ORIGINAL ARTICLE

Certification aspects of the fast gradient method for solving the dual of parametric convex programs

Stefan Richter · Colin Neil Jones · Manfred Morari

Abstract This paper examines the computational complexity certification of the fast gradient method for the solution of the dual of a parametric convex program. To this end, a lower iteration bound is derived such that for all parameters from a compact set a solution with a specified level of suboptimality will be obtained. For its practical importance, the derivation of the smallest lower iteration bound is considered. In order to determine it, we investigate both the computation of the worst case minimal Euclidean distance between an initial iterate and a Lagrange multiplier and the issue of finding the largest step size for the fast gradient method. In addition, we argue that optimal preconditioning of the dual problem cannot be proven to decrease the smallest lower iteration bound. The findings of this paper are of importance in embedded optimization, for instance, in model predictive control.

Keywords Fast gradient method · Certification · Lagrange relaxation

1 Introduction

Motivation In recent years, state-of-the-art optimization methods have changed numerous fields of engineering, primarily in the context of offline synthesis and

S. Richter (✉) · M. Morari
Automatic Control Laboratory, ETH Zürich, 8092 Zürich, Switzerland
e-mail: richters@ee.ethz.ch

M. Morari
e-mail: morari@ee.ethz.ch

C. N. Jones
Automatic Control Laboratory, EPFL Lausanne, 1015 Lausanne, Switzerland
e-mail: colin.jones@epfl.ch

 Springer

analysis. More recently, they have entered the area of embedded systems, where data is processed under real-time constraints on low-cost platforms. One of the challenges in embedded optimization is the *a priori* certification of the computational complexity, i.e. the derivation of an upper bound on the number of iterations that is valid for all data encountered during operation.

Certification is a crucial prerequisite in risk-of-loss applications, e.g. onboard trajectory planning in space missions (Blackmore et al. 2010), in safety-critical environments such as power systems (Fuchs et al. 2011), for a verified operation of engineering systems, for instance, in embedded model predictive control (Bleris and Kothare 2005) and real-time audio clipping (Defraene et al. 2012), as well as for the selection of appropriate hardware during the design phase. In control, certification is particularly challenging as solution times are restricted to lie within a sampling interval of the control loop. For systems with fast dynamics, sampling intervals are typically in the range of milli- and microseconds which underlines the need for a method that features both fast convergence and an upper iteration bound that is close to the practically observed number of iterations.

An upper iteration bound stems from a non-asymptotic convergence rate analysis of the solution method and is the very iteration count that ensures a pre-specified level of suboptimality. From here on, this upper bound is denoted a *lower iteration bound*, indicating that any greater iteration count will meet the suboptimality criterion. Note that an *upper iteration bound* is to be seen in the context of an actual implementation of the method, i.e. it puts a bound on the loop counter, whereas our preferred terminology of a *lower iteration bound* highlights the bound's origin in the convergence analysis of the method.

In this paper we restrict ourselves to convex programs since powerful solution methods and convergence results are available (See e.g. Nesterov 2004a). Interior point methods (Nesterov and Nemirovskii 1994) are popular methods for such problems and McGovern (2000) treats certification aspects of a short step, primal-dual path following method. The reported lower iteration bounds are up to two orders of magnitude off from the practically observed number of iterations which agrees with the widely accepted viewpoint that bounds for interior point methods lack expressiveness.

Active set methods (Bertsekas 1999) belong to the dominant solution methods for linearly constrained programs. These methods perform well in practice, but as their convergence rate is unknown, they are not eligible for certification. In contrast, gradient methods (Nesterov 2004a; Bertsekas 1999) come with a convergence rate analysis and provide expressive lower iteration bounds (e.g. see the example in Boyd and Vandenberghe (2004, §9.3.2) and compare the convergence rate with Polyak (1987, §3.1.2, Theorem 2).

In this paper, we will focus on the *fast* gradient method which was first derived in Nesterov (1983). Many variants of this method exist and we refer the reader to Tseng (2008) where an attempt is made to unify existing methods in the more general context of accelerated proximal gradient methods.

Problem setup and previous work In this paper, we discuss the computational aspects of deriving a lower iteration bound for a constant step size fast gradient method for the solution of the parametric convex program

$$f^*(b) \triangleq \min \ f(x) = \frac{1}{2} x^T H x + g^T x \tag{1}$$
$$\text{s.t. } Ax = b, \ x \in \mathbb{X},$$

via its dual.[1] In (1) the strongly convex quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ is minimized over the intersection of an affine set, given by $(A, b) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m$, and a nonempty closed convex set $\mathbb{X} \subseteq \mathbb{R}^n$. We assume set $\mathbb{X}$ to be 'simple' so that Euclidean projection can be evaluated at low computational cost. Simple sets include the Euclidean ball, simplex, the LP-, SOCP- and the SDP cone.

On embedded systems, problem (1) is solved with varying data, usually from measurements, which justifies denoting it a *parametric program*. In this work, we consider the right hand side b of the equality constraint as the only parameter. This important class of parametric programs is found, for instance, in linear quadratic model predictive control (Rawlings and Mayne 2009, §1.3).

In the authors' previous work (Richter et al. 2012), certification of (1) without equality constraints is considered when vector g of the objective is the parameter. For this setup, benefits from warm-starting could be quantified for model predictive control and system properties related to the lower iteration bound. The bounds derived for real-world problems were found to match the observed number of iterations within a factor of two to four. Recently, this work was extended in Richter et al. (2011) for parametric problems of the type addressed in this paper in the framework of Lagrange relaxation of the equality constraint (cf. Lemaréchal 2001). The reported bounds for the fast gradient method solving the dual problem are off by up to three orders of magnitude from the observed number of iterations; the main reason for this being the conservative, yet computationally tractable approach to bound the worst case Euclidean distance between the initial dual iterate and a Lagrange multiplier.

