

Alternating direction method of multipliers for strictly convex quadratic programs: Optimal parameter selection

Raghunathan, A.U.; Di Cairano, S.

TR2014-050 June 2014

Abstract

We consider an approach for solving strictly convex quadratic programs (QPs) with general linear inequalities by the alternating direction method of multipliers (ADMM). In particular, we focus on the application of ADMM to the QPs of constrained Model Predictive Control (MPC). After introducing our ADMM iteration, we provide a proof of convergence closely related to the theory of maximal monotone operators. The proof relies on a general measure to monitor the rate of convergence and hence to characterize the optimal step size for the iterations. We show that the identified measure converges at a Q-linear rate while the iterates converge at a 2-step Q-linear rate. This result allows us to relax some of the existing assumptions in optimal step size selection, that currently limit the applicability to the QPs of MPC. The results are validated through a large public benchmark set of QPs of MPC for controlling a four tank process.

American Control Conference (ACC), 2014

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Alternating Direction Method of Multipliers for Strictly Convex Quadratic Programs: Optimal Parameter Selection

Arvind U. Raghunathan and Stefano Di Cairano

Abstract—We consider an approach for solving strictly convex quadratic programs (QPs) with general linear inequalities by the alternating direction method of multipliers (ADMM). In particular, we focus on the application of ADMM to the QPs of constrained Model Predictive Control (MPC). After introducing our ADMM iteration, we provide a proof of convergence closely related to the theory of maximal monotone operators. The proof relies on a general measure to monitor the rate of convergence and hence to characterize the optimal step size for the iterations. We show that the identified measure converges at a Q-linear rate while the iterates converge at a 2-step Q-linear rate. This result allows us to relax some of the existing assumptions in optimal step size selection, that currently limit the applicability to the QPs of MPC. The results are validated through a large public benchmark set of QPs of MPC for controlling a four tank process.

I. INTRODUCTION

The Alternating Direction Method of Multipliers (ADMM) algorithm was first proposed by Gabay and Mercier [1] for the solution of variational inequalities that arise in the solution of partial differential equations and was developed in the 1970's in the context of optimization. In recent years, ADMM has emerged as a popular optimization algorithm for the solution of structured convex programs arising in several fields, see, e.g., the excellent introduction provided by [2].

The broad range of applications have also motivated the study of the convergence properties of ADMM. ADMM can be shown, under mild assumptions, to converge for all choices of the step-size [2]. Linear convergence of ADMM for strictly convex inequality constrained QPs was proved in [3], and relaxations were provided in [4], [5]. A first result on optimal ADMM step-size selection for strictly convex QPs with general inequality constraints was derived by [6]. The results of [6] require full row rank of the constraint matrix, making it inapplicable in several cases, for instance when some variables have upper and lower bounds.

Our interest in ADMM is especially motivated by its potential application to Model Predictive Control (MPC) algorithms. Model predictive control is a control algorithm for (constrained) dynamical systems that repeatedly solves a finite horizon optimal control formulated from the system dynamics, constraints, and a user specified cost function. For linear systems subject to linear constraints and with a quadratic cost function, the MPC finite horizon optimal control problem can be formulated as a QP [7]. Since in recent years MPC has been increasingly applied to systems

with fast dynamics and low computing power embedded processors [8], [9], [10] low complexity fast optimization algorithms have been investigated in the MPC context, see e.g., [11], [12], [13] and the references therein.

ADMM has been previously explored in the context of MPC in [14], [15], in particular for solving the QP via different decompositions. As mentioned earlier, the assumptions required by [6] for optimal ADMM step size selection do not hold in general for the QPs of MPC. The authors of [6] proposed a strategy remarking that it is heuristic.

In this paper we aim at solving by ADMM strictly convex QPs with general inequalities, which include those that are generated by constrained MPC. For this class of QPs, we aim at establishing the optimal step size for ADMM algorithms, noting that the optimal step size derived in the work of [6] is not applicable because the assumptions of [6] do not hold. Following [2], [14] we split the QP into two blocks, an unconstrained QP and a projection onto the general inequalities, and we provide a proof of 2-step Q-linear convergence that is based on a different method from [2], [3], [6], and that is closely related to the theory of maximal monotone operators introduced in [16]. In particular, we exploit the structure of the class of strictly convex QPs to provide short, self-contained proofs of convergence that identify the optimal step size.

The rest of the paper is organized as follows. Section II introduces the QP formulation and describes the ADMM algorithm. Convergence analysis of the algorithm is provided and the optimal selection for the step-size of ADMM is derived in Section III. Numerical performance on a simple QP is presented in Section IV, and in Section V we present some simulations results on MPC problems. Conclusions and future work are discussed in Section VI.

Notation: We will denote by \mathbb{R} the set of reals, by \mathbb{Z} the set of integers and by \mathbb{S}^n the set of symmetric $n \times n$ matrices. All vectors will be assumed to be column vectors. For a vector $x \in \mathbb{R}^n$, x^T will denote its transpose and for two vectors x, y the notation (x, y) will denote the stacking of the individual vectors. For a matrix $A \in \mathbb{R}^{n \times n}$, $\rho(A)$ denotes the spectral radius of A , $\lambda_i(A)$ denotes the eigenvalues of A and $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of the matrix A . For a matrix $A \in \mathbb{S}^n$, $A(\succeq) \succ 0$ will denote that matrix A is positive (semi-)definite. For a set \mathcal{Y} , $\text{int}(\mathcal{Y})$ will denote the strict interior of \mathcal{Y} . For a convex set $\mathcal{Y} \subset \mathbb{R}^n$, $\mathbb{P}_{\mathcal{Y}}(x)$ will denote the orthogonal projection of x onto the set. For a matrix $M \in \mathbb{R}^{n \times n}$, $M\mathbb{P}_{\mathcal{Y}}(x)$ will denote the product of matrix M and the vector resulting from the projection operation. We denote by I the identity

matrix and $(\mathbf{I} - \mathbb{P}_{\mathcal{Y}})(x)$ will denote $x - \mathbb{P}_{\mathcal{Y}}(x)$. The notation $\lambda \perp x \in \mathcal{Y}$ will denote the satisfaction of the following inequality $\lambda^T(x' - x) \geq 0 \forall x' \in \mathcal{Y}$. This is also called as the variational inequality. We use $\|\cdot\|$ to denote the 2-norm for vectors and matrices and $\|x\|_M^2 = x^T M x$ for some $M \succeq 0$. A sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* is said to converge at a Q-linear rate if $\|x^{k+1} - x^*\| \leq \alpha \|x^k - x^*\|$ where $0 < \alpha < 1$.

