the sequence $\{z_k\}$ is bounded. In addition, the sequence $\{p_k\}$ is also bounded.

20

Proof. From (37), we have

$$
\begin{aligned}
\lambda_{\min}(\hat{x})\mathrm{Tr}(z_k) &\leq \langle \hat{x},\, z_k \rangle = \langle \hat{x},\, \nabla q_k(x_k) - \mathcal{A}^* p_k - \delta_k \rangle \\
&= -\langle b,\, p_k \rangle + \langle \hat{x},\, \nabla q_k(x_k) \rangle - \langle \hat{x},\, \delta_k \rangle \\
&\leq \Delta + \langle \hat{x} - x_k,\, \nabla q_k(x_k) \rangle - \langle \hat{x},\, \delta_k \rangle \\
&= \Delta + \langle \hat{x} - x_k,\, \nabla f(y_k) + \mathcal{H}_k(x_k - y_k) \rangle - \langle \hat{x},\, \delta_k \rangle \\
&\leq \Delta + \|\hat{x} - x_k\| \|\nabla f(y_k) + \mathcal{H}_k(x_k - y_k)\| + \|\mathcal{H}_k^{1/2}\hat{x}\|\epsilon_k/(\sqrt{2}t_k) \\
&\leq \Delta + \|\hat{x} - x_k\| \|\nabla f(y_k) + \mathcal{H}_k(x_k - y_k)\| + \|\mathcal{H}_1^{1/2}\hat{x}\|\epsilon_1/\sqrt{2} \qquad (60)
\end{aligned}
$$

Since $\{x_k\}$ is bounded, it is clear that $\{y_k\}$ is also bounded. By the continuity of ∇f and that fact that $0 \preceq \mathcal{H}_k \preceq \mathcal{H}_1$, the sequence $\{\|\nabla f(y_k) + \mathcal{H}_k(x_k - y_k)\|\}$ is also bounded. From (60), we can now conclude that $\{z_k\}$ is bounded.

Next, we show that $\{p_k\}$ is bounded. Let $\mathcal{A}^\dagger = (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}$. Note that the matrix $\mathcal{A}\mathcal{A}^*$ is nonsingular since $\mathcal{A}$ is assumed to be surjective. From (37), we have $p_k = \mathcal{A}^\dagger(\nabla q_k(x_k) - z_k - \delta_k)$, and hence

$$
\|p_k\| \leq \|\mathcal{A}^\dagger\| \|\nabla q_k(x_k) - z_k - \delta_k\| \leq \|\mathcal{A}^\dagger\|\Big(\|\nabla f(y_k) + \mathcal{H}_k(x_k - y_k)\| + \|z_k\| + \|\delta_k\| \Big).
$$

Since $\mathcal{H}_k \preceq \mathcal{H}_1 \preceq \lambda_{\max}(\mathcal{H}_1)I$, we have $\|\delta_k\| \leq \sqrt{\lambda_{\max}(\mathcal{H}_1)}\|\mathcal{H}_k^{-1/2}\delta_k\| \leq \sqrt{\lambda_{\max}(\mathcal{H}_1)}\epsilon_1/\sqrt{2}$. By using the fact that the sequences $\{\|\nabla f(y_k) + \mathcal{H}_k(x_k - y_k)\|\}$ and $\{z_k\}$ are bounded, the boundedness of $\{p_k\}$ follows. $\qquad \square$

3.2 A semismooth Newton-CG method for solving the inner subproblem (35)

In Section 3, an inexact APG method (Algorithm 2) was presented for solving (P) with the desired convergent rate of $O(1/k^2)$. However, an important issue on how to efficiently solve the inner subproblem (35) has not been addressed.

