

# Towards Computational Complexity Certification for Constrained MPC Based on Lagrange Relaxation and the Fast Gradient Method

Stefan Richter, Manfred Morari and Colin N. Jones

**Abstract**—This paper discusses the certification of Nesterov’s fast gradient method for problems with a strongly convex quadratic objective and a feasible set given as the intersection of a parametrized affine set and a convex set. For this, we derive a lower iteration bound for the solution of the dual problem that is obtained from a partial Lagrange Relaxation and propose a new constant step-size rule that we prove to be optimal under mild assumptions. Finally, we apply the certification procedure to a constrained MPC problem and show that the new step-size rule improves performance significantly.

## I. INTRODUCTION

The motivation of this work comes from model predictive control (MPC). In MPC one utilizes a mathematical model of the plant to determine an optimal sequence of inputs over a finite prediction horizon according to a specified objective [1]. The key strength of MPC is that physical constraints on the plant’s state and the control inputs can be incorporated. For linear, discrete-time systems with convex constraints on both the inputs and the states and a convex objective, the MPC problem is a convex optimization problem. In a receding horizon control scheme, where only the first element of the sequence of optimal inputs is applied to the plant, a solution to this problem is required at every sampling instant when new state information is available. So, more specific the MPC problem can be considered a *multi-parametric* convex optimization problem with the state as the parameter.

Convex optimization problems are solved by iterative methods. This makes it challenging to deploy them in control applications where the sampling rate of the feedback loop imposes hard bounds on the solution time and only limited computational resources are available. This paper addresses the issue on how we can determine *a priori* a bound on the number of iterations needed by a first order method to return a solution with certified approximation error for *any* state within a compact set. We refer to this as the computational complexity certification problem.

In the authors’ previous work this problem was targeted for the case of a convex quadratic objective and input constraints only [2]. Therein an optimal first order method, called the *fast gradient method* by Nesterov [3, §2.2], was introduced and lower iteration bounds for cold- and warm-starting derived. As shown in [2] the bounds are practically relevant for many real-world problems and [4] reports a first application of this certification scheme for power converters.

S. Richter and M. Morari are with the Automatic Control Laboratory at ETH Zürich, e-mail: richters|morari@ee.ethz.ch. C.N. Jones is with the Automatic Control Laboratory at EPFL Lausanne, e-mail: colin.jones@epfl.ch.

In [5] the certification problem is considered for general conic convex programming using an interior point method. The derived iteration bounds are off from the practically observed number of iterations by two to three orders of magnitude. More details and a discussion of the certification issues of other methods can be found in the review in [2].

Since the fast gradient method relies on gradient information only, it can exploit sparsity in the problem data. In many cases it shows improved convergence over steepest descent while having the same complexity per iteration. The bottleneck of the method is the projection on the feasible set that has to be performed in every iteration. Sets for which this is viable with low computational cost are denoted as *simple* sets and include, for example, the Euclidean ball, simplex, box, hyperplane, halfspace, and some proper cones [6, §8.1].

MPC problems often come with simple input and state constraints making the fast gradient method a viable solution method although care must be taken of the equality constraints that stem from the state update equations. If only input constraints are considered, the equality constraints can be eliminated by expressing the states as a linear function of the initial state and the sequence of inputs (*condensing*) [2]. The situation is more involved if state constraints are imposed since the simplicity property of the feasible set – which now is the intersection of an affine set and a simple set – is lost and cannot be recovered by condensing.

Motivated by the previous discussion we will use Lagrange Relaxation [7] to make the fast gradient method applicable for state and input constrained MPC and then derive lower iteration bounds for the dual problem. Since we aim to keep the results in this paper not solely restricted to MPC, we will introduce a more general problem setup next.

### Problem Setup and Notation

This paper investigates the computational complexity certification problem for the multi-parametric convex problem

$$f^*(y) \triangleq \min_{z \in \mathbb{K}} f(z) = \frac{1}{2} z^T H z + g^T z \quad (1)$$
$$\text{s.t. } A z = b(y) ,$$

which encompasses both offset/non-offset free linear quadratic MPC regulation and tracking problems [1].

In (1) a convex quadratic function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is minimized over the feasible set given as the intersection of a parametrized affine set and a closed, convex and compact set  $\mathbb{K} \subseteq \mathbb{R}^n$ . The parameter  $y$  from the compact set  $\mathbb{Y} \subseteq \mathbb{R}^p$  changes the right hand side of the equality constraint via the map  $b : \mathbb{R}^p \rightarrow \mathbb{R}^m$ , while matrix  $A$  is in  $\mathbb{R}^{m \times n}$ .

## Contribution and Outline

The main contribution of this paper is the derivation and – to the best of our knowledge – *first* computation of lower iteration bounds for multi-parametric convex problems (1) in a Lagrange Relaxation framework when using the fast gradient method. In order to make the computation of the bounds tractable for high dimensions ( $n > m \gg 1$ ), we will resort to conservatism where required. In turn, this will result in lower iteration bounds that, as shown in the example in Section V, may be far off the observed number of iterations. Still, it is the first time that such bounds are reported and a starting point for further research in this field.

We start our investigations with a short review of Lagrange Relaxation and related work in Section II. Section III gives a precise definition of the computational complexity certification problem and, for the sake of clarity, states all the basic assumptions that hold throughout the paper. Section IV discusses the required steps to compute lower iteration bounds in detail. As an important ‘side product’ of this investigation we will improve an established step-size rule and show that the new rule is optimal under mild assumptions. Finally, Section V gives a road map for certifying constrained MPC problems and provides first numerical results which show that the new step-size rule can improve the performance of the fast gradient method significantly.