II. ADMM ALGORITHM FOR QUADRATIC PROGRAMS

Consider the QP:

$$\begin{aligned} \min_{\mathbf{y}} \quad & \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \mathbf{q}^T \mathbf{y} \\ \text{s.t. } \quad & \mathbf{y} \in \mathcal{Y} \end{aligned} \quad (1)$$

where $\mathbf{y}, \mathbf{q} \in \mathbb{R}^n$, $\mathbf{Q} \in \mathbb{S}^n \succ 0$ and \mathcal{Y} is a nonempty, polyhedron¹ $\mathcal{Y} = \{\mathbf{y} | \mathbf{A} \mathbf{y} \leq \mathbf{B}\}$. For example, \mathcal{Y} defines bounds on the components of \mathbf{y} and general inequalities formulated on \mathbf{y} . We make the following assumptions.

Assumption 1: The QP (1) is strictly feasible i.e. $\exists \mathbf{y}$ such that $\mathbf{y} \in \text{int}(\mathcal{Y})$ and the optimal value is finite.

The following result holds from strong duality [17].

Lemma 1: Suppose Assumption 1 holds. Then, there exists an optimal solution \mathbf{y}^* to (1) and multipliers $\boldsymbol{\lambda}^*$ satisfying,

$$\begin{aligned} \mathbf{Q} \mathbf{y}^* - \boldsymbol{\lambda}^* + \mathbf{q} &= 0 \\ \boldsymbol{\lambda}^* \perp \mathbf{y}^* &\in \mathcal{Y}. \end{aligned} \quad (2)$$

The last constraint in (2) is a variational inequality. In the case that $\mathcal{Y} = [\mathbf{y}_{\min}, \mathbf{y}_{\max}]$ the variational inequality signifies that $\lambda_i^* \geq 0$ if $y_i^* = \mathbf{y}_{\min}^i$, $\lambda_i^* \leq 0$ if $y_i^* = \mathbf{y}_{\max}^i$ and $\lambda_i^* = 0$ otherwise. In the case of one-sided inequalities such as $\mathcal{Y} = [\mathbf{y}_{\min}, \infty)$ the variational inequality reduces to the linear complementarity constraint $\boldsymbol{\lambda}^* \geq 0 \perp \mathbf{y}^* \geq \mathbf{y}_{\min}$.

Consider the following reformulation of the QP in (1) first proposed in [2] and also used in [14],

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{w}} \quad & \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \mathbf{q}^T \mathbf{y} \\ \text{s.t. } \quad & \mathbf{w} \in \mathcal{Y} \\ & \mathbf{y} = \mathbf{w} \end{aligned} \quad (3)$$

The advantage of (3) is that the inequalities are placed on different variables which are still coupled through $\mathbf{y} = \mathbf{w}$. The ADMM algorithm dualizes the constraints in the objective using multipliers $\boldsymbol{\lambda}$ and augments the objective with a penalty on the squared norm of the violation of the equality constraints. This can be represented as:

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{w}} \quad & L(\mathbf{y}, \mathbf{w}, \boldsymbol{\lambda}) := \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \mathbf{q}^T \mathbf{y} + \frac{\beta}{2} \|\mathbf{y} - \mathbf{w} - \frac{\boldsymbol{\lambda}}{\beta}\|^2 \\ \text{s.t. } \quad & \mathbf{w} \in \mathcal{Y} \end{aligned} \quad (4)$$

for $\beta > 0$. This results in a problem where the coupling now occurs only in the objective function. The steps of the

ADMM algorithm can be summarized as:

$$\begin{aligned} \mathbf{y}^{k+1} &= \arg \min_{\mathbf{y}} L(\mathbf{y}, \mathbf{w}^k, \boldsymbol{\lambda}^k) \\ &= \mathbf{M}(\mathbf{w}^k + \tilde{\boldsymbol{\lambda}}^k - \tilde{\mathbf{q}}) \\ \mathbf{w}^{k+1} &= \arg \min_{\mathbf{w}} L(\mathbf{y}^{k+1}, \mathbf{w}, \boldsymbol{\lambda}^k) \text{ s.t. } \mathbf{w} \in \mathcal{Y} \\ &= \mathbb{P}_{\mathcal{Y}}(\mathbf{y}^{k+1} - \tilde{\boldsymbol{\lambda}}^k) \\ \tilde{\boldsymbol{\lambda}}^{k+1} &= \tilde{\boldsymbol{\lambda}}^k + \mathbf{w}^{k+1} - \mathbf{y}^{k+1} \end{aligned} \quad (5)$$

where $\mathbf{M} := (\mathbf{Q}/\beta + \mathbf{I})^{-1}$ and $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}/\beta$, $\tilde{\mathbf{q}} = \mathbf{q}/\beta$. Substituting for \mathbf{y}^{k+1} in (5) and simplifying we obtain,

$$\begin{aligned} \mathbf{w}^{k+1} &= \mathbb{P}_{\mathcal{Y}}(\mathbf{M} \mathbf{w}^k + (\mathbf{M} - \mathbf{I}) \tilde{\boldsymbol{\lambda}}^k - \mathbf{M} \tilde{\mathbf{q}}) \\ \tilde{\boldsymbol{\lambda}}^{k+1} &= (\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(\mathbf{M} \mathbf{w}^k + (\mathbf{M} - \mathbf{I}) \tilde{\boldsymbol{\lambda}}^k - \mathbf{M} \tilde{\mathbf{q}}) \end{aligned} \quad (6)$$

Lemma 2 shows that the variational inequality in (2) holds between \mathbf{w}^{k+1} and $\tilde{\boldsymbol{\lambda}}^{k+1}$. The algorithm (5) can be viewed as attaining primal and dual feasibility in the limit.