In this section, we propose the use of a SSNCG method to solve (35) with warm-start using the iterate from the previous iteration. Note that the self-adjoint positive definite linear operator $\mathcal{H}_k$ can be chosen by the user. Suppose that at each iteration we are able to choose a linear operator of the form:

$$
\mathcal{H}_k := w_k \circledast w_k, \quad \text{where } w_k \in \mathcal{S}_{++}^n,
$$

such that $f(x) \leq q_k(x)$ for all $x \in \Omega$. (Note that we can always choose $w_k = \sqrt{L}I$ if there are no other better choices.) Then the objective function $q_k(\cdot)$ in (35) can equivalently be written as

$$
q_k(x) = \frac{1}{2}\|w_k^{1/2}(x - u_k)w_k^{1/2}\|^2 + f(y_k) - \frac{1}{2}\|w_k^{-1/2}\nabla f(y_k)w_k^{-1/2}\|^2,
$$

where $u_k = y_k - w_k^{-1} \nabla f(y_k) w_k^{-1}$. By dropping the last two constant terms in the above equation, we can equivalently write (35) as the following well-studied W-weighted semidefinite least squares problem

$$\min \left\{ \frac{1}{2} \| w_k^{1/2} (x - u_k) w_k^{1/2} \|^2 \; : \; \mathcal{A}(x) = b, x \succeq 0 \right\}. \tag{61}$$

Let $\bar{x} = w_k^{1/2} x w_k^{1/2}$, $\bar{u}_k = w_k^{1/2} u_k w_k^{1/2}$, and define the linear map $\bar{\mathcal{A}} : \mathcal{S}^n \to \mathcal{R}^m$ by

$$\bar{\mathcal{A}}(\bar{x}) = \mathcal{A}(w_k^{-1/2} \bar{x} \, w_k^{-1/2}).$$

Then (61) can equivalently be written as

$$\min \left\{ \frac{1}{2} \| \bar{x} - \bar{u}_k \|^2 \; : \; \bar{\mathcal{A}}(\bar{x}) = b, \; \bar{x} \succeq 0 \right\}, \tag{62}$$

whose Lagrangian dual problem is given by

$$\max \left\{ \theta(p) := b^T p - \frac{1}{2} \| \Pi_{\mathcal{S}_+^n} (\bar{u}_k + \bar{\mathcal{A}}^* p) \|^2 \mid p \in \mathcal{R}^m \right\} \tag{63}$$

where $\Pi_{\mathcal{S}_+^n}(u)$ denotes the metric projection of $u \in \mathcal{S}^n$ onto $\mathcal{S}_+^n$. The problem (63) is an unconstrained continuously differentiable convex optimization problem, and it can be efficiently solved by the SSNCG method developed in [10]. Note that once an approximate solution p_k is computed from (63), an approximate solution for (61) can be computed by $x_k = \Pi_{\mathcal{S}_+^n}(\bar{u}_k + \bar{\mathcal{A}}^* p_k)$ and its complementary dual slack variable is $z_k = \bar{u}_k + \bar{\mathcal{A}}^* p_k - x_k$.

Note that the problem (63) is an unconstrained continuously differentiable convex optimization problem which can also be solved by a gradient ascent method. In our numerical implementation, we use a gradient method to solve (63) during the initial phase of Algorithm 2 when high accuracy solutions are not required. When the gradient method encounters difficulty in solving the subproblem to the required accuracy or becomes inefficient, we switch to the SSNCG method to solve (63). We should note that approximate solution computed for the current subproblem can be used to warm start the SSNCG and gradient methods for solving the next subproblem. In fact, the strategy of solving a semidefinite least squares subproblem (35) in each iteration of our inexact APG algorithm is practically viable precisely because we are able to warm start the SSNCG or gradient methods when solving the subproblems. In our numerical experience, the SSNCG method would typically take less than 5 Newton steps to solve each subproblem with warm start.

To successfully apply the SSNCG method to solve (35), we must find a suitable symmetrized Kronecker product approximation of $\mathcal{Q}$. Note that for the H-weighted nearest correlation matrix problem where $\mathcal{Q}$ is a diagonal operator defined by $\mathcal{Q}(x) = (H \circ H) \circ x$, a positive definite symmetrized Kronecker product approximation for $\mathcal{Q}$ can be derived as follows. Consider a rank-one approximation dd^T of $H \circ H$, then $\mathrm{diag}(d) \circledast \mathrm{diag}(d)$ is a symmetrized Kronecker product approximation of $\mathcal{Q}$. For the vector $d \in \mathcal{R}^n$, one can simply take

$$d_j = \max \left\{ \epsilon, \max_{1 \le i \le n} \{ H_{ij} \} \right\}, \; j = 1, \ldots, n, \tag{64}$$

where $\epsilon > 0$ is a small positive number.

