52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Convergence Analysis of Primal Solutions in Dual First-order Methods

Jie Lu and Mikael Johansson

Abstract— **This paper studies primal convergence in dual first-order methods for convex optimization. Specifically, we consider Lagrange decomposition of a general class of inequality- and equality-constrained optimization problems with strongly convex, but not necessarily differentiable, objective functions. The corresponding dual problem is solved using a first-order method, and the minimizer of the Lagrangian computed when evaluating the dual function is considered as an approximate primal solution. We derive error bounds for this approximate primal solution in terms of the dual errors. Based on such error bounds, we show that the approximate primal solution converges to the primal optimum at a rate no worse than $O(1/\sqrt{k})$ if the projected dual gradient method is adopted and $O(1/k)$ if a fast gradient method is utilized, where k is the number of iterations. Finally, via simulation, we compare the convergence behavior of different approximate primal solutions in various dual first-order methods in the literature.**

I. Introduction

Lagrangian duality is a widely-used approach in large-scale optimization, especially when there are a few constraints that complicate an otherwise simple problem [1], [2]. Although many first-order methods can be applied to solve such problems directly in the primal space, the iteration cost can be very high since the projection onto the constraint set is often computationally difficult [3]. The corresponding dual problem has a more desirable structure: the dual constraint set has a simple form and the (sub)gradient of the dual function is relatively easy to evaluate. In addition, the dual function is often additive and suitable for distributed implementation, which has been exploited in a wide range of recent applications, including communication systems [4], [5], large-scale control [6], and multi-agent systems [7].

There are many practical and theoretical subtleties in using dual decomposition to generate optimal solutions to the engineering problems cited above. First, one needs to ensure that the dual optimal value agrees with the primal optimal value (*i.e.*, that there is no duality gap). Then, one needs to guarantee that the iterates generated by the dual optimization method converge to a dual optimum—for instance, the subgradient method with constant step-size achieves suboptimality only. Finally, in most engineering applications, we need to construct a primal solution from the dual iterates and ensure that its converges to a feasible and optimal point. Furthermore, we would also like to estimate how quickly the generated primal iterates converge. This motivates research on on-line construction of approximate primal solutions and studying their convergence properties.

Several results on the convergence properties of approximate primal solutions have been offered in the literature [6], [8]–[12]. At one extreme are results on dual decomposition of non-smooth convex problems, *e.g.* [10], where one typically has to form running averages of the iterates to construct a primal solution with guaranteed bounds on suboptimality and infeasibility at each iteration. At the other extreme are linearly constrained problems with strongly convex objective functions, *e.g.* [6], for which the dual function is Lipschitz continuous over the whole space. Not only does this simplify analysis, but it also allows the application of fast gradient methods [3], [13], [14] that achieve a convergence rate of $O(1/k^2)$ for the dual iterates.

In this paper, we consider a general class of convex optimization problem that covers the less explored middle ground between these two extremes. Specifically, we focus on a class of convex optimization problems with a strongly convex but not necessarily differentiable objective function and constraints defined by nonlinear inequalities, linear equalities, and set constraints. Notice that this class of optimization problems include those in [12] as a special case. We consider the unique minimizer of the Lagrangian for given dual variables as an approximate primal solution and study its convergence properties when the dual problem is solved using classical and fast gradient methods. We derive a number of error bounds for this approximate primal solution that relate its error in primal function value, distance to the primal optimum, and infeasibility to the error in dual function value and the distance between the dual variable and the dual optimal set. Then, by imposing assumptions on the smoothness of the inequality constraint functions, we construct a sufficient condition on the step-size to guarantee convergence of the dual iterates generated by the projected dual gradient method and prove that they converge sublinearly at a rate of order $O(1/k)$. It is worthwhile to mention that this convergence rate is not a known result, as the dual function only has *locally* Lipschitz continuous gradient *on the dual feasible set*. Further, this and the error bounds of the approximate primal solution lead to one of our main results, which says that the primal iterates converge to the primal optimum at a rate no worse than $O(1/\sqrt{k})$ in both Euclidean distance and primal function value. In addition, their infeasibility vanishes at a rate of the same order. We also consider a special case of the problem, where all the constraints are linear. In this case, the step-size is easier to choose, and we show that linear convergence of

J. Lu and M. Johansson are with ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: {jielu,mikaelj}@kth.se).

This work has been sponsored in part by the Swedish Research Council and the Swedish Foundation for Strategic Research.

both primal and dual iterates can be achieved under certain conditions. Finally, under the assumption of boundedness of the gradients of the nonlinear constraint functions, we analyze the convergence properties of the approximate primal solution when a fast gradient method is used to solve the dual problem. We demonstrate that the convergence rates of the primal iterates and their infeasibility are improved to $O(1/k)$. When all the constraint functions become affine, a simpler fast gradient method in [13], [14] can be applied, which yields the same dual and primal convergence rates.

The paper is organized as follows: Section II formulates the primal and dual problems. Section III provides the error bounds of the approximate primal solution and Section IV presents the convergence rates of the dual and primal iterates. In Section V, we compare different approximate primal solutions in various dual first-order methods in a numerical example. Finally, Section VI concludes the paper.

A. Notation

For a vector $q \in \mathbb{R}^n$, $q^{(i)} \in \mathbb{R}$, $i = 1, 2, \ldots, n$ denotes the ith element of q and $q^{(i:j)} \in \mathbb{R}^{j-i+1}$, $1 \le i < j \le n$ is the vector consisting of the ith, $(i+1)$th, ..., jth elements of q. We use $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ to represent the Euclidean norm, L_1 norm, and the infinity norm respectively. Furthermore, $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the largest and smallest eigenvalues of a real square matrix and $\sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^T A)}$ and $\sigma_{\min}(A) = \sqrt{\lambda_{\min}(A^T A)}$ are the maximum and minimum singular values of A. For any function $h : \mathbb{R}^n \to \mathbb{R}$, $\partial h(x) \subset \mathbb{R}^n$ denotes its subdifferential at $x \in \mathbb{R}^n$. If h is differentiable at x, then $\partial h(x) = \{\nabla h(x)\}$, where $\nabla h(x)$ is the gradient of h at x and its ith element is represented by $\nabla^{(i)} h(x)$. For any multi-variable function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $\partial_x h(x, y) \subset \mathbb{R}^n$ denotes its subdifferential at x with respect to the first argument. For any $Q \subset \mathbb{R}^n$, let $\operatorname{relint} Q$ be its relative interior, $\operatorname{conv} Q$ its convex hull, and $\mathcal{P}_Q$ the projection onto Q.