Lemma 2: At every iteration of the ADMM algorithm the updates $\mathbf{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}$ in (5) satisfy the variational inequality, i.e. $\tilde{\boldsymbol{\lambda}}^{k+1} \perp \mathbf{w}^{k+1} \in \mathcal{Y}$.

Proof: From the definition of projection operator, $\mathbb{P}_{\mathcal{Y}}(v) := \arg \min_{\theta \in \mathcal{Y}} \frac{1}{2} \|\theta - v\|^2$. From the convexity of \mathcal{Y} we have that at the solution any feasible direction is non-decreasing one for the objective. In other words,

$$\begin{aligned} (\mathbb{P}_{\mathcal{Y}}(v) - v)^T (v' - \mathbb{P}_{\mathcal{Y}}(v)) &\geq 0 \forall v' \in \mathcal{Y} \\ \implies (\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(v) \perp \mathbb{P}_{\mathcal{Y}}(v) &\in \mathcal{Y}. \end{aligned} \quad (7)$$

Thus, the variational inequality is satisfied by the operators $\mathbb{P}_{\mathcal{Y}}(v)$ and $(\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(v)$ for any vector v . Observe that the updates steps for $\mathbf{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}$ in (6) are precisely of this form and the claim holds. ■

The following result shows the equivalence between the fixed points of the iteration (5) and minimizer of (1).

Theorem 1: Suppose Assumption 1 holds. Then, if $(\mathbf{y}, \mathbf{w}, \tilde{\boldsymbol{\lambda}})$ is a fixed point of (5), then \mathbf{y} is a minimizer of (1) with multipliers $\beta \tilde{\boldsymbol{\lambda}}$ for the inequality constraints. Conversely, if $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a minimizer of (1) then $(\mathbf{x}^*, \mathbf{x}^*, \boldsymbol{\lambda}^*/\beta)$ is a fixed point of (5).

Proof: Suppose that $(\mathbf{y}, \mathbf{w}, \tilde{\boldsymbol{\lambda}})$ is a fixed point of (5). From the update for $\tilde{\boldsymbol{\lambda}}$ we obtain,

$$\tilde{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}} + \mathbf{w} - \mathbf{y} \implies 0 = \mathbf{w} - \mathbf{y} \implies \mathbf{y} \in \mathcal{Y}$$

where the second implication follows from the update equation for \mathbf{w} , since $\mathbf{w}^{k+1} \in \mathcal{Y}$. From Lemma 2, $\mathbf{w}, \tilde{\boldsymbol{\lambda}}$ satisfy the variational inequality in (2). Since $\mathbf{y} = \mathbf{w}$ and $\beta > 0$, we have that, $\beta \tilde{\boldsymbol{\lambda}} \perp \mathbf{y} \in \mathcal{Y}$. Also, from the update step for \mathbf{y} in (5) we have that,

$$\begin{aligned} (\mathbf{Q} + \beta \mathbf{I}) \mathbf{y} &= \beta \mathbf{w} + \beta \tilde{\boldsymbol{\lambda}} - \mathbf{q} = \beta \mathbf{y} + \beta \tilde{\boldsymbol{\lambda}} - \mathbf{q} \\ \implies \mathbf{Q} \mathbf{y} - \beta \tilde{\boldsymbol{\lambda}} + \mathbf{q} &= 0 \end{aligned}$$

which are the first order optimality conditions in (2). We have that $(\mathbf{x}, \beta \tilde{\boldsymbol{\lambda}})$ satisfy stationary conditions (2) which are also the sufficient conditions for a minimizer of (1). Thus, the first claim holds.

¹The analysis in this paper holds for general closed and convex sets \mathcal{Y} .

Next, suppose that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ solves (1). Hence, from (2)

$$\mathbf{Q}\mathbf{y}^* - \boldsymbol{\lambda}^* + \mathbf{q} = 0 \implies (\mathbf{Q} + \beta\mathbf{I})\mathbf{y}^* = \beta\mathbf{y}^* + \boldsymbol{\lambda}^* - \mathbf{q}$$

which is the fixed point of the update step for \mathbf{y} in (5) with $\mathbf{y}^{k+1} = \mathbf{w}^k = \mathbf{y}^*$, $\tilde{\boldsymbol{\lambda}}^k = \boldsymbol{\lambda}^*/\beta$. Further since from (2), $\boldsymbol{\lambda}^* \perp \mathbf{y}^* \in \mathcal{Y}$ and $\beta > 0$ we have that $\boldsymbol{\lambda}^*/\beta \perp \mathbf{y}^* \in \mathcal{Y}$ which by definition implies,

$$\begin{aligned} & (\boldsymbol{\lambda}^*/\beta)^T(v' - \mathbf{y}^*) \geq 0 \quad \forall v' \in \mathcal{Y} \\ \implies & (\mathbf{y}^* - \mathbf{y}^* + \boldsymbol{\lambda}^*/\beta)^T(v' - \mathbf{y}^*) \geq 0 \quad \forall v' \in \mathcal{Y}. \end{aligned}$$

Hence, $\mathbf{y}^* = \mathbb{P}_{\mathcal{Y}}(\mathbf{y}^* - \boldsymbol{\lambda}^*/\beta)$ since \mathbf{y}^* satisfies the first order optimality conditions for the convex orthogonal projection problem in (7). Thus, $\mathbf{y}^*, \boldsymbol{\lambda}^*/\beta$ are fixed points of the update step for \mathbf{w} in (5). The fixed point of the update equation in $\tilde{\boldsymbol{\lambda}}$ holds trivially and the second claim also holds. \blacksquare