For the convex QSDP problem (1) where the linear operator $\mathcal{Q}$ is defined by

$$\mathcal{Q}(x) = B \circledast I(x) = (Bx + xB)/2, \qquad (65)$$

where $B \in \mathcal{S}_+^n$, we propose the following strategy for constructing a suitable symmetrized Kronecker product approximation of $\mathcal{Q} = B \circledast I$. Suppose we have the eigenvalue decomposition $B = P\Lambda P^T$, where $\Lambda = \mathrm{diag}(\lambda)$ and $\lambda = (\lambda_1, \dots, \lambda_n)^T$ is the vector of eigenvalues of B. Then

$$\langle x, B \circledast I(x) \rangle = \frac{1}{2}\langle \hat{x}, \Lambda\hat{x} + \hat{x}\Lambda \rangle = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{x}_{ij}(\lambda_i + \lambda_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}\hat{x}_{ij}M_{ij},$$

where $\hat{x} = P^T x P$ and $M = \frac{1}{2}(\lambda e^T + e\lambda^T)$ with $e \in \mathcal{R}^n$ being the vector of all ones. For the choice of w_k, one may simply choose $w_k = \sqrt{\max(M)}I$, where $\max(M)$ is the largest element of M. However, if the matrix B is ill-conditioned, this choice of w_k may not work very well in practice since $\max(M)I \circledast I$ may not be a good approximation of $\mathcal{Q} = B \circledast I$. To find a better approximation of $\mathcal{Q}$, we propose to consider the following nonconvex minimization problem:

$$\min\left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \mid h_i h_j - M_{ij} \geq 0 \ \forall \ i,j = 1,\dots,n, \ h \in \mathcal{R}_+^n \right\}. \qquad (66)$$

Thus if $\hat{h}$ is a feasible solution to the above problem, then we have

$$\langle x, B \circledast I(x) \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n}\hat{x}_{ij}M_{ij} \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\hat{x}_{ij}\hat{h}_i\hat{h}_j = \langle x, w_k x w_k \rangle \quad \forall \ x \in \mathcal{S}^n$$

with $w_k = P\mathrm{diag}(\hat{h})P^T$. Note that the above strategy can also be used to get a suitable symmetrized Kronecker product approximation of the form $\mathrm{diag}(d) \circledast \mathrm{diag}(d)$ when $\mathcal{Q}$ is a diagonal operator.

To find a good feasible solution for (66), we consider the following strategy. Suppose we are given an initial vector $u \in \mathcal{R}_+^n$ such that $uu^T - M \geq 0$. For example, we may take $u = \sqrt{\max(M)}e$. Our purpose is to find a correction vector $\xi \in \mathcal{R}_+^n$ such that $h := u - \xi$ satisfies the constraint in (66) while the objective value is reduced. Note that we have

$$h_i h_j - M_{ij} = u_i u_j - M_{ij} - (u_i\xi_j + u_j\xi_i) + \xi_i\xi_j \geq u_i u_j - M_{ij} - (u_i\xi_j + u_j\xi_i).$$

Thus the constraints in (66) are satisfied if $\xi \leq u$ and

$$u_i\xi_j + u_j\xi_i \leq u_i u_j - M_{ij} \quad \forall \ i,j = 1,\dots,n.$$

Since $\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j = (e^T \xi)^2 - 2(e^T u)(e^T \xi) + (e^T u)^2$, and noting that $0 \leq e^T \xi \leq e^T u$, the objective value in (66) is minimized if $e^T \xi$ is maximized. Thus we propose to consider the following LP:

$$\max \left\{ e^T \xi \mid u_i \xi_j + u_j \xi_i \leq u_i u_j - M_{ij} \, \forall \, i, j = 1, \ldots, n, \, 0 \leq \xi \leq u \right\}. \tag{67}$$

Observe that the correction vector ξ depends on the given vector u. Thus if necessary, after a new u is obtained, one may repeat the process by solving the LP associated with the new u.

4   Numerical Experiments

In this section, we report the performance of the inexact APG algorithm (Algorithm 2) for large scale linearly constrained QSDP problems.

