Projection-free accelerated method for convex optimization

Max L.N. Gonçalves * Jefferson G. Melo * Renato D.C. Monteiro †

May 2, 2019 (revised: September 22, 2019)

Abstract

In this paper, we propose a projection-free accelerated method for solving convex optimization problems with unbounded feasible set. The method is an accelerated gradient scheme such that each projection subproblem is approximately solved by means of a conditional gradient scheme. Under reasonable assumptions, it is shown that an ε-approximate solution (concept related to the optimal value of the problem) is obtained in at most $\mathcal{O}(1/\sqrt{\varepsilon})$ gradient evaluations and $\mathcal{O}(1/\varepsilon)$ linear oracle calls. We also discuss a notion of approximate solution based on the first-order optimality condition of the problem and present iteration-complexity results for the proposed method to obtain an approximate solution in this sense. Finally, numerical experiments illustrating the practical behavior of the proposed scheme are discussed.

1 Introduction

The conditional gradient (CondG) method proposed by Frank and Wolfe [4, 5] is designed to solve the convex optimization problem

$$f^* := \min_{x \in X} f(x) \tag{1}$$

where X is a nonempty compact convex set and f is a differentiable convex function such that ∇f is Lipschitz continuous on X. Given $x^{k-1} \in X$, its k-th step first finds y^k as a minimum of the linear function $\langle \nabla f(x^{k-1}), \cdot \rangle$ over X and then set $x^k = (1-\alpha_k)x^{k-1} + \alpha_k y^k$ for some $\alpha_k \in [0,1]$. Its major distinguishing feature compared to other first-order algorithms such as the projected gradient (or accelerated gradient) method is that it replaces the usual projection onto X by a linear oracle which computes y^k as above. Since, for some relevant cases of X, the latter operation is considerably cheaper than the first one, the CondG method is competitive with first-order projection methods for solving large-scale instances of (1) and has received significant attention in different application areas (see [1, 2, 3, 6, 7, 8, 10, 12, 13, 14]). It has been shown that if α_k in the CondG method are properly chosen, then this algorithm can find an ε-solution of (1) (i.e., a point $\hat{x} \in X$ such that $f(\hat{x}) - f^* \leq \varepsilon$) in at most $\mathcal{O}(1/\varepsilon)$ iterations and gradient evaluations (see [6, 10, 13]).

Recently, there has also been a growing interest in the study of accelerated gradient methods for solving large-scale optimization problems; see, for instance, [9, 11, 15, 16, 17, 18]. These methods

*IME, Universidade federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mails: maxlng@ufg.br and jefferson@ufg.br). The work of these authors was supported in part by FAPEG/CNPq/PRONEM-201710267000532 and CNPq Grants 302666/2017-6 and 408123/2018-4.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grants CMMI-1300221 and ONR Grant ONR N00014-18-1-2077.

have optimal iteration-complexity, namely, an ε-approximate solution of (1) (with X not necessarily bounded) can be obtained in at most $\mathcal{O}(1/\sqrt{\varepsilon})$ gradient (as well as projections) evaluations. Motivated by the efficiency of the accelerated methods and taking into account that in some applications the associated projection subproblems are difficult to solve, Lan and Zhou [13] developed a projection-free accelerated gradient scheme for solving (1), called conditional gradient sliding (CGS) method, which uses the CondG method to approximately solve the associated projection subproblems. Note that in this case X needs to be bounded. They showed that the CGS method obtains an ε-approximate solution of (1) in at most $\mathcal{O}(1/\sqrt{\varepsilon})$ gradient evaluations and $\mathcal{O}(1/\varepsilon)$ linear oracle calls. The authors also presented some numerical experiments showing the advantages of the CGS algorithm over the standard CondG method applied directly to (1).

Our main goal in this paper is to develop and analyze a projection-free accelerated method for solving problem (1) when X is unbounded. The proposed scheme follows the same idea of the CGS method in the sense that it is an accelerated gradient method such that its projection subproblems are inexactly solved by means of the CondG method. In these subproblems, the unbounded set X is intercepted with a ball whose radius is iteratively and appropriately updated throughout the whole procedure. We show that an ε-approximate solution of (1) is obtained in at most $\mathcal{O}(1/\sqrt{\varepsilon})$ gradient evaluations and $\mathcal{O}(1/\varepsilon)$ linear oracle calls. We discuss a concept of approximate solution based on the first-order optimality condition for problem (1) and show that the iteration-complexity bounds to obtain such an approximate solution are basically the same as the ones stated above. An advantage of the latter concept is that it can easily be verified during the process of the method. It is worth mentioning that our accelerated scheme is an extension of a CGS variant developed here for solving (1) when X is bounded. Some numerical experiments are presented in order to illustrate the applicability of the general accelerated scheme to solve quadratic optimization problems with unbounded constraints.

The outline of this paper is as follows. Section 2 presents notations and reviews the conditional gradient method for solving a specific problem. This section also discusses a projection-free accelerated method for solving (1) and establishes its convergence rate. Section 3 develops and analyzes a general projection-free method for solving (1) when X is unbounded. A specific implementation of the latter method is studied in Section 4. Section 5 discusses a notion of approximate solution and presents a procedure for computing the initial radius of the general scheme. Section 6 contains some numerical experiments.

2 Notation and background materials

In this section, we present our notations and review the conditional gradient method for solving a specific problem. We also present an accelerated method for solving a convex optimization problem when the feasible set is bounded.

Throughout this paper, $\mathbb{E}$ denotes a finite-dimensional inner product real vector space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. We denote the sets of real numbers by $\mathbb{R}$, nonnegative numbers by $\mathbb{R}_+$ and positive numbers by $\mathbb{R}_{++}$.

Our problem of interest is the convex optimization problem

$$f^* := \min_{x \in X} f(x), \tag{2}$$

where X is a nonempty closed convex subset of $\mathbb{E}$ and $f : \mathbb{E} \to \mathbb{R}$ is a differentiable convex function

such that ∇f is L-Lipschitz continuous on X for some $L \geq 0$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad x,\ y \in X. \tag{3}$$

We assume that the set of optimal solutions X^* of (2) is nonempty. For a given $x^0 \in X$, let us denote by d_0 the distance from x^0 to X^* and

$$l_f(x, y) := f(y) + \langle \nabla f(y), x - y \rangle, \quad x, y \in X.$$

It follows from the convexity of f and (3) that

$$l_f(x, y) \leq f(x) \leq l_f(x, y) + \frac{L}{2}\|x - y\|^2 \quad y \in X. \tag{4}$$

The indicator function $\mathcal{I}_X : \mathbb{E} \to (-\infty, \infty]$ is defined as

$$\mathcal{I}_X(x) = \begin{cases} 0, & x \in X, \\ \infty, & x \notin X. \end{cases}$$

For a scalar $\varepsilon \geq 0$, the ε-subdifferential of a function $f : \mathbb{E} \to (-\infty, \infty]$ is the operator $\partial_\varepsilon f : \mathbb{E} \rightrightarrows \mathbb{E}$ defined as

$$\partial_\varepsilon f(x) = \{v \mid f(y) \geq f(x) + \langle y - x, v \rangle - \varepsilon,\ \forall y \in \mathbb{E}\}, \quad \forall x \in \mathbb{E}.$$

When $\varepsilon = 0$, the operator $\partial_\varepsilon f$ is simply denoted by ∂f and is referred to as the subdifferential of f. We also recall that a point $x^* \in X$ is a solution of (2) if and only if

$$0 \in \partial(f + \mathcal{I}_X)(x^*).$$

2.1 Conditional gradient method

In this section, we review the conditional gradient method (also known as the Frank-Wolfe method) for solving an optimization problem. We present its convergence rate which will play an important role in the derivation of iteration-complexity bounds for proposed accelerated schemes. As is well-known the conditional gradient method assumes that a linear optimization (LO) oracle is available, i.e., a routine which returns an optimal solution to the problem of minimizing a linear form over a nonempty compact convex set.

Let X be a nonempty compact convex subset of $\mathbb{E}$. Consider the following problem

$$\min_{u \in X} \left\{ \langle g, u \rangle + \frac{c}{2}\|u - x^0\|^2 \right\}, \tag{5}$$

where $c > 0$, $g \in \mathbb{E}$, and $x^0 \in X$. The conditional gradient method for solving (5) is formally described as follows.

Conditional gradient (CondG) method

Step 0 Let $\varepsilon > 0$ and $z^1 \in X$. Set $t = 1$.

Step 1 Use the LO oracle to compute an optimal solution u^t of

$$g^*_{t,c} = \min_{u \in X}\{\langle g + c(z^t - x^0), u - z^t \rangle\}.$$

Step 2 If $g^*_{t,c} \geq -\varepsilon$, then stop the algorithm and output z^t. Otherwise, set

$$z^{t+1} = z^t + \alpha_t(u^t - z^t),$$

where

$$\alpha_t := \min\left\{1, \frac{\langle g + c(z^t - x^0), z^t - u^t\rangle}{c\|u^t - z^t\|^2}\right\}. \tag{6}$$

Step 3 Set $t \leftarrow t + 1$, and go to step 1.

end

In the following, we consider an iteration-complexity result for the CondG method. Its proof can be founded, for example, in [13, Theorem 2.2(c)].

Proposition 2.1. *The total number of iterations performed by the CondG method for obtaining $z^t \in X$ such that $\langle g + c(z^t - x^0), u - z^t\rangle \geq -\varepsilon$ for all $u \in X$ is bounded by*

$$\left\lceil \frac{6cD_X^2}{\varepsilon} \right\rceil,$$

where D_X is the diameter of X.

2.2 Projection-free accelerated method

In this section, we present a projection-free accelerated method for solving (2) when the feasible set is bounded and discuss its convergence behavior. Such a method is related to the one proposed in [13]. The method is described as follows.

Projection-free accelerated (PFA) method

Step 0 Let $x^0 \in X$ be given, and set $A_0 = 0$, $\beta_0 > 0$, $y^0 = x^0$, function $\Gamma_0 : X \to \mathbb{R}$ as $\Gamma_0 \equiv 0$ and $k = 1$.