## II. LAGRANGE RELAXATION AND RELATED WORK

In Lagrange Relaxation one defines the *dual function* as

$$d(\lambda; y) \triangleq \min_{z \in \mathbb{K}} f(z) + \lambda^T (Az - b(y)) \quad , \quad (2)$$

with multiplier  $\lambda \in \mathbb{R}^m$ . Note that we explicitly leave the set constraint  $z \in \mathbb{K}$  in this definition, an approach called partial elimination [8, §4.2.2]. The *dual problem* to (1) is then to maximize the concave dual function

$$d^*(y) \triangleq \max_{\lambda \in \mathbb{R}^m} d(\lambda; y) \quad . \quad (3)$$

We denote the convex set of dual optimal solutions as

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

and refer to  $\lambda^*(y) \in \Lambda^*(y)$  as a Lagrange multiplier. If strong duality holds, we have  $f^*(y) = d^*(y)$  and the primal optimizer can be recovered by  $z^*(\lambda^*(y))$  where

$$z^*(\lambda) \in \arg \min_{z \in \mathbb{K}} f(z) + \lambda^T (Az - b(y)) \quad , \quad (5)$$

and  $Az^*(\lambda^*(y)) = b(y)$  [9, Prop. 5.3.3].

Thus, Lagrange Relaxation allows us to solve (1) via its dual (3), which we will solve with the fast gradient method in this paper. In order to do so, the gradient  $\nabla d(\lambda; y)$  will be obtained according to the next theorem.

*Theorem 1 ([8, Prop. 6.1.1]):* If  $z^*(\lambda)$  in (5) is unique for all  $\lambda \in \mathbb{R}^m$ , the dual function  $d(\lambda; y)$  is continuously differentiable with gradient  $\nabla d(\lambda; y) = Az^*(\lambda) - b(y)$ .

So, in order to compute the gradient  $\nabla d(\lambda; y)$ , we first need to solve problem (2), which is sometimes called the *inner problem* in view of the *outer problem* (3).

In [10] the computational complexity of a similar setup is investigated. For the case of a general smooth, convex function  $f$  and a simple set  $\mathbb{K}$  the authors derive lower iteration bounds for an Augmented Lagrangian approach that is used to ensure continuous differentiability of the dual function without the assumption made in Theorem 1 (see e.g. [8, §4.2]). They assume that the inner problem is solved by Nesterov’s fast gradient method whereas the outer problem is solved by standard steepest ascent. For this setup the derived bounds on the overall number of fast gradient iterations hold under inexact gradients that emerge from suboptimal solutions of the inner problem. A guess-and-check procedure is provided to circumvent the computation of the distance between the dual starting iterate and the set of Lagrange multipliers, which is an important entity for the computation of lower iteration bounds in this setting.

Another recent work on certification is [11] which investigates the fast gradient method for convex problems in finite and infinite dimensional spaces. It applies smoothing [12] to make the dual function continuously differentiable and adds another smoothing term rendering the dual function strongly concave. Based on this double smoothing the authors derive lower iteration bounds on the required fast gradient iterations to obtain a nearly primal feasible, suboptimal solution where the cost of solving the inner problems is neglected.

Note that neither [10] nor [11] consider multi-parametric problems, such as encountered in MPC, or actually compute lower iteration bounds. Also, the guess-and-check procedure in [10] is inappropriate in real-time environments.

## III. GENERAL ASSUMPTIONS AND DEFINITIONS

We will state the computational certification problem in terms of a dual  $\epsilon$ -solution and a lower iteration bound.

*Definition 1 (Dual  $\epsilon$ -Solution):* Fix any parameter  $y \in \mathbb{Y}$ . For a specified  $\epsilon > 0$ , a dual  $\epsilon$ -solution  $\lambda_\epsilon \in \mathbb{R}^m$  for the dual problem (3) satisfies  $d^*(y) - d(\lambda_\epsilon; y) \leq \epsilon$ .

*Definition 2 (Lower Iteration Bound):* We denote  $i_{\min}$  a lower iteration bound if for any number of iterations of an iterative solution method,  $i \geq i_{\min}$ , a dual  $\epsilon$ -solution is retrieved for every parameter  $y \in \mathbb{Y}$  and a common  $\epsilon > 0$ .

*Definition 3 (Computational Complexity Certification):*

Consists in finding a lower iteration bound  $i_{\min}$ .

*Remark 1:* We assume that an approximate primal solution is obtained from  $z^*(\lambda_\epsilon)$  according to (5). By nature of the dual scheme, this solution is primal infeasible with respect to the equality constraint in (1) in general.

Without exception we assume from here on for problem (1):

*Assumption 1:* Hessian  $H$  is positive definite.

*Assumption 2:* Matrix  $A$  has full row rank.

*Assumption 3:* For all  $y \in \mathbb{Y}$  a Lagrange multiplier  $\lambda^*(y)$  exists and strong duality holds.

*Assumption 4:* Inner problem (2) can be solved exactly.

*Remark 2:* Assumption 3 holds true, e.g., if set  $\mathbb{K}$  is a polytope and a feasible point exists [9, Prop. 5.3.6], whereas Assumption 4 is satisfied for many practical MPC problems, see Section V for an example. Also note that some of the results later in the paper require additional assumptions.

#### IV. COMPUTATION OF A LOWER ITERATION BOUND FOR THE FAST GRADIENT METHOD

In this section we will derive a lower iteration bound for the solution of the outer problem (3) when using the fast gradient method. We will discuss the computation of the entities that define this bound and reveal a new step-size rule that can considerably improve performance.

The considered fast gradient method in Algorithm 1 is a variant of the constant step-size scheme II in [3] where the Lipschitz constant  $L$  of the gradient  $\nabla d(\lambda; y)$  determines the step-size in line 2. It can be computed as pointed out next.

*Theorem 2 ([12, Theorem 1]):* Let  $\lambda_{\min}(H)$  be the smallest eigenvalue of the Hessian  $H$ . Then for each parameter  $y \in \mathbb{Y}$  it holds that

$$\|\nabla d(\lambda_1; y) - \nabla d(\lambda_2; y)\| \leq L \|\lambda_1 - \lambda_2\|, \quad (6)$$

for any  $\lambda_1, \lambda_2 \in \mathbb{R}^m$  with  $L = \|A\|^2 / \lambda_{\min}(H)$ , where  $\|A\|$  denotes the maximum singular value of matrix  $A$ .

The Lipschitz constant  $L$  is also a crucial entity in the computation of the lower iteration bound as shown next.