Contribution and outline This paper discusses the main certification issues for solving the parametric convex problem (1) in the framework of Lagrange relaxation using the fast gradient method and poses related open questions. We start in Sect. 2 with a review on Lagrange relaxation and the fast gradient method and refer to related work in the area of certification. Thereafter, Sect. 3 defines the certification problem of interest and summarizes basic assumptions that hold throughout the paper. In Sect. 4, we first define the smallest lower iteration bound and then proceed to discuss the computation of its defining entities. The first entity is the worst case minimal Euclidean distance between an initial iterate and a Lagrange multiplier whose importance beyond certification will be motivated (Sect. 4.1). Section 4.2 elaborates on computing the second important entity which is the smallest Lipschitz constant of the dual gradient (this section follows Richter et al. 2011, §IV.A). Since the Lipschitz constant determines the

---

[1] From a certification point of view, a constant step size is not a limitation as for gradient methods in convex optimization advanced step size rules, e.g. exact line search, do not exhibit better convergence rate results (Polyak 1987, §3.1.2).

step size in the fast gradient method, the latter investigation is of practical importance even when certification is not an issue. After that, in Sect. 4.3, we prove that preconditioning of the dual problem does not decrease the smallest lower iteration bound. Finally, we state some open questions with regard to the findings in this paper in Sect. 5.

2 Review and related work

Lagrange relaxation Assume that for a parameter b the feasible set of (1) is non-empty such that by strong convexity of f and closedness of the feasible set a unique minimizer $x^*(b)$ exists. In order to obtain it by Lagrange relaxation, we eliminate the complicating equality constraint, define the *dual function*

$$d(\lambda; b) \triangleq \min_{x \in \mathbb{X}} f(x) + \lambda^T (Ax - b) \ , \tag{2}$$

with multiplier $\lambda \in \mathbb{R}^m$ and solve the concave *dual problem*

$$d^*(b) \triangleq \sup_{\lambda \in \mathbb{R}^m} d(\lambda; b) \ . \tag{3}$$

If the supremum is attained (see Remark 1 in Sect. 3 for sufficient conditions), we denote the closed convex set of dual optimal solutions as

$$\Lambda^*(b) = \arg \max_{\lambda \in \mathbb{R}^m} d(\lambda; b) \tag{4}$$

and refer to any $\lambda^*(b) \in \Lambda^*(b)$ as a Lagrange multiplier. If strong duality holds, i.e. $d^*(b) = f^*(b)$, then by strong convexity of f and Rockafellar (1997, Corollary 28.1.1), the primal minimizer can be recovered from $x^*(\lambda^*(b))$ where

$$x^*(\lambda) = \arg \min_{x \in \mathbb{X}} f(x) + \lambda^T (Ax - b) \ . \tag{5}$$

In this paper, we use the fast gradient method to solve the dual problem (3); the required gradient $\nabla d (\lambda; b)$ is obtained according to the next theorem.

Theorem 1 *The dual function $d(\lambda; b)$ has a Lipschitz continuous gradient $\nabla d (\lambda; b) = Ax^*(\lambda) - b$, i.e. for each parameter b and any $\lambda_1, \lambda_2 \in \mathbb{R}^m$ we have*

$$\|\nabla d (\lambda_1; b) - \nabla d (\lambda_2; b)\| \leq L \|\lambda_1 - \lambda_2\| \tag{6}$$

with Lipschitz constant $L = \|A\|^2 / \lambda_{\min}(H)$, where $\|A\|$ denotes the maximum singular value of A and $\lambda_{\min}(H)$ is the smallest eigenvalue of H.

Proof The first statement follows from Danskin's Theorem in Bertsekas (1999, Proposition B.25) that applies if we modify the (potentially non-compact) feasible set $\mathbb{X}$ in (2) at every $\lambda \in \mathbb{R}^m$ by adding the convex set constraint $x \in \{x \in \mathbb{R}^n \mid f(x) + \lambda^T (Ax - b) \leq d(\lambda; b) + \epsilon'\}, \epsilon' > 0$, which contains $x^*(\lambda)$ and is compact by strong convexity of f. Lipschitz continuity of the gradient is provided by Nesterov (2004b, Theorem 1). □

It follows that for solving the dual or *outer* problem (3) using the fast gradient method, the *inner* problem (2) needs to be solved in every iteration in order to determine the dual gradient.

Fast gradient method We consider the constant step size scheme II in Nesterov (2004a, §2.2.1) which is given in Algorithm 2.1. The step size $1/L$ in line 2 is determined by the Lipschitz constant L of the gradient $\nabla d(\lambda; b)$, whereas the upper bound on the number of iterations is given in terms of the lower iteration bound $i_{\min}$ which will be formally defined in Sect. 3.

Algorithm 2.1 Fast Gradient Method for the Dual Problem (3)

Require: Initial iterate $\lambda_0 \in \mathbb{R}^m$, $\mu_0 = \lambda_0$, $\alpha_0 = (\sqrt{5} - 1)/2$, Lipschitz constant L
1: **for** $i = 0 \to (i_{\min} - 1)$ **do**
2:     $\lambda_{i+1} = \mu_i + 1/L \cdot \nabla d(\mu_i; b)$ {cf. Theorem 1}
3:     $\alpha_{i+1} = \frac{\alpha_i}{2}\left(\sqrt{\alpha_i^2 + 4} - \alpha_i\right)$
4:     $\beta_i = (\alpha_i(1 - \alpha_i))/(\alpha_i^2 + \alpha_{i+1})$
5:     $\mu_{i+1} = \lambda_{i+1} + \beta_i(\lambda_{i+1} - \lambda_i)$
6: **end for**
7: **return** $\lambda_{i_{\min}}$

Related work Lan and Monteiro (2009) investigate the certification of a problem similar to (1) with a general smooth convex function f. The authors derive a lower iteration bound for an augmented Lagrangian approach that ensures a smooth (augmented) dual function (see e.g. Bertsekas 1999, §4.2). It is assumed that the inner problem is solved by the fast gradient method whereas the outer problem is solved by the standard gradient method. The derived bound on the overall number of fast gradient iterations holds under inexact gradients obtained from suboptimal solutions of the inner problem. A guess-and-check procedure circumvents the computation of the distance between the initial dual iterate and the set of Lagrange multipliers which is an important entity for determining the lower iteration bound. Consequently, no *a priori* lower iteration bound as considered in this paper can be computed.

Devolder et al. (2012) propose to smooth the dual function and to add a strongly concave quadratic such that a lower iteration bound on the required fast gradient iterations to obtain a nearly primal feasible, suboptimal solution can be derived. The cost of solving the inner problems is thereby neglected.

In Doan et al. (2011) the relaxed constraints are linear inequalities. By constraint tightening and the theory developed in Nedić and Ozdaglar (2009), a lower iteration bound for obtaining a primal feasible iterate is derived. The bound depends on a Slater point and is valid for a projected gradient method solving the outer problem while the inner one is solved by conjugate gradients.

3 Definitions and assumptions

In this section, we define the considered certification problem in terms of a dual ϵ-solution and a lower iteration bound and state assumptions that implicitly hold throughout the paper. Let us start with the definition of some sets.