We measure the infeasibilities for the primal and dual problems (1) and (2) as follows:

$$R_P = \frac{\|b - \mathcal{A}(x)\|}{1 + \|b\|}, \quad R_D = \frac{\|\mathcal{Q}(x) + c - \mathcal{A}^* p - z\|}{1 + \|c\|}, \tag{68}$$

where x, p, z are computed from (63). In our numerical experiments, we stop the inexact APG algorithm when

$$\max\{R_P, R_D\} \leq \text{Tol}, \tag{69}$$

where Tol is a pre-specified accuracy tolerance. Unless otherwise specified, we set $\text{Tol} = 10^{-6}$ as the default tolerance. When solving the subproblem (63) at iteration k of our inexact APG method, we stop the SSNCG or gradient method when $\|\nabla \theta(p_k)\|/(1 + \|b\|) < \min\{1/t_k^{3.1}, 0.2\|\nabla f(x_{k-1}) - \mathcal{A}^* p_{k-1} - z_{k-1}\|/(1 + \|c\|)\}$.

4.1   Example 1

In this example, we consider the following H-weighted nearest correlation matrix problem

$$\min \left\{ \frac{1}{2} \|H \circ (x - u)\|^2 \mid \text{diag}(x) = e, x \succeq 0 \right\}. \tag{70}$$

We compare the performance of our inexact APG (IAPG) method and the augmented Lagrangian dual method (AL) studied by Qi and Sun in [11], whose MATLAB codes are available at http://www.math.nus.edu.sg/~matsundf. We consider the gene correlation matrices $\widehat{u}$ from [6]. For testing purpose we perturb $\widehat{u}$ to

$$u := (1 - \alpha)\widehat{u} + \alpha E,$$

where $\alpha \in (0, 1)$ and E is a randomly generated symmetric matrix with entries in $[-1, 1]$. We also set $u_{ii} = 1, i = 1, \ldots, n$. The weight matrix H is a sparse random symmetric

matrix with about 50% nonzero entries. The MATLAB code for generating H and E is as follows:

```
H = sprand(n,n,0.5); H = triu(H) + triu(H,1)'; H = (H + H')/2;
    E = 2*rand(n,n)-1; E = triu(E)+triu(E,1)'; E = (E+E')/2.
```

In order to generate a good initial point, we use the SSNCG method in [10] to solve the following unweighted nearest correlation matrix problem

$$\min\left\{\frac{1}{2}\|x - u\|^2 \mid \text{diag}(x) = e, x \succeq 0\right\}. \tag{71}$$

Due to the difference in stopping criteria for different algorithms, we set different accuracy tolerances for the IAPG and augmented Lagrangian methods. For the IAPG method, we set $\texttt{Tol} = 10^{-6}$. For the augmented Lagrangian method, its stopping criteria depends on a tolerance parameter $\texttt{Tol1}$ which controls the three conditions in the KKT system (31). We set $\texttt{Tol1} = 10^{-4}$.

Table 1 presents the numerical results obtained by the IAPG method and the augmented Lagrangian dual method (AL) for various instances of Example 1. We use the primal infeasibility, primal objective value and computing time to compare the performance of the two algorithms. For each instance in the table, we report the matrix dimension (n), the noise level (α), the number of outer iterations (iter), the total number of Newton systems solved (newt) the primal infeasibility (R_P), the dual infeasibility (R_D), the primal objective value (pobj) in (70), as well as the computation time (in the format hours:minutes:seconds) and the rank of the computed solution (rankX). We may observe from the table that the IAPG method can solve (70) very efficiently. For each instance, the IAPG method can achieve nearly the same primal objective value as the augmented Lagrangian method, and the former can achieve much better primal infeasibility while taking less than 50% of the time needed by the augmented Lagrangian method.

4.2   Example 2

We consider the same problem as in Example 1, but the weight matrix H is generated from a weight matrix H_0 used by a hedge fund company. The matrix H_0 is a 93×93 symmetric matrix with all positive entries. It has about 24% of the entries equal to 10^{-5} and the rest are distributed in the interval $[2, 1.28 \times 10^3]$. It has 28 eigenvalues in the interval $[-520, -0.04]$, 11 eigenvalues in the interval $[-5 \times 10^{-13}, 2 \times 10^{-13}]$, and the rest of 54 eigenvalues in the interval $[10^{-4}, 2 \times 10^4]$. The MATLAB code for generating the matrix H is given by $\texttt{tmp = kron(ones(25,25),H0); H = tmp([1:n],[1:n]); H = (H+H')/2}$.