*Theorem 3:* Assume that for all parameters  $y \in \mathbb{Y}$  the initial iterate is chosen as  $\lambda_0 = 0$ . A lower iteration bound according to Definition 2 for the fast gradient method in Algorithm 1 is given by

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

where  $\Delta_d$  is defined as

$$\Delta_d \triangleq \max_{y \in \mathbb{Y}} \min_{\lambda^*(y) \in \Lambda^*(y)} \|\lambda^*(y)\|. \quad (7)$$

*Proof:* Follows from Theorem 2.2.3 in [3]. ■

In view of Theorem 2 we conclude that the only missing entity to compute a lower iteration bound  $i_{\min}$  is  $\Delta_d$ . We will discuss this issue in detail in Section IV-B. Before we will improve the Lipschitz constant given by Theorem 2.

##### A. Computing a Tight Lipschitz Constant

We will show in this section that under mild assumptions on the problem data we can compute a Lipschitz constant such that for some pair  $(\bar{\lambda}_1, \bar{\lambda}_2)$  inequality (6) is tight. Now, the smaller the Lipschitz constant, the larger the step-size and the smaller the lower iteration bound (cf. Theorem 3). Hence, this investigation is crucial for the performance of Algorithm 1 but also from a certification point of view.

We start with an important observation: Let us define a change of variables for problem (2), i.e.  $z = Pw$  with invertible matrix  $P \in \mathbb{R}^{n \times n}$ . According to Theorem 1 we have  $\nabla d(\lambda; y) = APw^*(\lambda) - b(y) = Az^*(\lambda) - b(y)$ , however, the Lipschitz constant according to Theorem 2 changes, since in general we have

$$\frac{\|A\|^2}{\lambda_{\min}(H)} \neq \frac{\|AP\|^2}{\lambda_{\min}(P^T H P)}. \quad (8)$$

By minimizing the right hand side of (8) over all invertible matrices  $P$  we can get the smallest Lipschitz constant  $L^*$  under a linear change of variables. This problem can be

#### Algorithm 1 Fast Gradient Method for the Dual Problem (3)

**Require:** Initial iterate  $\lambda_0 \in \mathbb{R}^m$ ,  $\mu_0 = \lambda_0$ ,

$\alpha_0 = \frac{\sqrt{5}-1}{2}$ , Lipschitz constant of the gradient  $L$

- 1: **for**  $i = 1 \rightarrow i_{\min}$  **do**
- 2:  $\lambda_{i+1} = \mu_i + \frac{1}{L} \nabla d(\mu_i; y)$  {by Theorem 1}
- 3:  $\alpha_{i+1} = \frac{\alpha_i}{2} \left( \sqrt{\alpha_i^2 + 4} - \alpha_i \right)$
- 4:  $\beta_i = \frac{\alpha_i(1-\alpha_i)}{\alpha_i^2 + \alpha_{i+1}}$
- 5:  $\mu_{i+1} = \lambda_{i+1} + \beta_i(\lambda_{i+1} - \lambda_i)$
- 6: **end for**

cast as a convex semi-definite program following [13, §3.1]. But  $L^*$  can also be obtained analytically. In order to show this we require the following lemma.

*Lemma 1:* It holds that

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

*Proof:* For all invertible matrices  $P$  we have

$$\lambda_{\min}(PP^T) w^T A A^T w \leq w^T A P P^T A^T w, \quad \forall w \in \mathbb{R}^n. \quad (9)$$

This implies  $\lambda_{\min}(PP^T) \leq \|AP\|^2 / \|A\|^2$  and thus the lower bound  $\|A\|^2$  of the objective. It remains to note that  $P = I$  attains the lower bound. ■

*Theorem 4:* The smallest Lipschitz constant of the gradient  $\nabla d(\lambda; y)$  under a linear change of variables is

$$L^* = \left\| AH^{-\frac{1}{2}} \right\|^2.$$

*Proof:* Let  $P = H^{-\frac{1}{2}} S$ ,  $S$  invertible, and note that

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

which by Lemma 1 proves this theorem. ■

The next lemma indicates under which circumstance  $L^*$  can be smaller than Lipschitz constant  $L$  of Theorem 2.

*Lemma 2:* If  $L^*$  is the Lipschitz constant given by Theorem 4 and  $L$  the Lipschitz constant by Theorem 2, then

$$\frac{\|A\|^2}{\lambda_{\max}(H)} \leq L^* \leq L.$$

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

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

by using (9) and  $\lambda_{\min}(P^T H P) \leq \lambda_{\max}(H) \lambda_{\min}(P^T P)$ . ■

By Lemma 2 and the definition of Lipschitz constant  $L$  we infer 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 on the problem data as proven next.

*Theorem 5:* If there exists a  $\bar{\lambda} \in \mathbb{R}^m$  with  $z^*(\bar{\lambda}) \in \text{int } \mathbb{K}$ , then  $L^*$  given by Theorem 4 is a tight Lipschitz constant of the gradient  $\nabla d(\lambda; y)$  for every parameter  $y \in \mathbb{Y}$ .

*Proof:* We prove that there exists a subset of  $\mathbb{R}^m$  with nonempty interior on which the Lipschitz constant of the gradient of the dual function attains  $L^*$ .

Fix any  $y \in \mathbb{Y}$ . By the premise there exists a  $\delta > 0$  such that  $\mathbb{Z} = \{z \in \mathbb{R}^n \mid \|z - z^*(\bar{\lambda})\| \leq \delta\} \subseteq \mathbb{K}$ . Let set  $\mathbb{L}$  contain all multipliers  $\lambda$  with  $z^*(\lambda) \in \text{int } \mathbb{Z}$ . This means that set  $\mathbb{L}$  contains all multipliers for which the minimizer of (5) is free. In this case we can compute the minimizer explicitly in terms of the multiplier, i.e.  $z^*(\lambda) = -H^{-1}(g + A^T \lambda)$  for all  $\lambda \in \mathbb{L}$ , thus,  $\mathbb{L} = \{\lambda \in \mathbb{R}^m \mid \|H^{-1}A^T(\lambda - \bar{\lambda})\| < \delta\}$ . As set  $\{\lambda \in \mathbb{R}^m \mid \|H^{-1}A^T\| \|(\lambda - \bar{\lambda})\| < \delta\}$  is a full dimensional subset of  $\mathbb{L}$  we conclude that set  $\mathbb{L}$  has nonempty interior. Now, since the dual function defined over set  $\mathbb{L}$

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

is twice continuously differentiable, the Lipschitz constant of the gradient  $\nabla d(\lambda; z)$  is equal to  $\lambda_{\max}(AH^{-1}A^T)$  [3, Lemma 1.2.2]. But this is equivalent to  $L^*$  in Theorem 4. ■

*Remark 3:* A standard assumption in constrained MPC is that the origin is contained in the interior of the input and state sets [1]. Since in MPC there is no linear term in the objective of (1), the premise of Theorem 5 is satisfied with  $\bar{\lambda} = 0$  as  $z^*(\bar{\lambda}) = 0$ . So, for constrained MPC under standard assumptions a tight Lipschitz constant can be obtained.