Definition 1 (*set of admissible parameters*) The closed convex set of admissible parameters $\mathbb{B} \subseteq \mathbb{R}^m$ contains all right hand side vectors b of the equality constraint such that problem (1) is feasible, i.e. $b \in \mathbb{B} \iff f^*(b) < \infty$.

Definition 2 (*set of certified parameters*) The set of certified parameters $\mathbb{B}_c \subseteq \mathbb{B}$ contains all instances $b \in \mathbb{B}$ for which a lower iteration bound according to Definition 4 is to be derived.

Definition 3 (*dual ϵ-solution*) Let $b \in \mathbb{B}_c$. For a specified $\epsilon > 0$, a dual ϵ-solution $\lambda_\epsilon \in \mathbb{R}^m$ satisfies $d^*(b) - d(\lambda_\epsilon; b) \leq \epsilon$.

Definition 4 (*lower iteration bound*) We denote $i_{\min}$ a lower iteration bound if for any number of iterations of the fast gradient method, $i \geq i_{\min}$, a dual ϵ-solution is retrieved for every parameter $b \in \mathbb{B}_c$ and a common $\epsilon > 0$.

Definition 5 (*computational complexity certification*) Consists in determining a lower iteration bound $i_{\min}$ for a given set $\mathbb{B}_c$.

Assumption 1 For all parameters $b \in \mathbb{B}_c$, a Lagrange multiplier $\lambda^*(b)$ exists and strong duality holds.

Assumption 2 The inner problem (2) can be solved exactly.

Assumption 3 The certified set of parameters $\mathbb{B}_c$ is compact and convex.

Remark 1 Assumption 1 holds true if for every $b \in \mathbb{B}_c$ a feasible point $\bar{x}(b)$ in the relative interior of $\mathbb{X}$ exists, i.e. $A\bar{x}(b) = b$, $\bar{x}(b) \in ri(\mathbb{X})$ (cf. Bertsekas 2009, Prop. 5.3.3). A milder premise holds if $\mathbb{X}$ has a polyhedral structure (Bertsekas 2009, Prop. 5.3.6). Assumption 2 is satisfied for important problem instances of model predictive control (see Richter et al. 2011, §V), network resource allocation and others (cf. Nedić and Ozdaglar 2009, §2.2). See Devolder et al. (2011) and Schmidt et al. (2011) for convergence of the fast gradient method in case the inner problem cannot be solved exactly.

4 Obtaining the smallest lower iteration bound

In this section, we investigate the aspects related to the computation of a lower iteration bound in the sense of Definition 4 when using the variant of the fast gradient method given in Algorithm 2.1. For its practical importance, the focus will be laid on deriving the *smallest* lower iteration bound as defined next.

Theorem 2 *Let the initial iterate of the fast gradient method be determined by function* $\lambda_0 : \mathbb{R}^m \to \mathbb{R}^m$ *for every parameter* $b \in \mathbb{B}_c$, *and let* L^* *be the smallest Lipschitz constant of the gradient of the dual function. The smallest lower iteration bound for the fast gradient method in Algorithm 2.1 is given by*

$$i_{\min}^* = \max \left\{ \left\lceil 2\sqrt{\frac{L^* \Delta_d^2}{\epsilon}} - 2 \right\rceil, 0 \right\},$$

where $\Delta_d^2 \triangleq \sup_{b \in \mathbb{B}_c} h^*(b)$ and

$$h^*(b) \triangleq \min_{\lambda \in \Lambda^*(b)} \| \lambda - \lambda_0(b) \|^2 . \tag{7}$$

Proof Follows from Theorem 2.2.3 in Nesterov (2004a). □

The problem of determining Δ_d^2, which is the worst case minimal squared distance between an initial iterate and a Lagrange multiplier, is addressed in Sect. 4.1. In Sect. 4.2 the smallest Lipschitz constant L^* is derived under mild assumptions, whereas Sect. 4.3 investigates preconditioning of the dual problem in order to further decrease the smallest lower iteration bound.

Remark 2 Even if we are able to compute the smallest lower iteration bound, this does not necessarily imply that it is tight. Actually, it is unknown if there exists a problem instance for which the bound in Theorem 2 is tight.

4.1 Worst case distance between an initial iterate and a Lagrange multiplier

For the computation of Δ_d^2, we investigate the properties of function $h^*(b)$ in (7) based on Theorem 3 below. According to it, $h^*(b)$ is a closed convex function under certain assumptions, however, the satisfiability of these assumptions will be shown to depend on how the set of Lagrange multipliers $\Lambda^*(b)$ in the definition of $h^*(b)$ is represented. For a representation derived from a zero-duality gap formulation of the optimality conditions, the assumptions can provably never be met (Sect. 4.1.1), whereas this is not true for a representation based on support functions (Sect. 4.1.2). Section 4.1.3 finally elaborates on computational aspects as related to the previous findings.

Before stating the main theorem of this section, we note that knowing Δ_d^2 is also of interest for extending the approaches in Lan and Monteiro (2009), Devolder et al. (2012) to parametric problems as well as in the framework of exact penalty functions (Bertsekas 1999, §5.4.5). In order to illustrate the latter, let $\lambda_0(b) \equiv 0$ for all $b \in \mathbb{B}_c$ so that from the existence of Lagrange multipliers and the Minimax Theorem in Rockafellar (1997, Corollary 37.3.2)

$$f^*(b) = \max_{\|\lambda\| \le \Delta_d} \min_{x \in \mathbb{X}} f(x) + \lambda^T (Ax - b) = \min_{x \in \mathbb{X}} \max_{\|\lambda\| \le \Delta_d} f(x) + \lambda^T (Ax - b)$$

$$= \min_{x \in \mathbb{X}} f(x) + \Delta_d \cdot \|Ax - b\|, \quad b \in \mathbb{B}_c,$$

which can be solved, e.g. by the fast gradient method if smoothing (Nesterov 2004b) is applied to replace the nonsmooth norm by a smooth approximation.