Step 1 Set $(y, x, \Gamma) = (y^{k-1}, x^{k-1}, \Gamma_{k-1})$ and $(A, \beta) = (A_{k-1}, \beta_{k-1})$. Compute A^+, $\tilde{x}$ and $\Gamma^+ : X \to \mathbb{R}$ as

$$\begin{aligned} A^+ &:= A + \tfrac{\beta + \sqrt{\beta^2 + 4LA\beta}}{2L}, \\ \tilde{x} &:= \tfrac{A}{A^+} y + \left(1 - \tfrac{A}{A^+}\right) x, \\ \Gamma^+ &:= \tfrac{A}{A^+}\Gamma + \left(1 - \tfrac{A}{A^+}\right) l_f(\cdot, \tilde{x}). \end{aligned}$$

Step 2 Let $\eta^+ > 0$ and $\beta^+ \geq \beta$, and apply the CondG method with $g := A^+\nabla\Gamma^+$, $c := \beta^+$ and an arbitrary $z_1 \in X$ to obtain x^+ such that

$$\langle A^+\nabla\Gamma^+ + \beta^+(x^+ - x^0), u - x^+\rangle \geq -\eta^+ \qquad \forall u \in X. \tag{7}$$

Step 3 Compute

$$y^+ := \frac{A}{A^+}y + \left(1 - \frac{A}{A^+}\right)x^+.$$

Step 4 Let $(y^k, x^k, \tilde{x}^k, \Gamma_k) = (y^+, x^+, \tilde{x}, \Gamma^+)$ and $(A_k, \beta_k, \eta_k) = (A^+, \beta^+, \eta^+)$. Set $k \leftarrow k+1$ and go to step 1.

end

Let us make a few remarks about the PFA method. First, note that the update rule for A_k is equivalent to the identity

$$\frac{A_k}{(A_k - A_{k-1})^2} = \frac{L}{\beta_{k-1}} \qquad \forall k \geq 1. \tag{8}$$

Second, using the definition of Γ_k, we easily obtain by induction that $A_k\Gamma_k(u) = \sum_{i=1}^k (A_i - A_{i-1})l_f(u, \tilde{x}^i)$. Hence, in Step 2 of the PFA method, we are applying the conditional gradient method to find an approximate solution x^k for the subproblem

$$\min_{u \in X} \left\{ \sum_{i=1}^k (A_i - A_{i-1})l_f(u, \tilde{x}^i) + \frac{\beta_k}{2}\|u - x^0\|^2 \right\} \tag{9}$$

such that

$$\left\langle \sum_{i=1}^k (A_i - A_{i-1})\nabla f(\tilde{x}^i) + \beta_k(x^k - x^0), u - x^k \right\rangle \geq -\eta_k \qquad \forall u \in X. \tag{10}$$

Throughout this section, we let

$$\eta_0 = 0, \quad \text{and} \quad \bar{\eta}_k = \sum_{i=0}^k \eta_i.$$

The next proposition establishes the main property of the PFA method and, as a consequence, the convergence rate of $f(y^k) - f^*$.

Proposition 2.2. *For every $k \geq 0$, the following inequality holds*

$$A_k\Gamma_k(u) + \frac{\beta_k}{2}\|u - x^0\|^2 \geq A_k f(y^k) + \frac{\beta_k}{2}\|u - x^k\|^2 - \bar{\eta}_k, \qquad \forall\, u \in X. \tag{11}$$

As a consequence, for every $k \geq 1$, we have

$$f(y^k) - f^* \leq \frac{\beta_k}{2A_k}d_0^2 + \frac{\bar{\eta}_k}{A_k}.$$

Proof. Let us prove by induction that (11) holds for all $k \geq 0$. Since $\bar{\eta}_0 = A_0 = 0$, we have it trivially holds for $k = 0$. Suppose that (11) holds for $k-1$. Using the same notations of the PFA method and denoting $\bar{\eta} = \bar{\eta}_{k-1}$, the induction assumption becomes

$$A\Gamma(u) + \frac{\beta}{2}\|u - x^0\|^2 \geq Af(y) + \frac{\beta}{2}\|u - x\|^2 - \bar{\eta}, \qquad \forall\, u \in X. \tag{12}$$

Now, using the definition of Γ^+, $\beta_+ \geq \beta$ and (12) with $u = x^+$, we have

$$\begin{aligned}\Psi^+(x^+) := A^+\Gamma^+(x^+) + \tfrac{\beta^+}{2}\|x^+ - x^0\|^2 &\geq (A^+ - A)l_f(x^+, \tilde{x}) + A\Gamma(x^+) + \tfrac{\beta}{2}\|x^+ - x^0\|^2 \\ &\geq (A^+ - A)l_f(x^+, \tilde{x}) + Af(y) + \tfrac{\beta}{2}\|x^+ - x\|^2 - \bar{\eta}.\end{aligned}$$

Thus, since $l_f(\cdot, \tilde{x})$ is an affine function and minorizes f, we obtain

$$\Psi^+(x^+) \geq A^+ l_f\left(\frac{A}{A^+}y + \frac{A^+ - A}{A^+}x^+, \tilde{x}\right) + \frac{\beta}{2}\|x^+ - x\|^2 - \bar{\eta},$$

which, combined with the definitions of $\tilde{x}$ and y^+ and (8), yields

$$\begin{aligned}\Psi^+(x^+) &\geq A^+\left(l_f(y^+, \tilde{x}) + \tfrac{\beta A^+}{2(A^+ - A)^2}\|y^+ - \tilde{x}\|^2\right) - \bar{\eta} \\ &= A^+\left(l_f(y^+, \tilde{x}) + \tfrac{L}{2}\|y^+ - \tilde{x}\|^2\right) - \bar{\eta} \\ &\geq A^+ f(y^+) - \bar{\eta},\end{aligned} \tag{13}$$

where the last inequality is due to (4) and $A^+ > 0$. On the other hand, using Ψ^+ is a quadratic function and (7), we have

$$\Psi^+(u) = \Psi^+(x^+) + \langle \nabla\Psi^+(x^+), u - x^+\rangle + \frac{\beta^+}{2}\|u - x^+\|^2 \geq \Psi^+(x^+) - \eta^+ + \frac{\beta^+}{2}\|u - x^+\|^2.$$

Therefore, from the last inequality and (13), we obtain

$$\Psi^+(u) \geq A^+ f(y^+) - \bar{\eta} - \eta^+ + \frac{\beta^+}{2}\|u - x^+\|^2,$$

which, combined with the definitions of Ψ^+ and $\bar{\eta}$, concludes the induction proof. Now, letting $x^* \in X^*$ be such that $d_0 = \|x^0 - x^*\|$, the last statement of the proposition follows immediately from (11) with $u = x^*$ and the fact that $\Gamma \leq f$. □

Next corollary specializes the rate of convergence of $f(y^k) - f^*$ given in Proposition 2.2 by appropriately choosing the sequences $\{\beta_k\}$ and $\{\eta_k\}$. In particular, this instance implies complexity bound similar to the accelerated methods with projection step (see, e.g., [16, 17]). We state the results assuming the knowledge of an upper bound D_0 for the distance from x^0 to the solution set. Note that, since X is a compact set, we may consider D_0 as the diameter of X.

Corollary 2.3. *Let D_0 be an upper bound for the distance from x^0 to the solution set, and let $B > 0$. For every $k \geq 1$, set $\beta_k = k + 1$, $\eta_k = D_0^2/B$ and consider y^k be generated by the PFA method. Then,*

$$f(y^k) - f^* \leq \left(\frac{B + 2}{B}\right)\frac{9LD_0^2}{2k^2}, \qquad \forall\, k \geq 1. \tag{14}$$

As a consequence, the total number of outer and inner iterations of the PFA method for obtaining y^k such that $f(y^k) - f^* \leq \epsilon$ can be bounded, respectively, by

$$\mathcal{O}\left(\frac{D_0\sqrt{L}}{\sqrt{\epsilon}}\right) \qquad \text{and} \qquad \mathcal{O}\left(\frac{LD_X^2}{\epsilon}\right).$$

Proof. Lemma A.1 in Appendix A implies that $A_k \geq (k+1)k^2/(9L)$. Thus, the first statement of the corollary follows immediately from Proposition 2.2 and the definitions of $\{\beta_k\}$, $\{\eta_k\}$ and $\{\bar{\eta}_k\}$. It follows from (14) that the PFA method provides y^k such that $f(y^k) - f^* \leq \epsilon$ in at most $\mathcal{O}\left(D_0\sqrt{L/\epsilon}\right)$. On the other hand, Proposition 2.1 implies that the total number of LO oracle calls up to the k^{th} iteration of the PFA method can be bounded by

$$\sum_{i=1}^{k} \frac{6\beta_i D_X^2}{\eta_i} + 1 = \sum_{i=1}^{k} \frac{6(i+1)BD_X^2}{D_0^2} + 1 = \frac{6k(k+3)BD_X^2}{2D_0^2} + k \approx \mathcal{O}(k^2 D_X^2/D_0^2).$$

Since $k \approx \mathcal{O}(D_0\sqrt{L/\epsilon})$, the last claim follows from the previous estimate. □

Remark 2.4. *Since X is a compact set, as mentioned before, a possible estimate D_0 satisfying Corollary 2.3 is $D_0 = D_X$. However, if for a particular problem, an estimate D_0 much smaller than D_X is known, then Corollary 2.3 gives a sharper complexity result than taking $D_0 = D_X$. It is worth pointing out that if such a D_0 is known then problem (2) may be restricted to $X \cap B(x^0, D_0)$. In such a case, the number of LO oracle calls depends also on D_0 instead of D_X, and therefore, it may be smaller depending on how easy it is to solve the subproblems over the set $X \cap B(x^0, D_0)$.*

3 Adaptive projection-free accelerated method

In this section, we describe and analyze a projection-free method for solving problem (2) when X is unbounded. Since the diameter of X is not finite, the PFA method can not be directly applied. Therefore, we adapt the method in such a way that its subproblems are solved over the set $X \cap B(x^0, R_k)$, where the radius R_k is iteratively and carefully updated.

Adaptive projection-free accelerated (APFA) method

Step 0 Let $x^0 \in X$, $\tau > 1$ and $R_1 > 0$ be given, and set $A_0 = \bar{\eta}_0 = 0$, $\beta_0 > 0$, $y^0 = x^0$, function $\Gamma_0 : X \to \mathbb{R}$ as $\Gamma_0 \equiv 0$ and $k = 1$.

Step 1 Set $(y, x, \Gamma) = (y^{k-1}, x^{k-1}, \Gamma_{k-1})$ and $(A, \beta, \bar{\eta}) = (A_{k-1}, \beta_{k-1}, \bar{\eta}_{k-1})$.