II. Problem Formulation

We consider the following optimization problem with inequality, equality, and set constraints:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g^{(i)}(x) \le 0, \quad i = 1, 2, \ldots, m, \\ & Ax + b = 0, \\ & x \in X. \end{array} \tag{1}$$

Here, $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function, $g^{(i)} : \mathbb{R}^n \to \mathbb{R}$, $\forall i \in \{1, 2, \ldots, m\}$ represent nonlinear constraint functions, $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$ encode the linear equality constraints, and $X \subset \mathbb{R}^n$ is a closed and convex set. In addition, let the following assumption hold:

Assumption 1. Problem (1) satisfies the following:

(a) The objective function f is strongly convex with convexity parameter $\theta > 0$ over X.
(b) Each $g^{(i)}$, $i \in \{1, 2, \ldots, m\}$ is convex and Lipschitz continuous with Lipschitz constant $L_i > 0$ over X.
(c) There exists $\tilde{x} \in \operatorname{relint} X$ such that $g^{(i)}(\tilde{x}) < 0$ $\forall i \in \{1, 2, \ldots, m\}$ and $A\tilde{x} + b = 0$.
(d) The number of inequality and equality constraints is not zero, *i.e.*, $m + p \neq 0$. Also, if $m = 0$, at least one element of A is nonzero.

Assumptions 1(a), 1(b), and 1(c) guarantee that there is a unique optimal solution $x^\star$ to problem (1) and that the optimal value $f^\star = f(x^\star)$ is finite. In addition, they ensure that problem (1) has no duality gap, *i.e.*, $f^\star$ is equal to the optimal value $d^\star$ of the corresponding dual problem, and that the dual optimal set $D^\star$ is nonempty [2, Prop. 5.3.2].

To formulate the dual of (1), we first introduce the Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^{m+p} \to \mathbb{R}$ associated with (1):

$$\mathcal{L}(x, u) = f(x) + \sum_{i=1}^{m} u^{(i)} g^{(i)}(x) + \left(u^{(m+1:m+p)}\right)^T (Ax + b).$$

The dual function $d : \mathbb{R}^{m+p} \to \mathbb{R}$ can then be expressed as

$$d(u) = \min_{x \in X} \mathcal{L}(x, u) = f(\bar{x}(u)) + \sum_{i=1}^{m} u^{(i)} g^{(i)}(\bar{x}(u)) + \left(u^{(m+1:m+p)}\right)^T (A\bar{x}(u) + b),$$

where $\bar{x}(u) \in \arg\min_{x \in X} \mathcal{L}(x, u)$ and the dual problem is

$$\begin{array}{ll} \underset{u \in \mathbb{R}^{m+p}}{\text{maximize}} & d(u) \\ \text{subject to} & u \in D \triangleq \{u \in \mathbb{R}^{m+p} : u^{(i)} \ge 0 \\ & \qquad \forall i \in \{1, 2, \ldots, m\}\}. \end{array} \tag{2}$$

Since d is concave, the dual problem (2) is a convex optimization problem. Moreover, $\mathcal{L}(\cdot, u)$ is strongly convex over X for every $u \in D$, so $\bar{x}(u)$ exists and is unique. Furthermore, $\bar{x}(u) = x^\star$ if $u = u^\star$ for some dual optimal solution $u^\star \in D^\star$ [15, Prop. 6.1.1]. This makes $\bar{x}(u)$ a legitimate candidate for an *approximate primal solution* to (1) as long as $u \in D$.

The next two results establish boundedness of $\bar{x}(u)$ and differentiability of the dual function:

Lemma 1. *Consider problem* (1) *under Assumption 1. For any compact set $S \subset D$, the set $\{\bar{x}(u) : u \in S\}$ is bounded.*

Proposition 1. *Consider problem* (1) *under Assumption 1. Then, the dual function d is differentiable at every point in D. Moreover, for any $u \in D$,*

$$\nabla d(u) = [g^{(1)}(\bar{x}(u)), \cdots, g^{(m)}(\bar{x}(u)), (A\bar{x}(u) + b)^T]^T.$$

In this setting, the goal of this paper is to quantify how close the approximate primal solution $\bar{x}(u)$ is to the primal optimum $x^\star$ for any dual feasible point $u \in D$ and to derive the convergence rate of $\bar{x}(u)$ when the dual problem (2) is solved using common first-order methods. More specifically, we are interested in the primal optimality $|f(\bar{x}(u)) - f^\star|$, the distance $\|\bar{x}(u) - x^\star\|$ to the primal optimum $x^\star$, and the primal infeasibility

$$\Delta(\bar{x}(u)) = \left(\|A\bar{x}(u) + b\|^2 + \sum_{i=1}^{m} \left(\max\{0, g^{(i)}(\bar{x}(u))\}\right)^2\right)^{1/2}$$

of the approximate primal solution $\bar{x}(u)$.

III. Primal error bounds

In this section, we bound the errors of the approximate primal solution $\bar{x}(u) \in X$ in terms of errors of the dual variable $u \in D$. To this end, we introduce the following notation: for any $u \in D$, let

$$\gamma(u) = \frac{\max\{1, \sqrt{2m}\}}{\theta} \max\Big\{\sigma_{\max}(A), \sup_{q \in G(u)} \|q\|\Big\}, \quad (3)$$

where $G(u) = \cup_{i=1}^m \partial g^{(i)}(\bar{x}(u)) \subset \mathbb{R}^n$. Clearly, $\gamma(u) \geq 0$. From [15, Prop. 4.2.1] we know that $G(u)$ is a compact set, so $\sup_{q \in G(u)} \|q\| < \infty$, which implies that $0 \leq \gamma(u) < \infty$. Furthermore, if $\gamma(\tilde{u}) = 0$ for some $\tilde{u} \in D$, then it can be shown that $\bar{x}(u) = \bar{x}(\tilde{u})$ $\forall u \in D$. Due to this property, in the rest of the paper, we exclude this trivial case and assume without loss of generality that $\gamma(u) > 0$ $\forall u \in D$.