### B. Upper-Bounding the Norm of Lagrange Multipliers

Besides the Lipschitz constant the largest norm of a (minimum norm) Lagrange multiplier over all parameters,  $\Delta_d$ , is a crucial entity for the computation of a lower iteration bound (cf. Theorem 3). This section elaborates on computing an upper bound of  $\Delta_d$ . In order to do so we add another assumption that holds true from here on.

*Assumption 5:* For all  $y \in \mathbb{Y}$  the set of Lagrange multipliers  $\Lambda^*(y)$  is compact.

*Remark 4:* Assumption 5 is fulfilled if set  $\mathbb{Y}$  is such that

$$b(y) \in \text{int dom } p \quad \forall y \in \mathbb{Y} ,$$

with  $p(u) \triangleq \min \{f(z) \mid Az = u, z \in \mathbb{K}\}$  being the perturbation function of problem (1). In order to see this, note that  $-\Lambda^*(y)$  is the subdifferential of  $p$  at  $b(y)$  [8, §5.4.4]. Since the perturbation function  $p$  is convex and proper, the subdifferential and thus  $\Lambda^*(y)$  is compact [9, Prop. 5.4.1].

Under Assumption 5 the upper bound of  $\Delta_d$

$$\Delta_d \leq \max_{y \in \mathbb{Y}, \lambda^*(y) \in \Lambda^*(y)} \|\lambda^*(y)\| \triangleq \bar{\Delta}_d \quad (10)$$

is finite. An upper bound of  $\bar{\Delta}_d$  can be derived by exploiting a recent result in [11, Thm. 3 and Remark 4]. The following theorem is a special case of the aforementioned result.

*Theorem 6:* For any  $y \in \mathbb{Y}$  we have

$$\|\lambda^*(y)\| \leq \frac{v(y)}{r(y)} \quad \forall \lambda^*(y) \in \Lambda^*(y) , \quad (11)$$

with  $v(y), r(y)$  defined as

$$v(y) \triangleq \max_{z \in \mathbb{K}} (Hz^*(y) + g)^T (z - z^*(y)) , \quad (12)$$

$$r(y) \triangleq \max_{\mathbb{B}[b(y); r] \subseteq A\mathbb{K}} r \quad (13)$$

In (12) we let  $z^*(y) \triangleq z^*(\lambda^*(y))$  with a slight abuse of notation, whereas the set  $\mathbb{B}[b(y); r]$  in (13) is the two-norm ball in  $\mathbb{R}^m$ , i.e.  $\mathbb{B}[b(y); r] \triangleq \{w \in \mathbb{R}^m \mid \|w - b(y)\| \leq r\}$ .

Using Theorem 6 we are now ready to state an upper bound of  $\bar{\Delta}_d$  and thus of  $\Delta_d$  under additional assumptions.

*Theorem 7:* Let  $b : \mathbb{R}^p \rightarrow \mathbb{R}^m$  be affine. The upper bound

$$\Delta_d \leq \frac{\bar{v}_{\max}}{r_{\min}} ,$$

is finite with  $\bar{v}_{\max}$  and  $r_{\min}$  defined as

$$\bar{v}_{\max} \triangleq \sigma_{\mathbb{K}}(-g) + \sigma_{\mathbb{K}}(g) + \max_{z \in \mathbb{K}} \sigma_{\mathbb{K}}(Hz) , \quad (14)$$

$$r_{\min} \triangleq \min_{y \in \mathbb{Y}} r(y) , \quad (15)$$

where  $r(y)$  is a concave function. In (14)  $\sigma_{\mathbb{K}}(\cdot)$  denotes the convex support function of set  $\mathbb{K}$ , i.e.

$$\sigma_{\mathbb{K}}(z) \triangleq \max_{\bar{z} \in \mathbb{K}} z^T \bar{z} .$$

*Proof:* From (10) and (11) we conclude that

$$\Delta_d \leq \bar{\Delta}_d \leq \max_{y \in \mathbb{Y}} \frac{v(y)}{r(y)} \leq \frac{\max_{y \in \mathbb{Y}} v(y)}{\min_{y \in \mathbb{Y}} r(y)} \triangleq \frac{v_{\max}}{r_{\min}} .$$

Using (12) in the definition of  $v_{\max}$  and considering that Hessian  $H$  is positive definite, we obtain

$$\begin{aligned} v_{\max} &= \max_{y \in \mathbb{Y}} -z^*(y)^T Hz^*(y) - g^T z^*(y) + \sigma_{\mathbb{K}}(Hz^*(y) + g) \\ &\leq \max_{y \in \mathbb{Y}} -g^T z^*(y) + \max_{y \in \mathbb{Y}} \sigma_{\mathbb{K}}(Hz^*(y) + g) \\ &\leq \max_{z \in \mathbb{K}} -g^T z + \sigma_{\mathbb{K}}(g) + \max_{z \in \mathbb{K}} \sigma_{\mathbb{K}}(Hz) = \bar{v}_{\max} . \end{aligned}$$

By compactness of  $\mathbb{K}$  the value of  $\bar{v}_{\max}$  is finite. Let us look at  $r(y)$  next. Since  $\mathbb{K}$  is closed and convex, we have by [9, Proposition 1.4.13] that  $A\mathbb{K}$  is also closed and convex. Let

$$\mathbb{H} \triangleq \{(h, k) \in \mathbb{R}^m \times \mathbb{R} \mid h^T w \leq k, \forall w \in A\mathbb{K}\}$$

be the set of all pairs  $(h, k)$  that define closed halfspaces containing  $A\mathbb{K}$ . By [9, Proposition 1.5.4] we have

$$A\mathbb{K} = \{w \in \mathbb{R}^m \mid h^T w \leq k, \forall (h, k) \in \mathbb{H}\} . \quad (16)$$