Theorem 3 *Let $\lambda_0(b) = Kb + \hat{\lambda}_0$, where $K \in \mathbb{R}^{m \times m}$ is a symmetric matrix and $\hat{\lambda}_0 \in \mathbb{R}^m$. Furthermore, let $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a closed jointly convex function for which it holds that*

(i) $\phi(\cdot, \lambda)$ is strongly convex for every $\lambda \in \mathbb{R}^m$,
(ii) $\phi(x, \lambda) \ge -\lambda^T b$ for all $(x, \lambda) \in \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid Ax = b, x \in \mathbb{X}\}$,

and consider the convex program parametrized in $b \in \mathbb{B}_c$

$$p^*(b) = \min \|\lambda - \lambda_0(b)\|^2 \tag{8}$$

$$s.t. \ \phi(x, \lambda) + \lambda^T b \leq 0 \tag{IC}$$
$$Ax = b, \ x \in \mathbb{X} \ .$$

Assume that there exists a function $v^* : \mathbb{V} \to \mathbb{R}_+$, $\mathbb{B}_c \subseteq \mathbb{V} \subseteq \mathbb{R}^m$, which assigns to every parameter $b \in \mathbb{B}_c$ a nonnegative Lagrange multiplier $v^*(b)$ for the inequality constraint IC. If the supremum

$$v_c^* = \sup_{b \in \mathbb{B}_c} v^*(b) \tag{9}$$

exists, then

S1 *if $K \succeq v_c^*/4 \cdot I$, $p^*(b)$ is a closed convex function for all $b \in \mathbb{B}_c$,*
S2 *if $K \preceq v_c^*/4 \cdot I$, $p^*(b)$ is the sum of a concave quadratic function and a closed convex function for all $b \in \mathbb{B}_c$.*

Proof Choose any $b \in \mathbb{B}_c$. We denote the dual problem to (8) as

$$q^*(b) = \sup_{v \geq 0} \ \min_{\substack{Ax=b, \, x \in \mathbb{X} \\ \lambda \in \mathbb{R}^m}} \|\lambda - \lambda_0(b)\|^2 + v\left(\phi(x, \lambda) + \lambda^T b\right) \ , \tag{10}$$

and infer strong duality, i.e. $p^*(b) = q^*(b)$, from Gol'šteĭn (1972, Theorem 2) as the Lagrangian in (10) is strongly convex in (x, λ) (cf. Assumption (i)).

By Assumption (ii) there does not exist a Slater point for problem (8), hence by Gauvin's Theorem (Gauvin 1977) the set of Lagrange multipliers for the inequality constraint IC is either empty or nonempty but unbounded. Theorem 3 assumes the latter. By strong convexity and the assumption that the supremum v_c^* in (9) exists, it thus follows

$$p^*(b) = \min_{\substack{Ax=b, \, x \in \mathbb{X} \\ \lambda \in \mathbb{R}^m}} \|\lambda - \lambda_0(b)\|^2 + v_c^*\left(\phi(x, \lambda) + \lambda^T b\right), \quad b \in \mathbb{B}_c \ , \tag{11}$$

since by unboundedness of the set of Lagrange multipliers the scalar v_c^* is a viable Lagrange multiplier for *all* parametric problems with parameter $b \in \mathbb{B}_c$.

Note that the latter argument can be made alternatively via the theory of exact penalty functions (see e.g. Bertsekas 1999, §5.4.5).

In order to obtain Statement S1, we can verify using Schur's Lemma that $K \succeq v_c^*/4 \cdot I$ is necessary and sufficient for $\|\lambda - Kb - \hat{\lambda}_0\|^2 + v_c^*\lambda^T b$ being jointly convex in (λ, b) which in turn is sufficient for the objective in (11) to be jointly convex in (x, λ, b). Based on joint convexity of the objective and standard arguments (cf. Bertsekas 1999, §5.4.4), convexity of $p^*(b)$ follows.

Closeness follows from lower semicontinuity of $p^*(b)$ at every $b \in \mathbb{B}_c$ established by Bank et al. (1982, Theorem 4.3.4). In order for this to hold true, strong convexity

of the objective in (11) in (x, λ) for every $b \in \mathbb{B}_c$ and closedness of the convex set $\mathbb{X}$ are of importance as they imply boundedness of the set of minimizers and further that $\mathbb{X}$ can be represented as the intersection of all closed halfspaces containing it.

For the proof of Statement S2, we note that for every $b \in \mathbb{B}_c$

$$p^*(b) = \nu_c^* b^T \left(K - \tfrac{\nu_c^*}{4} \cdot I \right) b + \psi^*(b) \ , \tag{12}$$

where

$$\psi^*(b) \triangleq \nu_c^* \hat{\lambda}_0^T b + \min_{\substack{Ax=b,\, x \in \mathbb{X} \\ \lambda \in \mathbb{R}^m}} \left\| \lambda - \left(K - \tfrac{\nu_c^*}{2} \cdot I \right) b - \hat{\lambda}_0 \right\|^2 + \nu_c^* \phi(x, \lambda), \quad (13)$$

so that for $K \preceq \nu_c^*/4 \cdot I$ the (closed) quadratic term in (12) is negative semidefinite. Convexity and closedness of $\psi^*(b)$ follow from a similar reasoning as used in the proof of Statement S1. $\qquad\square$

Remark 3 For the case where neither $K \succeq \nu_c^*/4 \cdot I$ nor $K \preceq \nu_c^*/4 \cdot I$, it follows from the proof of Statement S2 that $p^*(b)$ is the sum of an indefinite quadratic function and a closed convex function.

Based on the next theorem, two representations of the set of Lagrange multipliers $\Lambda^*(b)$ will be derived in Sects. 4.1.1 and 4.1.2 so that problem (7) can be posed as (8). Interestingly enough, we will prove that only the latter representation allows one to validate the assumptions of Theorem 3.

Theorem 4 (Adapted from Bertsekas 2009, Prop. 5.3.3b) *For each parameter $b \in \mathbb{B}_c$ there holds $f^*(b) = d^*(b)$ and $(x^*(b), \lambda^*(b))$ is a primal/dual optimal solution pair if and only if $x^*(b)$ is primal feasible and*

$$x^*(b) = \arg\min_{x \in \mathbb{X}} f(x) + \lambda^*(b)^T (Ax - b) \ . \tag{14}$$

4.1.1 Zero-duality-gap-based representation of $\Lambda^*(b)$

This representation follows from the sufficiency condition of Theorem 4, i.e.

$$\Lambda^*(b) = \left\{ \lambda \in \mathbb{R}^m \mid \exists z \in \mathbb{X} \cap \{x \mid Ax = b\} : f(x) - d(\lambda; b) \le 0 \right\}, \ \ b \in \mathbb{B}_c \quad (\text{R1})$$

To render the constraints convex, the equality enforcing a zero-duality gap is replaced by an inequality inR1, legitimated by $f(x) \ge d(\lambda; b)$ for all primal/dual feasible pairs (x, λ). The next theorem proves that except for a trivial case the premise of Theorem 3 cannot be validated for representation R1.