We use the same implementation techniques as in Example 1. The stopping tolerance for the IAPG method is set to $\texttt{Tol} = 10^{-6}$ while the tolerance for the augmented Lagrangian method is set to $\texttt{Tol1} = 10^{-2}$. Table 2 presents the numerical results obtained

Algo.	problem	n	α	iter/newt	R_P	R_D	pobj	time	rankX
IAPG	Lymph	587	0.1	169/177	3.09e-11	1.75e-6	5.04289799e+0	2:35	179
			0.05	300/327	1.31e-10	2.04e-6	2.53103607e-1	4:16	207
AL	Lymph	587	0.1	12	4.13e-7	9.96e-7	5.04289558e+0	5:39	179
			0.05	12	2.96e-7	1.07e-6	2.53101698e-1	30:58	207
IAPG	ER	692	0.1	137/142	2.27e-10	2.43e-6	1.26095534e+1	3:10	189
			0.05	187/207	3.93e-11	9.54e-7	1.14555927e+0	3:40	220
AL	ER	692	0.1	12	3.73e-7	4.63e-7	1.26095561e+1	9:28	189
			0.05	12	3.21e-7	1.02e-6	1.14555886e+0	14:14	220
IAPG	Arabidopsis	834	0.1	115/123	3.28e-10	1.78e-6	3.46252363e+1	3:53	191
			0.05	131/148	2.41e-10	9.75e-7	5.50148194e+0	4:09	220
AL	Arabidopsis	834	0.1	13	2.28e-7	7.54e-7	3.46252429e+1	12:35	191
			0.05	12	2.96e-8	1.01e-6	5.50148169e+0	22:49	220
IAPG	Leukemia	1255	0.1	104/111	5.35e-10	7.97e-7	1.08939600e+2	9:24	254
			0.05	96/104	4.81e-10	9.31e-7	2.20789464e+1	8:35	276
AL	Leukemia	1255	0.1	12	3.06e-7	2.74e-7	1.08939601e+2	22:04	254
			0.05	11	2.90e-7	8.57e-7	2.20789454e+1	28:37	276
IAPG	hereditarybc	1869	0.1	67/87	2.96e-10	8.68e-7	4.57244497e+2	17:56	233
			0.05	64/85	9.58e-10	7.04e-7	1.13171325e+2	17:32	236
AL	hereditarybc	1869	0.1	13	2.31e-7	3.55e-7	4.57244525e+2	38:35	233
			0.05	11	2.51e-7	6.29e-7	1.13171335e+2	36:31	236

Table 1: Comparison of the inexact APG (IAPG) and augmented Lagrangian dual (AL) algorithms on (70) using sample correlation matrix from gene data sets. The weight matrix H is a sparse random matrix with about 50% nonzero entries.

Algo.	problem	n	α	iter/newt	R_P	R_D	pobj	time	rankX
IAPG	Lymph	587	0.1	72/159	1.76e-8	9.90e-7	8.92431024e+6	1:50	238
			0.05	60/148	3.81e-8	9.75e-7	1.69947194e+6	1:41	278
AL	Lymph	587	0.1	14	2.64e-5	1.06e-5	8.92425480e+6	56:07	260
			0.05	12	1.69e-4	4.41e-5	1.69925778e+6	29:15	286
IAPG	ER	692	0.1	62/156	2.48e-9	9.72e-7	1.51144194e+7	2:33	254
			0.05	56/145	3.58e-9	9.55e-7	3.01128282e+6	2:22	295
AL	ER	692	0.1	16	1.22e-5	5.80e-6	1.51144456e+7	2:05:38	288
			0.05	12	3.11e-5	6.29e-6	3.01123631e+6	53:15	309
IAPG	Arabidopsis	834	0.1	61/159	6.75e-9	9.98e-7	2.69548461e+7	4:01	254
			0.05	54/145	1.06e-8	9.82e-7	5.87047119e+6	3:41	286
AL	Arabidopsis	834	0.1	19	3.04e-6	3.94e-6	2.69548769e+7	4:49:00	308
			0.05	13	1.69e-5	6.76e-6	5.87044318e+6	1:28:59	328
IAPG	Leukemia	1255	0.1	65/158	8.43e-9	9.86e-7	7.17192454e+7	11:32	321
			0.05	55/143	1.19e-7	9.80e-7	1.70092540e+7	10:18	340
AL	Leukemia	1255	0.1	$*$	$*$	$*$	$*$	$*$	$*$
			0.05	13	3.19e-5	5.15e-6	1.70091646e+7	5:55:21	432
IAPG	hereditarybc	1869	0.1	48/156	2.08e-8	9.16e-7	2.05907938e+8	29:07	294
			0.05	49/136	6.39e-8	9.61e-7	5.13121563e+7	26:16	297
AL	hereditarybc	1869	0.1	$*$	$*$	$*$	$*$	$*$	$*$
			0.05	$*$	$*$	$*$	$*$	$*$	$*$