A. ADMM Approach of [6]

The authors in [6] proposed an ADMM algorithm for the class of QPs in (1), which is reformulated as

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{z}} & \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \mathbf{q}^T \mathbf{y} \\ \text{s.t.} & \mathbf{A} \mathbf{y} + \mathbf{z} = \mathbf{b}, \mathbf{z} \geq 0 \end{aligned} \quad (8)$$

The formulation of (8) for applying ADMM is,

$$\begin{aligned} \min_{\mathbf{y}, \mathbf{z}} & \widehat{L}(\mathbf{y}, \mathbf{z}, \boldsymbol{\nu}) := \frac{1}{2} \mathbf{y}^T \mathbf{Q} \mathbf{y} + \mathbf{q}^T \mathbf{y} + \frac{\beta}{2} \|\mathbf{A} \mathbf{y} + \mathbf{z} - \mathbf{b} - \frac{\boldsymbol{\nu}}{\beta}\|^2 \\ \text{s.t.} & \mathbf{z} \geq 0 \end{aligned} \quad (9)$$

and the ADMM algorithm proposed in [6] performs the following operations,

$$\begin{aligned} \mathbf{y}^{k+1} &= \arg \min_{\mathbf{y}} \widehat{L}(\mathbf{y}, \mathbf{z}^k, \boldsymbol{\nu}^k) \\ &= \widehat{\mathbf{M}} \mathbf{A}^T (-\mathbf{z}^k + \mathbf{b} + \frac{\boldsymbol{\nu}^k}{\beta} - \tilde{\mathbf{q}}) \\ \mathbf{z}^{k+1} &= \arg \min_{\mathbf{z}} \widehat{L}(\mathbf{y}^{k+1}, \mathbf{z}, \tilde{\boldsymbol{\lambda}}^k) \text{ s.t. } \mathbf{z} \geq 0 \\ &= \max(0, -\mathbf{A} \mathbf{y}^{k+1} + \mathbf{b} + \frac{\boldsymbol{\nu}^k}{\beta}) \\ \frac{\boldsymbol{\nu}^{k+1}}{\beta} &= \frac{\boldsymbol{\nu}^k}{\beta} - (\mathbf{A} \mathbf{y}^{k+1} + \mathbf{z}^{k+1} - \mathbf{b}) \end{aligned} \quad (10)$$

where $\widehat{\mathbf{M}} = (\mathbf{Q}/\beta + \mathbf{A} \mathbf{A}^T)^{-1}$. The main advantage of the approach in [6] is that the subproblem for update of \mathbf{z} is simple. On the other hand, the subproblem for \mathbf{w} update in (5) can be quite complicated depending on the structure of \mathcal{Y} . The theoretical results on optimal step size derived in [6] are only valid when \mathbf{A} is invertible or full row rank. We show in Section III that for the approach proposed in this paper the theoretically optimal value of the step size can be derived, regardless of the row rank of \mathbf{A} .

III. CONVERGENCE OF ALGORITHM

In this section we show that the ADMM algorithm converges to a solution of (1) for any choice of the parameter $\beta > 0$ and also characterize the optimal β that maximizes

the rate of convergence. First, we introduce some results on the projection operator.

Lemma 3: For any $v, v' \in \mathbb{R}^n$:

- (i) $(\mathbb{P}_{\mathcal{Y}}(v) - \mathbb{P}_{\mathcal{Y}}(v'))^T ((\mathbf{I} - \mathbb{P}_{\mathcal{Y}})(v) - (\mathbf{I} - \mathbb{P}_{\mathcal{Y}})(v')) \geq 0$
- (ii) $\|(\mathbb{P}_{\mathcal{Y}}(v), (\mathbf{I} - \mathbb{P}_{\mathcal{Y}})(v)) - (\mathbb{P}_{\mathcal{Y}}(v'), (\mathbf{I} - \mathbb{P}_{\mathcal{Y}})(v'))\| \leq \|v - v'\|$
- (iii) $\|(2\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(v) - (2\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(v')\| \leq \|v - v'\|$

Proof: The result follows by noting that $\mathbb{P}_{\mathcal{Y}}(v) := \arg \min_{\theta} \mathbb{I}_{\mathcal{Y}}(\theta) + \frac{1}{2} \|\theta - v\|^2$ where $\mathbb{I}_{\mathcal{Y}}(\theta)$ is the set membership indicator function being 0 when $\theta \in \mathcal{Y}$ and ∞ otherwise. Thus, $\mathbb{P}_{\mathcal{Y}} = (\mathbf{I} + \partial \mathbb{I}_{\mathcal{Y}})^{-1}$ where $\partial \mathbb{I}_{\mathcal{Y}}$ is the subgradient of the extended real-valued convex function, $\mathbb{I}_{\mathcal{Y}}(\cdot)$ and hence, $\mathbb{P}_{\mathcal{Y}}(\cdot)$ is a maximal monotone operator [16]. The claims follow from Proposition 1 of [16]. \blacksquare

The next step in the proof is to show that the iterates generated by (6) are bounded. Prior to that we state a result on the spectral radius of \mathbf{M} .

Lemma 4: Suppose $\mathbf{Q} \succ 0$. Then, the spectral radius of \mathbf{M} satisfies $\rho(\mathbf{M}) < 1$ and $\mathbf{M} \succ 0$.