Theorem 5 *Consider representation R1 of the closed convex set of Lagrange multipliers $\Lambda^*(b)$. Let λ_0 be any function that maps $\mathbb{R}^m$ into $\mathbb{R}^m$. If $\lambda_0(b) \in \Lambda^*(b)$ for*

every $b \in \mathbb{B}_c$, *we have* $\nu_c^* = 0$ *trivially, else, the premise of Theorem 3 cannot be validated.*

Proof We identify function $\phi(x, \lambda)$ in Theorem 3 as $\phi(x, \lambda) = f(x) - \hat{d}(\lambda)$, where

$$\hat{d}(\lambda) \triangleq \min_{x \in \mathbb{X}} f(x) + \lambda^T A x$$

is continuously differentiable according to Theorem 1. Assumptions (i) and (ii) on $\phi(x, \lambda)$ hold since f is strongly convex and the relation $f(x) \geq \hat{d}(\lambda) - \lambda^T b$ holds for all primal/dual feasible pairs (x, λ). If $\lambda_0(b) \in \Lambda^*(b)$ for every $b \in \mathbb{B}_c$, then $p^*(b) \equiv 0$ which implies $\nu^*(b) \equiv 0$ for all $b \in \mathbb{B}_c$, so $\nu_c^* = 0$.

On the other hand, let there be a $\bar{b} \in \mathbb{B}_c$ with $\lambda_0(\bar{b}) \notin \Lambda^*(\bar{b})$. For the sake of contradiction, assume that there exists a Lagrange multiplier $\nu^* = \nu^*(\bar{b}) \geq 0$ for the inequality constraint IC. Then by strong convexity of the Lagrangian in (x, λ) and Rockafellar (1997, Corollary 28.1.1) we have for the pair of minimizers $(x^*(\bar{b}), \lambda^{**}(\bar{b}))$ of (8), that we denote as (x^*, λ^{**}) below,

$$\left(x^*, \lambda^{**}\right) = \arg \min_{\substack{Ax=b, \, x \in \mathbb{X} \\ \lambda \in \mathbb{R}^m}} \left\| \lambda - \lambda_0(\bar{b}) \right\|^2 + \nu^* \left(f(x) - \hat{d}(\lambda) + \lambda^T \bar{b} \right),$$

or equivalently (by differentiability of f and $\hat{d}$ (cf. Bertsekas 2009, Prop. 1.1.8))

$$\nu^* \nabla f(x^*)^T \left(x - x^*\right) + \left(2 \left(\lambda^{**} - \lambda_0(\bar{b})\right) - \nu^* \left(\nabla \hat{d}(\lambda^{**}) - \bar{b}\right)\right)^T \left(\lambda - \lambda^{**}\right) \geq 0$$

for all $(x, \lambda) \in \left\{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid Ax = \bar{b}, \, x \in \mathbb{X}\right\}$ and $x^* \in \mathbb{X} \cap \left\{x \mid Ax = \bar{b}\right\}$. For the latter inequality to hold, we have $2(\lambda^{**} - \lambda_0(\bar{b})) - \nu^*(\nabla \hat{d}(\lambda^{**}) - \bar{b}) = 0$, but as $\nabla \hat{d}(\lambda^{**}) - \bar{b} = Ax^* - \bar{b} = 0$ (cf. Theorem 1 and dual optimality), we end up with $\lambda^{**} - \lambda_0(\bar{b}) = 0$ which contradicts $\lambda_0(\bar{b}) \notin \Lambda^*(\bar{b})$. $\square$

4.1.2 Support-function-based representation of $\Lambda^*(b)$

This representation is based on the necessary condition of Theorem 4.

Lemma 1 *For each parameter* $b \in \mathbb{B}_c$, *the convex set of Lagrange multipliers* $\Lambda^*(b)$ *can be represented as*

$$\Lambda^*(b) = \Big\{ \lambda \in \mathbb{R}^m \mid \exists x \in \mathbb{X} \cap \{x \mid Ax = b\} : \tag{R2}$$
$$x^T H x + g^T x + \sigma_{\mathbb{X}} \left(-Hx - g - A^T \lambda\right) + \lambda^T b \leq 0 \Big\},$$

where $\sigma_{\mathbb{X}}(y) = \sup_{x \in \mathbb{X}} y^T x$ *denotes the closed convex support function of* $\mathbb{X}$.

Proof By Proposition 1.1.8 in Bertsekas (2009) we have the equivalence

$$(14) \iff x^*(b) \in \mathbb{X} \quad \text{and} \quad 0 \leq \left(Hx^*(b) + g + A^T\lambda^*(b) \right)^T \left(x - x^*(b) \right) \quad \forall x \in \mathbb{X}$$

$$\iff x^*(b) \in \mathbb{X} \quad \text{and} \quad 0 \leq \inf_{x \in \mathbb{X}} \left(Hx^*(b) + g + A^T\lambda^*(b) \right)^T \left(x - x^*(b) \right),$$

which using the definition of $\sigma_{\mathbb{X}}(\cdot)$ and primal feasibility proves the lemma. $\qquad\square$

Before illustrating with an example that representation R2, as opposed to R1, is meaningful with respect to Theorem 3, we prove that this cannot be expected for *every* parametric problem.

Theorem 6 *Consider representation R2 of the closed convex set of Lagrange multipliers $\Lambda^*(b)$. There exist parametric problems of type (1) with $b \in \mathbb{B}_c$ for which the premise of Theorem 3 cannot be validated.*