Table 2: Same as Table 1, but with a "bad" weight matrix H.

by the IAPG and augmented Lagrangian dual (AL) methods. In the table, " $*$ " means that the augmented Lagrangian method cannot achieve the required tolerance of 10^{-2} in 24 hours. As we can see from Table 2, the IAPG method is much more efficient than the augmented Lagrangian method, and it can achieve much better primal infeasibility. For the last gene correlation matrix of size 1869, the IAPG method can find an good approximate solution within half an hour. For the augmented Lagrangian method, because the map $\mathcal{Q}$ associated with the weight matrix H is highly ill-conditioned, the CG method has great difficulty in solving the ill-conditioned linear system of equations obtained by the semismooth Newton method.

4.3 Example 3

In this example, we report the performance of the inexact APG on the linearly constrained QSDP problem (1). The linear operator $\mathcal{Q}$ is given by $\mathcal{Q}(x) = \frac{1}{2}(Bx + xB)$ for a given $B \succ 0$, and the linear map $\mathcal{A}$ is given by $\mathcal{A}(x) = \text{diag}(x)$. We generate a positive definite matrix X and set $b = \mathcal{A}(x)$. Similarly we can generate a random vector $p \in \mathcal{R}^m$ and a positive definite matrix z and set $c = \mathcal{A}^*(p) + z - \mathcal{Q}(x)$. The MATLAB code for generating the matrix B is given by `randvec = 1+ 9*rand(n,1); tmp = randn(n,ceil(n/4)); B = diag(randvec)+(tmp*tmp')/n; B = (B+B')/2`. Note that the matrix B generated is rather well conditioned.

n	m	cond(B)	iter/newt	R_P	R_D	pobj	dobj	time
500	500	9.21e+0	9/9	3.24e-10	9.70e-7	-4.09219187e+4	-4.09219188e+4	13
1000	1000	9.43e+0	9/9	3.68e-10	9.28e-7	-8.41240999e+4	-8.41241006e+4	1:13
2000	2000	9.28e+0	9/9	3.16e-10	8.53e-7	-1.65502323e+5	-1.65502325e+5	8:49
2500	2500	9.34e+0	9/9	3.32e-10	8.57e-7	-2.07906307e+5	-2.07906309e+5	16:15
3000	3000	9.34e+0	9/9	2.98e-10	8.13e-7	-2.49907743e+5	-2.49907745e+5	29:02

Table 3: Numerical results of the inexact APG algorithm on (1), where the positive definite matrix B for the linear operator $\mathcal{Q}$ is well-conditioned.

As discussed in section 3.2, we are able to find a good symmetrized Kronecker product approximation $w \circledast w$ of $\mathcal{Q}$. By noting that

$$\frac{1}{2}\langle x, w \circledast w(x)\rangle + \langle c, x\rangle = \frac{1}{2}\|w^{1/2}(x-u)w^{1/2}\|^2 - \frac{1}{2}\|w^{-1/2}cw^{-1/2}\|^2,$$

where $u = -w^{-1}cw^{-1}$, and dropping the constant term, we propose to solve the following problem to generate a good initial point for the inexact APG method:

$$\min\left\{\frac{1}{2}\|w^{1/2}(x-u)w^{1/2}\|^2 \mid \mathcal{A}(x) = b, \ x \succeq 0\right\},$$

which can be efficiently solved by the the SSNCG method in [10].