Using representation (16) of set  $A\mathbb{K}$ , we observe that

$$r(y) = \min_{(h, k) \in \mathbb{H}} \frac{k - h^T b(y)}{\|h\|} , \quad (17)$$

cf. [6, §4.3.1]. Since  $b(y)$  is affine by assumption,  $r(y)$  as the minimum of affine functions is concave. Last, the value of  $r_{\min}$  is positive as  $r(y) > 0$  for all  $y \in \mathbb{Y}$  and set  $\mathbb{Y}$  is compact. So the bound given by the theorem is finite. ■

*Remark 5:* In MPC we often encounter input and state sets that are boxes, so that  $\mathbb{K} = \{z \in \mathbb{R}^n \mid \|z\|_{\infty} \leq 1\}$  if we assume unit boxes. An upper bound of the expression  $\max_{z \in \mathbb{K}} \sigma_{\mathbb{K}}(Hz)$  in (14) is  $n\|H\|_1$ , where  $\|\cdot\|_1$  denotes the induced 1-norm. We conclude that an upper bound on  $\bar{v}_{\max}$  can be easily computed in this case, even for  $n \gg 1$ .

On the contrary, computation of  $r_{\min}$  in (15) is a challenging problem in high dimensions as discussed next. However, a lower bound of  $r_{\min}$  is sufficient for deriving a lower iteration bound  $i_{\min}$ . A computationally tractable lower bound of  $r_{\min}$  will be derived in the next section for polytopic sets  $\mathbb{Y}, \mathbb{K}$ .

### Computing a Lower Bound of $r_{\min}$ for Polytopic Sets $\mathbb{Y}, \mathbb{K}$

The following corollary of Theorem 7 explains why it is convenient to restrict to a polytopic set  $\mathbb{Y}$ .

*Corollary 1:* Let  $b : \mathbb{R}^p \rightarrow \mathbb{R}^m$  be affine. For a polytopic set  $\mathbb{Y}$  the value of  $r_{\min}$  can be computed by considering only the vertices of  $\mathbb{Y}$ , i.e. let  $\mathbb{V}_{\mathbb{Y}}$  contain all vertices of  $\mathbb{Y}$ , then

$$r_{\min} = \min_{y \in \mathbb{V}_{\mathbb{Y}}} r(y) .$$

*Proof:* This follows from concavity of  $r(y)$ . ■

For the computation of  $r(y)$  it is useful to restrict to a polytopic set  $\mathbb{K}$  too, since the image  $A\mathbb{K}$  (cf. (13)) is again a polytope as indicated by the next lemma and the computation of  $r(y)$  according to (17) becomes straightforward then.

*Lemma 3:* Let  $\mathbb{K} = \{z \in \mathbb{R}^n \mid Fz \leq f\}$  be a polytopic set where matrix  $F \in \mathbb{R}^{q \times n}$  and vector  $f \in \mathbb{R}^q$ . Set  $A\mathbb{K}$  is polytopic and obtained from a projection, i.e.

$$A\mathbb{K} = \{v \in \mathbb{R}^m \mid \exists w \in \mathbb{R}^{n-m} : FA^\dagger v + FN_A w \leq f\} ,$$

where  $A^\dagger$  denotes the pseudo-inverse of  $A$  and the columns of matrix  $N_A$  span the nullspace of  $A$ .

*Proof:* Let  $A[V_1, V_2] = U[S, 0]$  be the singular value decomposition of  $A$ . The projection formulation comes from

$$\begin{aligned} A\mathbb{K} &= \{v \in \mathbb{R}^m \mid v = Az, Fz \leq f, z = V_1 w_1 + V_2 w_2\} \\ &= \{v \mid v = USw_1, FV_1 w_1 + FV_2 w_2 \leq f\} , \end{aligned}$$

and eliminating  $w_1$  above. As  $A^\dagger = V_1 S^{-1} U^T$  and  $N_A = V_2$  the result follows. Also, since  $A\mathbb{K}$  is the projection of a polytope, it is also a polytope [14]. ■

The projection of a polytope can be computed, e.g. by Fourier-Motzkin Elimination [14], as implemented in the Matlab toolbox MPT [15]. However, computation is tractable only for dimensions  $m < n \lesssim 10$ . The next theorem provides a lower bound on  $r(y)$  that comes without explicitly computing the projection. For its proof the following remark, which is easy to verify, is important.

*Remark 6:* If  $\mathbb{P}$  is a polytope in  $\mathbb{R}^n$  with the origin contained in its interior, it can be represented as  $\mathbb{P} = \{z \in \mathbb{R}^n \mid Ez \leq \mathbf{1}\}$ , where  $E$  is an appropriate matrix and  $\mathbf{1}$  denotes the vector of ones.

*Theorem 8:* Let all of the assumptions of Lemma 3 hold and define the polytope

$$\mathbb{P} \triangleq \{(v, w) \in \mathbb{R}^n \mid FA^\dagger v + FN_A w \leq f\} .$$

Let the translation of  $\mathbb{P}$  by a vector  $(b(y), w(y)) \in \text{int } \mathbb{P}$  be

$$\mathbb{P}(y) = \mathbb{P} - \begin{bmatrix} b(y) \\ w(y) \end{bmatrix} ,$$

represented as  $\mathbb{P}(y) = \{(v, w) \in \mathbb{R}^n \mid C(y)v + D(y)w \leq \mathbf{1}\}$ . Then for all parameters  $y \in \mathbb{Y}$  it holds that

$$r(y) \geq \tilde{r}(y) ,$$

where  $\tilde{r}(y)$  is given as

$$\tilde{r}(y) = \left( \max_{i=1, \dots, q} \|C(y)_i^T\| \right)^{-1} ,$$

and vector  $C(y)_i^T$  denotes the  $i$ th row of matrix  $C(y)$ .