If $k = 1$, set $A^+ = \beta_0/(2L)$, $\tilde{x} = x^0$ and $R^+ = R_1$; else, let the unique $A^+ > A$ solution of

$$\frac{1}{2}\left(\frac{A^+}{A^+ - A}\right)^2 - \frac{LA^+}{\beta} = \tau^2, \tag{15}$$

set

$$\tilde{x} := \frac{A}{A^+}y + \left(1 - \frac{A}{A^+}\right)x, \tag{16}$$

and compute R^+ as

$$R^+ := \frac{2\tau}{\tau - 1}\|x^0 - \tilde{x}\| + \frac{\tau + 1}{\tau - 1}\sqrt{\frac{2\bar{\eta}}{\beta}}. \tag{17}$$

Step 2 Let $\eta^+ > 0$ and $\beta^+ \geq \beta$ and apply the CondG method with $g := A^+\nabla\Gamma^+$, $c := \beta^+$ and an arbitrary $z_1 \in X \cap B(x^0, R_+)$ to obtain x^+ satisfying

$$\langle A^+\nabla\Gamma^+ + \beta^+(x^+ - x^0), u - x^+\rangle \geq -\eta^+ \quad \forall u \in X \cap B(x^0, R_+), \tag{18}$$

where

$$\Gamma^+ := \frac{A}{A^+}\Gamma + \left(1 - \frac{A}{A^+}\right) l_f(\cdot, \tilde{x}), \tag{19}$$

set $\bar{\eta}^+ := \bar{\eta} + \eta^+$ and

$$y^+ := \frac{A}{A^+}y + \left(1 - \frac{A}{A^+}\right) x^+. \tag{20}$$

Step 3 Let $(y^k, x^k, \tilde{x}^k, \Gamma_k) = (y^+, x^+, \tilde{x}, \Gamma^+)$ and $(A_k, \beta_k, \eta_k, \bar{\eta}_k, R_k) = (A^+, \beta^+, \eta^+, \bar{\eta}^+, R^+)$. Set $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks about the above method. First, a fundamental property for the convergence analysis of the method is that the ball $B(x^0, R_k)$ contains the unknown exact solution of the subproblem (9) over the unbounded set X. Such a property will be a priori required for $k = 1$ and, as it will be seen in Lemma 3.3, it holds for all $k \geq 2$. Corollary 4.3 presents a simple choice for R_1 satisfying the required condition and in Section 5 a procedure to refine this choice is described. Second, note that for updating A_k, it is necessary to find a solution of a third degree polynomial equation which can be easily obtained by root-finding algorithms like Newton method. Third, similarly to the PFA method, we see that $A_k\Gamma_k(u) = \sum_{i=1}^{k}(A_i - A_{i-1})l_f(u, \tilde{x}^i)$ and the conditional gradient method is applied to obtain an approximate solution x^k for the subproblem (9) satisfying (10) with X replaced by $X \cap B(x^0, R_k)$.

Before establishing the convergence rate for the APFA method, we need some technical results.

Lemma 3.1. *Let $A \geq 0$, $\beta > 0$, $x, y \in X$ and an affine function $\Gamma \leq f$ be given, and define*

$$\chi = \chi(A, \Gamma, \beta, x, y) := Af(y) - \min_{u \in X}\left\{A\Gamma(u) + \frac{\beta}{2}\|u - x^0\|^2 - \frac{\beta}{4}\|u - x\|^2\right\}.$$

Then, the following statements hold:

a) for any $x^* \in X^*$, we obtain

$$A(f(y) - f^*) \leq \frac{\beta}{2}\|x^* - x^0\|^2 + \chi;$$

b) for any $A^+ > A$ and $u \in X$ we have

$$A^+\Gamma^+(u) + \frac{\beta}{2}\|u - x^0\|^2 \geq A^+\left(l_f(\tilde{u}(u), \tilde{x}) + \frac{\beta A^+}{4(A^+ - A)^2}\|\tilde{u}(u) - \tilde{x}\|^2\right) - \chi, \tag{21}$$

where $\tilde{x}$ and Γ^+ are as defined in (16) and (19), respectively, and

$$\tilde{u}(u) := \frac{A}{A^+}y + \left(1 - \frac{A}{A^+}\right) u. \tag{22}$$

Proof. It follows from the definition of χ that

$$A\Gamma(x^*) + \frac{\beta}{2}\|x^* - x^0\|^2 \geq Af(y) - \chi \quad \forall x^* \in X^*,$$

which, combined with $\Gamma \leq f$, implies the first statement. Now, using the definitions of Γ^+ and χ, we have

$$\begin{aligned} A^+\Gamma^+(u) + \tfrac{\beta}{2}\|u - x^0\|^2 &= (A^+ - A)l_f(u, \tilde{x}) + A\Gamma(u) + \tfrac{\beta}{2}\|u - x^0\|^2 \\ &\geq (A^+ - A)l_f(u, \tilde{x}) + Af(y) + \tfrac{\beta}{4}\|u - x\|^2 - \chi. \end{aligned}$$

Thus, since $l_f(\cdot, \tilde{x})$ is an affine function and minorizes f, we obtain

$$\begin{aligned} A^+\Gamma^+(u) + \tfrac{\beta}{2}\|u - x^0\|^2 &\geq A^+ l_f\left(\tfrac{A}{A^+}y + \tfrac{A^+ - A}{A^+}u, \tilde{x}\right) + \tfrac{\beta}{4}\|u - x\|^2 - \chi \\ &= A^+ l_f\left(\tilde{u}(u), \tilde{x}\right) + \tfrac{\beta}{4}\|u - x\|^2 - \chi, \end{aligned} \tag{23}$$

where the last equality is due to the definition of $\tilde{u}(u)$. On the other hand, the definition of $\tilde{u}(u)$ and $\tilde{x}$ implies

$$u - x = \frac{A^+}{A^+ - A}\left(\tilde{u}(u) - \tilde{x}\right).$$

Hence, statement (b) now follows from (23) and the last equality. □

Lemma 3.2. *Let $A \geq 0$, $\beta > 0$, $x, y \in X$ and an affine function $\Gamma \leq f$ be given, and let $\chi := \chi(A, \Gamma, \beta, x, y)$, where $\chi(\cdot,\cdot,\cdot,\cdot,\cdot)$ is defined in Lemma 3.1. Also, let $A^+ > A$ satisfying*

$$\frac{A^+}{2(A^+ - A)^2} \geq \frac{L}{\beta} \tag{24}$$

be given, and consider $\tilde{x}$ and Γ^+ as in (16) *and* (19)*, respectively. Moreover, assume that x^+ satisfies* (18) *for some $\eta^+ \geq 0$, $\beta^+ \geq \beta$ and $R_+ \geq \|x_r^+ - x^0\|$, where x_r^+ is the unique minimizer of*

$$\min_{u \in X}\left\{\Psi^+(u) := A^+\Gamma^+(u) + \frac{\beta^+}{2}\|u - x^0\|^2\right\}, \tag{25}$$

and consider y^+ as in (20)*. Then, $\chi^+ := \chi(A^+, \Gamma^+, \beta^+, x^+, y^+)$ satisfies*

$$\chi^+ \leq \chi + \eta^+.$$

Proof. Since $\beta^+ \geq \beta$, it follows from the definition of Ψ^+ in (25) that

$$\Psi^+(x^+) \geq A^+\Gamma^+(u) + \frac{\beta}{2}\|u - x^0\|^2,$$

which, combined with Lemma 3.1(b) with $u = x^+$ and the definition of y^+, yileds

$$\begin{aligned} \Psi^+(x^+) &\geq A^+\left(l_f(y^+, \tilde{x}) + \tfrac{\beta A^+}{4(A^+ - A)^2}\|y^+ - \tilde{x}\|^2\right) - \chi \\ &\geq A^+\left(l_f(y^+, \tilde{x}) + \tfrac{L}{2}\|y^+ - \tilde{x}\|^2\right) - \chi \\ &\geq A^+ f(y^+) - \chi, \end{aligned} \tag{26}$$

where the second and third inequalities are due to (24) and (4), respectively. Using the fact that Ψ^+ is a quadratic function and x_r^+ is the unique minimizer of (25), we have

$$\Psi^+(u) \geq \Psi^+(x_r^+) + \frac{\beta^+}{2}\|u - x_r^+\|^2 \quad \forall u \in X. \tag{27}$$

Since by assumption $R_+ \geq \|x_r^+ - x^0\|$, it follows by using the Taylor expansion of Ψ^+ at x^+ and (18) that

$$\begin{aligned} \Psi^+(x_r^+) &= \Psi^+(x^+) + \langle \nabla\Psi^+(x^+), x_r^+ - x^+\rangle + \tfrac{\beta^+}{2}\|x_r^+ - x^+\|^2 \\ &\geq \Psi^+(x^+) - \eta^+ + \tfrac{\beta^+}{2}\|x_r^+ - x^+\|^2. \end{aligned}$$

In view of (27) and the last inequality, we have

$$\begin{aligned} \Psi^+(u) &\geq \Psi^+(x^+) - \eta^+ + \tfrac{\beta^+}{2}\|u - x_r^+\|^2 + \tfrac{\beta^+}{2}\|x_r^+ - x^+\|^2 \\ &\geq \Psi^+(x^+) - \eta^+ + \tfrac{\beta^+}{4}\|u - x^+\|^2 \\ &\geq A^+ f(y^+) - \chi - \eta^+ + \tfrac{\beta^+}{4}\|u - x^+\|^2, \end{aligned}$$

where the second inequality is due to the convexity of $\|\cdot\|^2$, and the last one is due to (26). It follows from the previous estimate and the definition of Ψ^+ that

$$A^+\Gamma^+(u) + \frac{\beta^+}{2}\|u - x^0\|^2 - \frac{\beta^+}{4}\|u - x^+\|^2 - A^+ f(y^+) \geq -\chi - \eta^+ \quad \forall u \in X.$$

Combining the definition of χ^+ with the last inequality, the proof of the lemma follows. □

Lemma 3.3. *Let $A > 0$, $\beta > 0$, $x, y \in X$ and an affine function $\Gamma \leq f$ be given, and let $\chi := \chi(A, \Gamma, \beta, x, y)$, where $\chi(\cdot,\cdot,\cdot,\cdot,\cdot)$ is defined in Lemma 3.1. Assume that $A^+ > A$ satisfies (15) for some given $\tau > 1$ and define $\tilde{x}$ and Γ^+ as in (16) and (19), respectively. Then, for any $\beta^+ \geq \beta$, the unique minimizer x_r^+ of (25) satisfies*

$$\|x_r^+ - x^0\| \leq R(\chi) := \frac{2\tau}{\tau - 1}\|x^0 - \tilde{x}\| + \frac{\tau + 1}{\tau - 1}\sqrt{\frac{2\max\{\chi, 0\}}{\beta}}.$$