The performance results of our IAPG method on convex QSDP problems are given in Table 3, where "pobj" and "dobj" are the primal and dual objective values for QSDP, respectively. We may see from the table that the IAPG method can solve all the five instances of QSDP problems very efficiently with very good primal infeasibility.

4.4 Example 4

We consider the same problem as Example 3 but the linear map $\mathcal{A}$ is generated by using the first generator in [8] with order $p = 3$. The positive definite matrix B is generated by using MATLAB's built-in function: `B = gallery('lehmer',n)`. The condition number cond(B) of the generated Lehmer matrix B is within the range $[n, 4n^2]$. For this example, the simple choice of $w = \sqrt{\lambda_{\max}(B)}I$ in the symmetrized Kronecker product $w \circledast w$ for approximating $\mathcal{Q}$ does not work well. In our numerical implementation, we employ the strategy described in section 3.2 to find a suitable w.

Table 4.4 presents the performance results of our IAPG method on convex QSDP problems where the matrix B is very ill-conditioned. As observed from the table, the condition numbers of B are large. We may see from the table that the IAPG method can solve the problem very efficiently with very accurate approximate optimal solution.

n	m	cond(B)	iter/newt	R_P	R_D	pobj	dobj	time
500	10000	2.67e+5	51/102	3.02e-8	9.79e-7	-9.19583895e+3	-9.19584894e+3	1:29
1000	50000	1.07e+6	62/115	2.43e-8	9.71e-7	-1.74777588e+4	-1.74776690e+4	11:46
2000	100000	4.32e+6	76/94	5.24e-9	5.28e-7	-3.78101950e+4	-3.78101705e+4	1:14:04
2500	100000	6.76e+6	80/96	4.62e-9	5.64e-7	-4.79637904e+4	-4.79637879e+4	2:11:01

Table 4: Same as Table 3, but the matrix B for the linear operator $\mathcal{Q}$ is ill-conditioned and the linear map $\mathcal{A}$ is randomly generated as in [8].

References

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

[2] V. Chandrasekaran, B. Recht, P.A. Parrilo, and A.S. Willsky, *The convex geometry of linear inverse problems*, preprint, December, 2010.

[3] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[4] O. Devolder, F. Glineur, and Y. Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, preprint, 2011.

[5] N.J. Higham, *Computing the nearest correlation matrix—a problem from finance*, IMA Journal of Numerical Analysis, 22 (2002), 329–343.

[6] L. Li and K.C. Toh, *An inexact interior point method for L1-regularized sparse co-variance selection*, Mathematical Programming Computation, 2 (2010), 291–315.

[7] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM Journal on Matrix Analysis and Applications, 26 (2004), 272–284.

[8] J. Malick, J. Povh, F. Rendl, and A. Wiegele, *Regularization methods for semidefinite programming*, SIAM Journal on Optimization, 20 (2009), 336–356.

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

[10] H. Qi and D.F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM J. on Matrix Analysis and Applications, 28 (2006), 360–385.

[11] H. Qi and D.F. Sun, *An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem*, IMA Journal of Numerical Analysis, 31 (2010), 491–511.

[12] K.C. Toh, *An inexact primal-dual path-following algorithm for convex quadratic SDP*, Mathematical Programming, 112 (2008), 221–254.

[13] K.C. Toh, R.H. Tutuncu, and M.J. Todd, *Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems*, Pacific Journal of Optimization, 3 (2007), 135–164.

[14] K.C. Toh and S.W. Yun,*An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, Pacific Journal of Optimization 6 (2010), 615–640.

[15] P. Tseng, *On Accelerated Proximal Gradient Methods for Convex-Concave Optimization*, submitted to SIAM Journal on Optimization, 2008.

[16] C.J. Wang, D.F. Sun, and K.C. Toh, *Solving log-determinant optimization problems by a Newton-CG primal proximal point algorithm*, SIAM Journal on Optimization, 20 (2010), 2994–3013.

[17] K.G. Woodgate, *Efficient stiffness matrix estimation for elastic structures*, Computers and Structures, 69 (1998), pp. 79–84.

[18] X.Y. Zhao, *A Semismooth Newton-CG Augmented Lagrangian Method for Large Scale Linear and Convex Quadratic SDPs*, PhD thesis, National University of Singapore, 2009.