*Proof:* Denote  $\pi_v \mathbb{P}$  as the projection of set  $\mathbb{P}$  onto the  $v$ -space, i.e.  $\pi_v \mathbb{P} = \{v \in \mathbb{R}^m \mid \exists w \in \mathbb{R}^{n-m} : (v, w) \in \mathbb{P}\}$ . Then  $\pi_v \mathbb{P} = A\mathbb{K}$  in view of Lemma 3. Starting from the definition of  $r(y)$  in (13) we obtain

$$\begin{aligned} r(y) &= \max_{\mathbb{B}[b(y); r] \subseteq A\mathbb{K}} r = \max_{\mathbb{B}[0; r] \subseteq A\mathbb{K} - b(y)} r \\ &= \max_{\mathbb{B}[0; r] \subseteq \pi_v \mathbb{P} - b(y)} r = \max_{\mathbb{B}[0; r] \subseteq \pi_v \mathbb{P}(y)} r . \end{aligned}$$

Since  $(b(y), w(y)) \in \text{int } \mathbb{P}$  we have  $0 \in \text{int } \mathbb{P}(y)$ . Thus, by Remark 6, polytope  $\mathbb{P}(y)$  can be represented by a finite number of inequalities with all ones on the right hand side. From the Projection Lemma [16] we have that

$$\pi_v \mathbb{P}(y) = \{v \in \mathbb{R}^m \mid u^T C(y)v \leq 1, \forall u \in \mathbb{U}(y)\} ,$$

where

$$\mathbb{U}(y) \triangleq \{u \in \mathbb{R}^q \mid D(y)^T u = 0, u \geq 0, \mathbf{1}^T u = 1\} .$$

By this characterization of the projection  $\pi_v \mathbb{P}(y)$  we can derive the following equivalences

$$\begin{aligned} \mathbb{B}[0; r] \subseteq \pi_v \mathbb{P}(y) &\iff \max_{v \in \mathbb{B}[0; r]} u^T C(y)v \leq 1 \quad \forall u \in \mathbb{U}(y) \\ &\iff r \|C(y)^T u\| \leq 1 \quad \forall u \in \mathbb{U}(y) . \end{aligned}$$

The latter inequality is tight at the maximum  $r(y)$ , so

$$\begin{aligned} r(y) &= \left( \max_{u \in \mathbb{U}(y)} \|C(y)^T u\| \right)^{-1} \\ &\geq \left( \max_{u \geq 0, \mathbf{1}^T u = 1} \|C(y)^T u\| \right)^{-1} . \end{aligned}$$

Since in the previous problem the objective is convex, the maximum is attained at one of the vertices of the feasible set. But this set is the unit simplex in  $\mathbb{R}^q$  with vertices  $u_i \in \mathbb{R}^q$ ,  $i = 1, \dots, q$ , where  $u_i$  is the zero vector having a '1' at the  $i$ th component. This proves the theorem. ■

*Remark 7:* In Theorem 8 a vector  $w(y) \in \mathbb{R}^{n-m}$  can be computed, for instance, by computing the Chebyshev center of polytope  $\mathbb{P}$  with the first coordinates of the center fixed to  $b(y)$ . See e.g. [6, §4.3.1] for a reformulation of this problem as a linear program.

Let us bring together the findings of this section next.

*Corollary 2:* Let sets  $\mathbb{Y}, \mathbb{K}$  be polytopic and map  $b : \mathbb{R}^p \rightarrow \mathbb{R}^m$  be affine. It holds that

$$r_{\min} \geq \tilde{r}_{\min} \triangleq \min_{y \in \mathbb{V}_{\mathbb{Y}}} \tilde{r}(y) ,$$

where  $\tilde{r}(y)$  is defined in Theorem 8 and set  $\mathbb{V}_{\mathbb{Y}}$  contains all vertices of  $\mathbb{Y}$ . Furthermore, we have

$$\Delta_d \leq \frac{\bar{v}_{\max}}{\tilde{r}_{\min}} ,$$

with  $\bar{v}_{\max}$  given by (14).

*Proof:* Follows from Theorem 8 and Corollary 1. ■

In the context of linear MPC with polytopic input and state sets, Corollary 2 provides a practical way to derive an upper bound on the value of  $\Delta_d$ . However, the lower iteration bounds based on this corollary might be subject to a high degree of conservatism as shown in the next section.

## V. COMPLEXITY CERTIFICATION FOR MPC

In this section we will summarize the steps required to certify a constrained MPC problem according to the results obtained so far. We assume that the dimension of the problem does not allow the evaluation of the projection in Lemma 3 which is true for most practical MPC problems. Finally, we will certify MPC for a ball on plate system.

Let us consider the problem of regulating a linear, time-invariant system to the origin. The corresponding constrained MPC problem at the current state  $x \in \mathbb{R}^{n_x}$  is formulated as

$$\begin{aligned} J^*(x) &\triangleq \min \frac{1}{2} x_N^T Q_N x_N + \frac{1}{2} \sum_{k=0}^{N-1} x_k^T Q x_k + u_k^T R u_k \quad (18) \\ \text{s.t. } &x_{k+1} = A_d x_k + B_d u_k, \quad k = 0 \dots N-1 \\ &(u_k, x_k) \in \mathbb{U} \times \mathbb{X}, \quad k = 0 \dots N-1 \\ &x_N \in \mathbb{X}_f, x_0 = x. \end{aligned}$$

In accordance with the general problem formulation in (1) we consider a quadratic objective in the states  $x_k \in \mathbb{R}^{n_x}$  and inputs  $u_k \in \mathbb{R}^{n_u}$ , where  $N$  denotes the horizon length. The penalty matrices  $Q$ ,  $Q_N \in \mathbb{R}^{n_x \times n_x}$  and  $R \in \mathbb{R}^{n_u \times n_u}$  are assumed to be positive definite. Both the input set  $\mathbb{U} \subset \mathbb{R}^{n_u}$  and the state set  $\mathbb{X} \subset \mathbb{R}^{n_x}$  are compact convex sets, same as the terminal set  $\mathbb{X}_f \subset \mathbb{R}^{n_x}$ . Under these assumptions, (18) is a multi-parametric convex optimization problem with the current state  $x \in \mathbb{X}_0 \subset \mathbb{R}^{n_x}$  as the parameter.