Proof. First, as $\Gamma \leq f$ and $l_f(\cdot, \tilde{x}) \leq f$, it follows trivially from (19) that

$$\Gamma^+ \leq f. \tag{28}$$

Now, let us define an auxiliary point $\tilde{y}$ as

$$\tilde{y} := \operatorname*{argmin}_{z \in X} \left\{ l_f(z, \tilde{x}) + \frac{\beta A^+}{4(A^+ - A)^2}\|z - \tilde{x}\|^2 \right\}. \tag{29}$$

Since $\tilde{u}(u) \in X$ for all $u \in X$ (see (22)), it follows by Lemma 3.1(b) and (29) that

$$\begin{aligned} A^+\Gamma^+(u) + \tfrac{\beta}{2}\|u - x^0\|^2 &\geq A^+\left(l_f(\tilde{u}(u), \tilde{x}) + \tfrac{\beta A^+}{4(A^+ - A)^2}\|\tilde{u}(u) - \tilde{x}\|^2\right) - \chi \\ &\geq A^+\left(l_f(\tilde{y}, \tilde{x}) + \tfrac{\beta A^+}{4(A^+ - A)^2}\|\tilde{y} - \tilde{x}\|^2\right) - \chi \\ &= A^+\left(l_f(\tilde{y}, \tilde{x}) + \tfrac{L}{2}\|\tilde{y} - \tilde{x}\|^2\right) + \tfrac{\beta\tau^2}{2}\|\tilde{y} - \tilde{x}\|^2 - \chi \\ &\geq A^+ f(\tilde{y}) + \tfrac{\beta\tau^2}{2}\|\tilde{y} - \tilde{x}\|^2 - \chi, \end{aligned} \tag{30}$$

where the equality is due to (15) and the last inequality is due to (4). Since $\beta^+ \geq \beta$, the latter inequality with $u = x_r^+$ implies that

$$A^+\Gamma^+(x_r^+) + \frac{\beta^+}{2}\|x_r^+ - x^0\|^2 \geq A^+ f(\tilde{y}) - \chi. \tag{31}$$

From the fact that x_r^+ is the unique minimizer of (25), we have

$$\begin{aligned} A^+\Gamma^+(\tilde{y}) + \tfrac{\beta^+}{2}\|\tilde{y} - x^0\|^2 &\geq A^+\Gamma^+(x_r^+) + \tfrac{\beta^+}{2}\|x_r^+ - x^0\|^2 + \tfrac{\beta^+}{2}\|\tilde{y} - x_r^+\|^2 \\ &\geq A^+ f(\tilde{y}) - \chi + \tfrac{\beta^+}{2}\|\tilde{y} - x_r^+\|^2, \end{aligned}$$

where the last inequality is due to (31). Using (28) and rearranging the last inequality, we obtain

$$\|\tilde{y} - x_r^+\| \leq \|\tilde{y} - x^0\| + \sqrt{\frac{2\max\{\chi, 0\}}{\beta^+}}.$$

Using the triangle inequality and the fact that $\beta_+ \geq \beta$, we have

$$\|x_r^+ - x^0\| \leq 2\|\tilde{y} - x^0\| + \sqrt{\frac{2\max\{\chi, 0\}}{\beta}}. \tag{32}$$

On the other hand, inequality (30) with $u = \tilde{y}$ yields

$$A^+\Gamma^+(\tilde{y}) + \frac{\beta}{2}\|\tilde{y} - x^0\|^2 \geq A^+ f(\tilde{y}) + \frac{\beta\tau^2}{2}\|\tilde{y} - \tilde{x}\|^2 - \chi.$$

Thus, using (28) and simple algebraic manipulations, we have

$$\tau\|\tilde{y} - \tilde{x}\| \leq \sqrt{\|\tilde{y} - x^0\|^2 + \frac{2\max\{\chi, 0\}}{\beta}},$$

and then

$$\tau\left(\|\tilde{y} - x^0\| - \|x^0 - \tilde{x}\|\right) \leq \|\tilde{y} - x^0\| + \sqrt{\frac{2\max\{\chi, 0\}}{\beta}}.$$

Rearranging the last inequality, we obtain

$$(\tau - 1)\|\tilde{y} - x^0\| \leq \tau\|x^0 - \tilde{x}\| + \sqrt{\frac{2\max\{\chi, 0\}}{\beta}}.$$

Hence, using the assumption $\tau > 1$, we have

$$\|\tilde{y} - x^0\| \leq \frac{\tau}{\tau - 1}\|x^0 - \tilde{x}\| + \frac{1}{\tau - 1}\sqrt{\frac{2\max\{\chi, 0\}}{\beta}}.$$

Therefore, the result follows from the last inequality and (32). □

The next proposition provides an essential inequality from which follows our main convergence rate result for the APFA method.

Proposition 3.4. *Let x_r^1 be the unique minimizer of*

$$\min_{u\in X}\left\{A_1\Gamma_1(u)+\frac{\beta_1}{2}\|u-x^0\|^2\right\}. \tag{33}$$

If the input R_1 in the APFA method satisfies $R_1\geq\|x_r^1-x^0\|$, then the following inequality holds, for every $k\geq 1$,

$$A_kf(y^k)+\frac{\beta_k}{4}\|u-x^k\|^2\leq A_k\Gamma_k(u)+\frac{\beta_k}{2}\|u-x^0\|^2+\bar{\eta}_k\quad\forall u\in X. \tag{34}$$

Proof. The proof of (34) is done by induction on k. It clearly holds for $k=0$ due to fact that $A_0=\bar{\eta}_0=0$. Assume then that (34) holds for some $k\geq 0$ and let us show that it holds for $k+1$. This induction assumption is clearly equivalent to

$$\chi_k:=\chi(A_k,\Gamma_k,\beta_k,x^k,y^k)\leq\bar{\eta}_k,$$

due to the definition of $\chi(\cdot,\cdot,\cdot,\cdot,\cdot)$ in Lemma 3.1. Also, letting

$$\begin{aligned}(A,\Gamma,\beta,x,y)&=(A_k,\Gamma_k,\beta_k,x^k,y^k),\\ (A^+,\Gamma^+,\beta^+,x^+,y^+)&=(A_{k+1},\Gamma_{k+1},\beta_{k+1},x^{k+1},y^{k+1}),\quad R^+=R_{k+1},\quad\eta^+=\eta_{k+1},\end{aligned}$$

it follows from the APFA method and Lemma 3.3 (or the assumption that $R_1\geq\|x_r^1-x^0\|$) that the hypothesis of Lemma 3.2 holds. Therefore, the latter lemma implies that

$$\chi_{k+1}\leq\chi_k+\eta^+\leq\bar{\eta}_k+\eta_{k+1}=\bar{\eta}_{k+1},$$

and hence that (34) holds for $k+1$. We have thus proved that (34) holds for every $k\geq 1$. □

The following result for the APFA method establishes the convergence rate of $f(y^k)-f^*$ and the boundedness of the sequences $\{x^k\}$, $\{y^k\}$ and $\{\tilde{x}^k\}$.

Theorem 3.5. *Assume that the input R_1 in the APFA method satisfies $R_1\geq\|x_r^1-x^0\|$, where x_r^1 is the unique minimizer of* (33)*. Then, for any $k\geq 1$,*

$$f(y^k)-f^*\leq\frac{\beta_k}{2A_k}d_0^2+\frac{\bar{\eta}_k}{A_k}, \tag{35}$$

$$\max\{\|x^k-x^*\|,\|y^k-x^*\|,\|\tilde{x}^{k+1}-x^*\|\}\leq\sqrt{2}d_0+2\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}}, \tag{36}$$

where $x^*\in X^*$ is such that $d_0=\|x^0-x^*\|$.

Proof. It follows from Proposition 3.4 with $u=x^*$ and $\Gamma_k\leq f$ that

$$A_kf(y^k)+\frac{\beta_k}{4}\|x^k-x^*\|^2\leq A_kf^*+\frac{\beta_k}{2}d_0^2+\bar{\eta}_k\quad\forall k\geq 1. \tag{37}$$

Since $A_k>0$ for any $k\geq 1$, inequality (37) clearly implies that (35) holds. From (37) we also obtain

$$\|x^k-x^*\|^2\leq 2d_0^2+\frac{4\bar{\eta}_k}{\beta_k}\quad\forall k\geq 1,$$

and thus,

$$\|x^k - x^*\| \leq \sqrt{2}d_0 + 2\sqrt{\frac{\bar{\eta}_k}{\beta_k}} \quad \forall k \geq 1. \tag{38}$$

On the other hand, since $A_0 = 0$ and $A_k = \sum_{j=1}^{k}(A_j - A_{j-1})$, the definition of y^k in (20) gives

$$y^k - x^* = \frac{1}{A_k}\sum_{j=1}^{k}(A_j - A_{j-1})(x^j - x^*),$$

which, combined with (38), implies that

$$\|y^k - x^*\| \leq \max_{1\leq j\leq k}\|x^j - x^*\| \leq \sqrt{2}d_0 + 2\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}} \quad \forall k \geq 1. \tag{39}$$

Now, combining the definition of $\tilde{x}_k$ in (16) with (38) and (39), we have, for any $k \geq 1$,

$$\|\tilde{x}^{k+1} - x^*\| \leq \frac{A_k}{A_{k+1}}\|y^k - x^*\| + \left(1 - \frac{A_k}{A_{k+1}}\right)\|x^k - x^*\| \leq \sqrt{2}d_0 + 2\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}}.$$

Therefore, (36) follows from (38), (39) and last inequality. □

To end this section, we present a result on the radius sequence $\{R_k\}$ which will be important to analyze a particular instance of the APFA method.

Corollary 3.6. *Let $\{R_k\}$ be generated by the APFA method and assume that $R_1 \geq \|x_r^1 - x^0\|$, where x_r^1 is the unique minimizer of* (33)*. Then*

$$R_k \leq \frac{(4+\sqrt{2})\tau + \sqrt{2}}{\tau - 1}\left(\frac{2\tau(\sqrt{2}+1)}{(4+\sqrt{2})\tau + \sqrt{2}}d_0 + \max_{1\leq j\leq k-1}\sqrt{\frac{\bar{\eta}_j}{\beta_j}}\right) \quad \forall k \geq 2.$$

Proof. Let $x^* \in X^*$ be such that $d_0 = \|x^0 - x^*\|$. It follows from (36) and the triangle inequality that

$$\|x^0 - \tilde{x}^k\| \leq (\sqrt{2}+1)d_0 + 2\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_{j-1}}{\beta_{j-1}}} \quad \forall k.$$

Therefore, since $\tau > 1$, for any $k \geq 2$, the definition of R_k in (17) implies that

$$\begin{aligned} R_k &= \tfrac{2\tau}{\tau-1}\|x^0 - \tilde{x}^k\| + \tfrac{\tau+1}{\tau-1}\sqrt{\tfrac{2\bar{\eta}_{k-1}}{\beta_{k-1}}} \\ &\leq \tfrac{2\tau}{\tau-1}\left((\sqrt{2}+1)d_0 + 2\max_{1\leq j\leq k}\sqrt{\tfrac{\bar{\eta}_{j-1}}{\beta_{j-1}}}\right) + \tfrac{\tau+1}{\tau-1}\sqrt{\tfrac{2\bar{\eta}_{k-1}}{\beta_{k-1}}}, \end{aligned}$$

which implies the desired inequality. □

4 An instance of the APFA method

In this section, we study the convergence rate of an instance of the APFA method. Basically, in this special version we specify the choices of the constant τ and the sequences $\{\beta_k\}$ and $\{\eta_k\}$.