Next, consider the following lemma:

Lemma 2. *Consider problem* (1) *under Assumption 1. Then, for any $u_1, u_2 \in D$,*

$$\|\bar{x}(u_1) - \bar{x}(u_2)\| \leq \min\{\gamma(u_1), \gamma(u_2)\}\|u_1 - u_2\|,$$

where $\gamma(u_2) \in (0, \infty)$ is defined in (3).

Lemma 2 allows to relate the error $\|\bar{x}(u) - x^\star\|$ of the approximate primal solution to the error $\|u - u^\star\|$ of the dual variable. The next theorem bounds $\|\bar{x}(u) - x^\star\|$ by virtue of the error $d^\star - d(u)$ in the dual function value.

Theorem 1. *Consider problem* (1) *under Assumption 1. Then, for any $u \in D$,*

$$\|\bar{x}(u) - x^\star\| \leq \gamma(u^\star)\|u - u^\star\|, \quad \forall u^\star \in D^\star,$$

$$\|\bar{x}(u) - x^\star\| \leq \sqrt{\frac{d^\star - d(u)}{\theta}},$$

where $\gamma(u^\star) \in (0, \infty)$ is defined in (3).

Having derived error bounds on the Euclidean distance between $\bar{x}(u)$ and $x^\star$, we turn our attention to the error in primal function value, *i.e.*, $f(\bar{x}(u)) - f^\star$. Since $\bar{x}(u)$ may not be primal feasible, $f(\bar{x}(u)) - f^\star$ can be negative. Therefore, both upper and lower bounds on $f(\bar{x}(u)) - f^\star$ are necessary. For convenience, for any compact subset $S \subset D$, define

$$L(S) = \sup_{u \in S} \gamma(u)\Big(\sigma_{\max}^2(A) + \sum_{i=1}^m L_i^2\Big)^{1/2} > 0. \quad (4)$$

From Lemma 1 and [15, Prop. 4.2.3], the boundedness of S implies that the set $\cup_{u \in S} G(u)$ is bounded. Consequently, $L(S) < \infty$. With $L(S)$ defined, we provide the following:

Proposition 2. *Consider problem* (1) *under Assumption 1. Then, ∇d is locally Lipschitz continuous over D, i.e., for any compact set $S \subset D$, $\|\nabla d(u_1) - \nabla d(u_2)\| \leq L(S)\|u_1 - u_2\|$ $\forall u_1, u_2 \in S$, where $L(S) \in (0, \infty)$ is defined in* (4). *Moreover, if S is convex, then*

$$d(u_2) - d(u_1) - \nabla d(u_1)^T(u_2 - u_1) \geq -\frac{L(S)}{2}\|u_1 - u_2\|^2.$$

The local Lipschitz continuity of ∇d established in Proposition 2 allows to guarantee upper bounds on the primal error $f(\bar{x}(u)) - f^\star$ and the primal infeasibility $\Delta(\bar{x}(u))$ of any $\bar{x}(u)$ with u in some compact subset of D:

Theorem 2. *Consider problem* (1) *under Assumption 1. Let $S \subset D$ be compact and define*

$$\Phi(S) = \mathrm{conv}\Big(\Big\{\mathcal{P}_D[u' + \beta\nabla d(u')] : u' \in S,\ \beta \in [0, \tfrac{1}{L(S)}]\Big\}\Big). \quad (5)$$

Then, for any $u \in S$,

$$f(\bar{x}(u)) - f^\star \leq \Big(\|u\|_\infty \sqrt{2L(\Phi(S))(m+p)} + 2\sqrt{d^\star - d(u)}\Big)\sqrt{d^\star - d(u)},$$

$$f(\bar{x}(u)) - f^\star \geq -\|u^\star\|\sqrt{2L(\Phi(S))(d^\star - d(u))},$$

$$\Delta(\bar{x}(u)) \leq \sqrt{2L(\Phi(S))(d^\star - d(u))},$$

where $L(\Phi(S)) \in (0, \infty)$ is defined in (4) *and $u^\star \in D^\star$ is any dual optimal solution.*

The bounds above depend on the compact set $\Phi(S)$. Thus, unlike Theorem 1, the results in Theorem 2 only hold *locally*, which stems from the fact that ∇d is locally Lipschitz continuous over D. Nevertheless, under the assumption below, similar conclusions can be established *globally* over D:

Assumption 2. The set $\cup_{u \in D} G(u)$ is bounded.

Assumption 2 is satisfied when, for instance, the constraint set X is compact or the constraint functions $g^{(i)}$ $\forall i \in \{1, 2, \ldots, m\}$ are affine. Under Assumption 2, $\sup_{q \in \cup_{u \in D} G(u)} \|q\| < \infty$, which leads to the Lipschitz continuity of ∇d over D and the following error bounds:

Corollary 1. *Consider problem* (1) *under Assumptions 1 and 2. Then, ∇d is Lipschitz continuous over D with constant*

$$\tilde{L} = \sup_{u \in D} \gamma(u)\Big(\sigma_{\max}^2(A) + \sum_{i=1}^m L_i^2\Big)^{1/2} \in (0, \infty),$$

i.e., $\|\nabla d(u_1) - \nabla d(u_2)\| \leq \tilde{L}\|u_1 - u_2\|$ $\forall u_1, u_2 \in D$. Moreover, for any $u \in D$, the bounds in Theorem 2 hold with $L(\Phi(S))$ replaced by $\tilde{L}$.

In the final part of this section, we study a special case of (1) where all constraints are linear and derive sharper and more explicit error bounds. Specifically, we consider

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & A'x + b' \leq 0, \\ & Ax + b = 0, \\ & x \in X, \end{array} \quad (6)$$

where $A' \in \mathbb{R}^{m \times n}$ and $b' \in \mathbb{R}^m$. For convenience, let $\tilde{A} = [(A')^T, A^T]^T \in \mathbb{R}^{(m+p) \times n}$ and $\tilde{b} = [(b')^T, b^T]^T \in \mathbb{R}^{m+p}$.