Proof For representation R2 we identify function $\phi(x, \lambda)$ in Theorem 3 as

$$\phi(x, \lambda) = x^T Hx + g^T x + \sigma_{\mathbb{X}} \left(-Hx - g - A^T\lambda \right) ,$$

which is strongly convex in x as $H \succ 0$, so meets Assumption (i), and satisfies Assumption (ii) since for every $(x, \lambda) \in \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid Ax = b, \ x \in \mathbb{X}\}$

$$\phi(x, \lambda) \geq x^T Hx + g^T x - x^T Hx - g^T x - \lambda^T Ax = -\lambda^T b$$

by definition of the support function $\sigma_{\mathbb{X}}(\cdot)$. For the sake of contradiction, assume that the premise of Theorem 3 can be validated for a problem with

$$H = I, \ g = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \ A = \begin{bmatrix} -1 \ 1 \end{bmatrix}, \ \mathbb{X} = \left\{ x \in \mathbb{R}^2 \mid \|x\|_\infty \leq 1 \right\} , \qquad (15)$$

for which the set of admissible parameters is $\mathbb{B} = [-2, 2]$. Let $\mathbb{B}_c = \mathbb{B}$ and $K = 0$, $\hat{\lambda}_0 = 0$. Since $\mathbb{B}_c$ is a closed interval of the real line and $\psi^*(b)$ in (13) is closed convex by Statement S2 in Theorem 3, it follows from Bertsekas (2009, Proposition 1.3.12) that $\psi^*(b)$ is continuous on $\mathbb{B}_c$. But this implies that $h^*(b)$ must be continuous on $\mathbb{B}_c$, however, by basic calculations we find

$$h^*(b) = \begin{cases} \frac{1}{4} (4 - b)^2 , & \text{for } b \in [-2, 2) , \\ 0 , & \text{for } b = 2 , \end{cases} \qquad (16)$$

which is closed but not continuous. $\qquad\square$

Let us revisit the example (15) in the previous proof in order to understand which part of the assumptions of Theorem 3 cannot be validated. From this insight, we will then alter the example appropriately so that Theorem 3 applies.

We start with defining map $\nu^*(b)$ as the one that returns for every parameter $b \in \mathbb{B}_c$ the smallest Lagrange multiplier for inequality constraint IC. Note that here $\sigma_{\mathbb{X}}(\cdot) = \|\cdot\|_1$, so that for $b \in [-2, 2)$ we obtain this map from

$$\nu^*(b) = \min_{\nu \geq 0} \nu \tag{17}$$

$$\text{s.t. } \left(x^*(b), \lambda^{**}(b)\right) = \arg \min_{\substack{Ax=b, \, x \in \mathbb{X} \\ \lambda \in \mathbb{R}}} \lambda^2 + \nu \left(x^T x + g^T x + \lambda b + \|x + g + A^T \lambda\|_1\right),$$

by Theorem 4 and Rockafellar (1997, Cor. 28.1.1). For problem (15) we obtain

$$x^*(b) = \begin{bmatrix} -2 \\ 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix} (4-b), \quad \lambda^{**}(b) = \frac{1}{2}(4-b), \quad b \in [-2, 2),$$

such that after some calculation, the function in (17) is found to be

$$\nu^*(b) = \begin{cases} 1 + \frac{2}{2-b}, & \text{for } b \in [-2, 2), \\ 0, & \text{for } b = 2. \end{cases}$$

Now, $\nu^*(b)$ is defined everywhere on $\mathbb{B}_c$, however, the supremum ν_c^* in (9) does not exist, which explains why Theorem 3 does not apply. Differently, the theorem applies if $\mathbb{B}_c = [-2, 2-\delta]$, $\delta \in (0, 4]$, since ν_c^* then exists.

4.1.3 Computational aspects

The investigation of function $h^*(b)$ based on Theorem 3 depends on the existence of the supremum ν_c^* in (9) for representation R2. Let us assume for the moment that the supremum exists and is available. If the set of certified parameters $\mathbb{B}_c$ is contained in the relative interior of the admissible set of parameters $\mathbb{B}$, i.e. $\mathbb{B}_c \subset \text{ri } \mathbb{B}$, then $h^*(b)$ is continuous on $\mathbb{B}_c$ (this follows from Statements S1 and S2 in Theorem 3 and Rockafellar (1997, Theorem 10.4), thus by Weierstrass' Theorem the value of Δ_d^2 in (7) is attained.

If Statement S1 applies, the supremum is attained at some extreme point since $\mathbb{B}_c$ is assumed convex (Rockafellar 1997, Corollary 32.3.2). For instance, if $\mathbb{B}_c$ is a polytope, then it suffices to evaluate $h^*(b)$ at its vertices. Although Statement S2 is weaker, it can be used to get an upper bound on Δ_d^2 by omitting the nonpositive quadratic term in (12) and maximizing $\psi^*(b)$, i.e.

$$\Delta_d^2 \leq \sup_{b \in \mathbb{B}_c} \psi^*(b).$$

Note that this includes the case $K = 0$, $\hat{\lambda}_0 = 0$, which is the problem of determining an upper bound on the largest squared norm Lagrange multiplier.

Evaluating $h^*(b)$ pointwise corresponds to solving a single convex program if the dual function $d(\lambda; b)$ or the support function of set $\mathbb{X}$ can be represented conveniently.

In this respect, alternative R1 works, e.g. for quadratic programming ($\mathbb{X}$ is a polyhedron) if the linear inequality constraints defining set $\mathbb{X}$ are relaxed too (cf. Bertsekas 2009, Example 5.3.1), and linear programming ($H = 0$, $\mathbb{X}$ is a polyhedron), whereas R2 is a viable representation if set $\mathbb{X}$ is a 1-, 2- or ∞-norm ball, a simplex, ellipsoid, proper cone or a Cartesian product, Minkowski sum and/or union of them (see Rockafellar 1997, §13).

Remark 4 It is standard to characterize the set of Lagrange multipliers $\Lambda^*(b)$ by the Karush–Kuhn–Tucker (KKT) conditions (Bertsekas 1999, §3.3.1). This approach requires additional constraint qualifications to hold and set $\mathbb{X}$ to be representable as the intersection of finitely many level sets of closed convex functions $s_j(x)$, i.e. $\mathbb{X} = \{x \in \mathbb{R}^n \,|\, s_j(x) \leq 0, \ j = 1, \ldots, l\}$. However, the nonconvex complementary slackness conditions, as part of the KKT conditions, complicate the analysis of $h^*(b)$ and also prevent one from evaluating it by convex programming—despite the convexity of set $\Lambda^*(b)$.

4.2 Computation of the smallest Lipschitz constant

The smallest Lipschitz constant of the gradient, L^*, is the other crucial entity in the computation of the smallest lower iteration bound (cf. Theorem 2). We will show that L^* can indeed be computed under mild assumptions so that for some $\bar{\lambda}_1, \bar{\lambda}_2$ inequality (6) is tight. This result is also crucial for practical performance of the fast gradient method as $1/L^*$ is the implemented step size.