Specialized adaptive projection-free accelerated (S-APFA) method

---

This method is an instance of the APFA method with the constant τ and the sequences $\{\beta_k\}$ and $\{\eta_k\}$ defined by

$$\tau = \sqrt{3}, \quad \beta_k = k+1 \quad \forall k \geq 0, \quad \text{and} \quad \eta_k = \frac{R_k^2}{B} \quad \forall k \geq 1, \tag{40}$$

where the positive constant B satisfies

$$B > p := \frac{(4\sqrt{3}+\sqrt{6}+\sqrt{2})^2}{(\sqrt{3}-1)^2} \approx 218. \tag{41}$$

Clearly, there exist various options of choices for the constant τ and the sequences $\{\beta_k\}$ and $\{\eta_k\}$. The choices provided in the S-APFA method lead to optimal complexity bounds on the total number of outer and inner iterations for problem (2).

In view of the S-APFA method, we can prove a finer estimate of the radius sequence $\{R_k\}$.

Proposition 4.1. *Consider the sequence $\{R_k\}$ generated by the S-APFA method and assume that $R_1 \geq \|x_r^1 - x^0\|$, where x_r^1 is the unique minimizer of* (33)*. Then,*

$$R_k \leq \frac{\sqrt{Bp}}{\sqrt{2}(\sqrt{B}-\sqrt{p})}\left[\left(\sqrt{\frac{2}{p}}+1\right)d_0 + \frac{R_1}{\sqrt{B}}\right] \quad \forall k \geq 2, \tag{42}$$

Proof. Let us prove by induction that (42) holds. First, from the S-APFA method we have

$$\sqrt{B} > \sqrt{p}, \quad \bar{\eta}_j = \sum_{i=1}^{j} \eta_i = \sum_{i=1}^{j} \frac{R_i^2}{B}. \tag{43}$$

Thus, it follows from Corollary 3.6 and definitions of τ and p in the S-APFA method that

$$R_2 \leq \sqrt{\frac{p}{2}}\left[\left(\sqrt{\frac{2}{p}}+1\right)d_0 + \frac{R_1}{\sqrt{B}}\right],$$

which, combined with the fact $1 < \sqrt{B}/(\sqrt{B}-\sqrt{p})$, implies that (42) trivially holds for $k = 2$. Now, assume that (42) holds any $j \in \{2, \ldots, k\}$ for some $k \geq 2$. Since $\beta_j = j+1$, (43) implies that

$$\frac{\bar{\eta}_j}{\beta_j} = \frac{1}{B(j+1)}\sum_{i=1}^{j} R_i^2 \leq \frac{R_1^2}{2B} + \frac{1}{B(j+1)}\sum_{i=2}^{j} R_i^2 \leq \frac{R_1^2}{2B} + \frac{\theta_j^2}{B} \quad \forall j \geq 2,$$

where $\theta_j = \max_{i\in\{2,\ldots,j\}} R_i$. Hence, from Corollary 3.6 and definitions of τ and p in the S-APFA method, we have

$$\begin{aligned} R_{k+1} &\leq \sqrt{\tfrac{p}{2}}\left[\left(\sqrt{\tfrac{2}{p}}+1\right)d_0 + \tfrac{R_1}{\sqrt{B}} + \sqrt{\tfrac{2}{B}}\max_{j\in\{2,\ldots,k\}}\theta_j\right] \\ &= \sqrt{\tfrac{p}{2}}\left[\left(\sqrt{\tfrac{2}{p}}+1\right)d_0 + \tfrac{R_1}{\sqrt{B}} + \sqrt{\tfrac{2}{B}}\theta_k\right]. \end{aligned}$$

Observing that the induction assumption implies that θ_k is majorized by the right hand side of (42), it follows from the last inequality that

$$\begin{aligned} R_{k+1} &\leq \sqrt{\tfrac{p}{2}}\left[\left(\sqrt{\tfrac{2}{p}}+1\right)d_0 + \tfrac{R_1}{\sqrt{B}} + \tfrac{\sqrt{p}}{(\sqrt{B}-\sqrt{p})}\left(\left(\sqrt{\tfrac{2}{p}}+1\right)d_0 + \tfrac{R_1}{\sqrt{B}}\right)\right] \\ &= \sqrt{\tfrac{p}{2}}\left(\left(\sqrt{\tfrac{2}{p}}+1\right)\tfrac{\sqrt{B}d_0}{(\sqrt{B}-\sqrt{p})} + \tfrac{R_1}{(\sqrt{B}-\sqrt{p})}\right). \end{aligned}$$

Hence, the induction proof follows easily from the last inequality.

□

In what follows, we present our main result on convergence rates of the S-APFA method.

Theorem 4.2. *Consider the sequence $\{y^k\}$ generated by the S-APFA method and assume that $R_1 \geq \|x_r^1 - x^0\|$, where x_r^1 is the unique minimizer of (33). Then*

$$f(y^k) - f^* \leq \left[1 + 2\left(\frac{\sqrt{2}+\sqrt{p}}{\sqrt{B}-\sqrt{p}}\right)^2\right]\left(d_0^2 + \frac{R_1^2}{B}\right)\frac{25L}{k^2} \quad \forall k \geq 1. \tag{44}$$

As consequence, if $B = 4p$ then

$$f(y^k) - f^* \leq \left(d_0^2 + \frac{R_1^2}{4p}\right)\frac{86L}{k^2} \quad \forall k \geq 1.$$

Proof. From the S-APFA method and Proposition 4.1, we have

$$\begin{aligned} \bar{\eta}_k = \textstyle\sum_{i=1}^k \frac{R_i^2}{B} &\leq \tfrac{R_1^2}{B} + \tfrac{p}{2(\sqrt{B}-\sqrt{p})^2}\left[\left(1+\tfrac{\sqrt{2}}{\sqrt{p}}\right)d_0 + \tfrac{R_1}{\sqrt{B}}\right]^2(k-1) \\ &\leq \tfrac{R_1^2}{B} + \tfrac{p}{(\sqrt{B}-\sqrt{p})^2}\left[\left(1+\tfrac{\sqrt{2}}{\sqrt{p}}\right)^2 d_0^2 + \tfrac{R_1^2}{B}\right](k-1), \end{aligned}$$

where the last inequality is due to $(a+b)^2 \leq 2(a^2+b^2)$ for all $a, b \in \mathbb{R}$. Hence, using Theorem 3.5 and $\beta_k = k+1$, we conclude that

$$\begin{aligned} f(y^k) - f^* &\leq \tfrac{1}{2A_k}\left[(k+1)d_0^2 + 2\bar{\eta}_k\right] \\ &\leq \tfrac{1}{2A_k}\left[(k+1)d_0^2 + \tfrac{2R_1^2}{B} + \tfrac{2p}{(\sqrt{B}-\sqrt{p})^2}\left(\left(1+\tfrac{\sqrt{2}}{\sqrt{p}}\right)^2 d_0^2 + \tfrac{R_1^2}{B}\right)(k-1)\right] \\ &\leq \tfrac{k+1}{2A_k}\left[1 + 2\left(\tfrac{\sqrt{2}+\sqrt{p}}{\sqrt{B}-\sqrt{p}}\right)^2\right]\left(d_0^2 + \tfrac{R_1^2}{B}\right), \end{aligned}$$

which, combined with the estimate of A_k in Lemma A.2, implies (44). The second inequality of the proposition is an immediate consequence of first one and $B = 4p$. □

Next result, we present a possible choice of R_1 which depends on $\nabla f(x^0)$. Moreover, we specialize Theorem 4.2 for this choice of R_1.

Corollary 4.3. *Let x_r^1 be the unique minimizer of (33). Then, the following inequality holds*

$$\|x_r^1 - x^0\| \leq \frac{1}{4L}\|\nabla f(x^0)\|. \tag{45}$$

As a consequence, the S-APFA method with $R_1 = \|\nabla f(x^0)\|/4L$ and $B = 4p$ generates a sequence $\{y^k\}$ satisfying

$$f(y^k) - f^* \leq \left(d_0^2 + \frac{\|\nabla f(x^0)\|^2}{64pL^2}\right)\frac{86L}{k^2} \quad \forall k \geq 1.$$

Proof. The first statement is proved in lemma B.1(a) of Appendix B. The second part follows directly from Theorem 4.2 and definition of R_1. □

The following result establishes iteration-complexity bounds for the S-APFA method to obtain an approximate solution y^k, i.e., $f(y^k) - f^* \leq \varepsilon$, where $\varepsilon > 0$ is a given tolerance.

Corollary 4.4. *For a given tolerance $\varepsilon > 0$, the S-APFA method with $R_1 = \|\nabla f(x^0)\|/4L$ and $B = 4p$ generates a point y^k satisfying $f(y^k) - f^* \leq \varepsilon$ in at most*

$$\mathcal{O}\left(\sqrt{Ld_0^2 + \frac{\|\nabla f(x^0)\|^2}{L}}\frac{1}{\sqrt{\varepsilon}}\right), \quad \mathcal{O}\left(\left(Ld_0^2 + \frac{\|\nabla f(x^0)\|^2}{L}\right)\frac{1}{\varepsilon}\right) \tag{46}$$

outer iterations and LO oracle calls, respectively.

Proof. The first bound in (46) follows immediately from the last statement of Corollary 4.3. Now, by Proposition 2.1, we obtain that the total number of LO oracle calls up to the k^{th} outer iteration of the S-APFA method can be bounded by

$$\sum_{i=1}^{k}\left(\frac{24\beta_i R_i^2}{\eta_i} + 1\right) = \sum_{i=1}^{k}[24B(i+1)+1] = 12Bk(k+3) + k = \mathcal{O}(k^2).$$

Therefore, the second bound in (46) follows from the first one and the last conclusion.