If f is convex over the whole $\mathbb{R}^n$, X is a polyhedral set, and the constraint set of problem (6) is nonempty, then Assumption 1(c) can be removed [2, Prop. 5.2.1]. Also, Assumption 2 is automatically satisfied for this problem due to the linearity of the constraints. The Lagrangian and the dual functions of problem (6) reduce to $\mathcal{L}(x,u) = f(x) + u^T(\tilde{A}x + \tilde{b})$ and $d(u) = f(\bar{x}(u)) + u^T(\tilde{A}\bar{x}(u) + \tilde{b})$ respectively. Besides, $\bar{x}(u)$ exists and is unique for any $u \in \mathbb{R}^{m+p}$ and d is differentiable over $\mathbb{R}^{m+p}$.

Following Lemma 2, we show in the corollary below that the distance between approximate primal solutions is proportional to that between the dual variables:

Corollary 2. *Consider the linearly constrained problem* (6) *under Assumption 1. Then, for any $u_1, u_2 \in \mathbb{R}^{m+p}$,*

$$\|\bar{x}(u_1) - \bar{x}(u_2)\| \leq \frac{\sigma_{\max}(\tilde{A})}{\theta}\|u_1 - u_2\|.$$

Since the inequality constraints are linear in (6), the bound provided in Corollary 2 is independent of u_1 and u_2 and tighter than that in Lemma 2, *i.e.*, $\sigma_{\max}(\tilde{A})/\theta \leq \gamma(u_2)$. When there are no inequality constraints, *i.e.*, $m = 0$, the two bounds are equal. The linearity of the constraints also allows us to show that the dual function has Lipschitz continuous gradient over the whole space and provide global bounds on the errors in primal function value and primal infeasibility:

Proposition 3. *Consider the linearly constrained problem* (6) *under Assumption 1. Then ∇d is Lipschitz continuous with Lipschitz constant $\sigma^2_{\max}(\tilde{A})/\theta$ over $\mathbb{R}^{m+p}$. Moreover, for any $u \in D$ and any $u^\star \in D^\star$,*

$$\begin{aligned} f(\bar{x}(u)) - f^\star &\leq \|u\|\sigma_{\max}(\tilde{A})\sqrt{\frac{2(d^\star - d(u))}{\theta}}, \\ f(\bar{x}(u)) - f^\star &\geq -\|u^\star\|\sigma_{\max}(\tilde{A})\sqrt{\frac{2(d^\star - d(u))}{\theta}} \\ \Delta(\bar{x}(u)) &\leq \sigma_{\max}(\tilde{A})\sqrt{\frac{2(d^\star - d(u))}{\theta}}. \end{aligned} \tag{7}$$

Note that the upper bound in (7) is not specialized from and is tighter than the corresponding bound in Theorem 2.

IV. Convergence of Primal Iterates

In this section, we use the connections between primal and dual errors that are built in Section III to analyze the convergence properties of the primal iterates when the projected gradient method and some fast gradient methods are employed to solve the dual problem.

A. Primal convergence in the projected dual gradient method

We first consider the projected gradient method. Let $(u_k)_{k=0}^{\infty} \subset D$ be a sequence generated by

$$u_{k+1} = \mathcal{P}_D[u_k + \alpha\nabla d(u_k)], \tag{8}$$

from an arbitrary initial point $u_0 \in D$. To derive the convergence rate of the primal iterates $(\bar{x}(u_k))_{k=0}^{\infty}$, we impose the following assumption:

Assumption 3.

(a) The constraint functions $g^{(i)}$ $\forall i \in \{1, 2, \ldots, m\}$ are differentiable at every point in X.

(b) There exists $\tilde{u} \in \mathbb{R}^{m+p}$ such that $\tilde{u}^{(i)} < 0$ $\forall i \in \{1, 2, \ldots, m\}$ and $\mathcal{L}(\cdot, \tilde{u})$ is strongly convex over X.

To satisfy Assumption 3(b), it suffices that each $\nabla g^{(i)}$ is Lipschitz continuous over X. Using the $\tilde{u}$ in Assumption 3, we define the set

$$\tilde{D} = \{u \in \mathbb{R}^{m+p} : u^{(i)} \geq \tilde{u}^{(i)} \ \forall i \in \{1, 2, \ldots, m\}\} \supset D.$$

For any $u \in \tilde{D}$, $\mathcal{L}(\cdot, u)$ is strongly convex over X and thus $\bar{x}(u)$ uniquely exists.

Before presenting our main result on the dual and primal convergence rates of the projected dual gradient method (8), we also need to introduce some additional notation. First, for any convex and compact set $S \subset D$, let

$$\hat{\eta}(S) = \sup_{v_1, v_2 \in S} \max_{i \in \{1,2,\ldots,m\}} \frac{-|\nabla^{(i)} d(v_1) - \nabla^{(i)} d(v_2)|}{\tilde{u}^{(i)}}.$$

If $m = 0$, $\hat{\eta}(S) = -\infty$. Otherwise, since ∇d is Lipschitz continuous over S, $0 \leq \hat{\eta}(S) < \infty$. Based on $\hat{\eta}(S)$, define

$$\eta(S) = \begin{cases} \hat{\eta}(S), & \text{if } \hat{\eta}(S) > 0, \\ \frac{\sigma^2_{\max}(A)}{\theta}, & \text{otherwise.} \end{cases} \tag{9}$$

Clearly, $\eta(S) \in [0, \infty)$. If $\eta(S) = 0$, then by Assumption 1, A is a zero matrix and $\nabla d(u)$ is constant over S.

In addition, given $S \subset D$, we define a compact set $\Psi(S) \subset \tilde{D}$ in (10). If $\eta(S) > 0$, then for any $\eta \geq \eta(S)$, $\eta^{-1}(\nabla^{(i)} d(u) - \nabla^{(i)} d(v)) \geq \tilde{u}^{(i)}$, $\forall i \in \{1, 2, \ldots, m\}$ $\forall u, v \in S$. Thus, $S \subset \Psi(S) \subset \tilde{D}$. Moreover, $\Psi(S)$ is compact and convex.