Let us start with an important observation. If we define a change of variables for problem (2), i.e. $x = Pw$ with invertible matrix $P \in \mathbb{R}^{n \times n}$, then from Theorem 1 we obtain $\nabla d\,(\lambda; b) = APw^*(\lambda) - b = Ax^*(\lambda) - b$, however, the Lipschitz constant according to the same theorem changes, since in general

$$\|A\|^2/\lambda_{\min}(H) \neq \|AP\|^2/\lambda_{\min}(P^T H P) \ . \tag{18}$$

By minimizing the right hand side of (18) over all invertible matrices P we obtain the smallest Lipschitz constant L^* under a linear change of variables. Whereas this problem can be cast as a convex semidefinite program (see Boyd et al. 1994, §3.1) it can also be solved analytically based on the next lemma.

Lemma 2 *It holds that*

$$\min_{P \text{ invertible}} \|AP\|^2/\lambda_{\min}(P^T P) = \|A\|^2 \ .$$

Proof For all invertible matrices P we have

$$\lambda_{\min}\,(PP^T)\, w^T A A^T w \leq w^T A P P^T A^T w, \quad \forall w \in \mathbb{R}^n. \tag{19}$$

This implies $\lambda_{\min}\,(PP^T) \leq \|AP\|^2 \,/\, \|A\|^2$ and thus the lower bound $\|A\|^2$ of the objective. But choosing $P = I$ attains this lower bound.   □

Theorem 7 *The smallest Lipschitz constant of the dual gradient under a linear change of variables is $L^* = \|AH^{-\frac{1}{2}}\|^2$.*

Proof Let $P = H^{-\frac{1}{2}}S$, S invertible, and apply Lemma 2 to the r.h.s. of

$$\min_{P \text{ invertible}} \|AP\|^2/\lambda_{\min}(P^T H P) = \min_{S \text{ invertible}} \|AH^{-\frac{1}{2}}S\|^2/\lambda_{\min}(S^T S) \ .$$

□

Let us investigate when $L^* < L$, where L is from Theorem 1.

Lemma 3 *If L^* is the Lipschitz constant from Theorem 7 and L the one from Theorem 1, then $\|A\|^2/\lambda_{\max}(H) \leq L^* \leq L$.*

Proof L is an upper bound of L^* by definition. Also,

$$\|AP\|^2/\lambda_{\min}(P^T H P) \geq \lambda_{\min}(P^T P) \, \|A\|^2/\lambda_{\min}(P^T H P) \geq \|A\|^2/\lambda_{\max}(H)$$

by using 19 and $\lambda_{\min}(P^T H P) \leq \lambda_{\max}(H) \, \lambda_{\min}(P^T P)$.   □

So, we deduce that $L^* < L$ only if $\lambda_{\max}(H) > \lambda_{\min}(H)$ which is true whenever Hessian H is not a positive multiple of the identity matrix. Also, L^* is a tight Lipschitz constant under a mild assumption as shown next.

Theorem 8 *If there exists a $\bar{\lambda} \in \mathbb{R}^m$ with $x^*(\bar{\lambda}) \in \text{int } \mathbb{X}$, then L^* from Theorem 7 is a tight Lipschitz constant of the dual gradient.*

Proof We prove that there exists a subset of $\mathbb{R}^m$ with nonempty interior on which the Lipschitz constant of the dual gradient attains L^*. By the premise, there exists a $\delta > 0$ such that $\overline{\mathbb{X}} = \left\{ x \in \mathbb{R}^n \mid \|x - x^*(\bar{\lambda})\| < \delta \right\} \subseteq \mathbb{X}$. Let set $\mathbb{M}$ contain all multipliers λ with $x^*(\lambda) \in \overline{\mathbb{X}}$, or equivalently, for all $\lambda \in \mathbb{M}$ the minimizer of (5) is free. In this case, we can compute the minimizer explicitly, i.e. $x^*(\lambda) = -H^{-1}(g + A^T\lambda)$, thus $\mathbb{M} = \left\{ \lambda \in \mathbb{R}^m \mid \left\| H^{-1}A^T(\lambda - \bar{\lambda}) \right\| < \delta \right\}$.

Since $\left\{ \lambda \in \mathbb{R}^m \mid \left\| H^{-1}A^T \right\| \left\| (\lambda - \bar{\lambda}) \right\| < \delta \right\}$ is an m-dimensional open subset of $\mathbb{M}$, we conclude that $\mathbb{M}$ has nonempty interior. The dual function defined over $\mathbb{M}$ is

$$d(\lambda; b) = -1/2 \left(g^T + \lambda^T A \right) H^{-1} \left(g + A^T\lambda \right) - \lambda^T b \ ,$$

which is twice continuously differentiable, so the Lipschitz constant of its gradient is $\lambda_{\max}(AH^{-1}A^T)$ (Nesterov, 2004a, Lemma 1.2.2), which is L^*.   □

Remark 5 In model predictive control, the interior assumption of Theorem 8 is a standard assumption (cf. Rawlings and Mayne 2009, §1.2)). So, for this class of problems, a tight Lipschitz constant can be obtained from Theorem 7.

Let us illustrate the findings of this section with the example given by (15). Consider a change of variables for its dual problem (2) according to

$$x = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} w, \qquad 0 < p_1 \leq p_2 \ .$$

From Theorem 7, the smallest Lipschitz constant is given by $L^* = 2$ for any $p_1, p_2 > 0$. As $x^*(2) = 0 \in \operatorname{int} \mathbb{X}$, it is also tight (cf. Theorem 8). On the contrary, the Lipschitz constant from Theorem 1 is

$$L(p_1, p_2) = 1 + (p_2/p_1)^2 \geq L^* \, ,$$

which can be arbitrarily larger than L^* if p_1 and p_2 are chosen accordingly.

4.3 Optimal preconditioning of the dual problem

Theorem 2 states the smallest lower iteration bound for Algorithm 2.1. Yet, we might get a better bound by considering the dual problem in a different basis. For a *strongly* concave problem the lower iteration bound can be improved if the preconditioner decreases the condition number [see Bertsekas (1999), §1.3.2) for heuristic and Boyd et al. (1994), §3.1) for optimal preconditioners]. The dual function in (2) lacks strong concavity, leaving the condition number undefined. In this section, we propose to take the smallest lower iteration bound in Theorem 2 as an alternative selection criterion for an optimal preconditioner of the dual function. It turns out that under the computationally tractable approximate reformulation of this problem introduced below, it cannot be ensured that the obtained optimal preconditioner gives a strictly better smallest lower iteration bound than the original one.