In order to comply with the problem formulation (1) we define  $z \triangleq (x_0, x_1, \dots, x_N, u_0, \dots, u_{N-1})$ , parameter  $y \triangleq x$  and set  $\mathbb{Y} \triangleq \mathbb{X}_0$  which leads to the matrices

$$H \triangleq \text{blkdiag}(Q, \dots, Q, Q_N, R, \dots, R),$$

$$g \triangleq 0,$$

$$A \triangleq - \begin{bmatrix} -I & 0 & \dots & \dots & 0 & 0 & \dots & \dots & 0 \\ A_d & -I & 0 & \dots & 0 & B_d & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots & 0 & B_d & \ddots & \vdots \\ \vdots & \ddots & A_d & -I & 0 & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A_d & -I & 0 & \dots & 0 & B_d \end{bmatrix}.$$

The parameter changes the right hand side of the equality constraint by  $b(y) \triangleq B y$  with  $B \triangleq (I, 0, \dots, 0)^T$ . The set constraint is given as  $\mathbb{K} \triangleq \mathbb{X} \times \dots \times \mathbb{X} \times \mathbb{X}_f \times \mathbb{U} \times \dots \times \mathbb{U}$ .

The following properties of the MPC problem data ensure the validity of Assumptions 1-5.

- Assumption 1 is satisfied as penalty matrices  $Q, Q_N$  and  $R$  are assumed positive definite.
- Assumption 2 is fulfilled by definition of  $A$  before.
- If  $\mathbb{X}_0 \subseteq \text{int} \{x \mid (18) \text{ feasible at } x\}$  then Assumptions 3 and 5 hold true.
- Assumption 4 is satisfied if  $Q, R$  are positive definite diagonal penalty matrices and sets  $\mathbb{U}, \mathbb{X}$  are box constraints. Additionally, the pair  $(Q_N, \mathbb{X}_f)$  must comply with one of the following cases:
  - (i)  $\mathbb{X}_f$  is a level set of the terminal penalty function, i.e.  $\mathbb{X}_f = \{x_N \in \mathbb{R}^{n_x} \mid \frac{1}{2} x_N^T Q_N x_N \leq c\}, c > 0$ .
  - (ii) Matrix  $Q_N$  is diagonal and  $\mathbb{X}_f$  is a box constraint.

In both cases the inner problem (2) can be solved exactly by a series of  $2N$  projections on boxes and one projection on either the two-norm ball or a box, all of which can be easily computed.

*Remark 8:* Note that it is common practice to avoid a formulation with an elliptic terminal set  $\mathbb{X}_f$  as in case (i) since this leads to a quadratically constrained QP (QCQP) which requires different solver implementations than a standard QP. However, an elliptic terminal set is a more practical choice than a polytopic approximation to the maximum positively invariant set which is hard to compute in high dimensional state spaces and might introduce a larger number of constraints (see e.g. [5, §6.1] for details).

In the following we will report first certification results for an MPC problem that meets the requirements of case (ii).

*Example: Complexity Certification for a Ball on Plate System*

In the ball on plate system a plate is tilted around two axes to control the position of a ball. For small tilt angles the dynamics of the system can be decoupled and each axis controlled independently. We will consider the MPC control of a single axis from here on. Assuming a sampling time of  $T_s = 0.01s$  we obtain for the ball on plate system at the Automatic Control Lab of ETH Zürich described in [17]

$$A_d = \begin{bmatrix} 1 & 0.01 \\ 0 & 1 \end{bmatrix}, \quad B_d = \begin{bmatrix} -0.0004 \\ -0.0701 \end{bmatrix},$$

with the ball position and velocity along a single axis as the states and the tilt angle as the input. The weight matrices are  $Q = \begin{bmatrix} 100 & 0 \\ 0 & 10 \end{bmatrix}$ ,  $R = 1$ ,  $Q_N = Q$ , and we assume input and state confined to  $\mathbb{U} = \{u \in \mathbb{R} \mid -0.0524 \leq u \leq 0.0524\}$  and  $\mathbb{X} = \{x \in \mathbb{R}^2 \mid \begin{bmatrix} -0.2 \\ -0.1 \end{bmatrix} \leq x \leq \begin{bmatrix} 0.01 \\ 0.1 \end{bmatrix}\}$  respectively. We notice that the state set  $\mathbb{X}$  has an upper limit on the ball's position close to the origin. The MPC regulation problem in (18) naturally takes this constraint into account when regulating the ball to the origin.

We will now certify the fast gradient method in Algorithm 1 for the solution of the ball on plate MPC problem. For this we assume a required accuracy of  $\epsilon = 10^{-2}$  and horizon lengths from  $N = 5$  to  $N = 15$ . For any horizon length we proceed as follows:

- 1) Compute Lipschitz constant  $L^*$  (Theorem 4) and, for comparison reasons, Lipschitz constant  $L$  (Theorem 2).
- 2) Compute an upper bound of  $\bar{v}_{\max}$  following Remark 5.
- 3) Compute an upper bound of  $\Delta_d$  (Corollary 2).
- 4) Get lower iteration bounds from Theorem 3, using  $L^*$  and  $L$  from 1)

The obtained lower iteration bounds are more descriptive, if they are used to derive an expected solution time. Since the problem matrices are sparse and structured, an iteration of the fast gradient method in Algorithm 1 amounts to

$$n_{\text{flops}} = 4Nn_x^2 + 4Nn_x n_u + 9Nn_x + 2Nn_u + 6n_x$$

floating point operations (flops), which is linear in the horizon. In the following we assume a computing performance of 1 Gflops/s, for which the expected solution time follows from the lower iteration bounds and  $n_{\text{flops}}$ .

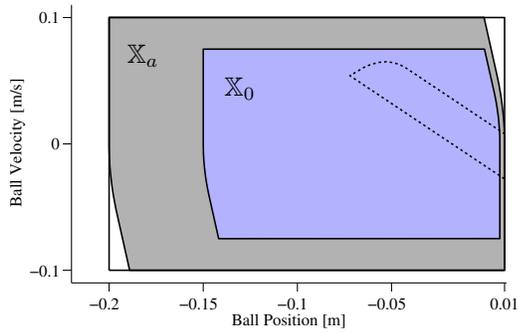


Fig. 1. Illustration of sets for the certification of the ball on plate system.

The set of initial states  $\mathbb{X}_0$  is obtained from scaling the polytopic maximum admissible set for  $N = 15$ ,  $\mathbb{X}_a$ , as depicted in Fig. 1. The dotted region is the set of initial states for which no constraint is active at the solution.