Under Assumption 3, the definitions of $G(u)$ and $L(S)$ in Section III can be extended to hold for any $u \in \tilde{D}$ and any compact set $S \subset \tilde{D}$. Also, Lemma 1 still holds when D is replaced by $\tilde{D}$, which implies that $0 < L(\Psi(S)) < \infty$. With the above definitions, we provide the following theorem:

Theorem 3. *Consider problem* (1) *under Assumptions 1 and 3. Let $(u_k)_{k=0}^{\infty} \subset D$ be a sequence generated by the projected dual gradient method* (8)*. Also, let $u^\star \in D^\star$ and $D_0 = \{u \in D : \|u - u^\star\| \leq \|u_0 - u^\star\|\} \subset D$. Moreover, let $\Phi(D_0) \subset D$ be defined in* (5)*, $L(\Phi(D_0)) \in (0, \infty)$ in* (4)*, $\eta(D_0) \in [0, \infty)$ in* (9)*, $\Psi(D_0) \subset \tilde{D}$ in* (10)*, and $L(\Psi(D_0)) \in (0, \infty)$ in* (4)*. If*

$$0 < \alpha < \begin{cases} \frac{2}{L(\Psi(D_0))}, & \textit{if } L(\Psi(D_0)) > \eta(D_0), \\ 4\left(\frac{1}{\eta(D_0)} - \frac{L(\Psi(D_0))}{2\eta(D_0)^2}\right), & \textit{otherwise}, \end{cases} \tag{11}$$

then for any $k \geq 0$,

$$d^\star - d(u_k) \leq \frac{R_0}{1 + kR_0\delta\rho^{-1}}, \tag{12}$$

$$\|\bar{x}(u_k) - x^\star\| \leq \left(\frac{R_0\theta^{-1}}{1 + kR_0\delta\rho^{-1}}\right)^{1/2}, \tag{13}$$

$$\begin{aligned} &f(\bar{x}(u_k)) - f^\star \leq (\|u^\star\| + \|u_0 - u^\star\|) \\ &\cdot \left(\frac{2(m+p)L(\Phi(D_0))R_0}{1 + kR_0\delta\rho^{-1}}\right)^{1/2} + \frac{2R_0}{1 + kR_0\delta\rho^{-1}}, \end{aligned} \tag{14}$$

$$\Psi(S) = \begin{cases} \mathrm{conv}\Big(\Big\{u + \beta(\nabla d(u) - \nabla d(v)) : u, v \in S,\ \beta \in [0, \frac{1}{\eta(S)}]\Big\}\Big), & \text{if } \eta(S) > 0, \\ S & \text{otherwise.} \end{cases} \tag{10}$$

$$f(\bar{x}(u_k)) - f^\star \geq -\|u^\star\| \left(\frac{2L(\Phi(D_0))R_0}{1 + kR_0\delta\rho^{-1}}\right)^{1/2}, \tag{15}$$

$$\Delta(\bar{x}(u)) \leq \Big(\frac{2L(\Phi(D_0))R_0}{1 + kR_0\delta\rho^{-1}}\Big)^{1/2}, \tag{16}$$

where $R_0 = d^\star - d(u_0) \in (0, \infty)$, $\rho = (\sup_{u \in D_0} \|\nabla d(u)\| + \|u_0 - u^\star\|/\alpha)^2 \in (0, \infty)$, *and* $\delta = 1/\alpha - L(\Psi(D_0))/2 \in (0, \infty)$.

Theorem 3 says that under Assumptions 1 and 3 as well as a proper step-size choice (11), the dual function value at the dual iterates $(u_k)_{k=0}^\infty$ converges to $d^\star$ at a rate $O(1/k)$. Note that this result extends earlier analyses of the projected gradient method for functions with globally Lipschitz continuous gradient (e.g., [16]) to a class of functions with locally Lipschitz continuous gradient over a closed and convex set. Furthermore, this result implies that the primal iterates $(\bar{x}(u_k))_{k=0}^\infty$ converge at a rate no worse than $O(1/\sqrt{k})$ in primal optimality, distance to $x^\star$, and primal infeasibility.

When problem (1) reduces to problem (6), Assumption 3(b) holds for every $\tilde{u} \in \mathbb{R}^{m+p}$ with $\tilde{u}^{(i)} < 0$ $\forall i \in \{1, 2, \ldots, m\}$. By taking $\tilde{u}^{(i)} < 0$ $\forall i \in \{1, 2, \ldots, m\}$ sufficiently small, we can make $\hat{\eta}$ equal to $\sigma_{\max}^2(A)/\theta$. Thus, $\eta(D_0) \leq \sigma_{\max}^2(A)/\theta \leq \sigma_{\max}^2(\tilde{A})/\theta$. This, along with the fact that ∇d is Lipschitz continuous with constant $\sigma_{\max}^2(\tilde{A})/\theta$, implies that $0 < \alpha < 2\theta/\sigma_{\max}^2(\tilde{A})$ satisfies (11). This step-size condition coincides with the standard one used to guarantee the convergence of the gradient methods when the objective function has globally Lipschitz continuous gradient [3]. The following corollary presents convergence rates of the dual and primal iterates for problem (6) with such step-sizes:

Corollary 3. *Consider the linearly constrained problem* (6) *under Assumption 1. Let* $(u_k)_{k=0}^\infty \subset D$ *be a sequence generated by the projected dual gradient method* (8) *and let* $u^\star \in D^\star$. *If* $0 < \alpha < 2\theta/\sigma_{\max}^2(\tilde{A})$, *then for any* $k \geq 0$, (12)–(16) *hold with* $L(\Phi(D_0)) = \sigma_{\max}^2(\tilde{A})/\theta$.

It is known that the projected gradient method is able to converge linearly when the objective function is strongly convex. Theorem 1 thus suggests that primal iterates achieve linear convergence if the dual function is strongly concave. Next, we impose an assumption on problem (6) and show that it guarantees strong concavity of the dual function.