In order to see this, define $d_C(\upsilon; b) \triangleq d(C\upsilon; b)$ as the preconditioned dual function where $C \in \mathbb{R}^{m \times m}$ is an invertible preconditioner. In order to find a preconditioner that minimizes the smallest lower iteration bound for the preconditioned problem we need to minimize $L^*(C)\,\Delta_d^2(C)$ over all invertible matrices C (cf. Theorem 2), where

$$L^*(C) = \|C^T A H^{-\frac{1}{2}}\|^2 \, , \quad \Delta_d^2(C) = \sup_{b \in \mathbb{B}_c} \min_{\lambda \in \Lambda^*(b)} \|C^{-1} (\lambda - \lambda_0(b))\|^2 \, .$$

Minimizing $L^*(C)\,\Delta_d^2(C)$ directly is hard in view of $\Delta_d^2(C)$, however, a tractable formulation can be obtained from the upper bound

$$\min_{C \text{ inv.}} L^*(C)\Delta_d^2(C) \leq \min_{C \text{ inv.}} L^*(C)\|C^{-1}\|^2\Delta_d^2 = \min_{C \text{ inv.}} \frac{\|C^T A H^{-\frac{1}{2}}\|^2}{\lambda_{\min}(C^T C)} \Delta_d^2 \, .$$

Assume that a preconditioner C^* is obtained from solving the upper bound. Then by the previous inequality and Lemma 2, we have

$$L^*(C^*)\Delta_d^2(C^*) \leq L^*(C^*)\|C^{*-1}\|^2\Delta_d^2 = L^*\Delta_d^2 \, ,$$

which implies that the original smallest lower iteration bound is not guaranteed to be strictly improved. Also, if matrix A is sparse, then $C^{*T}A$ might not be sparse thus rendering an iteration of the fast gradient method more expensive.

5 Discussion

Let us pose some open questions based on the findings in this paper.

As shown in Sect. 4.1, characterization of function $h^*(b)$ critically depends on how the set of Lagrange multipliers is represented. The example in Sect. 4.1.2 illustrates that with representation R2 it is possible to reveal that $h^*(b)$ is the sum of a concave and a convex term (which in this example is convex), however, this cannot be concluded from representation R1. Thus, it would be interesting to have an analysis independent from the representation.

Another issue concerns the computation of v_c^* in (9) and, as a prerequisite, verifying if $v^*(b)$ is defined on $\mathbb{B}_c$. For a meaningful lower iteration bound, it is also important to find an affine initialization map $\lambda_0(b)$ that makes Δ_d^2 'small'.

References

Bank B, Guddat J, Klatte D, Kummer B, Tammer K (1982) Non-linear parametric optimization. Akademie-Verlag, Berlin

Bertsekas DP (1999) Nonlinear programming, 2nd edn. Athena Scientific, Massachusetts

Bertsekas DP (2009) Convex Optimization theory, 1st edn. Athena Scientific, Massachusetts

Blackmore L, Açikmese B, Scharf D (2010) Minimum-landing-error powered-descent guidance for mars landing using convex optimization. J Guid Control Dyn 33(4):1161–1171

Bleris L, Kothare M (2005) Real-time implementation of model predictive control. In: American control conference, vol 6, pp 4166–4171

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Boyd S, Ghaoui LE, Feron E, Balakrishnan V (1994) Linear matrix inequalities in system & control theory. Soc Ind Appl Math, vol 15. SIAM studies in Applied Mathematics, Philadelphia, PA

Defraene B, van Waterschoot T, Ferreau HJ, Diehl M, Moonen M (2012) Real-time perception-based clipping of audio signals using convex optimization. IEEE Trans Audio Speech Lang Process 20(10):2657–2671

Devolder O, Glineur F, Nesterov Y (2011) First-order methods of smooth convex optimization with inexact oracle. Math Program (submitted). Available at http://www.optimizationonline.org/DBFILE/2010/12/2865.pdf

Devolder O, Glineur F, Nesterov Y (2012) Double smoothing technique for large-scale linearly constrained convex optimization. SIAM J Optim 22(2):702–727

Doan M, Keviczky T, De Schutter B (2011) A dual decomposition-based optimization method with guaranteed primal feasibility for hierarchical MPC problems. In: 18th IFAC world congress

Fuchs A, Mariéthoz S, Larsson M, Morari M (2011) Grid stabilization through VSC-HVDC using wide area measurements. In: IEEE powertech. Power System Technology, Norway

Gauvin J (1977) A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. Math Program 12:136–138

Gol'šteĭn EG (1972) Theory of convex programming, vol 36. American Mathematical Society, Providence

Lan G, Monteiro RD (2009) Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Math Program (submitted). Available at http://www.optimizationonline.org/DB_HTML/2009/05/2294.html

Lemaréchal C (2001) Lagrangian relaxation. In: Junger M, Naddef D (eds) Computational combinatorial optimization. Lecture notes in computer science, vol 2241. Springer, Berlin, Heidelberg, pp 112–156

McGovern LK (2000) Computational analysis of real-time convex optimization for control systems. Thesis, Massachusetts Institute of Technology

Nedić A, Ozdaglar A (2009) Approximate primal solutions and rate analysis for dual subgradient methods. SIAM J Optim 19(4):1757–1780

Nesterov Y (1983) A method for solving a convex programming problem with convergence rate $1/k^2$. Soviet Math Dokl 27(2):372–376

Nesterov Y (2004a) Introductory lectures on convex optimization. Springer, Berlin

Nesterov Y (2004b) Smooth minimization of non-smooth functions. Math Program 103(1):127–152

Nesterov Y, Nemirovskii A (1994) Interior-point polynomial algorithms in convex programming. Soc Ind Appl Math, SIAM studies in Applied Mathematics, Philadelphia, PA, p 405

Polyak BT (1987) Introduction to optimization. Optimization Software

Rawlings JB, Mayne DQ (2009) Model predictive control theory and design. Nob Hill Pub, Madison

Richter S, Morari M, Jones CN (2011) Towards computational complexity certification for constrained MPC based on Lagrange relaxation and the fast gradient method. In: Conference on decision and control (CDC), Orlando

Richter S, Jones CN, Morari M (2012) Computational complexity certification for real-time MPC with input constraints based on the fast gradient method. IEEE Trans Autom Control 57(6):1391–1403

Rockafellar RT (1997) Convex analysis. Princeton University Press, Princeton

Schmidt M, Roux NL, Bach F (2011) Convergence rates of inexact proximal-gradient methods for convex, optimization. arXiv:11092415

Tseng P (2008) On accelerated proximal gradient methods for convex-concave optimization. SIAM J Optim (submitted)