Proof: By definition, $\mathbf{M} = (\mathbf{Q}/\beta + \mathbf{I})^{-1}$. Thus, the eigenvalues of \mathbf{M} are $\beta/(\beta + \lambda(\mathbf{Q}))$ which lie in $(0, 1)$ for all $\beta > 0, \lambda(\mathbf{Q}) > 0$. The claim holds. \blacksquare

To characterize the convergence rate, consider the sequence $\{\mathbf{v}^k\}$, where

$$\begin{aligned} \mathbf{v}^{k+1} &= \mathbf{M} \mathbf{w}^{k+1} + (\mathbf{M} - \mathbf{I}) \tilde{\boldsymbol{\lambda}}^{k+1} - \mathbf{M} \tilde{\mathbf{q}} \\ \mathbf{v}^* &= \mathbf{M} \mathbf{w}^* + (\mathbf{M} - \mathbf{I}) \tilde{\boldsymbol{\lambda}}^* - \mathbf{M} \tilde{\mathbf{q}} \end{aligned} \quad (11)$$

where $\tilde{\boldsymbol{\lambda}}^* = \boldsymbol{\lambda}^*/\beta$. An obvious reason for using \mathbf{v}^k to measure convergence of the iteration in (6) is \mathbf{v}^k appears in the update equations for both \mathbf{w} and $\tilde{\boldsymbol{\lambda}}$ in (6). More importantly, the choice of \mathbf{v}^k is motivated by Lemma 3(ii) from which we have that,

$$\begin{aligned} & \|(\mathbf{w}^{k+1}, \tilde{\boldsymbol{\lambda}}^{k+1}) - (\mathbf{w}^*, \tilde{\boldsymbol{\lambda}}^*)\| \\ &= \|(\mathbb{P}_{\mathcal{Y}}(\mathbf{v}^k), (\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(\mathbf{v}^k)) - (\mathbb{P}_{\mathcal{Y}}(\mathbf{v}^*), (\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(\mathbf{v}^*))\| \\ &\leq \|\mathbf{v}^k - \mathbf{v}^*\|. \end{aligned} \quad (12)$$

Consequently, convergence of $\{\mathbf{v}^k\}$ ensures convergence to a fixed point of (6) since $\mathbf{w}^k, \tilde{\boldsymbol{\lambda}}^k$ are uniquely determined from \mathbf{v}^k .

Simplifying the operators in (11) we obtain,

$$\begin{aligned} \mathbf{v}^{k+1} &= \mathbf{M} \mathbb{P}_{\mathcal{Y}}(\mathbf{v}^k) + (\mathbf{M} - \mathbf{I})(\mathbb{P}_{\mathcal{Y}} - \mathbf{I})(\mathbf{v}^k) - \mathbf{M} \tilde{\mathbf{q}} \\ &= ((2\mathbf{M} - \mathbf{I})\mathbb{P}_{\mathcal{Y}} + \mathbf{I} - \mathbf{M})(\mathbf{v}^k) - \mathbf{M} \tilde{\mathbf{q}} \\ &= \left((2\mathbf{M} - \mathbf{I})\mathbb{P}_{\mathcal{Y}} - \frac{2\mathbf{M} - \mathbf{I}}{2} + \frac{\mathbf{I}}{2} \right) (\mathbf{v}^k) - \mathbf{M} \tilde{\mathbf{q}} \\ &= \left(\frac{2\mathbf{M} - \mathbf{I}}{2} (2\mathbb{P}_{\mathcal{Y}} - \mathbf{I}) + \frac{\mathbf{I}}{2} \right) (\mathbf{v}^k) - \mathbf{M} \tilde{\mathbf{q}}. \end{aligned}$$

Lemma 5: Suppose Assumption 1 holds. Then,

$$\|\mathbf{v}^k - \mathbf{v}^*\| \leq \|(\mathbf{w}^k, \tilde{\boldsymbol{\lambda}}^k) - (\mathbf{w}^*, \tilde{\boldsymbol{\lambda}}^*)\|. \quad (13)$$

Proof: Squaring the left hand side of (13),

$$\begin{aligned}
\|v^k - v^*\|^2 &= \|M(w^k - w^*) + (I - M)(\tilde{\lambda}^k - \tilde{\lambda}^*)\|^2 \\
&= \|w^k - w^*\|_{M^2}^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|_{(I-M)^2}^2 \\
&\quad + 2(w^k - w^*)^T M(I - M)(\tilde{\lambda}^k - \tilde{\lambda}^*) \\
&\leq \|w^k - w^*\|_{M^2}^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|_{(I-M)^2}^2 \\
&\quad + \|w^k - w^*\|_{M(I-M)}^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|_{M(I-M)}^2 \\
&\leq \|w^k - w^*\|_M^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|_{(I-M)}^2 \\
&\leq \|w^k - w^*\|^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|^2
\end{aligned} \tag{14}$$

where the equality is from (11), the second equality is a simple expansion of the terms, the first inequality follows from

$$\begin{aligned}
M(I - M) &\succeq 0 \\
\implies \|w^k - w^* - (\tilde{\lambda}^k - \tilde{\lambda}^*)\|_{M(I-M)}^2 &\geq 0 \\
\implies \|w^k - w^*\|_{M(I-M)}^2 + \|\tilde{\lambda}^k - \tilde{\lambda}^*\|_{M(I-M)}^2 \\
&\geq 2(w^k - w^*)^T M(I - M)(\tilde{\lambda}^k - \tilde{\lambda}^*).
\end{aligned}$$

The second inequality in (14) follows by collecting terms and the final inequality holds since $\rho(M) < 1$, $M \succ 0$ (Lemma 4). Hence, the claim holds. ■

We can now state the linear convergence result.

Theorem 2: Suppose Assumption 1 holds. Then, the sequence $\{v^k\}$ is Q-linearly convergent and the sequence $\{(w^k, \tilde{\lambda}^k)\}$ is 2-step Q-linearly convergent.