Assumption 4. The subgradients of f satisfy a Lipschitz condition with constant $M > 0$, *i.e.*, for any $x_1, x_2 \in X$, $\|\tilde{\nabla} f(x_1) - \tilde{\nabla} f(x_2)\| \leq M\|x_1 - x_2\|$, $\forall\ \tilde{\nabla} f(x_1) \in \partial f(x_1)$, $\forall\ \tilde{\nabla} f(x_2) \in \partial f(x_2)$. Also, the matrix $\tilde{A}$ in problem (6) has full row rank and $X = \mathbb{R}^n$.

Proposition 4. *Consider the linearly constrained problem* (6) *under Assumptions 1 and 4. Then,* d *is strongly concave with concavity parameter* $-\theta\sigma_{\min}^2(A)/M^2 < 0$.

Based on Proposition 4, the linear primal and dual convergence rates of the projected dual gradient method (8) for solving problem (6) can now be established:

Proposition 5. *Consider the linearly constrained problem* (6) *under Assumptions 1 and 4. Let* $(u_k)_{k=0}^\infty \subset D$ *be a sequence generated by* (8) *with* $0 < \alpha \leq 2M^2\theta/(\theta^2\sigma_{\min}^2(\tilde{A}) + M^2\sigma_{\max}^2(\tilde{A}))$. *Then, for any* $k \geq 0$,

$$\|u_k - u^\star\| \leq q^k\|u_0 - u^\star\|,$$

$$\|\bar{x}(u_k) - x^\star\| \leq q^k\frac{\sigma_{\max}(\tilde{A})}{\theta}\|u_0 - u^\star\|,$$

where

$$q = \Big(1 - \frac{2\alpha\theta\sigma_{\min}^2(\tilde{A})\sigma_{\max}^2(\tilde{A})}{\theta^2\sigma_{\min}^2(\tilde{A}) + M^2\sigma_{\max}^2(\tilde{A})}\Big)^{1/2} \in (0, 1)$$

and $u^\star$ *is the unique dual optimizer. Moreover,* q *reaches its minimum value for* $\alpha = 2M^2\theta/(\theta^2\sigma_{\min}^2(\tilde{A}) + M^2\sigma_{\max}^2(\tilde{A}))$.

B. Primal convergence in dual fast gradient methods

In this subsection, we move on to dual fast gradient methods. To the best of the authors' knowledge, all the existing fast gradient methods require Lipschitz continuity of the objective function over at least the feasible region in order to reach a convergence rate of $O(1/k^2)$. Hence, through out this subsection, we assume that Assumptions 1 and 2 hold, so that d is Lipschitz continuous over D with constant $\tilde{L}$ defined in Corollary 1. We also assume that (an upper bound on) $\sup_{u \in D} \gamma(u)$ and thus $\tilde{L}$ are known.

We consider the 1-memory fast gradient method from [13] for solving the dual problem (2). To start with, define the following: Let $h : \mathbb{R}^{m+p} \to \mathbb{R}$ be differentiable on an open set containing D and let $Q(u, v) = h(u) - h(v) - \nabla h(v)^T(u - v)$ $\forall u \in \mathbb{R}^{m+p}$ $\forall v \in D$. Assume h is strictly convex and satisfies $Q(u, v) \geq \|u - v\|^2/2$ $\forall u, v \in D$. Also, let $\ell_{-d}(u, v) = -d(v) - \nabla d(v)^T(u - v)$ $\forall u, v \in D$. For completeness, we provide the algorithm below:

Algorithm 1 (Algorithm 1, [13])**.**
Initialization: Before $k = 0$:
1) Let $\beta_0 = 1$ and choose $u_0, w_0 \in D$.
Operation: At each time $k \geq 0$:
2) Choose a closed convex set $U_k \subset \mathbb{R}^{m+p} : U_k \cap D^\star \neq \emptyset$.
3) Let $v_k = (1 - \beta_k)u_k + \beta_k w_k$,
$w_{k+1} = \arg\min_{u \in U_k \cap D} \ell_{-d}(u, v_k) + \beta_k\tilde{L}Q(u, w_k)$,
$\hat{u}_{k+1} = (1 - \beta_k)u_k + \beta_k w_{k+1}$.
4) Choose u_{k+1} be such that $\ell_{-d}(u_{k+1}, v_k) + \frac{\tilde{L}}{2}\|u_{k+1} - v_k\|^2 \leq \ell_{-d}(\hat{u}_{k+1}, v_k) + \frac{\tilde{L}}{2}\|\hat{u}_{k+1} - v_k\|^2$.

5) Choose $\beta_{k+1} \leq 2/(k+3)$. ∎

In Algorithm 1, the variables u_k, v_k, and w_k remain in D all the time. Also, different ways of choosing U_k and u_k in Steps 2 and 4 can be found in [13]. Note that the above algorithm is indeed specialized from the more general Algorithm 1 in [13], in order to achieve the primal and dual convergence rates in the following proposition:

Proposition 6. *Consider problem* (1) *under Assumptions 1 and 2. Let $(u_k)_{k=0}^{\infty} \subset D$ be a sequence generated by Step 4) in Algorithm 1 and $u^\star \in D^\star$. Then, for any $k \geq 1$,*

$$
\begin{aligned}
d^\star - d(u_k) &\leq \frac{4\tilde{L}Q(u^\star, w_0)}{(k+1)^2},\\
\|\bar{x}(u_k) - x^\star\| &\leq \frac{(4\tilde{L}Q(u^\star, w_0)/\theta)^{1/2}}{k+1},\\
f(\bar{x}(u_k)) - f^\star &\leq \frac{\tilde{L}\|u_k\|_\infty \big(8(m+p)Q(u^\star, w_0)\big)^{1/2}}{k+1}\\
&\quad + \frac{8\tilde{L}Q(u^\star, w_0)}{(k+1)^2}, \qquad (17)\\
f(\bar{x}(u_k)) - f^\star &\geq -\frac{\tilde{L}\|u^\star\|(8Q(u^\star, w_0))^{1/2}}{k+1},\\
\Delta(\bar{x}(u_k)) &\leq \frac{\tilde{L}(8Q(u^\star, w_0))^{1/2}}{k+1}.
\end{aligned}
$$

where $\tilde{L}$ is defined in Corollary 1.