□

5 Approximate solution and a procedure for obtaining R_1

In this section, we discuss a notion of approximate solution that naturally generalizes the concept of solution of an optimization problem. We show that the S-APFA method generates an approximate solution, which can easily be verified during the process of the method. We also present a procedure which is capable of controlling the size of $\nabla f(x^0)$ when compared to d_0, which is an interesting property used to obtain a better estimate of our iteration-complexity results.

Definition 5.1. *Given $\rho, \epsilon > 0$, we say that $z \in X$ is a (ρ, ϵ)–solution of problem (2) if and only if there exists $v \in \partial_\epsilon(f + \mathcal{I}_X)(z)$ such that $\|v\| \leq \rho$.*

Note that any solution of problem (2) is a (ρ, ϵ)-solution of (2) for any $\rho, \epsilon > 0$. Therefore, the above definition is consistent with the concept of approximate solution.

Proposition 5.2. *Consider the APFA method and assume that $R_1 \geq \|x_r^1 - x^0\|$, where x_r^1 is the unique minimizer of (33). Then, for every $k \geq 1$, y^k is a (ρ_k, ϵ_k)-solution of (2), where (ρ_k, ε_k) is computed as*

$$\rho_k := \frac{\beta_k}{A_k}\left[\left(\frac{\sqrt{2}}{2}+1\right)\|y^k - x^0\| + \frac{1}{2}\|y^k - x^k\| + \sqrt{\frac{\bar{\eta}_k}{\beta_k}}\right],$$

$$\varepsilon_k := \frac{\beta_k}{2A_k}\|y^k - x^0\|^2 - \frac{\beta_k}{4A_k}\|y^k - x^k\|^2 + \frac{\bar{\eta}_k}{A_k}.$$

Proof. First, let the auxiliary quadratic convex function $q_k : \mathbb{E} \to \mathbb{R}$ be defined as

$$q_k(u) = A_k(\Gamma_k(u) - f(y^k)) + \frac{\beta_k}{2}\|u - x^0\|^2 - \frac{\beta_k}{4}\|u - x^k\|^2 + \bar{\eta}_k \quad \forall u \in X, \tag{47}$$

and let w^k be its minimizer over X. Hence, we obtain for any $u \in X$,

$$\begin{aligned} A_k(\Gamma_k(u) - f(y^k)) &= q_k(u) + \tfrac{\beta_k}{4}\|u - x^k\|^2 - \tfrac{\beta_k}{2}\|u - x^0\|^2 - \bar{\eta}_k \\ &\geq q_k(w^k) + \tfrac{\beta_k}{4}\|u - w^k\|^2 + \tfrac{\beta_k}{4}\|u - x^k\|^2 - \tfrac{\beta_k}{2}\|u - x^0\|^2 - \bar{\eta}_k \\ &\geq g_k(u) := \tfrac{\beta_k}{4}\|u - w^k\|^2 + \tfrac{\beta_k}{4}\|u - x^k\|^2 - \tfrac{\beta_k}{2}\|u - x^0\|^2 - \bar{\eta}_k, \end{aligned} \tag{48}$$

where the second inequality is due to the fact that $q_k(w^k) \geq 0$ (see Proposition 3.4 with $u = w^k$). Now, let

$$v^k := \frac{\nabla g_k(y^k)}{A_k}, \quad \mu_k := -\frac{g_k(y^k)}{A_k}.$$

Hence, since $\Gamma_k \leq f$ and g_k is affine, it follows from (48) and the first order Taylor expansion of g_k at y^k that

$$f(u) - f(y^k) \geq \langle v^k, u - y^k\rangle - \mu_k \quad \forall u \in X,$$

which implies that

$$v^k \in \partial_{\mu_k}(f + \mathcal{I}_X)(y^k). \tag{49}$$

Now, observe that from (48) with $u = y^k$ and $\Gamma_k \leq f$, we easily have

$$A_k\varepsilon_k \geq A_k\mu_k = \frac{\beta_k}{2}\|y^k - x^0\|^2 - \frac{\beta_k}{4}\|y^k - w^k\|^2 - \frac{\beta_k}{4}\|y^k - x^k\|^2 + \bar{\eta}_k \geq 0.$$

Hence,

$$\|y^k - w^k\|^2 \leq 2\|y^k - x^0\|^2 - \|y^k - x^k\|^2 + \frac{4\bar{\eta}_k}{\beta_k},$$

which implies that

$$\|y^k - w^k\| \leq \sqrt{2}\|y^k - x^0\| + 2\sqrt{\frac{\bar{\eta}_k}{\beta_k}}. \tag{50}$$

On the other hand, from the definition of v^k it is immediate to see that

$$\|v^k\| = \frac{\beta_k}{A_k}\left\|\frac{1}{2}(y^k - x^k + y^k - w^k) + x^0 - y^k\right\| \leq \frac{\beta_k}{A_k}\left(\frac{1}{2}\|y^k - x^k\| + \frac{1}{2}\|y^k - w^k\| + \|y^k - x^0\|\right),$$

which, combined with (50), implies that $\|v^k\| \leq \rho_k$. Therefore, since $\mu_k \leq \epsilon_k$, the lemma follows from (49) and definition 5.1. □

In the following theorem, we present an iteration-complexity result for the S-APFA method .

Theorem 5.3. *Consider the S-APFA method with some $R_1 \geq \|x_r^1 - x^0\|$, where x_r^1 is the unique minimizer of (33). For a given tolerance pair $(\rho, \varepsilon) \in \mathbb{R}^2_{++}$, the S-APFA method certifies that an iterate y^k is a (ρ, ε)-solution of (2) in at most*

$$\mathcal{O}\left(\sqrt{L}\max\left\{\frac{\sqrt{d_0+R_1}}{\sqrt{\rho}}, \frac{d_0+R_1}{\sqrt{\varepsilon}}\right\}\right), \quad \mathcal{O}\left(L\max\left\{\frac{d_0+R_1}{\rho}, \frac{(d_0+R_1)^2}{\varepsilon}\right\}\right) \tag{51}$$

outer iterations and LO oracle calls, respectively.

Proof. Let $x^* \in X^*$ be such that $d_0 = \|x^0 - x^*\|$, and ρ_k and ε_k as defined in Proposition 5.2. Note that, if $\rho_k \leq \rho$ and $\varepsilon_k \leq \varepsilon$, then combining Proposition 5.2 and definition 5.1, we have y^k is a (ρ, ε)-solution of (2). Therefore, let us first estimate ρ_k and ε_k. Using the triangle inequality and (36), we obtain

$$\|x^0 - y^k\| \leq d_0 + \|y^k - x^*\| \leq (\sqrt{2}+1)d_0 + 2\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}}$$

and

$$\|y^k - x^k\| \leq \|y^k - x^*\| + \|x^* - x^k\| \leq 2\sqrt{2}d_0 + 4\max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}},$$

which imply that

$$\rho_k \leq \frac{7\beta_k}{A_k}\left(d_0 + \max_{1\leq j\leq k}\sqrt{\frac{\bar{\eta}_j}{\beta_j}}\right) \tag{52}$$

$$\varepsilon_k \leq \frac{\beta_k}{2A_k}\|y^k - x^0\|^2 + \frac{\bar{\eta}_k}{A_k} \leq \frac{7\beta_k}{A_k}\left(d_0^2 + \max_{1\leq j\leq k}\frac{\bar{\eta}_j}{\beta_j}\right). \tag{53}$$

From the definition of $\bar{\eta}_j$, $\beta_j = j+1$ and Propostion 4.1 we see that there exists $c_1 > 0$ such that

$$\frac{\bar{\eta}_j}{\beta_j} = \frac{1}{B(j+1)}\sum_{i=1}^{j} R_i^2 \leq \frac{R_1^2}{2B} + \frac{1}{B(j+1)}\sum_{i=2}^{j} R_i^2 \leq c_1(R_1^2 + d_0^2).$$

Combining the last inequality, (52), (53) and Lemma A.2, we obtain $c_2 > 0$

$$\rho_k \leq \frac{Lc_2}{k^2}\left(d_0 + R_1\right), \quad \varepsilon_k \leq \frac{Lc_2}{k^2}\left(d_0^2 + R_1^2\right),$$

which implies that in at most

$$\mathcal{O}\left(\sqrt{L}\max\left\{\frac{\sqrt{d_0+R_1}}{\sqrt{\rho}}, \frac{d_0+R_1}{\sqrt{\varepsilon}}\right\}\right),$$

outer iteration, the iterate y^k is a (ρ, ε)-solution of (2). Now, note that Proposition 2.1 implies that the total number of LO oracle calls up to the k^{th} outer iteration of the S-APFA method can be bounded by

$$\sum_{i=1}^{k}\left(\frac{24\beta_i R_i^2}{\eta_i} + 1\right) = \sum_{i=1}^{k}(24B(i+1)+1) = 12Bk(k+3) + k = \mathcal{O}(k^2).$$

Therefore, the second bound in (51) follows from the first one, which concludes the proof. □

In the remaining part of this section, we focus our attention on refining the initial radius $R_1 = \|\nabla f(x^0)\|/(4L)$ given in Corollary 4.3. Our goal is to keep R_1/d_0 relatively small. For this, we present a procedure which needs the following set

$$Z(R) := \left\{ z \in X \cap B(x^0, R) : \left\langle \frac{\nabla f(x^0)}{2L} + 2(z - x^0), u - z \right\rangle \geq -\frac{R^2}{M_1} \quad \forall u \in X \cap B(x^0, R) \right\}, \quad (54)$$

where $R, M_1 > 0$.

Remark 5.4. *Observe that, in view of Proposition 2.1, the CondG method applied for solving problem* (5) *with* $g = \nabla f(x^0)/(2L)$ *and* $c = 2$ *computes a point* $z \in Z(R)$ *by performing no more than* $\mathcal{O}(M_1)$ *LO oracle calls.*

Procedure 5.5. *Let scalars* $\rho, \epsilon > 0$ *and* $M_1 > 2$ *be given and set*

$$\delta := \min\left\{ \frac{\rho}{4L}, \frac{\epsilon}{\|\nabla f(x^0)\|} \right\}, \quad R := \frac{\|\nabla f(x^0)\|}{8L}. \quad (55)$$

Step 1 *Compute* $z \in Z(R)$ *using the CondG method, where* $Z(R)$ *is defined in* (54)*; if* $\|z - x^0\| \geq R/2$*, output* $R_1 := 2R$ *and* **stop***; else, if* $R \leq \delta$*, declare "*x^0 *is a* (ρ, ϵ)*-solution of* (2)*" and* **stop***;*

Step 2 *set* $R := R/2$ *and go to Step 1.*

end

Next proposition shows that our goal to refine R_1 depending on d_0 is achieved, apart from the unlikely situation in which x^0 be already a (ρ, ϵ)-solution of problem (2). The proof of the next proposition will be presented in Appendix B.