Proof: The convergence rate of the iteration can be deduced from,

$$\begin{aligned}
\|v^{k+1} - v^*\| &= \left\| \left(\frac{2M - I}{2} (2\mathbb{P}y - I) + \frac{I}{2} \right) (v^k) \right. \\
&\quad \left. - \left(\frac{2M - I}{2} (2\mathbb{P}y - I) + \frac{I}{2} \right) (v^*) \right\| \\
&\leq \left(\frac{\|2M - I\|}{2} + \frac{1}{2} \right) \|v^k - v^*\|
\end{aligned}$$

where the inequality above follows from Lemma 3(iii). Since $M \succ 0$ and $\rho(M) < 1$ (Lemma 4) we have that $\frac{\|2M - I\|}{2} + \frac{1}{2} < 1$ and hence, $\{v^k\}$ converges Q-linearly. Consider,

$$\begin{aligned}
\|(w^{k+2}, \tilde{\lambda}^{k+2}) - (w^* - \tilde{\lambda}^*)\| &\leq \|v^{k+1} - v^*\| \\
&\leq \alpha \|v^k - v^*\| \leq \alpha \|(w^k, \tilde{\lambda}^k) - (w^* - \tilde{\lambda}^*)\|
\end{aligned}$$

where the first inequality follows from (12), second from linear convergence for $\{v^k\}$ for some $\alpha \in (0, 1)$ and the last inequality from (13). This proves the claim. ■

The optimal β^* should be chosen to minimize $\frac{\|2M - I\|}{2} + \frac{1}{2}$ where the eigenvalues of M satisfy $\lambda(M) = \beta/(\beta + \lambda(Q))$. Thus, the optimal choice for the step size is given by,

$$\beta^* = \arg \min_{\beta > 0} \max_i \left(\left| \frac{\beta}{\beta + \lambda_i(Q)} - \frac{1}{2} \right| + \frac{1}{2} \right).$$

We can easily rearrange the right hand side to obtain,

$$\beta^* = \arg \min_{\beta > 0} \max_i \left(\left| \frac{\beta/\lambda_i(Q)}{\beta/\lambda_i(Q) + 1} - \frac{1}{2} \right| + \frac{1}{2} \right). \tag{15}$$

The eq. (15) is identical in form to that obtained in eq. (36) of [6], which is given by,

$$\beta^{*,[6]} = \arg \min_{\beta > 0} \max_i \left(\left| \frac{\beta \lambda_i(AQ^{-1}A^T)}{\beta \lambda_i(AQ^{-1}A^T) + 1} - \frac{1}{2} \right| + \frac{1}{2} \right).$$

In essence, the optimal step-size in our approach depends on $\lambda(Q^{-1})$ while the approach in [6] depends on $\lambda(AQ^{-1}A^T)$. Thus, the functional expression for the optimal step-size can be readily seen to be identical to the one in [6], where Q^{-1} is substituted for $AQ^{-1}A^T$. Hence, the analysis used to obtain the optimal step-size in [6] holds with such a substitution.

Theorem 3: Suppose Assumption 1 holds and $Q \succ 0$. Then, the optimal step-size for the class of strictly convex QPs in (1) is

$$\beta^* = \sqrt{\lambda_{\min}(Q)\lambda_{\max}(Q)} \tag{16}$$

and the corresponding convergence rate is,

$$\frac{1/\lambda_{\min}(Q)}{1/\lambda_{\min}(Q) + 1/\left(\sqrt{\lambda_{\min}(Q)\lambda_{\max}(Q)}\right)}. \tag{17}$$

Proof: The proof is similar to that of Theorem 4 in [6], with Q^{-1} substituted for $AQ^{-1}A^T$, and hence not repeated. ■

IV. SIMPLE QP

In order to verify the performance of the proposed optimal step-size selection on the class of strictly convex QPs (1) we consider first a simple strictly convex QP, that was proposed in [6], where

$$\begin{aligned}
Q &= \begin{pmatrix} 40.513 & 0.069 \\ 0.069 & 40.389 \end{pmatrix}, q = 0, \\
A &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ 0.1151 & 0.9934 \end{pmatrix}, b = \begin{pmatrix} 6 \\ 6 \\ -0.3422 \end{pmatrix}.
\end{aligned}$$

In [6] the authors exploit this QP to demonstrate the inapplicability of their optimal step-size selection $\beta^{*,[6]}$, since A is not full row rank. The authors in [6] propose to use the smallest non-zero eigenvalue of $AQ^{-1}A^T$ instead of the actual smallest eigenvalue $\lambda_{\min}(AQ^{-1}A^T)$, which is 0. For the considered QP problem, the eigenvalues of $AQ^{-1}A^T$ are (0, 0.0247, 0.0495). The heuristic step-size value is $\beta^{\text{heur},[6]} = 1/\sqrt{0.0247 \times 0.0495} = 28.6$.

On the other hand the approach proposed in this paper can be applied to the considered QP to find the optimal value of step-size, resulting in $\beta^* = 40.4509$. Figure 1 reports the iterations required to attain convergence for different values of the step-size parameter β for our ADMM algorithm (in red) and the one proposed in [6] (in blue). The initial guess for the ADMM iterations in (10) was set to: $z^0 = \max(0, b - A(-Q^{-1}q))$, $\nu^0 = 0$ while for our approach (5) the initial guess was set to $w^0 = -Q^{-1}q$, $\lambda^0 = 0$. The algorithm proposed in this paper terminates when the 2-norm of the primal and dual infeasibility, defined respectively as $\|\tilde{\lambda}^{k+1} - \tilde{\lambda}^k\|$ and $\|\beta(w^{k+1} - w^k)\|$, are both below 10^{-6} . The iterations in (10) terminate when $\beta\|z^{k+1} - z^k\|$ and

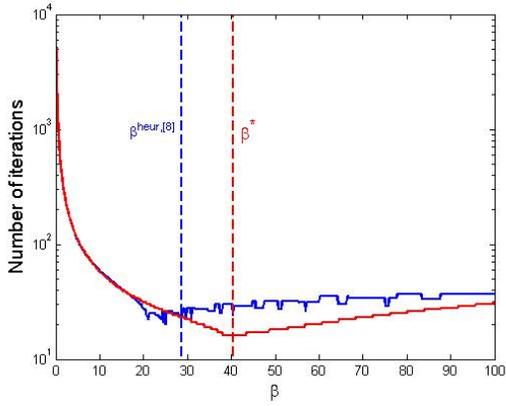


Fig. 1. Plot of number of iterations for convergence against the step-size β for the ADMM algorithm in [6] in blue and our approach in red. The dotted blue line indicates the heuristic value proposed in [6] while the red dotted line indicates the β^* from (16).