Proposition 6 says that Algorithm 1 yields $O(1/k^2)$ convergence rate of $(u_k)_{k=1}^{\infty}$ in dual optimality. In addition, the distance between $\bar{x}(u_k)$ and $x^\star$ as well as the primal infeasibility of $\bar{x}(u_k)$ vanishes at a rate no worse than $O(1/k)$. As Algorithm 1 does not guarantee that $(u_k)_{k=1}^{\infty}$ is bounded, (17) says nothing about the convergence rate of $f(\bar{x}(u_k))$. Nevertheless, if problem (1) has only inequality constraints, the dual optimal set is bounded [10] and so is $(u_k)_{k=1}^{\infty}$. This leads to the following proposition, which states that in the absence of equality constraints, $f(\bar{x}(u_k))$ converges to $f^\star$ at a rate $O(1/k)$ after some finite time:

Proposition 7. *Consider problem* (1) *under Assumptions 1 and 2. Suppose $p = 0$. Let $(u_k)_{k=0}^{\infty} \subset D$ be a sequence generated by Step 4) in Algorithm 1. Also, let $\tilde{x} \in \mathbb{R}^n$ satisfy Assumption 1(c), $u^\star \in D^\star$, and $\bar{u} \in D \backslash D^\star$. Then,*

$$
\begin{aligned}
f(\bar{x}(u_k)) - f^\star &\leq \frac{2L(d(\bar{u}) - f(\tilde{x}))\big(2(m+p)Q(u^\star, w_0)\big)^{1/2}}{(k+1)(\max_{i\in\{1,2,\ldots,m\}} g^{(i)}(\tilde{x}))}\\
&\quad + \frac{8LQ(u^\star, w_0)}{(k+1)^2}, \quad \forall k > \Big(\frac{4LQ(u^\star, w_0)}{d^\star - d(\bar{u})}\Big)^{1/2}.
\end{aligned}
$$

Remark 1. In addition to Algorithm 1, *i.e.*, the 1-memory fast gradient method in [13], the dual problem (2) can also be solved by the ∞-memory fast gradient method in [13], which would produce similar primal and dual convergence rates as those in Propositions 6 and 7. Due to space limitation and since such analysis is very similar to that in Propositions 6 and 7, we omit this algorithm in the paper.

Recall that for the linearly constrained problem (6), the dual function d is differentiable and has Lipschitz continuous gradient with constant $\sigma_{\max}^2(\tilde{A})/\theta$ over the entire space. In this case, the following algorithm, which has a simpler form than Algorithm 1, can be adopted to solve the dual problem:

Algorithm 2 (Algorithm 2, [13])**.**
Initialization: Before $k = 0$:
1) Let $\beta_0 = \beta_{-1} = 1$ and choose $u_0 = u_{-1} \in D$.

Operation: At each time $k \geq 0$:
2) Choose a closed convex set $U_k \subset \mathbb{R}^{m+p} : U_k \cap D^\star \neq \emptyset$.
3) Let $v_k = u_k + \beta_k(1/\beta_{k-1} - 1)(u_k - u_{k-1})$ and
$u_{k+1} = \arg\min_{u\in U_k \cap D} \ell_{-d}(u, v_k) + \frac{\sigma_{\max}^2(\tilde{A})}{2\theta}\|u - v_k\|^2$.
4) Choose $\beta_{k+1} \leq 2/(k+3)$. ∎

If we pick $U_k = \mathbb{R}^{m+p}$ and $\beta_{k+1} = (\sqrt{\beta_k^4 + 4\beta_k^2} - \beta_k^2)/2$, then Algorithm 2 becomes the fast iterative shrinkage-thresholding algorithm (FISTA) in [14]. It can be shown that the primal and dual convergence rates of Algorithm 2 are very similar to those of Algorithm 1 in Propositions 6 and 7.

The above dual fast gradient methods are capable of increasing the primal convergence rate $O(1/\sqrt{k})$ produced by the projected dual gradient method (8) to $O(1/k)$. However, it is more complicated to implement these algorithms. Moreover, the projected dual gradient method is suitable for distributed implementation, while the dual fast gradient methods can hardly do that.

V. NUMERICAL EXAMPLE

In this section, we compare the dual and primal convergence performance of the projected dual gradient method (8), the dual fast gradient method in Algorithm 2, and the double smoothing method [11] in a numerical example.

We consider the following MPC problem, which has a very similar form as the one formulated in [6]:

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) = \tfrac{1}{2}x^T Hx + t^T x + \gamma\|Px - s\|_1\\
\text{subject to} \quad & A_1 x + b_1 \leq 0,\\
& A_2 x + b_2 = 0,\\
& x \in \{y \in \mathbb{R}^n : |y^{(i)}| \leq r_i \ \forall i = 1, 2, \ldots, n\},
\end{aligned}
$$

where $H \in \mathbb{R}^{n\times n}$ being positive definite, $t \in \mathbb{R}^n$, $\gamma > 0$, $P \in \mathbb{R}^{q\times n}$, $s \in \mathbb{R}^q$, $A_1 \in \mathbb{R}^{m\times n}$, $b_1 \in \mathbb{R}^m$, $A_2 \in \mathbb{R}^{p\times n}$, $b_1 \in \mathbb{R}^p$, $r_i > 0$ are randomly generated with $n = 10$, $q = 5$, $m = 3$, and $p = 2$.