Proposition 5.6. *Assume that Procedure 5.5 computes all the iterates* $z \in Z(R)$ *via the CondG method. Then, in at most*

$$\mathcal{O}\left(M_1 \left[1 + \log_2 \frac{\|\nabla f(x^0)\|}{8L\delta} \right] \right) \quad (56)$$

LO oracle calls, Procedure 5.5 either certifies that the iterate x^0 *is a* (ρ, ϵ)*-solution of problem* (2) *or outputs* R_1 *satisfying*

$$\|x_r^1 - x^0\| \leq R_1 \leq \frac{8\sqrt{2M_1}}{\sqrt{2M_1} - 2} d_0, \quad (57)$$

where x_r^1 *is the unique minimizer of* (33) *and* δ *is defined in* (55)*.*

Next theorem summarizes the complexity results of the S-APFA method when Procedure 5.5 is used to compute R_1. We also show that it generates and certifies an approximate solution.

Theorem 5.7. *Let a tolerance pair* $(\rho, \varepsilon) \in \mathbb{R}^2_{++}$ *be given and* δ *as defined in* (55)*. If the S-APFA method use Procedure 5.5 to compute* R_1*, then an iterate* y^k *is certified to be a* (ρ, ε)*-solution in at most*

$$\mathcal{O}\left(\sqrt{L} \max\left\{ \frac{\sqrt{d_0}}{\sqrt{\rho}}, \frac{d_0}{\sqrt{\varepsilon}} \right\} \right), \quad \mathcal{O}\left(L \max\left\{ \frac{d_0}{\rho}, \frac{d_0^2}{\varepsilon} \right\} + M_1 \left(1 + \log_2 \frac{\|\nabla f(x^0)\|}{8L\delta} \right) \right)$$

outer iterations and LO oracle calls, respectively.

Proof. First, note that if the second stopping criterion of Procedure 5.5 is satisfied, then the result trivially follows from Proposition 5.6 and the fact that $y^0 = x^0$. Otherwise, Procedure 5.5 output R_1 and then the proof follows directly from Theorem 5.3 and Proposition 5.6. □

6 Numerical experiments

The main purpose of this section is to illustrate the performance of the S-APFA to solve quadratic programming (QP) problems over the cone of $n \times n$ symmetric positive semidefinite matrices S_n^+. More specifically, let a linear operator $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$ be given, the problem of our interest is

$$\min_{x\in S_n^+} \frac{1}{2}\|\mathcal{A}x - b\|_2^2. \tag{58}$$

The computational results were obtained using MATLAB R2016b on a 3.5 GHz intel Core i5 computer with 8GB of RAM and OS X system. In our experiment, the linear operator $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^m$ is sparse with entries uniformly distributed over $[0, 1]$, and the total number of nonzero entries is specified by the density parameter d. The vector $b \in \mathbb{R}^m$ was obtained as $b = \mathcal{A}s$, where $s := c^T c$ and the entries of the matrix $c \in \mathbb{R}^{n\times n}$ is uniformly distributed over $[0, 1]$. Thus, for each instance, the optimal value is 0 and s is a solution. We specify the four problems studied in Table 1.

Table 1: Instances for the QP problems

Instances	n	m	d	Instances	n	m	d
QP 1	2500	500	1e-4	QP 3	7500	1500	1e-5
QP 2	5000	1000	1e-4	QP 4	10000	2000	1e-5

In the implementation of the S-APFA method, we set $x^0 = 0$, $R_1 = \|\nabla f(x^0)\|/4L$ (see Corollary 4.3) and $B = 220$. The Lipschitz constant L was estimated using the power iteration method. Note that (18) is equivalent to

$$\langle w^+ - x^+, u - x^+\rangle \le \eta^+/\beta^+ \quad \forall u \in X \cap B(x^0, R_+), \tag{59}$$

where $w^+ = x^0 - A^+\nabla\Gamma^+/\beta^+$, i.e., x^+ is an approximate projection of w^+ on $X \cap B(x^0, R_+)$. At each outer iteration, the right-hand side of the inequality in (59) was replaced by $\max\{1, \eta^+/\beta^+\}$ to avoid an excessive number of inner iterations, and z_1 as in Step 2 of the APFA method was set as

$$z^1 = \frac{R_+(w^+ - \lambda_{min}I)}{\max\{R_+, \|w^+ - \lambda_{min}I\|_F\}},$$

where λ_{min} is the smallest eigenvalue of w^+ and $\|\cdot\|_F$ denotes the Frobenius Norm. Note that z_1 is the projection of the positive semidefinite matrix $\tilde{w}^+ := w^+ - \lambda_{min}I$ onto $B(x^0, R_+)$.

For comparison purpose, we run the accelerated projected gradient (APG) method corresponding to the variant of the PFA method in which projections onto the feasible set X are computed exactly, i.e., the iterate x^+ in Step 2 of the PFA method is replaced by the exact solution of (9). Note that, in this case, the assumption on the boundedness of X required in Section 2.2 is not necessary. We are also interested in analyzing the behavior of the S-APFA method to obtain a (ρ, ϵ)-solution (see Definition 5.1). For this, we considered (ρ_k, ϵ_k) as defined in Proposition 5.2.

Table 2 shows the performance of S-APFA and APG methods for solving the four QP problems described in Table 1. In Table 2, "Outer" and "Inner" are the total numbers of outer and inner iterations, respectively, "Time" is the CPU time in seconds and "$f(y^k)$" is the value of the objective function at the final iterate y^k.

Table 2: Perfomance of the S-APFA and APG methods for solving QPs 1-4.

		$f(y^k) - f^* \leq 10^{-2}$					$\max\{\rho_k/\rho_0, \epsilon_k/\epsilon_0\} \leq 10^{-1}$			
		S-APFA			APG		S-APFA			
Prob.	$f(y^0)$	Outer	Inner	Time	Outer	Time	Outer	Inner	Time	$f(y^k)$
QP 1	59.26	65	184	41.90	28	56.56	87	250	62.50	4.2e-3
QP 2	16.59	57	114	297.10	22	341.77	62	124	352.00	4.0e-3
QP 3	33.38	80	223	409.61	26	1338.21	86	241	487.74	5.2e-3
QP 4	15.32	61	122	837.82	21	2301.90	70	140	940,41	3.9e-3

From Table 2, we can see that the S-APFA method performed well for all instances of QP problems considered and our stopping criterion based on (ρ, ϵ)-solution is reliable and suitable if there is no knowledge of the optimal value f^*. Moreover, a relatively low accuracy in the (ρ, ϵ)-stopping criterion implied a considerable accuracy for the primal gap $f(y^k) - f^*$ (recall that $f^* = 0$). We also see that the S-APFA method required more outer iterations than the APG method to approximately solve (58). However, the latter method was much more time consuming than the former. The main reason is that at each iteration of the APG method demands to compute the exact solution of the projection subproblem (9), which requires the computation of the complete eigenvalue decomposition of a large matrix, whereas each subproblem of the CondG method requires to compute only its leading singular vector. The latter requirement is usually much less computationally expensive (see, for example, [10, 8] for more details). Therefore, we can conclude that the S-APFA method is suitable for solving large scale instances of the QP problem (58), being very competitive with other accelerated gradient schemes.

Appendix

A Properties of the sequences $\{A_k\}$

In this section, we establish some properties on the sequences $\{A_k\}$ defined in the PFA and S-APFA methods.

Lemma A.1. *Consider the sequence $\{A_k\}$ generated by the PFA method, then*

$$A_k \geq \frac{(k+1)k^2}{9L}. \tag{60}$$

Proof. It is easy to check that (60) holds for $k = 1$. Assume that (60) holds for some $k \geq 1$, we will now prove by induction that this estimate also holds for $k + 1$. From the update formula for A_{k+1}, the induction assumption and $\beta_k = k + 1$ we obtain

$$\begin{aligned} 9LA_{k+1} &= 9LA_k + \tfrac{9}{2}\beta_k + \tfrac{9}{2}\sqrt{\beta_k^2 + 4L\beta_k A_k} \\ &\geq (k+1)\left(k^2 + \tfrac{9}{2} + 3k\right) \\ &\geq (k+1)^2(k+2) \end{aligned}$$

It follows from the previous estimates that (60) holds for $k + 1$, completing the proof. □

Lemma A.2. *Consider the sequence $\{A_k\}$ generated by the S-APFA method, then*

$$A_k \geq \frac{(k+1)k^2}{50L} \quad \forall k \geq 1. \tag{61}$$

Proof. Using a root-finding algorithm, we obtain the first 10 terms of the sequence $\{A_k\}$ in the S-APFA method multiplied by L are 0.5, 0.8106, 1.3104, 2.1021, 3.3328, 5.2040, 7.9761, 11.9669, 17.5428 and 25.1050, which trivially imply that (61) holds for $k = 1, \ldots, 10$. Now, assume that (61) holds for some $k-1$ with $k \geq 10$, we are going to prove by induction that this estimate also holds for k. Let us define

$$P(x) = \frac{1}{2}\left(\frac{x}{x - A_{k-1}}\right)^2 - \frac{Lx}{k}, \qquad \bar{A} = \frac{(k+1)k^2}{50L}.$$

By using simple calculus, we obtain

$$P(\bar{A}) \geq \frac{1}{2}\left(\frac{(k+1)k}{3k-1}\right)^2 - \frac{k(k+1)}{50} \geq \frac{1}{18}(k+1)^2 - \frac{1}{50}k(k+1).$$

Since the last term above is greater than 3 for all $k \geq 9$, we have $P(\bar{A}) \geq 3$. From the updating formula of $\{A_k\}$ in the S-APFA method, we see that A_k is the unique solution of $P(x) = 3$ in $]A_{k-1}, \infty[$. Now, from the fact that P is decreasing in this interval and $P(\bar{A}) \geq 3$, we obtain $A_k \geq \bar{A}$ which concludes the induction proof.

□

B Proof of Proposition 5.6

Our goal in this section is to establish Proposition 5.6. For this, let us first observe that for the S-APFA method the auxiliary point x_r^1 as defined in Proposition 3.4 becomes

$$x_r^1 = \operatorname*{argmin}_{x \in X} \left\{ \Psi_1(u) := \frac{1}{2L} l_f(u, x^0) + \|u - x^0\|^2 \right\}. \tag{62}$$

In the following, we present two auxiliary results.