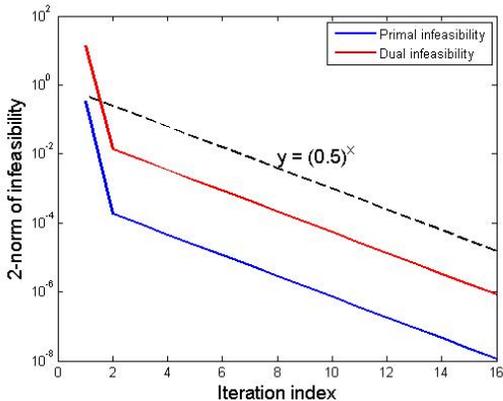


Fig. 2. Plot of norm of the primal and dual infeasibility against the iteration index and graph of the linear function defining the convergence rate according to (17).

$\|v^{k+1} - v^k\|/\beta$ are both below 10^{-6} . In our numerical simulations the lowest iteration count of 20 was achieved for $\beta = 25.0$ for the ADMM algorithm in [6] while for the heuristic value of $\beta^{\text{heur},[6]} = 28.6$ the algorithm required 24 iterations to attain convergence. The ADMM algorithm proposed in this paper attains the lowest iteration count of 16 for all values of $\beta \in [38.0, 43.4]$ which includes the optimal step-size chosen according to the approach described in this paper, $\beta^* = 40.4509$. According to (17) the optimal rate of convergence is 0.501. Figure 2 reports the 2-norm of the primal and dual infeasibility respectively, versus the iteration index. Also Figure 2 reports the graph of the function $y = 0.5^x$. Hence, we can clearly see that the primal and dual infeasibility decrease at the rate predicted by (17), thus validating the theory.

V. MODEL PREDICTIVE CONTROL

In this section we apply the results developed in this paper to the QP of Model Predictive Control (MPC). Consider the

discrete-time linear system,

$$x(k+1) = Ax(k) + Bu(k) + Fr(k) \quad (18)$$

where $x \in \mathbb{R}^{n_x}$ is the state vector, $u \in \mathbb{R}^{n_u}$ is the control input vector, $r \in \mathbb{R}^{n_r}$ is the reference vector, $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $F \in \mathbb{R}^{n_x \times n_r}$ are the system, control and reference transfer matrices, respectively. At every discrete time step $k \in \mathbb{Z}$, $k \geq 0$, given the future sequence of references $\{r(k+i)\}_{i=0}^{N-1}$ the MPC algorithm solves the optimization problem

$$\begin{aligned} \min_{\{x_t\}_{t=0}^{N-1}, \{u_t\}_{t=0}^{N-1}} & \frac{1}{2} \sum_{t=0}^{N-1} (x_t^T Q_x x_t + u_t^T R u_t) + \frac{1}{2} x_N^T P x_N \\ \text{s.t. } & x_{t+1} = Ax_t + Bu_t + Fr_t \\ & (x_{t+1}, u_t) \in \mathcal{X} \times \mathcal{U} \\ & r_t = r(k+t) \\ & x_0 = x(k) \end{aligned} \quad (19)$$

where $x(k)$ is the state of the system at current time instant, $Q_x, P \in \mathbb{R}^{n_x \times n_x}$ are stage and terminal matrices on the state, respectively, $R \in \mathbb{R}^{n_u \times n_u}$ is the cost matrix for the controls and \mathcal{X}, \mathcal{U} are polyhedral sets defining feasible region for the states and controls. Typically, $Q_x \succeq 0$, $P \succ 0$ and $R \succ 0$. The MPC algorithm solves problem (19) to find the optimal input sequence $\{u_t^*\}_{t=0}^{N-1}$, and then applies to the system the control input $u(k) = u_0^*$. In (19), the state variables can be eliminated by exploiting the equations of the system dynamics, thus obtaining a strictly convex QP (1), where the optimization vector contains only the input sequence $\mathbf{y} = [u_0^T \dots u_{N-1}^T]^T$. If the MPC problems has only lower and upper bounds, similarly to [11], on the inputs, the update for \mathbf{w} entails a simple projection, while in the other cases, a projection on general polyhedra is required.

To verify the performance of the proposed optimal step size selection we consider the MPC problems for the quadruple-tank process [18] which has $n_x = 4$ states and $n_u = 2$ controls. The authors of [6] have made publicly available [19] the data of 170 QPs corresponding to such MPC problems for different initial conditions. The problems have a horizon of $N = 5$, which leads to QP where $\mathbf{y} \in \mathbb{R}^{10}$ once the states are eliminated from the MPC problem. The constraint matrix $\mathbf{A} \in \mathbb{R}^{40 \times 10}$ specifies lower and upper bounds on the controls and on some states.

As in the case of the simple QP in Section IV, the MPC problem results in the constraint matrix \mathbf{A} which does not have full row rank. Consequently, the optimal step-size analysis of [6] does not hold, and the heuristic approach previously described has to be applied. For the given data [19], the heuristic step-size from [6] recommends $\beta^{\text{heur},[6]} = 24.35$. For the ADMM algorithm presented in this paper the optimal step-size can be applied and it results in $\beta^* = 50.03$. Figure 3 reports the iterations to attain convergence for different values of the step size parameter β for the ADMM algorithm described here (in red) and the iterations of the algorithm proposed in [6] (in blue) for the first problem instance among the MPC problems provided

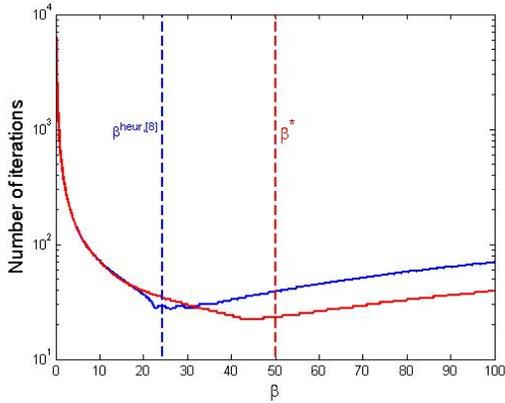


Fig. 3. Plot of number of iterations for convergence against the step size β for the ADMM algorithm in [6] in blue and our approach in red for the first instance of MPC problem in [19]. The dotted blue line indicates the heuristic value proposed in [6] while the red dotted line indicates the β^* from (16).