For the projected dual gradient method (8), we choose the step-size $\alpha = 2\lambda_{\min}(H)/\lambda_{\max}(A_1^T A_1 + A_2^T A_2) \times 99\%$, which satisfies the step-size condition in Corollary 3. For the fast gradient method in Algorithm 2, we choose the parameters U_k and β_k to be such that this method reduces to the fast iterative shrinkage-thresholding algorithm (FISTA) in [14]. For the double smoothing method [11], since the linear constraint of the problem class that this method can handle is in the form of $\mathcal{A}x \in T$ with $\mathcal{A}$ being a linear operator and T being a compact set, we put $\mathcal{A} = A_2$, $T = \{b_2\}$, and view $\{y \in \mathbb{R}^n : A_1 y + b_1 \leq 0,\ |y^{(i)}| \leq r_i \ \forall i = 1, 2, \ldots, n\}$ as its set constraint. Moreover, since one smoothing parameter in

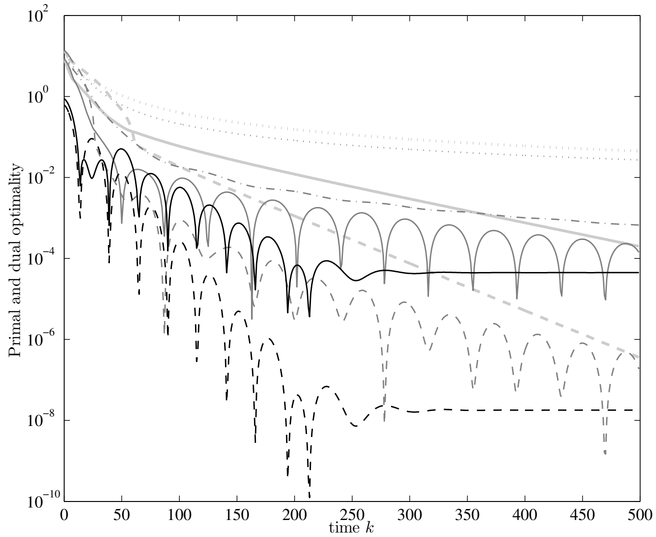

Fig. 1. Primal optimality of approximate primal solutions and dual optimality of dual iterates (The light grey, grey, and black dashed curves represent the dual optimality $d^\star - d(u_k)$ of u_k in the projected dual gradient method, FISTA, and the double smoothing method respectively. The light grey, grey, and black solid curves represent the primal optimality $|f(\bar{x}(u_k)) - f^\star|$ of $\bar{x}(u_k)$ in the projected dual gradient method, FISTA, and the double smoothing method respectively. The light grey and grey dotted curves represent the primal optimality $|f(\tilde{x}_k) - f^\star|$ of $\tilde{x}_k$ in the projected dual gradient method and FISTA respectively. The grey dash-dotted curve represent the primal optimality $|f(\hat{x}_k)-f^\star|$ of $\hat{x}_k$ in FISTA.).

the double smoothing method relies on an upper bound on some dual optimum, we adopt its practical implementation version in [11], which starts with an initial guess of this upper bound and repeatedly applying the method to a sequence of doubly smoothed dual functions with increasing guess on the upper bound until a correct guess is achieved. We choose the desired accuracy ϵ of the method to be 0.05.

In addition to the approximate primal solution $\bar{x}(u_k)$ studied in this paper, we also consider in the simulation the average $\tilde{x}_k \triangleq \sum_{\ell=0}^{k} \bar{x}(u_\ell)/k$ of the primal iterates as in [10] for the projected dual gradient method and FISTA, and the running average $\hat{x}_k \triangleq (\sum_{\ell=0}^{k} \beta_\ell^{-1}\bar{x}(u_\ell))/(\sum_{\ell=0}^{k} \beta_\ell^{-1})$ with the weights being $1/\beta_\ell$ as in [12] for FISTA. For the double smoothing method, u_k is a sequence generated by a fast gradient method in [3, Sec. 2.2.1] applied to the smoothed dual; the approximate primal solution $\bar{x}(u_k)$ for this specific problem is the unique minimizer of the Lagrangian of the original problem.

Figure 1 compares the convergence of the dual iterates generated by the three methods mentioned above, alongside with various choices for the approximate primal solution. Generally speaking, the dual convergence rate is faster than the primal in all of the three methods. Also, the primal iterates have faster convergence than their averages in the projected dual gradient method and FISTA. The projected dual gradient method converges slower than the other two methods in the dual optimality, primal optimality, distance to $x^\star$, and infeasibility. FISTA and the double smoothing method have comparable performance, but the double smoothing method has the drawback that it only guarantees a prespecified target accuracy and does not ensure asymptotic convergence.

VI. CONCLUSION

This paper studies the primal convergence properties of common dual first-order methods for solving a class of convex optimization problems with a strongly convex objective function and inequality, equality, and set constraints. We select the unique minimizer of the Lagrangian with the Lagrange multipliers given by the current dual iterate as the approximate primal solution. A number of error bounds for this approximate primal solution are provided. Moreover, both the dual and primal convergence rates for the projected dual gradient method and some dual fast gradient methods are derived. Finally, the convergence of different choices of approximate primal solution in various dual first-order methods are compared in a numerical example.

REFERENCES

[1] L. S. Lasdon, *Optimization Theory For Large Systems*. New York, NY: Macmillan, 1970.

[2] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

[3] Y. Nesterov, *Introductory lectures on Convex Optimization: A Basic Course*. Norwell, MA: Kluwer Academic Publishers, 2004.

[4] S. H. Low and D. E. Lapsley, "Optimization flow control, i: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[5] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Transactions on Communications*, vol. 52, no. 7, pp. 1136–1144, 2004.

[6] P. Giselsson, M. D. Doan, T. Keviczky, B. Schutter, and A. Rantzer, "Accelerated gradient methods and dual decomposition in distributed model predictive control," *Automatica*, vol. 49, no. 3, pp. 829–833, 2013.

[7] A. Nedić and A. Ozdaglar, "Cooperative distributed multi-agent optimization," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds. Cambridge University Press, 2010, pp. 340–386.

[8] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods," in *Lecture Notes for EE364B*. Stanford, CA: Stanford University, 2006-2007.

[9] I. Necoara and J. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Transactions on Automatic Control*, vol. 53, no. 11, pp. 2674–2679, 2008.

[10] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.

[11] O. Devolder, F. Glineur, and Y. Nesterov, "Double smoothing technique for large-scale linearly constrained convex optimization," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 702–727, 2012.

[12] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," to appear in *IEEE Transactions on Automatic Control*, 2013.

[13] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," submitted to *SIAM Journal on Optimization*.

[14] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[15] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[16] E. S. Levitin and B. T. Polyak, "Constrained minimization problems," *USSR Computational Mathematics and Mathematical Physics*, vol. 6, pp. 1–50, 1966, english version in Zh. Vychisl. Mat. mat. Fiz., vol.6, pp.787-823, 1966.