Lemma B.1. *Let x_r^1 be as defined in* (62)*. Then, the following hold*

a) $\|x_r^1 - x^0\| \leq \min\left\{2d_0, \frac{1}{4L}\|\nabla f(x^0)\|\right\}$;

b) $4L(x^0 - x_r^1) \in \partial_{\mu_1}(f + \mathcal{I}_X)(x^0)$, *where* $\mu_1 := \langle \nabla f(x^0), x^0 - x_r^1 \rangle$.

Proof. Using the first order optimality condition for problem (62), we have

$$\langle \frac{1}{2L}\nabla f(x^0) + 2(x_r^1 - x^0), u - x_r^1 \rangle \geq 0 \qquad \forall u \in X. \tag{63}$$

By taking $u = x^0$ in the previous inequality, we obtain $\|x_r^1 - x^0\|^2 \leq \langle \nabla f(x^0), x^0 - x_r^1 \rangle / 4L$, which, combined with Cauchy-Schwarz inequality, yields

$$\|x_r^1 - x^0\| \leq \|\nabla f(x^0)\|/(4L). \tag{64}$$

Now, let $x^* \in X^*$ be such that $d_0 = \|x^0 - x^*\|$. Since Ψ_1 is strongly convex and x_r^1 is its minimizer over X, we have

$$\begin{aligned} \Psi_1(x^*) &\geq \Psi_1(x_r^1) + \|x^* - x_r^1\|^2 \\ &= \tfrac{1}{2L}\left(l_f(x_r^1, x^0) + \tfrac{L}{2}\|x_r^1 - x^0\|^2\right) + \|x^* - x_r^1\|^2 \\ &\geq \tfrac{1}{2L} f(x_r^1) + \|x^* - x_r^1\|^2, \end{aligned} \tag{65}$$

where the last inequality is due to (4). On the other hand, since $l_f(\cdot, x^0) \leq f$, it follows from the definition of Ψ_1 that $\Psi_1(x^*) \leq f(x^*)/2L + d_0^2$, and then (65) implies that

$$\frac{1}{2L} f(x^*) + d_0^2 \geq \frac{1}{2L} f(x_r^1) + \|x^* - x_r^1\|^2.$$

Thus, since $f(x^*) \leq f(x_r^1)$ we obtain $\|x^* - x_r^1\| \leq d_0$, which implies by the triangle inequality that $\|x_r^1 - x^0\| \leq 2d_0$. Therefore, (a) follows by combining the last inequality with (64).

Now let us prove that (b) holds. It follows from (63) that, for any $u \in X$,

$$\langle \nabla f(x^0) + 4L(x_r^1 - x^0), u - x^0 \rangle \geq \langle \nabla f(x^0), x_r^1 - x^0 \rangle + 4L\|x_r^1 - x^0\|^2,$$

which implies from the definition of μ_1 and by letting $v := 4L(x^0 - x_r^1)$ that

$$\langle \nabla f(x^0), u - x^0 \rangle \geq \langle v, u - x^0 \rangle - \mu_1 \quad \forall u \in X.$$

Therefore, from the previous estimate and the convexity of f, we obtain

$$f(u) - f(x^0) \geq \langle v, u - x^0 \rangle - \mu_1 \quad \forall u \in X,$$

which trivially implies that (b) holds. □

Lemma B.2. *Let $M_1 > 2$, $R > 0$ and $z \in Z(R)$ be given, where $Z(R)$ is defined in* (54)*. Consider x_r^1 as defined in* (62)*. Then,*

a) if $\|z - x^0\| \geq R/2$ then $R \leq \frac{4\sqrt{2M_1}}{\sqrt{2M_1}-2} d_0$;

b) if $\|z - x^0\| \leq R/2$ then $x_r^1 \in B(x^0, R)$.

Proof. First, let us consider the auxiliary point

$$z_R := \operatorname*{argmin}_{u \in X \cap B(x^0, R)} \Psi_1(u), \tag{66}$$

where Ψ_1 is defined in (62). Since Ψ_1 is a quadratic function and $z \in Z(R)$, we have

$$\begin{aligned} \Psi_1(z_R) &\geq \Psi_1(z) - \tfrac{R^2}{M_1} + \|z_R - z\|^2 \\ &\geq \Psi_1(z_R) + \|z_R - z\|^2 - \tfrac{R^2}{M_1} + \|z_R - z\|^2, \end{aligned}$$

where in the last inequality we also applied the first order optimality condition for z_R. Hence, from the previous estimate, we obtain

$$\|z_R - z\| \leq R/\sqrt{2M_1}. \tag{67}$$

For proving (a), first note that using $\|z - x^0\| \geq R/2$ and (67), we obtain

$$\|z_R - x^0\| \geq \|z - x^0\| - \|z_R - z\| \geq \frac{R}{2} - \frac{R}{\sqrt{2M_1}} = \frac{(\sqrt{2M_1} - 2)R}{2\sqrt{2M_1}},$$

which implies that

$$R \leq \frac{2\sqrt{2M_1}}{\sqrt{2M_1} - 2}\|z_R - x^0\|.$$

On the other hand, from the definition of x_r^1 and z_R we easily see that

$$\|z_R - x^0\| \leq \|x_r^1 - x^0\|.$$

Therefore, statement (a) follows now from the last two inequalities and Proposition B.1(a).

Now, let us prove (b). Combining (67) with $\|z - x^0\| \leq R/2$ and $M_1 > 2$, we have

$$\|z_R - x^0\| \leq \|z_R - z\| + \|z - x^0\| \leq \frac{R}{\sqrt{2M_1}} + \frac{R}{2} < R.$$

Thus, z_R is an interior point of convex set $X \cap B(x^0, R)$. Hence, since Ψ_1 is convex, it follows from the definition of x_r^1 and z_R that $x_r^1 = z_R$, which implies that $x_r^1 \in B(x^0, R)$. □

We are now ready to prove Proposition 5.6.

Proof of Proposition 5.6. It is easy to see that the number of execution of the CondG method in Procedure 5.5 is at most

$$1 + \log_2 \frac{\|\nabla f(x^0)\|}{8L\delta}.$$

Therefore, (56) follows now from Proposition 2.1 (see also Remark 5.4).

Let us prove that if the second stopping criterion of Procedure 5.5 holds, then in fact x^0 is a (ρ, ϵ)-solution of (2). Indeed, the last computed R and $z \in Z(R)$ are such that $\|z - x^0\| \leq R/2$ and $R \leq \delta$. As a consequence, it follows from Lemma B.2(b) that $\|x_r^1 - x^0\| \leq \delta$. Hence, letting $v := 4L(x^0 - x_r^1)$ and $\mu_1 := \langle \nabla f(x^0), x^0 - x_r^1 \rangle$, Lemma B.1(b) implies that $v \in \partial_{\mu_1}(f + \mathcal{I}_X)(x^0)$. Therefore, since $\|v\| \leq 4L\delta \leq \rho$ and $\mu_1 \leq \|\nabla f(x^0)\|\delta \leq \epsilon$, the statement of the Proposition about x^0 follows. Now, assume that Procedure 5.5 output R_1. It is easy to see that the computed $z \in Z(R_1/2)$ satisfies $\|z - x^0\| \geq R_1/4$, and then the second inequality in (57) follows from Lemma B.2(a) with $R = R_1/2$. On the other hand, if Procedure 5.5 stop in the first iteration, then the first inequality in (57) follows from Lemma B.1(a), otherwise, as the computed $z \in Z(R_1)$ satisfies $\|z - x^0\| \leq R_1/2$, the desired inequality follows from Lemma B.2(b) with $R = R_1$. □

References

[1] Bach, F. Duality between subgradient and conditional gradient methods. *SIAM Journal on Optimization 25*, 1 (2015), 115–129.

[2] Beck, A., and Teboulle, M. A conditional gradient method with linear rate of convergence for solving convex linear systems. *Mathematical Methods of Operations Research 59*, 2 (2004), 235–247.

[3] Bredies, K., Lorenz, D., and Maass, P. A generalized conditional gradient method and its connection to an iterative shrinkage method. *Computational Optimization and Applications 42*, 2 (2009), 173–193.

[4] Dunn, J. C. Convergence rates for conditional gradient sequences generated by implicit step length rules. *SIAM Journal on Control and Optimization 18*, 5 (1980), 473–487.

[5] Frank, M., and Wolfe, P. An algorithm for quadratic programming. *Naval Research Logistics Quarterly 3*, 1-2 (1956), 95–110.

[6] Freund, R., and Grigas, P. New analysis and results for the FrankWolfe method. *Mathematical Programming* (2014), 1–32.

[7] Guzmán, C., and Nemirovski, A. On lower complexity bounds for large-scale smooth convex optimization. *Journal of Complexity 31*, 1 (2015), 1 – 14.

[8] Harchaoui, Z., Juditsky, A., and Nemirovski, A. Conditional gradient algorithms for norm-regularized smooth convex optimization. *Mathematical Programming 152*, 1–2 (2015), 75–112.

[9] He, Y., and Monteiro, R. An accelerated hpe-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM Journal on Optimization 26*, 1 (2016), 29–56.

[10] Jaggi, M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)* (2013), vol. 28, pp. 427–435.

[11] Kolossoski, O., and Monteiro, R.D.C. An accelerated non-euclidean hybrid proximal extragradient-type algorithm for convexconcave saddle-point problems. *Optimization Methods and Software 32*, 6 (2017), 1244–1272.

[12] Lan, G. The complexity of large-scale convex programming under a linear optimization oracle. *Available on http://www.optimization-online.org*.

[13] Lan, G., and Zhou, Y. Conditional gradient sliding for convex optimization. *SIAM Journal on Optimization 26*, 2 (2016), 1379–1409.

[14] Luss, R., and Teboulle, M. Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint. *SIAM Review 55*, 1 (2013), 65–98.

[15] Monteiro, R.D.C., Ortiz, C., and Svaiter, B.F. An adaptive accelerated first-order method for convex optimization. *Computational Optimization and Applications 64*, 1 (2016), 31–73.

[16] Nemirovski, A., and Yudin, D. *Problem complexity and method efficiency in optimization.* Wiley, 1983.

[17] Nesterov, Y.E. *Introductory lectures on convex optimization: a basic course.* Kluwer Academic Publ., Boston, 2004.

[18] Nesterov, Y.E. Smooth minimization of nonsmooth functions. *Mathematical Programming 103* (2005), 127–152.