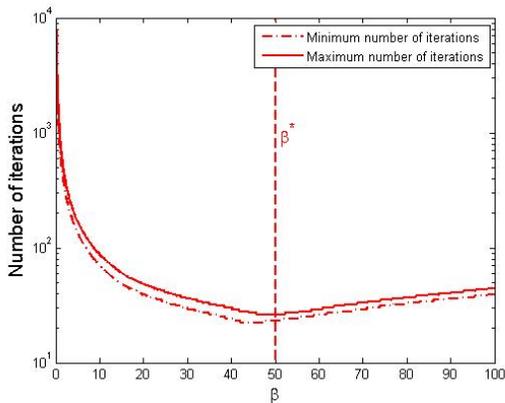


Fig. 4. Plot of minimum and maximum number of iterations to attain convergence over the set of 170 QPs for different values of the step size β for the algorithm proposed in this paper.

in [19]. The initial guess and termination criterion were identical to that used in Section IV. The lowest iteration count of 27 is attained for the ADMM algorithm in [6] for $\beta \in [26.0, 26.6]$ and the choice of $\beta = 24.35$ takes 29 iterations for convergence. For the algorithm proposed in this paper, the lowest iteration count of 22 is achieved for $\beta = [42.2, 47.3]$ and the choice of $\beta = 50.03$ takes 23 iterations for convergence.

Figure 4 reports the minimum and maximum number of iterations over the 170 QP instances for each value of the step size β . The optimal step size β^* selected by the approach proposed in this paper coincides with the value of the step size that minimizes the maximum number of iterations.

VI. CONCLUSION

We have presented an alternating direction method of multipliers for strictly convex QPs with particular focus on

the QPs arising from MPC. We proposed a convergence proof and a method of analysis that allowed us to select the optimal step-size. We have validated our approach for choosing the by a publicly available benchmark set of problems. A drawback of our approach is that, the subproblem for updating w may not be simple to solve, depending on the constraints. In the future we will extend our analytical framework to an algorithm that always solve simple subproblems.

REFERENCES

- [1] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems in finite-element approximations," *Computers and Mathematics with Applications*, vol. 2, pp. 17–40, 1976.
- [2] S. Boyd, N. Parikh, E. Chu, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [3] D. Boley, "Linear convergence of the alternating direction method of multipliers," Dept. of Comp. Sci. and Eng., U. Minnesota, Tech. Rep. TR 12-009, 2012.
- [4] M. Hong and Z.-Q. Luo, "On the Linear Convergence of the Alternating Direction Method of Multipliers," Tech. Rep. arXiv:1208.3922, 2013.
- [5] R. D. C. Monteiro and B. F. Svaiter, "Iteration-complexity of block-decomposition algorithms and the alternating minimization augmented lagrangian method," *SIAM Journal on Optimization*, vol. 23, no. 1, pp. 475–507, 2013.
- [6] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems," Tech. Rep. arXiv:1306.2454v1, 2013.
- [7] S. Di Cairano and A. Bemporad, "Model predictive control tuning by controller matching," *IEEE Tr. Automatic Control*, vol. 55, no. 1, pp. 185–190, 2010.
- [8] S. Di Cairano, "An industry perspective on MPC in large volumes applications: Potential Benefits and Open Challenges," in *Proc. 4th IFAC Nonlinear Model Predictive Control Conference*, Noordwijkerhout, The Netherlands, 2012, pp. 52–59.
- [9] S. Di Cairano, H. Tseng, D. Bernardini, and A. Bemporad, "Vehicle yaw stability control by coordinated active front steering and differential braking in the tire sideslip angles domain," *IEEE Tr. Contr. Sys. Technology*, vol. 21, no. 4, pp. 1236–1248, 2013.
- [10] S. Di Cairano, H. Park, and I. Kolmanovsky, "Model predictive control approach for guidance of spacecraft rendezvous and proximity maneuvering," *Int. J. Rob. Nonlinear Control*, 2012.
- [11] S. Richter, C. Jones, and M. Morari, "Real-time input-constrained mpc using fast gradient methods," in *Proc. 48th IEEE Conf. on Dec. and Control*, Shanghai, China, 2009, pp. 7387–7393.
- [12] I. Necoara and J. Suykens, "Application of a smoothing technique to decomposition in convex optimization," *IEEE Tr. Automatic Control*, vol. 53, no. 11, pp. 2674–2679, 2008.
- [13] S. Di Cairano, M. Brand, and S. Bortoff, "Projection-free parallel quadratic programming for linear model predictive control," *Int. J. Control*, 2013, in press, available online at www.tandfonline.com.
- [14] B. O'Donoghue, G. Stathopoulos, and S. Boyd, "A splitting method for optimal control," *IEEE Trans. Control Systems Technology* (to appear), 2013.
- [15] G. Stathopoulos, T. Keviczky, and Y. Wang, "A hierarchical time-splitting approach for solving finite-time optimal control problems," in *European Control Conference*, 2013.
- [16] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Opt.*, vol. 14, pp. 877–898, 1976.
- [17] O. L. Mangasarian, *Nonlinear Programming*. SIAM, 1987.
- [18] K. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, 2000.
- [19] (2013). [Online]. Available: http://www.dropbox.com/s/x2w74mpbezejbee/MPC_QP-quadtank_170_Np5_SxQ.mat