The certification results in terms of lower iteration bounds (solid) and expected solution times (dashed) are depicted in Fig. 2. For validating the quality of the lower iteration bounds the figure also shows the observed number of iterations from sampling ( $10^3$  samples) as a mean-max curve (dotted). The minimum iteration count was ‘1’ in all scenarios while Gurobi was chosen as the reference solver in Matlab [18] to compute  $d^*(y)$ .

In Fig. 2 the certification results are organized in two groups, each consisting of three curves. The first group (black, *square*) illustrates the results when using the original Lipschitz constant  $L$  from Theorem 2, the second group (gray, *triangle*) depicts the corresponding results for the Lipschitz constant  $L^*$  from Theorem 4.

Depending on the horizon, we obtain  $L \in [3.97, 3.99]$  and  $L^* \in [0.38, 0.40]$  in this example. As the Lipschitz constants differ by a factor of about 10, the lower iteration bounds corresponding to  $L^*$  are smaller by a factor of about  $\sqrt{10}$  than the bounds obtained with  $L$  (cf. Theorem 3). Interestingly, this is also the speedup in the observed number of iterations. So, computing the Lipschitz constant according to Theorem 4 not only gives smaller iteration bounds but also improves the actual performance of the fast gradient method.

Unfortunately, the derived lower iteration bounds are off by more than three orders of magnitude from the observed number of iterations. The main reason for this discrepancy is the conservative upper bound of  $\Delta_d$  given in Corollary 2. This claim is justified by determining  $\Delta_d$  *approximately* by sampling  $\|\lambda^*(y)\|$  ( $10^3$  samples). The sampled bounds are off by about one order of magnitude from the observed number of iterations only (not shown in Fig. 2).

Note that in all scenarios the growth rate of the lower iteration bounds is similar to the growth rate of the observed number of iterations. We conclude that the fast gradient method shows satisfying *practical* performance as it requires a maximum number of 90–245 iterations (depending on the horizon), if the tight Lipschitz constant  $L^*$  is chosen. This amounts to an expected solution time of 20–165  $\mu$ s.

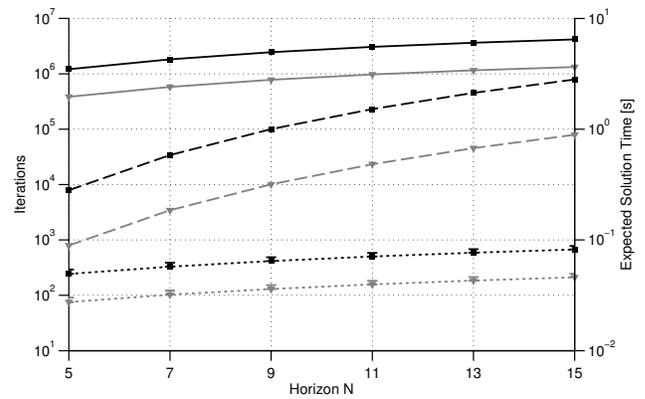


Fig. 2. Certification of constrained MPC for a ball on plate system: Lower iteration bound (solid), expected solution time assuming 1 Gflops/s (dashed) and observed number of iterations from sampling ( $10^3$  samples) as mean-max curve (dotted). The group of black curves (*square*) illustrates the results for Lipschitz constant  $L$ , whereas the group of gray curves (*triangle*) the ones for Lipschitz constant  $L^*$ .

## REFERENCES

- [1] J. B. Rawlings and M. D. Q., *Model Predictive Control Theory and Design*. Nob Hill Pub, Aug. 2009.
- [2] S. Richter, C. N. Jones, and M. Morari, “Computational Complexity Certification for Real-Time MPC with Input Constraints Based on the Fast Gradient Method,” *Automatic Control, IEEE Transactions on*, 2011, accepted for publication.
- [3] Y. Nesterov, *Introductory lectures on convex optimization*. Springer, 2004.
- [4] S. Richter, S. Mariétoz, and M. Morari, “High-Speed Online MPC Based on a Fast Gradient Method Applied to Power Converter Control,” in *American Control Conference*, Baltimore, USA, June 2010.
- [5] L. K. McGovern, “Computational analysis of real-time convex optimization for control systems,” Thesis, Massachusetts Institute of Technology, Dept. of Aeronautics and Astronautics, 2000.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.
- [7] C. Lemaréchal, “Lagrangian relaxation,” in *Computational Combinatorial Optimization*, 2001, pp. 112–156.
- [8] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, Massachusetts: Athena Scientific, 1999.
- [9] —, *Convex Optimization Theory*, 1st ed. Athena Scientific, June 2009.
- [10] G. Lan and R. D. Monteiro, “Iteration-complexity of first-order augmented lagrangian methods for convex programming,” Georgia Institute of Technology, School of Industrial & Systems Engineering, Tech. Rep., 2009.
- [11] O. Devolder, F. Glineur, and Y. Nesterov, “Double Smoothing Technique for Convex Optimization Problems in Hilbert Spaces with Applications to Optimal Control,” Université catholique de Louvain (UCL), Belgium, Center for Operations Research and Econometrics (CORE), Tech. Rep., Jan. 2011.
- [12] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2004.
- [13] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System & Control Theory*. Society for Industrial and Applied Mathematics, July 1994.
- [14] G. M. Ziegler, *Lectures on polytopes*. Springer, 1995.
- [15] M. Kvasnica, P. Grieder, and M. Baotić, “Multi-Parametric Toolbox (MPT),” 2004. [Online]. Available: <http://control.ee.ethz.ch/~mpt/>
- [16] S. Cernikov, “Contraction of finite systems of linear inequalities,” *Dokl. Akad. Nauk SSSR*, vol. 152, no. 5, pp. 1075–1078, 1963, english translation in Soc. Math. Dokl. 4(5), 15201524 (1963).
- [17] R. Waldvogel, “MPC control of a ball on plate system,” Master’s thesis, ETH Zurich, Automatic Control Lab, Sept. 2010. [Online]. Available: <http://www.youtube.com/watch?v=wqmP-y-a2qY>
- [18] Y. Wotao, “Gurobi mex: A matlab interface for gurobi,” 2011. [Online]. Available: <http://www.convexoptimization.com/wikimization/index.php/Gurobi%5Fmex>